Advanced statistical learning method for multi-physics NDT-NDE

S Ahmed¹, P Calmon¹, R Miorelli¹, C Reboud¹ and A Massa²,³

¹CEA LIST, Centre de Saclay, F-91191 Gif-sur-Yvette, France
²ELEDIA Research Center (ELEDIA@UniTN - University of Trento)
Via Sommarive 9, I-38123 Trento, Italy
³ELEDIA Research Center (ELEDIA@L2S - UMR8506)
3 rue Joliot-Curie 91192 Gif-sur-Yvette, France
E-mail: roberto.miorelli@cea.fr

Abstract. This work presents an innovative multi-physics (MP) Learning-by-Examples (LBE) inversion methodology for real-time non-destructive testing (NDT). Eddy Current Testing (ECT) and Ultrasonic Testing (UT) data are effectively combined to deal with the localization and characterization of a crack inside a conductive structure. An adaptive sampling strategy is applied on ECT-UT data in order to build an optimal (i.e., having minimum cardinality and highly informative) training set. Support vector regression (SVR) is exploited to obtain a computationally-efficient and accurate surrogate model of the inverse operator and, subsequently, to perform real-time inversions on previously-unseen measurements provided by simulations. The robustness of the proposed MP-LBE approach is numerically assessed in presence of synthetic noisy test set and compared to single-physic (i.e., ECT or UT) inversion.

1. Introduction

Real time accurate inversion solution becomes the main priority in non-destructive testing and evaluation (NDT-NDE) applications. Among different iterative [1, 2, 3] and non-iterative [4, 5, 6, 7] inversion solutions, Learning by examples (LBE) strategy is getting more attention for having quasi-real time inversion capabilities. In this work, LBE has been adopted for a NDE problem where a narrow crack is occurred around a fastener (e.g., bore hole) within an inspected medium [8]. This is an important problem for the aging aircraft NDE community and Eddy Current Testing (ECT) is widely applied while the structure under test (SUT) thickness is thin. However, the penetration depth of the induced currents is limited by the skin depth. This makes the detection and resolution of defects more difficult as the depth increases. Whereas ultrasound testing (UT) NDT inspection is suitable for high resolution, but the inspection is affected by the surface roughness of the inspected medium [9]. That means, each of these NDT methods has some pros and cons according to their own physics. Moreover, for the mentioned problem at hand, ECT signal is mostly affected for the presence of fastener. Due to the significant probe impedance variation, the area of the fastener is acting as a circular defect within the inspected medium. The impedance variation due to the fastener is much stronger than narrow crack, thus, when the crack is placed deeper inside the SUT, the ECT signals contribution due to the presence of narrow crack becomes weaker. Conversely, UT signals are stronger for subsurface crack compare to the crack placed at the top surface of the SUT. As a consequence,
different impacts on the crack characterization and localization performance are expected based on ECT and UT methods. Thus, multi-physics (MP) data fusion (ECT-UT) has been applied to maximize the inversion performance for crack characterization and localization.

In general, LBE is a two-phases approach. During the preliminary phase (so called offline phase), a fast and accurate inverse/trained model is built based on a training set made of input-output (I/O) pairs by learning algorithm. The developed (trained) model from offline phase is then used to predict the output associated to an unknown test sample during the second phase (online phase). Within the framework of LBE, an adaptive sampling strategy combining Partial Least Square (PLS) [10] feature extraction and modified version of output space filling (OSF) [11] (i.e., PLS-OSF sampling [4]) has been adopted for obtaining optimal training sets by both ECT and UT methods separately. An updated version of PLS-OSF sampling algorithm has also been illustrated for dealing with ECT-UT data. Support vector regression (SVR) [12] is used to obtain accurate training model and perform real time inversion. Finally, the performance of the MP-LBE inversion schema for crack characterization and localization is compared to single-physic (i.e., ECT and UT) inversion on noisy data.

2. Mathematical formulation of forward and inverse problem

Let us consider a homogeneous plate made by aluminium 2024 alloy of thickness 6 mm, density $2.77 \text{ g.cm}^{-3}$, has been investigated by both ECT and UT NDT methods. The plate consists of a fastener (bore hole) of radius 3.75 mm and 6.00 mm height. The plate is affected by a single notch (e.g., narrow crack) of volume $\Omega$ having fixed width 0.01 mm and height 2 mm (Fig. 1) which is attached with the fastener. The crack is characterized by total $Q = 3$ descriptors of length ($l_c$), ligament ($\delta_c$) and angular distance ($\varphi_c$) (i.e., $p = (l_c, \delta_c, \varphi_c)$).

![Figure 1. Examples of studied (a) ECT and (b) UT configuration](image)

2.1. ECT treatment

The plate is inspected by a single coil working in absolute mode of frequency 1 kHz with lift off 1 mm. The coil impedance variation due to presence of the crack measured at the $k$–th ($k = 1, ..., K$) scanning position with respect to the flawless region is given by [13].

$$\chi^\text{ECT}_k = \frac{1}{I^2} \int_{\Omega} E^{\text{inc}}(r|r_k) \cdot \rho(r|r_k) d r$$

$I$ is the current flowing inside the coil while $E^{\text{inc}}(r|r_k)$ is the incident field generated at position $r$ in the flawless plate ($r_k = (x_k, y_k)$ represents the $k$–th coil position within the plate). $\rho(r|r_k)$ is the unknown induced current dipole density, which models the presence of the crack and is related to the total field, $E^{\text{tot}}(r|r_k)$ that can be expressed by $\rho(r|r_k) = \sigma(r) - \sigma \cdot E^{\text{tot}}(r|r_k)$. CIVA simulator [14] has been used as a forward operator $\Phi^\text{ECT}$ to generate ECT coil signals $\chi^\text{ECT}$. Due to the complex nature of the ECT signals, ECT signals are represented by $X^\text{ECT} = \left\{ \Re \{\chi^\text{ECT}_k\} ; \Im \{\chi^\text{ECT}_k\} \right\}_{k = 1, ..., K}$ of $F^\text{ECT} = 2K$ ECT features.
2.2. Ultrasound testing treatment
The plate has been investigated by a ray probe by using water coupling medium (i.e., density 1 g.cm\(^{-3}\)). The probe is acting both for transmitting and receiving UT signals. More details of the treated problem and probe definition are available in [15]. CIVA uses a hybrid model known as Physical Theory of Diffraction (PTD), based on Kirchhoff approximation and high-frequency Geometrical Theory of Diffraction (GTD) for generating diffraction/scattering waves from planner-like defects. The scattered field, for the PTD model can be expressed by [16]

\[
u^\text{Scat}(\text{PTD})(r) = \nu^\text{Rayleigh}(r) + \sum_{\beta} \left( D_\beta^{\text{GTD}}(r) - D_\beta^{\text{KA}}(r) \right) \frac{e^{i\lambda_\beta S_\beta}}{\sqrt{\lambda_\beta L_\beta}} e_\beta(r). \tag{2}
\]

where, \(\alpha = L, TV, \text{or} \ TH\) (Longitudinal, Transverse Vertical or Transverse Horizontal, respectively) incident type wave vector and \(\beta = L, TV, \text{or} \ TH\) is the scatter type wave vector. \(S_\beta\) is the distance between the diffraction point \(r_\beta^0\) and the observation point \(r\). \(L_\beta\) is a parameter distance. \(D_\beta^{\text{GTD}}\) is the GTD diffraction coefficient and \(D_\beta^{\text{KA}}\) is the Kirchoff edge diffraction coefficient. \(\nu^\text{Kir}(r)\) is the displacement scattered field at the observation \(r\), and the Rayleigh field \(\nu^\text{Rayleigh}(r)\) comprises the surface waves. The reflections/scatters wave have been collected through C-Scan (e.g., maximum ray amplitude available at each inspection point) and represented by \(\chi^\text{UT}\). CIVA simulator [14] has been utilized as a forward operator \(\Phi^\text{UT}\{\cdot\}\) in order to generate \(\overline{U^T}\) data. Unlike, ECT signals, UT signals contain only the real data, hence UT signals are represented by \(\overline{U^T} = \{\chi^\text{UT}_k; k = 1, ..., K\}\) of \(F^\text{UT} = K\) UT features.

2.3. Data fusion using ECT and UT
In this case, ECT signals and UT signals are generated separately by their own forward solver (i.e., \(\Phi^\text{ECT}\{\cdot\}\) and \(\Phi^\text{UT}\{\cdot\}\)) and both of these data sets are fused by concatenating ECT and UT data. The obtained fused ECT-UT data are represented by \(\overline{\chi}^\text{ECT-UT} = \{\chi^\text{ECT}; \chi^\text{UT}\}\) of \(F^\text{ECT-UT} = 3K\) ECT-UT features.

2.4. Adaptive sampling through feature extraction
The main goal of the adaptive sampling (i.e., PLS-OSF) is to apply PLS feature extraction for reducing the dimension of the actual features (e.g., \(F^\text{ECT}\), \(F^\text{UT}\) and \(F^\text{ECT-UT}\)) by projecting into extracted feature space. After-which, adaptive sampling is performed directly in the extracted feature space to build suitable I/O pairs for building optimal training model by using lowest number of training samples during offline phase. Though, PLS-OSF [4] has been directly applied for ECT and UT data separately, a modified version of PLS-OSF is needed for dealing with ECT-UT data. The following steps describe the updated PLS-OSF sampling strategy.

i Initialization- Generate \(N_0\) number of initial samples by uniform GRID (i.e., full factorial grid) sampling approach. A matrix of defect parameters \(p = [p^{(n)}; n = 1, ..., N_0]\) having \((N_0 \times Q)\) dimension is formed, where \(p^{(n)}\) is the \(n\)-th row of \(p\). By using \(\Phi^\text{ECT}\{\cdot\}\) and \(\Phi^\text{UT}\{\cdot\}\) generate ECT and UT data respectively, and fill the \((N_0 \times F^\text{ECT-UT})\) feature matrix \(\chi^\text{ECT-UT} = \{(\chi^\text{ECT-UT}^{(n)}; n = 1, ..., N_0\}\).

ii PLS Feature Extraction- In this step, the \(F^\text{ECT-UT} = 3K\) dimensional ECT-UT data are reduced to \(J\) number of extracted features where, \(J << F^\text{ECT-UT}\). A \((N_0 \times F^\text{ECT-UT})\) matrix \(\chi^\text{ECT-UT}'\) is built by subtracting to each \(f\)-th \((f = 1, ..., F^\text{ECT-UT})\) column of \(\chi^\text{ECT-UT}\) its mean value \(\mu_f\). Similarly, a \((N_0 \times Q)\) matrix \(p'\) is built by subtracting to each
At this stage, an adaptive sampling step adds new sample iteratively until N_{iterative} is maximized \( \text{i.e., } T_\text{opt} \) through Latin Hypercube Sampling (LHS) strategy where \( v = 1, ..., V \). An estimation of the J-dimensional set of extracted features corresponding to each \( v \)-th candidate, \( T_{\text{cand}}^{(v)} \) retrieved by applying a multi-dimensional linear interpolator on \( \tilde{D}_{N_{iterative}} \). Select the optimal \( v = v_{\text{opt}} \) candidate (i.e., \( p_{\text{cand},q}^{(v_{\text{opt}})} \)) from \( V \) such that the minimum distance between the obtained extracted features \( T_{\text{cand}}^{(v_{\text{opt}})} \) and all the available extracted features \( T_{\text{cand}}^{(n)} \) \( n = 1, ..., N_{iterative} \) is maximized (i.e., \( v_{\text{opt}} = \text{arg} (\max_{v=1,...,V} \{ \min_{n=1,...,N} |d_{vn}| \}) \). \( d_{vn} \) is the Euclidean distance between \( T_{\text{cand}}^{(v_{\text{opt}})} \) and \( T_{\text{cand}}^{(n)} \), which can be described by:

\[
\sum_{j=1}^{J} \left( \frac{T_{\text{cand},j}^{(v_{\text{opt}})}}{T_{\text{cand},j}^{(n)}} - 1 \right)^2 . \]

The set of extracted features is obtained by \( T_{\text{cand}}^{(v_{\text{opt}})} = \left( \frac{T_{\text{cand}}^{(v_{\text{opt}})}}{T_{\text{cand}}^{(n)}} \right) \times W \). Update the training set \( \tilde{D}_{N_{iterative}+1} = \tilde{D}_{N_{iterative}} \cup \left( \frac{T_{\text{cand}}^{(v_{\text{opt}})}}{T_{\text{cand}}^{(v_{\text{opt})}} \times W} \right) \) with \( N_{iterative} = N_{iterative} + 1 \).

iv Stop Criterion - The adaptive sampling step adds new sample iteratively until \( N_{iterative} = N \) (N is desired/feasible training size).

At this stage, an \( \varepsilon \)-SVR [12] has been utilized to train separately \( q \)-th set of I/O pairs \( \tilde{D}_{N,q} = \left[ \left( \frac{T_{\text{cand}}^{(n)}}{p_{q}^{(n)}} \right) : n = 1, ..., N \right] \) on the generated training set for each \( q \)-th parameter \( (q = 1, ..., Q) \) of the crack. The \( m \)-th test sample \( \left( \frac{T_{\text{cand}}^{(m)}}{p_{(m)}} \right) \) of \( F_{\text{cand}} \) ECT-UT features associated to a previously-unseen crack parameter configuration \( p_{(m)} \) is projected.
through $\mathbf{W}$ into the $J$-dimensional PLS-extracted features space [i.e., $(F^{\text{ECT-UT}})^{(m)} = \left\{ \mathbf{X}^{\text{ECT-UT}}^{(m)} \right\}' \times \mathbf{W}$. Finally, the $q$-th crack parameter associated to $(F^{\text{ECT-UT}})^{(m)}$ is estimated (i.e., $\tilde{p}_q^{(m)}$) by the corresponding trained model in online phase.

3. Numerical validation

The ECT probe and the UT probe collect their corresponding NDT data from 81 positions along $X$ directions with a step size of 0.5 mm and from 41 positions along $Y$ directions with a step size of 1 mm, respectively through a raster scan. Therefore, ECT (i.e., impedance variation signal) and UT signals (i.e., reflected rays) are collected from $K = 81 \times 41 = 3321$ number of inspected points. Therefore, for a single crack configuration (i.e., sample), $F^{\text{ECT}} = 2K = 6642$, $F^{\text{UT}} = K = 3321$ and $F^{\text{ECT-UT}} = 9963$ actual features are treated. For having a valid comparison, $J = 20$ most significant features are extracted from each of these higher dimensional data sets by PLS feature extraction strategy. Different training sets (corresponding to three different crack parameters) have been created by changing the crack dimensions within the range $l_c \in [3.00, 10.00]$ mm, $\delta_c \in [0, 4.00]$ mm and $\varphi_c \in [0, 90]$ deg. by PLS-OSF sampling approach through utilizing ECT, UT and ECT-UT data separately. The initial and maximum number of samples are chosen for $N_0 = 27$ and $N = 216$, respectively. Figure 2 represents the distribution of the resultant training samples in the parameter space as well as in the extracted feature space (for imaging purpose, the first 2 extracted features are considered) of all the data sets (i.e., ECT, UT and ECT-UT). $M = 1000$ unknown samples for 3 crack parameters have been generated by using LHS design and the corresponding 3 test sets of ECT, UT and ECT-UT actual features are obtained. $J = 20$ features are extracted by projecting the test sets into extracted feature space through the PLS weight matrix (e.g., obtained from the corresponding ECT, UT and ECT-UT methods during training phase) for each test set. To partially consider noise effects, Additive White Gaussian Noise (AWGN) has been imposed for different signal to noise ratio ($SNR$) (e.g., $[10, 20, 30, 40]$ [dB]) for blurring ECT, UT signals separately, obtain the corrupted ECT-UT features and project to the extracted feature space. Normalized mean error

unsatisfied
(NME) described in [4] has been utilized for evaluating the inversion performance.

![Prediction Error Plots](image)

**Figure 3.** NME vs. SNR representation for crack (a) length $l_c$, (b) ligament $\delta_c$, and (c) angular position $\varphi_c$ estimation for $N = 216$, $J = 20$, $M = 1000$ through ECT, UT and ECT-UT data.

![Predicted Length Plots](image)

**Figure 4.** Actual vs. predicted plots for $M=1000$ test configurations at $SNR = 20 \text{ [dB]}$, for $N = 216$, $J = 20$, $M = 1000$ for crack (a)-(c) length $l_c$, (d)-(f) ligament $\delta_c$ and (g)-(i) angular position $\varphi_c$ through ECT, UT and ECT-UT.

ECT signals are mostly corrupted for imposing noise and by combining ECT and UT signals, we can improve the overall inversion performance. Fig. 3 (a) shows that $l_c$ estimation by using ECT suffers on noisy data than UT, while on Noiseless test set by both ECT and UT have shown similar prediction accuracy. On the other hand, crack ligament distance $\delta_c$ estimation is showing lower prediction error for adopting ECT signals than UT signals for both noisy and Noiseless test set (Fig. 3 (b)). Whereas, UT data shows lower NME than ECT data for
angular distance $\phi_c$ estimation (Fig. 3 (c)). By combining both ECT and UT signals, ECT-UT data fusion contains both information from ECT and UT signals. Whereas, applying PLS-OSF sampling, we can retrieve most significant information from ECT-UT. As a consequence, it improves the learning ability of SVR during training model development. Hence, ECT-UT data fusion has shown higher prediction accuracy than ECT and UT data for all the cases on noisy and Noiseless test sets. Fig. 4 shows the scatter plots of true vs. predicted crack parameters obtained for $N = 216$ for noisy test set ($SNR = 20$ [dB]). Qualitatively, ECT-UT data fusion provides better $l_c$, $\delta_c$ and $\phi_c$ estimation than ECT and UT signals. Concerning real time solution, it takes 0.03s for testing 1000 samples during online phase.

4. Conclusions
In this work, we have shown an innovative MP-LBE inversion strategy for crack dimension and position estimation. Within the framework of LBE, PLS-OSF/SVR strategy has been applied for solving a NDE problem by utilizing ECT signals, UT signals and ECT-UT data fusion. By combining two different NDT methods, we first retrieved the variation of actual ECT and UT signals for changing crack parameters. Applying adaptive sampling through PLS feature extraction retrieves most significant information from the actual ECT-UT features that improves the learning ability of SVR. ECT-UT shows better prediction accuracy than ECT and UT methods separately for performing inversion on both noisy and noiseless synthetic test set.

References
[1] Salucci M, Poli L, Anselmi N and Massa A 2017 IEEE Transactions on Geoscience and Remote Sensing 55 1305–1317
[2] Salucci M, Poli L and Massa A 2017 Signal Processing 132 306 – 318 ISSN 0165-1684
[3] Salucci M, Oliveri G and Massa A 2015 IEEE Transactions on Geoscience and Remote Sensing 53 6573–6592 ISSN 0196-2892
[4] Ahmed S, Salucci M, Miorelli R, Anselmi N, Oliveri G, Calmon P, Reboud C and Massa A 2017 Journal of Physics: Conference Series 904
[5] Ahmed S, Miorelli R, Salucci M and Massa A 2017 Studies in Applied Electromagnetics and Mechanics: Electromagnetic Nondestructive Evaluation (XX) 42 228–235
[6] Salucci M, Vrba J, Merunka I and Massa A 2017 Microwave and Optical Technology Letters 59 2796–2799 ISSN 1098-2760
[7] Massa A, Oliveri G, Salucci M, Anselmi N and Rocca P 2018 Journal of Electromagnetic Waves and Applications 32 516–541
[8] Pipis K, Skarlatos A, Theodoulidis T and Lesselier D 2016 IEEE Transactions on Magnetics 52 1–8 ISSN 0018-9464
[9] Liu Z, Forsyth D S, Komorowski J P, Hanasaki K and Kirubarajan T 2007 IEEE Transactions on Instrumentation and Measurement 56 2435–2451
[10] Wold S, Sjostrom M and Eriksson L 2001 Chemometrics and Intelligent Laboratory Systems 58 109–130
[11] Biecz S, Lambert M and Gmytho S 2010 Inverse Problems 26 074012
[12] Smola A J and Schölkopf B 2004 Statistics and Computing 14 199–222 ISSN 0960-3174
[13] Lesselier D and Razek A 2002 Scattering: Scattering and Inverse Scattering in Pure and Applied Science: Part 1 – Scattering of Waves by Macroscopic Targets, (Academic Press) pp 186–507
[14] CEA-LIST CIVA: Simulation and Analysis for NDT http://www-civa.cea.fr/en/ [Online; accessed Mar. 2018]
[15] Hopkins D, Datuin M, Aldrin J, Warchol M, Warchol L and Forsyth D 2017 AIP Conference Proceedings 1806 090007
[16] Zernov V, Fradkin L and Darmon M 2012 UltraSonics 52 830–835
[17] De Jong S 1993 Chemometrics and Intelligent Laboratory Systems 18 251–263

Acknowledgments
This work has been partially supported by the SIRENA project (2014-2017) funded by DIGITEO (France) under the Call for Chairs 2014, the French project ANR-ByPASS, and by the Italian Ministry of Foreign Affairs and International Cooperation, Directorate General for Cultural and Economic Promotion and Innovation within the SNATCH Project (2017-2019).