A deep learning framework for efficient global sensitivity analysis and SHAP values calculations applied to eddy current testing problems

Gerardo Granados, Roberto Miorelli, Filippo Gatti, Didier Clouteau

▶ **To cite this version:**

Gerardo Granados, Roberto Miorelli, Filippo Gatti, Didier Clouteau. A deep learning framework for efficient global sensitivity analysis and SHAP values calculations applied to eddy current testing problems. QNDE, Jul 2023, Austin, TX, United States. 10.1115/QNDE2023-118352 . hal-04250590

HAL Id: hal-04250590
https://hal.science/hal-04250590

Submitted on 23 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A deep learning framework for efficient global sensitivity analysis and SHAP values calculations applied to eddy current testing problems

GE Granados[1], R Miorelli[1], F Gatti[2], D Clouteau[2]

[1] *Université Paris-Saclay, CEA, List, F-91120 Palaiseau, France*
gerardo.granados@cea.fr, roberto.miorelli@cea.fr
[2] *Laboratoire de Mécanique Paris-Saclay, CentraleSupélec, ENS Paris-Saclay, CNRS, Gif-sur-Yvette, France*
filippo.gatti@centralesupelec.fr, didier.clouteau@centralesupelec.fr

14 november 2023

Abstract

In the context of the nondestructive testing and evaluation research community, global sensitivity analysis (GSA) methods are widespread tools for quantifying the sensitivity of measurements concerning the variation of inputs over the whole design space. For parameter ranking purposes, GSA methods have been historically preferred by NDT&E scholars compared to feature importance (FI) techniques such as random forest-based, SHAP values, among others. For practical applications, GSA and FI can face limitations when the number of evaluations of the physical model is very high. The main issues that one needs to face with GSA and FI in practical problems are the low computing time efficiency of the numerical solver and/or the high cardinality (i.e., the number of inputs) of the problem considered. This paper targets two main goals. First, we propose to tackle the problem of an efficient GSA and FI procedure relying on a tailored deep neural network to be employed as a metamodel (or surrogate model) to replace the less efficient numerical model. Second, we compare GSA (i.e., Sobol' indices and δ-sensitivity measure) indices and FI (i.e., SHAP) method for parameter ranking purposes. In particular, we describe a generative deep neural network framework to be straightforwardly applied to GSA and FI studies. The numerical experiments in this communication correspond to an eddy current testing inspection problem where multiple arbitrarily oriented cracks lie in a conductive planar multilayered structure.

Keywords: Global sensitivity analysis, SHAP, deep neural network, deep generative models, Sobol' indices, delta-importance measure, eddy current testing, metamodels

1 Introduction

In nondestructive testing & evaluation (NDT&E) framework, the vast improvement in numerical simulation tools in terms of efficiency is mainly due to the increases of computational power on standard PCs and through the use of distributed and cloud computing resources. Nevertheless, this progress has just partially mitigated the computational efficiency issues that one faces in performing very demanding statistical studies such as global sensitivity analysis (GSA) [1, 2], model-assisted probability of detection (MAPOD) [3, 4, 5], stochastic optimization [6], inversion[7, 8], etc. In order to decrease the computational burden without degrading the quality of the results, metamodels (also known as surrogate models) are employed to replace the "true" physics-based model for a given set of parameters (or factors). Loosely speaking, a metamodel can be defined as a mathematical function mapping a parameter space versus the measures space, where measurements can be scalar or vector-valued quantities. In the context of machine learning and statistical methods, the most employed metamodels rely on shallow learning methods (i.e., kernel machines) [7], statistical methods (i.e., Gaussian process, polynomial chaos expansion, etc.) [9, 10], ensemble methods (i.e., random forest, extreme gradient boosting, etc.) or deep neural network (DNN) architectures [11][12]. All these methods exploit a supervised learning framework where the inputs correspond to a labelled target (or, equivalently, the measurement output).

That is, the use of a pre-computed dataset (or database), which contains a collection of input parameters and output signals, is generated during an off-line phase based on physics-based simulations, and

a specific ML model is fit on the data available. In the second phase, referred to as the online phase, a metamodel is used to generate output signals on a set of unseen input parameters. The metamodel acts as a black-box quasi-real-time replacement of the complete forward solver. In this way, it can be plugged in a transparent way within any kind of algorithm involving the use of a physical model for speeding-up considerably the computational efficiency.

In the NDT&E community, the use of metamodels applied to GSA [2][13] and MAPOD [3] has been extensively investigated in the past based on different configurations. Nevertheless, using metamodels based on DNN remains marginal compared to kernel machines or statistical methods. One of the main reasons for considering DNN architectures over the aforementioned approaches is the higher scaling efficiency in large and complex datasets. Therefore, DNN-based methods are among the most suitable regressor to be applied when the cardinality of the inputs is very large.

This paper focuses on an eddy current testing (ECT) inspection scenario driven by twelve parameters used to describe probe, specimen and defects' positional or geometric characteristics. Simulations have been performed into the CIVA-DS application (i.e., the database generation and ML-focused package of the CIVA commercial platform) [14], efficiently addressing the generation of datasets for a wide range of NDT&E problems regardless of the method considered. A DNN is trained over this dataset to later be used to efficiently compute the GSA and the FI.

This communication is structured as follows. In Section 2, we present the supervised framework adopted in this paper with a particular emphasis on the DNN architecture developed. In Section 3, we introduce the GSA methods employed in this work along with the Shapely additive explanation (SHAP) [15] used to analyze the impact/importance of the parameters on the model outputs, named the FI method in this work. Subsequently, the problem studied in this paper is described, the DNN performance is assessed in Section 4, and the GSA and FI results are discussed in Section 5. The paper ends with the conclusion and future perspectives.

2 Supervised DNN regression schema applied to EDT signals

In the last decade, fast regression models (i.e., a metamodel or surrogate model) based on pre-computed databases made of homogeneous collections of NDT inspection signals (e.g., A-scan, B-scan, C-scan etc. signals) and/or engineered extracted features from inspection signals (e.g., peak values, time of flight, etc.) [16] and the associated inspection parameters have been increasingly employed to speed-up the computational time of practical studies for assessing the performance of the inspection procedure such as the MAPOD framework. More recently, the use of surrogate models has been applied to enable the almost real-time application of global sensitivity analysis studied based on statistical distributions. Nevertheless, the use of metamodels based on DNN architectures is not deeply studied in the context of NDT&E based on ECT signals if applied to the efficient calculation of GSA indexes or feature importance ranking based on SHAP values.

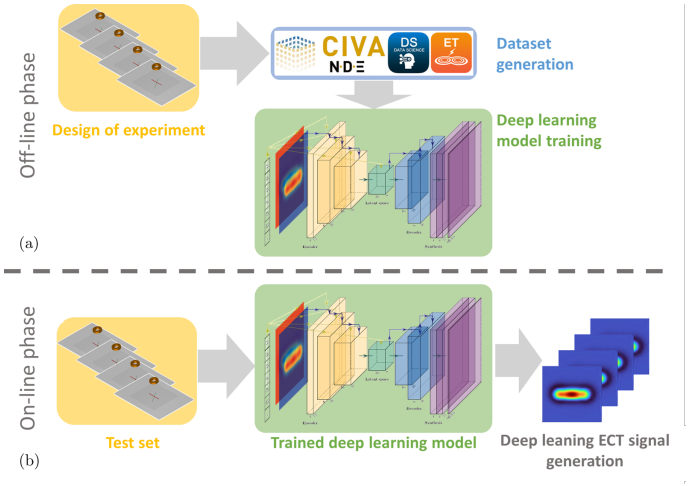


Figure 1: Sketch of the supervised deep neural network schema employed divided as (a) offline and (b) online phases.

$$\mathcal{M}\left\{\mathbf{x}_n, \mathbf{Y}_n\right\} = (\mathcal{S}_3 \circ \mathcal{S}_2 \circ \mathcal{S}_1 \circ \mathcal{D}_2 \circ \mathcal{D}_1 \circ \mathcal{LS} \circ \mathcal{E}_3 \circ \mathcal{E}_2 \circ \mathcal{E}_1)(\mathbf{x}_n, \mathbf{Y}_n) \tag{1}$$

$$\begin{gathered}
\mathcal{E}_i\left\{\mathbf{x}_n, \mathbf{Y}_{\mathbf{f_{i-1}}n}\right\} = \mathcal{L}_{pool} \circ \sigma(\mathcal{L}_{\mathrm{p}}(\mathbf{x}_n)) \circ \sigma(\mathcal{L}_{\mathrm{p}}(\mathbf{x}_n, \mathbf{Y}_{\mathbf{f_{i-1}}n})) + \mathcal{L}_{pool} \circ \mathcal{L}_{conv}(\mathbf{Y}_{\mathbf{f_{i-1}}n}) \\
\mathcal{L}_{\mathrm{p}}\left\{\mathbf{x}_n, \mathbf{Y}_{\mathbf{f_{i-1}}n}\right\} = \mathcal{L}_{pST}(\mathbf{x}_n) \circ \mathcal{L}_{FiLM}(\mathbf{x}_n) \circ \mathcal{L}_{conv}(\mathbf{Y}_{\mathbf{f_{i-1}}n}) \\
\mathcal{LS}\left\{\mathbf{Y}_{\mathbf{f_{i-1}}n}\right\} = \sigma(\mathcal{L}_{\mathrm{p}}(\mathbf{x}_n)) \circ \sigma(\mathcal{L}_{\mathrm{p}}(\mathbf{x}_n, \mathbf{Y}_{\mathbf{f_{i-1}}n})) + \mathcal{L}_{conv}(\mathbf{Y}_{\mathbf{f_{i-1}}n}) \\
\mathcal{D}_i\left\{\mathbf{Y}_{\mathbf{f_{i-1}}n}\right\} = \mathcal{L}_{up} \circ \sigma(\mathcal{L}_{IN} \circ \mathcal{L}_{conv})(\mathbf{Y}_{\mathbf{f_{i-1}}n}) + \mathcal{L}_{up} \circ \mathcal{L}_{conv}(\mathbf{Y}_{\mathbf{f_{i-1}}n}) \\
\mathcal{S}_i\left\{\mathbf{Y}_{\mathbf{f_{i-1}}n}\right\} = \sigma_{TanH}(\mathcal{L}_{IN} \circ \mathcal{L}_{conv})(\mathbf{Y}_{\mathbf{f_i}n})
\end{gathered} \tag{2}$$

2.1 Database and DNN metamodel generation

We define a database $\mathbb{D}$ containing a set of N Input/Output (I/O) couples (or samples) as

$$\mathbb{D} = \left[(\mathbf{x}_1, \mathbf{Y}_1), (\mathbf{x}_2, \mathbf{Y}_2), ..., (\mathbf{x}_N, \mathbf{Y}_N)\right],$$

where the vector associated with the i-th sample writes as $\mathbf{x}_i = [x_1, x_2, ..., x_D]$ with D input parameters size such that $\mathbf{x}_i \in \mathbb{R}^D$ and the corresponding target vector made by M element is

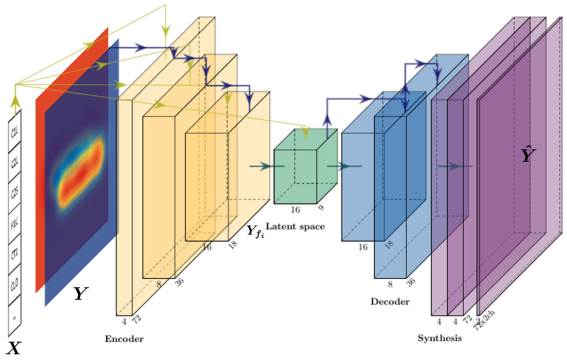
$$\mathbf{Y}_i = [y_{11}, y_{12}, ..., y_{1M_y}; y_{21}, y_{22}, ..., y_{2M_y}; ...; y_{N1}, y_{N2}, ..., y_{M_xM_y}].$$

These target or output values are obtained by applying a deterministic forward operator $\mathcal{F}$ on the set of input parameters $\mathbf{x}_i$, i.e., $\mathbf{Y}_i = \mathcal{F}\{\mathbf{x}_i\}$. More generally, we can define $\mathcal{F} : \mathbb{R}^{1\times \mathrm{D}} \mapsto \mathbb{C}^{\mathrm{M_x}\times \mathrm{M_y}}$ and therefore $\mathbf{Y}_i \in \mathbb{C}^{\mathrm{M_x}\times \mathrm{M_y}}$ where $\mathcal{F}\{\mathbf{x}_i\}$ is obtained via a CIVA solver simulation based on the integral approach [17]. Different database-building strategies have been developed in the literature; some of them rely on fixed sampling schema of the parameter space, while others aim to increase the parsimony in terms of the number of calls to the forward solver without degradation of the accuracy in metamodel results. In this work, we employ a database sampling-based one-shot strategy where the parameter space has been sampled based on Latin hyper-cube sampling.

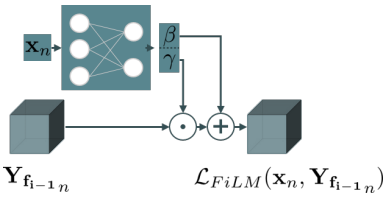
The metamodel ($\mathcal{M}$) employed in this work is based on a DNN architecture [11] tailored for complex-valued ECT signals based on a C-scan inspection procedure. Based on the tight analogy of ECT signals with images (indeed, up to some extent, ECT measurements can be seen as hyperspectral images), we developed a specific encoder-decoder [or autoencoder (AE)] architecture based on $2D$-convolutional layers alternated to PReLU activation function [18] and average pooling layers. Furthermore, to enable the regression capability based on the variation of input parameters only (i.e., see (Eq. 1)), we accounted for the feature-wise linear modulation (FiLM) [19] and a modified spatial transformer (ST) layers [20], name here as parametric ST (pST).

The AE's architectures in the bibliography may vary regarding the objective of the DNN and the type of data used to train it. For AE-type neural networks, the layers before the middle latent space (LS) are tailored to the expected task during the on-line phase (Fig. 1 a)); e.g., variational AEs rely on fully connected layers to infer distribution parameters, long short-term memory encoders are used to embedded the useful information to infer futures states of an input. In our case, the feature extraction procedure performed by the architecture based on $\mathbf{Y}_i$ conditioned by $\mathbf{x}_i$ at every layer of the encoder. This leads to an LS structured by the parameters used during the data generation. The expected behaviour during the on-line phase is a coherent ECT generation piloted by $\mathbf{x}_i$. It is worth to be noticed that a structured LS based on the knowledge of input parameters enhance a better insight into the inner working mechanisms of the NN architecture as well as assures that the underlying physics is injected and preserved into the architecture in the different layers. A sketch of the architecture developed in this work is represented in Fig. 2. The encoder, bottleneck and decoder blocks are highlighted in yellow ($\mathcal{E}_i$), green ($\mathcal{LS}$) and blue ($\mathcal{D}_i$), respectively. Additional synthesis layers ($\mathcal{S}_i$), displayed in purple, are added to improve reconstruction accuracy. The encoder and decoders count with Res-Net-like skip-connections to help the convergence for the training. Spatial dropout layers are added at every block to promote the independence of per-channel features.

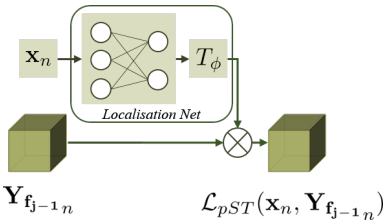
We denote with the subscript n a sample that is forwarded in metamodel $\mathcal{M}$ (Eq. 1). $\mathbf{Y_f}$ is the features extracted by the layers in $\mathcal{M}$. i is the number of the layers in $\mathcal{M}$ whose input is the $\mathbf{Y_f}$ from the precedent layer $i-1$, together with the $\mathbf{x}_n$ corresponding to the sample $\mathbf{Y}_\mathrm{n}$. The inner layers operators at each layer (Eq. 2) are identified as $\mathcal{L}_\mathrm{p}$the parametric layer triplet (composed by the parametric spatial transformer$\mathcal{L}_{pST}$, the feature-wise linear modulator $\mathcal{L}_{FiLM}$ and a 2D convolution $\mathcal{L}_{conv}$. σ denotes the PReLU activation, while σ_{TanH} is a tangent hyperbolic activation. $\mathcal{L}_{up}$and $\mathcal{L}_{pool}$ represent the up-sampling and average pooling operators. The encoder normalizes its convolution output by an instance normalization $\mathcal{L}_{IN}$ [21].



a) $\mathcal{M}$block disposition



b)$\mathcal{L}_{FiLM}$ operator



c)$\mathcal{L}_{pST}$ operator

Figure 2: Overview of the DNN architecture used is given in a). In b) and c), the schematic FiLM and ST blocks are provided, respectively.

A Mean Square Error (MSE) loss is used to train the architecture. The neural networks optimized the reconstruction of $\mathbf{Y}_n$ (2-channel input) while learning how the FiLM and ST transformations during the encoding conditioned by $\mathbf{x}_n$. Those two layers are schematically represented in 2*b*) and 2c, where $\mathbf{Y}_{\mathbf{f_{j-1}}n}$are the extracted feature by any precedent layer ($\mathcal{L}$). For the pST, a dense layer infers $\boldsymbol{\phi} \in \mathbb{R}^{6\times C_k}$ from $\mathbf{x_n}$ to built $T_{\boldsymbol{\phi}} \in \mathbb{R}^{3\times 3\times C_k}$, C_k are the number of channel (or filters) in the block. $\boldsymbol{\phi}$ and $T_{\boldsymbol{\phi}}$ for each C_k are built as

$$\begin{gathered}\boldsymbol{\phi}_{C_k}(\mathbf{x}_n) = [\phi_{1_{C_k}}(\mathbf{x}_n),\ \phi_{2_{C_k}}(\mathbf{x_n}),\ \phi_{3_{C_k}}(\mathbf{x}_n),\\ \phi_{4_{C_k}}(\mathbf{x}_n),\ \phi_{5_{C_k}}(\mathbf{x}_n),\ \phi_{6_{C_k}}(\mathbf{x}_n)],\\ T_{\boldsymbol{\phi}_{C_k}}(\boldsymbol{\phi}_{C_k}) = \begin{bmatrix} \phi_{1C_k} & \phi_{2C_k} & \phi_{3C_k} \\ \phi_{4C_k} & \phi_{5C_k} & \phi_{6C_k} \\ 1 & 1 & 1 \end{bmatrix},\end{gathered} \tag{3}$$

to be applied as an affine transformation per C_k-map as follows,

$$\mathcal{L}_{pST}(\mathbf{x}_n, \mathbf{Y}_{\mathbf{f_{j-1}}n}) = T_{\boldsymbol{\phi}}(\boldsymbol{\phi}(\mathbf{x}_n)) \otimes \mathbf{Y}_{\mathbf{f_{j-1}}n}\ . \tag{4}$$

Analogously to pST, the FiLM uses a dense layer to infer $\beta \in \mathbb{R}^{C_k}$ and $\gamma \in \mathbb{R}^{C_k}$ from $\mathbf{x}_n$ to apply, then

$$\mathcal{L}_{FiLM}(\mathbf{x}_n, \mathbf{Y}_{\mathbf{f_{i-1}}n}) = \gamma(\mathbf{x}_n) \cdot \frac{\mathbf{Y}_{\mathbf{f_{j-1}}} - \mathbb{E}[\mathbf{Y}_{\mathbf{f_{j-1}}n}]}{\sqrt{\sigma^2[\mathbf{Y}_{\mathbf{f_{j-1}}n}] + \epsilon}} + \beta(\mathbf{x}_n), \tag{5}$$

where $\mathbb{E}[\cdot]$ and $\sigma^2[\cdot]$ represent the empirical average and variance, respectively.

In order to use the AE as a parametric regressor, we fix the input $\mathbf{Y}_{\text{n}}$once $\mathcal{M}$ is trained. This is, during the online phase, the hyper-spectral image input becomes $\mathbf{Y}_{\text{fix}}$ for any $\mathbf{x}_{\text{n}}$, and $\mathbf{Y}_{\text{fix}}$ is the mean of the training set samples for $\mathbf{Y}$.

Behind the choice of this architecture is the underlying assumption that a set of spatial transformations (Eq. 3) exists to generate the ECT images from a fixed input, and these transformations depend on the simulation parameters. Similarly, the distribution of each ECT image depends on the input parameters; this is learned by Eq. 5. Both transformations are a function of the simulation parameters. For the first assumption, the coin tilt is an example of a spatial transformation required to produce different coherent outputs from a fixed image input but a changing parameter input ($\mathbf{Y}_{\text{fix}}, \mathbf{x}_{\mathbf{var}}$). For instance, the main feature in Fig. 4-I.b-right, in contrast to Fig. 4-I.b-left, presents a spatial scale learned by the pST, among others. Similarly, the two images present different distributions learned by the FiLM.

2.2 Metamodel validation metrics

To quantitatively evaluate the metamodel prediction accuracy of multivariate ECT signals, we shall employ three different metrics. The first one is the normalized root mean squared error (NRMSE), defined as

$$NRMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\frac{\left(\mathcal{F}\left\{\mathbf{x}_n\right\} - \mathcal{M}\left\{\mathbf{x}_n\right\}\right)^2}{\mathbf{Y}_{max} - \mathbf{Y}_{min}}}, \tag{6}$$

where $\mathbf{Y}_n$ stands for the n-th true value to be estimated and $\mathcal{M}\left\{\mathbf{x}_n\right\}$ is the associated estimation. $\mathbf{Y}_{max}$ and $\mathbf{Y}_{min}$ correspond to the maximum and minimum ECT signals values among Nvalues considered, respectively. To estimate the fit of the predicted ECT signals against the ground truth (GT), we employ the correlation coefficient (R^2) defined as

$$R^2 = 1 - \sum_{n=1}^{N}\frac{\left(\mathcal{F}\left\{\mathbf{x}_n\right\} - \mathcal{M}\left\{\mathbf{x}_n\right\}\right)^2}{\left(\mathcal{F}\left\{\mathbf{x}_n\right\} - \bar{\mathbf{Y}}\right)^2}, \tag{7}$$

where $\bar{\mathbf{Y}}$ is the mean value of the ECT signals among the N samples considered. Lastly, the mean normalized Frobenius norm ($\|\cdot\|_F^2$) error (MNFE) has been calculated as

$$MNFE = \frac{1}{N}\sum_{n=1}^{N}\frac{\|\mathcal{F}\left\{\mathbf{x}_n\right\} - \mathcal{M}\left\{\mathbf{x}_n\right\}\|_F^2}{\|\mathcal{F}\left\{\mathbf{x}_n\right\}\|_F^2}. \tag{8}$$

3 DNN metamodel-based SA and FI studies applied to ECT signals

In real-case scenarios, the agreement between experimental data and simulated results depends on the capability to master the whole acquisition chain (i.e., probe position, dimensions, knowledge of specimen characteristics, measurement noise, etc.). That is, in operative conditions, the measurement signals are impacted by "hidden" factors that cannot be directly measured. In order to have a better insight into the impacts of these factors on measurement, one can consider them along with the driving factors that are supposed to be "known", i.e., flaw size, position, etc., in a sensitivity analysis or feature importance ranking framework.

The main goal of global sensitivity analysis [1] consists in identifying and ranking the set of parameters that impacts the model output variability across the entire dataset variability. In a nutshell, the underlying concept is that the bigger the sensitivity of a parameter is, the higher its influence on the output is. In this section, we introduce the GSA method variance decomposition based on Sobol' indices [22] and the moment-independent GSA method based on δ-sensitivity measure (or indexes), which is commonly employed for ranking and screening purposes [1]. In the ML research community, the ranking of the most important feature is performed through feature importance methods. In this paper, we focused on the use of SHAP values[15] as a feature importance (FI) method to be applied to the analysis of most important factors driving the ECT signals measurements.

Regardless of the use of GSA or FI methods for ranking purposes, the computational burden associated with their calculation makes the use of metamodels mandatory for performing the sensitivity analysis in an acceptable amount of time. Referring to GSA, one can show that the first-order Sobol' and the total order indices of the i-th factor are given by [22, 23] $S_i = \frac{V_i}{\mathbb{V}ar(\mathcal{M}\{\mathbf{X}\})}$ and $S_{T_i} = \frac{\mathbb{V}ar_{\mathbf{X}_{\sim i}}(\mathbb{E}(\mathcal{M}\{\mathbf{X}\}|\mathbf{X}_{\sim i}))}{\mathbb{V}ar(\mathcal{M}\{\mathbf{X}\})}$,

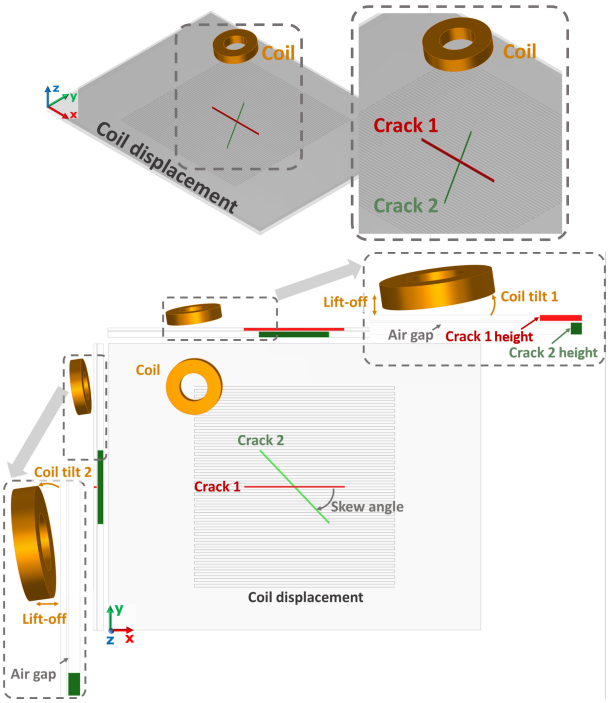


Figure 3: ECT inspection problem studied. On the left is the $3D$ view, and on the right, the associated three main orthogonal projections.

receptively. S_i and S_{T_i} represent the main effects of the i-th variable alone and the effects of the i-th variable and its interactions with the other variables, respectively. Where $\mathbb{V}ar(\cdot)$ represent the variance, $\mathbb{E}(\cdot)$ the expectation with $V_i = \mathbb{V}ar_{x_{\sim i}}\left(\mathbb{E}_{\mathbf{X}_{\sim i}}\left(\mathcal{M}\{\mathbf{X}\}|x_i\right)\right)$ with the subscript "$\sim$" identifies the left out index factor in the calculations [23]. It is worth to be mentioned that Sobol' indices capture the overall behaviour of uncertainties when the variance of the outputs represents sufficiently (i.e., it is a good proxy for its estimation). δ-sensitivity measure is used to compute GSA indices based on the variation of conditional and unconditional probability density functions. This sensitivity analysis method is particularly suitable in the presence of a correlation between parameter and when the distribution of the outputs are highly skewed or multi-modal. δ-importance (or δ-sensitivity) measure for the i-th factor is defined as [24] $\delta_i = \frac{1}{2}\mathbb{E}_{X_i}\left[s\left(X_i\right)\right]$ with $\mathbb{E}_{X_i}\left[s\left(X_i\right)\right] = \int f_{X_i}\left[\int \left|f_Y\left(\mathcal{M}\{\mathbf{X}\}\right) - f_{Y|X_i}\left(\mathcal{M}\{\mathbf{X}\}|x_i\right)\right| dy dx_i\right]$ with, called inner statistic or inner separation, represents is the area enclosed between the conditional $(f_{Y|X_i})$ and unconditional (f_Y) model output densities obtained for a particular value of X_i. This means that in the case of $f_{Y|X_i}$ is equal to f_Y, removing the uncertainty on x_i does not affect the distribution of the output; thus, the i-th input does not impact the output of $\mathcal{M}\{\mathbf{X}\}$.

Feature importance methods can be used in order to rank the most impacting inputs onto the model outputs. Among tha wide set of FI methods developed by the ML scientific community, in this work, we adopt SHAP [15], which is a method to explain predictions based on the use of Shapley values in the case of coalition game theory. Furthermore, SHAP values enable access to both global estimations of feature importance as well as to a qualitative assessment of the impact of inputs onto the model output. Loosely speaking, Shapley values express how the predictions are homogeneously distributed among the features (i.e., the parameters) and in the framework of SHAP, Shapley values are calculated as an additive feature attribution method as $g(z') = \phi_0 + \sum_{j=1}^{M} \phi_j z'_j$with $g(\cdot)$ being the explanation mode, z' z' is the so-called coalition vector (i.e., the features/inputs considered), ϕ_j the Shapley value for the j-th feature and $\phi_0 = \mathbb{E}\left[f(z)\right]$ with $f(\cdot) = \mathcal{M}\{\mathbf{X}\}|z$.

4 Numerical assessment

In this section, we analyze the results obtained for the proposed DNN-based metamodel schema once applied to ECT signals based on an inspection problem parameterized by coil, specimen and crack(s) parameters. The synthesis of ECT signals has been applied to the fast calculation of GSA indices and FI calculations.

4.1 Generative DNN-based metamodel performance

Referring to the inspection case shown in Fig. 3, a suitable database has been built accounting for twelve parameters involving the cracks, probe and specimen parametrization. The forward solver simulations, based on the integral equation method [17], have been performed by CIVA software [14]. More into detail, crack 1 length (l_{C1}) was made varying [25.0; 35.0] mm, and its width (w_{C1}) takes values in the interval [0.05; 0.5] mm. Crack 2 length (l_{C2}), width (w_{C2}) and height (h_{C2}) were varied in the ranges of [25.0; 35.0] mm,[0.05; 0.5] mm and [0.25; 2.0] mm, respectively. The skew angle between the two cracks (ϕ_{12}) takes values between [0.0; 110.0] deg. The coil lift-off (lo), tilt in the xz-plane (θ_{xz}), tilt in the yz-plane (θ_{yz}) assumed values in the range of [0.05; 0.8] mm, [0.0; 8.0] deg and [0.0; 8.0] deg, respectively. Concerning the specimen, both plates' conductivity ($\sigma_1$and σ_2) varied between [16.0; 22.0] MS/m along with the air gap thickness between the two plates (t) ranging as [0.03; 0.3] mm.

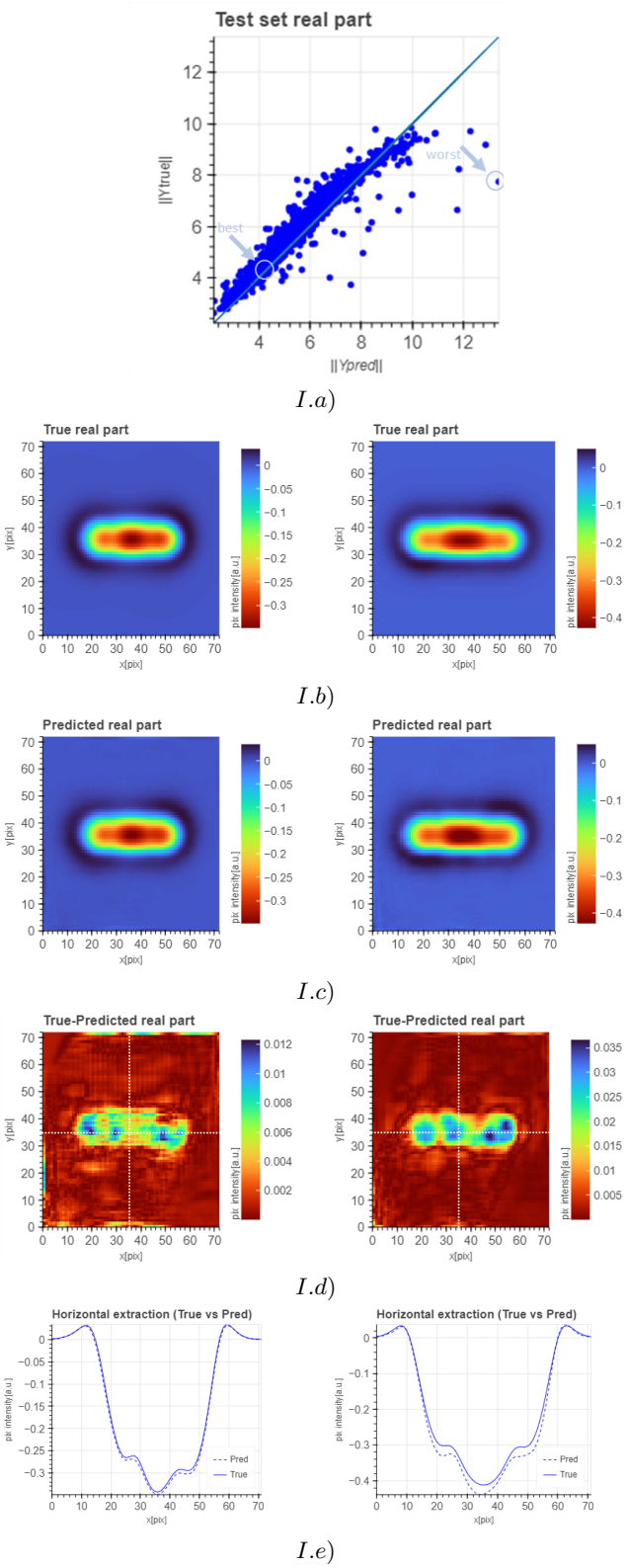
For simulating the inspection procedure, a C-scan (i.e., a map) made by 72×72 points centered on the cracks zone has been considered for simulations. Therefore each sample within the database contains 5040 complex-valued measurement points corresponding to the coil impedance variation signal (ΔZ). The database has been sampled based on a Latin hyper-cube sampling schema made of 5 k items and then split into training, validation and test sets by choosing 1.25k, 625 and 2.5k samples, respectively. The generative DNN architecture introduced in Section 2 has been trained with a learning rate equal to $1e^{-3}$ based on ADAM optimizer [25] with a batch size equal to 128. The training procedure ended after 5 k epochs in about 4 h on a GPU cluster equipped with one NVIDIA HGX A100 graphic card.

The DNN metamodel performance has been assessed on the whole set of 2.5k test samples accordingly to the metrics provided in equations 6, 7 and 8, obtaining error values as 0.0083, 0.887 and $3.7e-9$, respectively. Qualitative comparisons of results based on the test set samples are given for the real and imaginary parts of the signal in Fig. 4 and 5. A good agreement has been observed between DNN prediction and ground truth (GT) all over the whole set of test. Therefore, one can conclude that the errors introduced by metamodel predictions of ECT signal are negligible for its exploitation for ranking purposes by GSA indices and FI calculation. Furthermore, from the point of view of computational performance, the test set predictions based on the whole test set take about 1.25 s on a PC equipped with an Intel Xeon CPU @3.70 GHz and a QUADRO RTX GPU 6000, which is a non-negligible speed-up in computational time efficiency if it is compared to the about 40 seg needed to compute one forward solver calculation.

5 Sobol' indices, δ-importance measure and SHAP results analysis

Based on the DNN metamodel validated in Subsection 4.1, we applied it to the calculation of GSA indices and FI calculation for parameter ranking purposes, as introduced in Section 3. We compute 25k samples for the GSA calculation routine of the SALib Python-based sensitivity analysis library [26], where the twelve driving parameters have been varied by Latin hyper-cube sampling. That is, both Sobol' and δ-importance indexes were calculated efficiently in about 370 seconds, and the results are shown in Figure 6 and Fig. 7, respectively. In Fig. 8, the FI ranking relies on SHAP values analysis, where the calculations were based on *KernelExplainer* within the Python-based library called SHAP [15].

From the cross-comparisons between the two GSA indices, one can notice that there is an overall high correlation between Sobol' and δ-importance indices regardless of the ECT components considered (i.e., absolute, real or imaginary). On the other hand, the comparisons between GSA and SHAP show overall good agreement in the order of indices with complete concordance in the real part. The biggest discordance between the analysis is in the index of the for l_{C2} in the absolute value, being the more important feature for SHAP while its Sobol' and δ-importance index is low. Such a particular behaviour needs further investigation since it seems to be associated with the real part of the ECT signal (which is embedded into the absolute value too). Unlike both the GSA methods, through SHAP values analysis, one may have deeper information on how the input parameters impact the variability of ECT signals collected by the coil. More specifically, thanks to the so-called "bee-swarm" plots, we can analyze the distribution of SHAP values for each of the twelve considered parameters as shown in Fig. 9. Looking at Fig. 9, one can notice that the coil position parameters (tilt and lift-off) and both the crack 1 parameters (namely the length and the width) have a symmetric impact. h_{C2} has a more asymmetric impact for high and low values. This behaviour can be justified by considering the underlying physics. Indeed, the penetration of eddy currents into the medium diminishes exponentially accordingly to the skin depth.As a consequence, a very minor impact of crack 2 on the variation of coil impedance is expected for smaller crack heights



I.a)

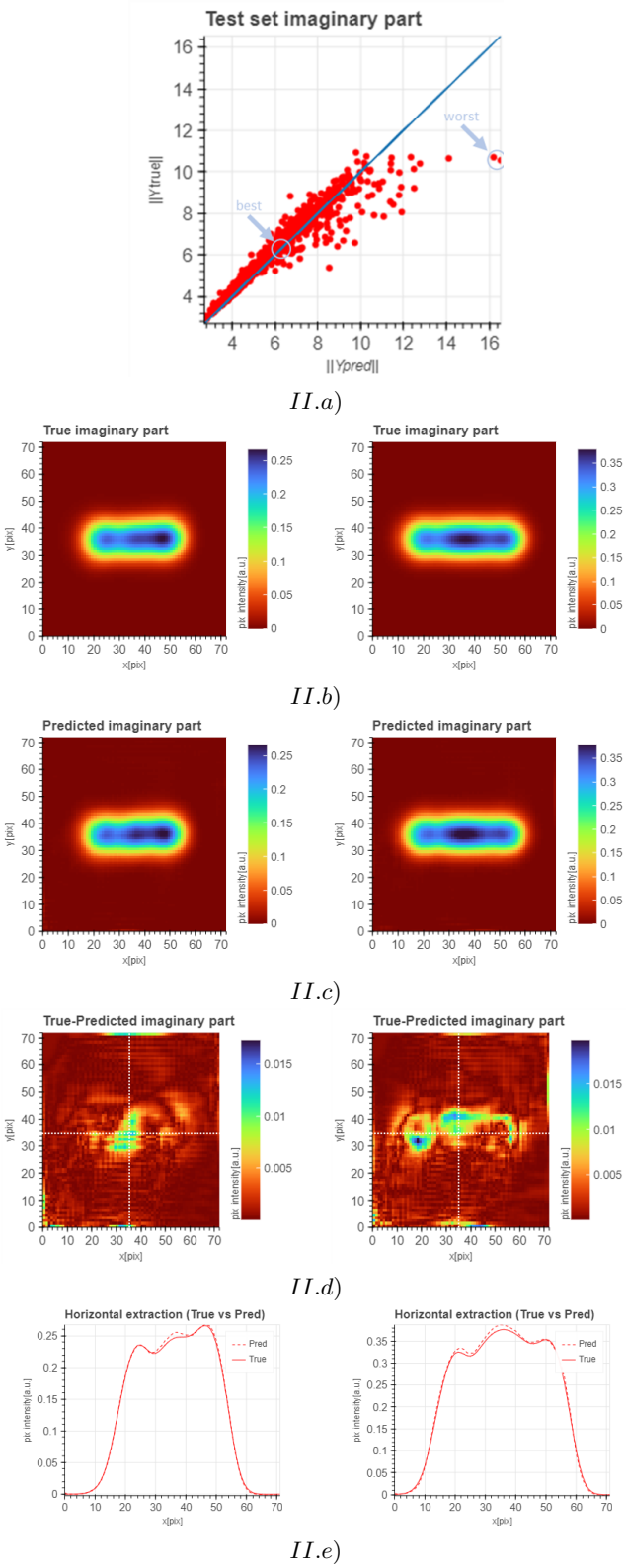
I.b)

I.c)

I.d)

I.e)

Figure 4: Qualitative assessment of the DNN metamodel regressor for the test set prediction of the real part (I) of ECT signals ($\Re\{\Delta Z\}$). In a) the GT versus predicted plot based on the Frobenius norm of the ECT C-scan. In b) and c), the top row represents the best and the bottom one the worst predictions agreement, respectively. In c), the differences between the C-scans of GT and predictions are given for the best and worst scenarios. e) represents the scan extractions along the horizontal cut sketched on d) plots.



II.a)

II.b)

II.c)

II.d)

II.e)

Figure 5: Qualitative assessment of the DNN metamodel regressor for the test set prediction of the imaginary part (II) of ECT signals ($\Im\{\Delta Z\}$). In a) the GT versus predicted plot based on the Frobenius norm of the ECT C-scan. In b) and c), the top row represents the best and the bottom one the worst predictions agreement, respectively. In c), the differences between the C-scans of GT and predictions are given for the best and worst scenarios. e) represents the scan extractions along the horizontal cut sketched on d) plots.

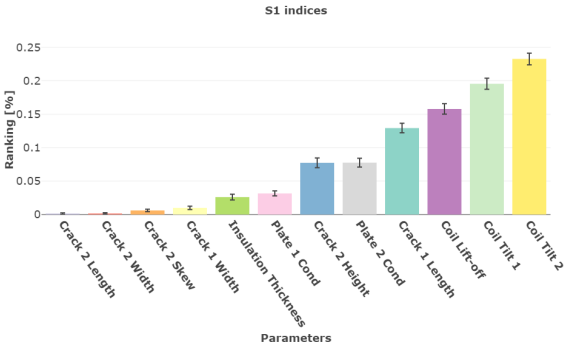
compared to the larger one. Furthermore, the relation between the variation of coil impedance and h_{C2} is linear when crack 2 approaches the second plate's upper surface, and it becomes non-linear when crack 2 breaks the second plate's upper surface. Based on these results, it is believed that the richer information provided by "bee-swarm" -like plots is a valuable tool to deeply analyze both GSA and FI results in many realistic inspection cases (not only linked to ECT inspection signals).

5.1 A deeper insight into the generative DNN metamodel procedure

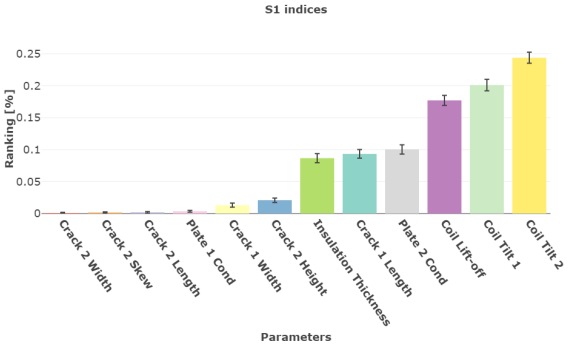
Hereafter we provide some insight into how the DNN learns to generate data. We explored the latent space block's output o have access to the analysis of the operation of the encoder. Toward this end, the t-SNE manifold projection [27] was used to obtain a compact 2D visual representation of the latent space across the entire dataset (i.e., the whole 5k samples).

Fig.10 represents a scatter plot of the LS space. Every point is located on the 2D manifold by its coordinates (LS_1,LS_2) based on the t-SNE with an initialization method relying on the principal component analysis (PCA). As a result, each point can be connected to a parameter value and properly visualized based on the "colour bar dimension". In the analysis of the manifold, one can clearly appreciate that hierarchical arrangements appear as shown in Fig.10 $a) - b)$, whereas a less pronounced order is observed for the lift-off parameter. The observations demonstrated that the entire data set is structured by the DNN and that the generative model developed focused first (i.e., encoded or extracted features) on the spatial information (i.e., the ECT flaw signature on the C-scan) associated with ϕ_{12} and l_{C1} and then to lo. Such behaviour is possible since the input vector can produce both previously unseen samples as well as other samples found in the data collection. Furthermore, the projection demonstrates how the encoder features are altered in a coherent direction during the generation by regression.

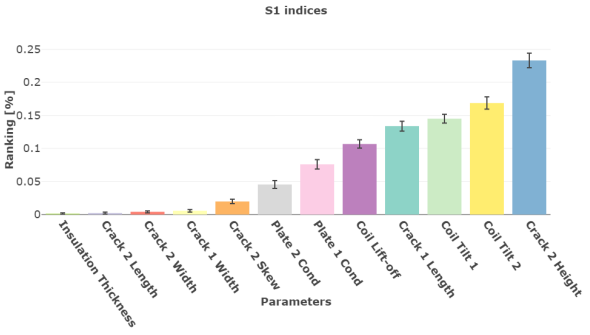
A well-structured latent space does not implicate a good reconstruction, but if the hierarchy found in the projection is coherent with the training dataset structure, it may indicate that the DNN has captured correctly the spatial features of the input images and so the empty space can be filled in a coherent manner. In other words, the DNN who makes a regression over this structured space is expected to generate correct new data. The interest in this architecture is that even if the distribution of the represented latent space is unknown (arbitrary since it is not imposed), the regression is possible in a controlled way since we have access parameters for the generation as a DNN input, a fact that is no always true in other AE architectures.



a)

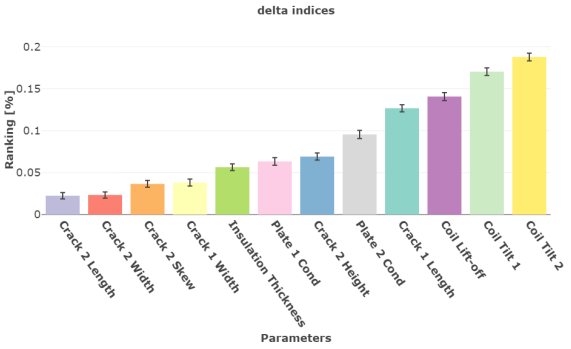


b)

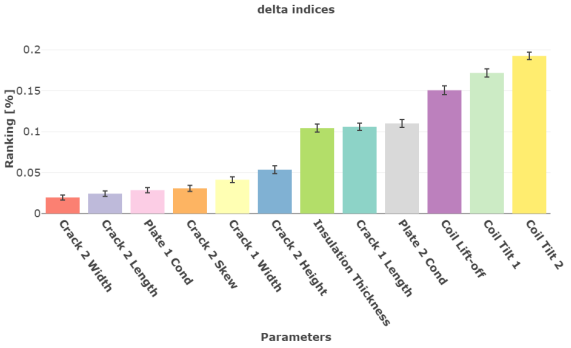


c)

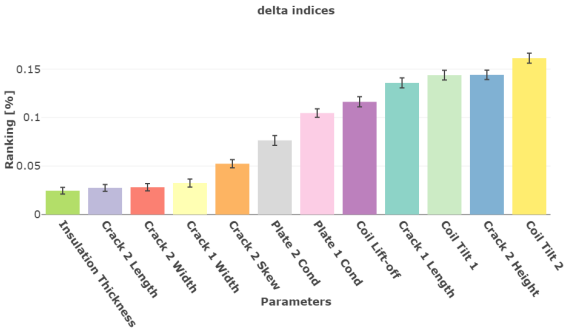
Figure 6: Results of global sensitivity analysis through Sobol' first-order indexes. In (a), the results obtained from the extraction based on the $L2$-norm calculation performed on the absolute value of C-scan ECT signals. In $b)$ and $c)$ the results display the same calculation performed in $a)$, based on real and imaginary part of ΔZ, respectively.



a)

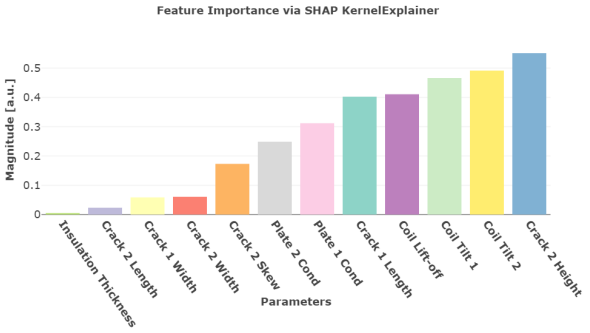


b)

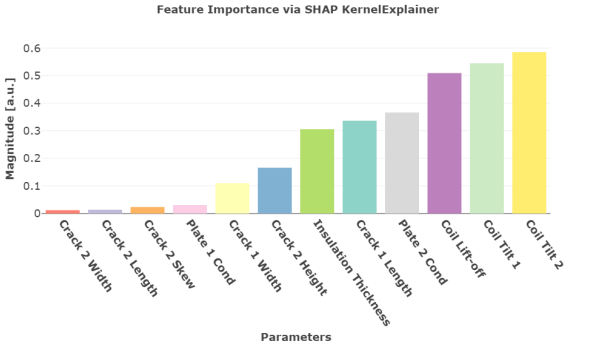


c)

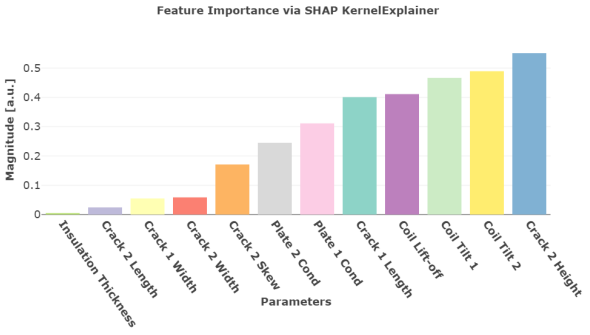
Figure 7: Results of global sensitivity analysis through δ-importance index. In (a), the results obtained from the extraction based on the $L2$-norm calculation performed on the absolute value of C-scan ECT signals. In $b)$ and $c)$ the results display the same calculation performed in $a)$, based on real and imaginary part of ΔZ, respectively.



a)



b)



c)

Figure 8: Results feature importance ranking based on SHAP values. In (a), the results obtained from the extraction based on the $L2$-norm calculation performed on the absolute value of C-scan ECT signals. In $b)$ and $c)$ the results display the same calculation performed in $a)$, based on real and imaginary part of ΔZ, respectively.

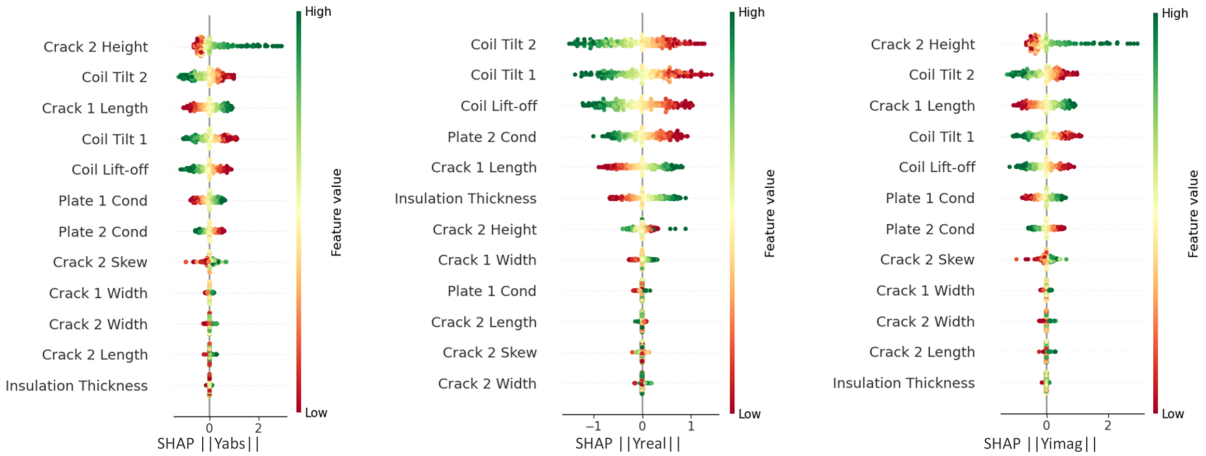


Figure 9: Bees-warm SHAP plots values calculation applied to the DNN metamodel are given (from right to left) for the absolute, real and imaginary parts of ECT signals.

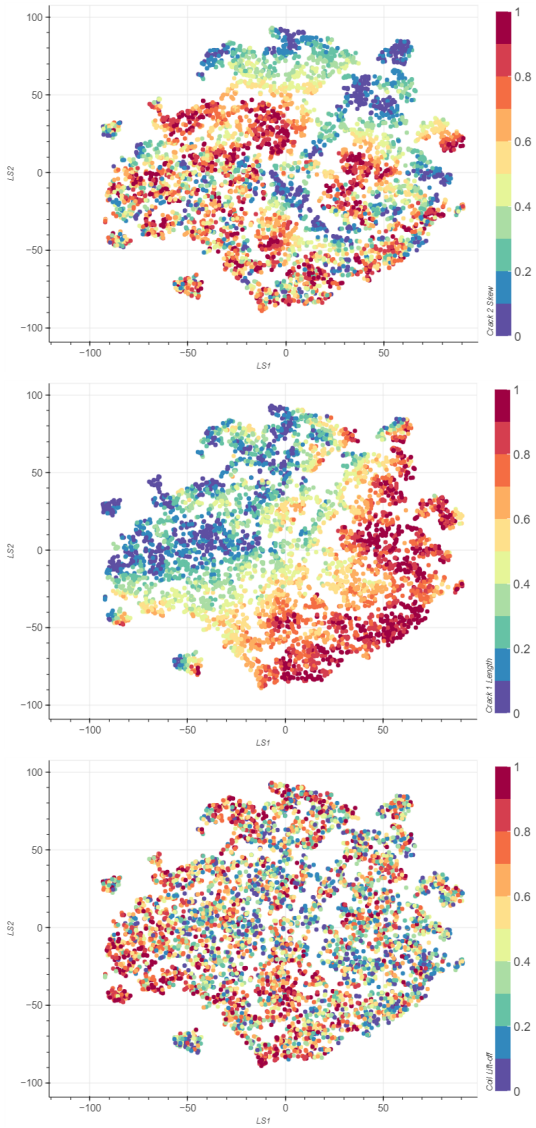


Figure 10: DNN bottleneck layer representation based on $2D$-projection t-SNE projection. From top to bottom, points are colored accordingly to ϕ_{12}, $l_{C1}$and lo, respectively.

6 Conclusion and perspectives

The present work proposes a tailored generative deep learning framework for very efficient simulation in NDT&E for ECT signals analysis. Our work shown how quantifying the sensitivity and establishing a feature importance ranking in parametric simulations that are often limited by the intrinsic computational burden associated with numerical simulations, particularly when high cardinality problems (i.e., the size of input parameters) are considered. The present framework proposes a fast metamodel to overcome this limitation to perform GSA and FI computation.

A DNN regressor applied to an ECT inspection problem, where multiple arbitrarily oriented cracks are lying in a metallic multilayered planar structure, described by twelve parameters like specimen configuration, coil position, and crack geometries, has been studied. The DNN regression model has shown good prediction accuracy and high computational efficiency. Its application to global sensitivity and SHAP analysis has permitted to carry out of the complete study in a negligible amount of time if compared to the intensive use of the forward solver. Additionally, The analysis presented in this paper focuses on a comparative study between Sobol' indexes δ-importance measure and SHAP values. A satisfactory correlation of parameter ranking has been obtained regardless of the method considered. Furthermore, a deeper analysis of SHAP (i.e., bee-swarm plots) has permitted us to retrieve some physics-rooted behaviour on the way that SHAP ranks the parameters. Furthermore, the analysis of inner-working mechanisms of the DNN architecture via the t-SNE manifold projection allowed to join the feature extraction procedure performed by the DNN model with the underlying physics linked to the use of a model-driven dataset generation.

It is worth mentioning that this framework is not limited in terms of the problem cardinality and the problem addressed, as shown in some preliminary studies elsewhere [12]. This is, more parameters can be considered by paying more time to train bigger DNNs with more data with a similar computational evaluation efficiency as a result. In the perspective of this work, we expect to explore the power of our AE backbone architecture for inverse problems, given that structure of the latent space seems promising for this task.

7 Acknowledgments

This project has received funding from the European Union's
Horizon 2020 research and innovation program under grant agreement
No 800945-NUMERICSH2020-MSCA-COFUND-2017.

References

[1] F. Pianosi, K. Beven, J. Freer, J. W. Hall, J. Rougier, D. B Stephenson, and T. Wagener. Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environ. Modell. Software*, 79:214–232, 2016.

[2] J. Nagawkar and L. Leifsson. Efficient global sensitivity analysis of model-based ultrasonic nondestructive testing systems using machine learning and sobol' indices. *J. Nondestr. Eval. Diagn. Progn. Eng. Syst.*, 4(4), 2021.

[3] J. C. Aldrin, E. K. Oneida, E. B. Shell, H. A. Sabbagh, E. Sabbagh, R. K. Murphy, S. Mazdiyasni, E. A. Lindgren, and R. D. Mooers. Model-based probe state estimation and crack inverse methods addressing eddy current probe variability. In *AIP Conf. Proc.*, volume 1806, page 110013, 2017.

[4] J. C Aldrin, E. A Lindgren, and D. S. Forsyth. Intell. augmentation in nondestr. eval. In *AIP Conf. Proc.*, volume 2102, page 020028, 2019.

[5] X. Du, L. Leifsson, W. Meeker, P. Gurrala, J. Song, and R. Roberts. Efficient model-assisted probability of detection and sensitivity analysis for ultrasonic testing simulations using stochastic metamodeling. *J. of Nondestr. Eval., Diagnog. and Progn. of Eng. Syst.*, 2(4), 2019.

[6] C. Cai, R. Miorelli, M. Lambert, T. Rodet, D. Lesselier, and PE Lhuillier. Metamodel-based markov-chain-monte-carlo parameter inversion applied in eddy current flaw characterization. *NDT & E Int.*, 99:13–22, 2018.

[7] S. Bilicz, M. Lambert, and Sz Gyimóthy. Kriging-based generation of optimal databases as forward and inverse surrogate models. *Inverse Probl.*, 26(7):074012, 2010.

[8] S. Ahmed, R. Miorelli, M. Salucci, and A. Massa. Real-time flaw characterization through learning-by-examples techniques: A comparative study applied to ect. In *Electromagn. Nondestr. Eval.*, pages 228–235. IOS Press, 2017.

[9] J. Nagawkar and L. Leifsson. Applications of polynomial chaos-based cokriging to simulation-based analysis and design under uncertainty. In *Int. Des. Eng. Tech. Conf. and Comput. and Inf. in Eng. Conference*, 2020.

[10] X. Du and L. Leifsson. Multifidelity modeling by polynomial chaos-based cokriging to enable efficient model-based reliability analysis of ndt systems. *J. Nondestr. Eval.*, 39(1):13, 2020.

[11] I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. 2016.

[12] G. E. Granados, R. Miorelli, G. Gatti, S. Robert, and D. Clouteau. Towards a multi-fidelity deep learning framework for a fast and realistic generation of ultrasonic images in complex geometries with the multi-modal total focusing method. *NDT & E Int.*, 2023.

[13] A. Bingler and S. Bilicz. Sensitivity analysis using a sparse grid surrogate model in electromagnetic nde. In *22nd International Workshop on Electromagnetic Nondestructive Evaluation (ENDE 2017)*, France, September 6–8 2019.

[14] Extende. Civa platform. `https://www.extende.com/civa-in-a-few-words`, 2023. Accessed: April 2023.

[15] S. Lundberg, S. M .and Lee. A unified approach to interpreting model predictions. *Adv. in neural Inf. Process. Syst.*, 30, 2017.

[16] V. Nerlikar, O. Mesnil, R. Miorelli, and O. D'almeida. Damage detection with ultrasonic guided waves using machine learning and aggregated baselines. *Struct. Health Monit.*, April 2023 [accepted].

[17] R. Miorelli, C. Reboud, T. Theodoulidis, N. Poulakis, and D. Lesselier. Efficient modeling of ECT signals for realistic cracks in layered half-space. *IEEE Trans. Magn.*, 49(6):2886–2892, June 2013.

[18] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *CoRR*, abs/1502.01852, 2015.

[19] E. Perez, F. Strub, H. De Vries, V. Dumoulin, and A. Courville. Film: Visual reasoning with a general conditioning layer. In *AAAI Conf. Proc. on Artif. Intell.*, volume 32, 2018.

[20] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu. Spatial transformer networks. *Adv. in Neural Inf. Proc. Syst.*, 2015-Janua:2017–2025, 2015.

[21] D. Ulyanov, A. Vedaldi, and V. S. Lempitsky. Instance normalization: The missing ingredient for fast stylization. *CoRR*, abs/1607.08022, 2016.

[22] Ilya M. Sobol. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Math. Comput. Simul*, 55(1-3):271–280, 2001.

[23] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global sensitivity analysis: the primer*. John Wiley & Sons, 2008.

[24] E. Borgonovo et al. A new uncertainty importance measure. *Reliab. Eng. Syst. Saf.*, 92(6):771–784, 2007.

[25] D. P. Kingma and Jimmy B. Adam: A method for stochastic optimization, 2017.

[26] T. Iwanaga, W. Usher, and J. Herman. Toward SALib 2.0: Advancing the accessibility and interpretability of global sensitivity analyses. *Socio-Environ. Sys. Modell.*, 4:18155, 2022.

[27] L. Van Der Maaten and G. Hinton. Visualizing Data using t-SNE. *J. Mach. Learn. Res.*, 9:2579–2605, 2008.