



**HAL**  
open science

## Robust Estimation of Hidden Corrosion Parameters using an Eddy Current Technique

Marie-Eve Davoust, Laurent Le Brusquet, Gilles Fleury

► **To cite this version:**

Marie-Eve Davoust, Laurent Le Brusquet, Gilles Fleury. Robust Estimation of Hidden Corrosion Parameters using an Eddy Current Technique. *Journal of Nondestructive Evaluation*, 2010, 29 (3), pp.155-167. hal-00515603

**HAL Id: hal-00515603**

**<https://centralesupelec.hal.science/hal-00515603>**

Submitted on 7 Sep 2010

**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.

# Robust Estimation of Hidden Corrosion Parameters using an Eddy Current Technique

Marie-Ève Davoust, Laurent Le Brusquet, and Gilles Fleury\*

## Abstract

An eddy current technique is used to inspect the interface between air and a conductive material such as aluminum, which can be covered with a non-conductive material. Hidden corrosion may appear inside the conductive material. This corrosion leads to flaws whose shape varies greatly depending of the flaw. The proposed methodology addresses this problem by considering the potential shapes as realizations of a random process. The goal of the proposed approach is not to find the exact shape of the corrosion flaw but to estimate some of its dimensional parameters. The area and the dimension ratio of the shape have been chosen because they depict the importance of the corrosion damage.

The estimation of the area and the dimension ratio is achieved in a Nondestructive Evaluation context: An alternating magnetic field is created in the air above the inspected material and the magnetic field near the air-aluminum interface is measured. It is a typical inverse measurement problem. Due to the complexity of the shape and of the physical phenomena, no algebraic model exists to solve this inverse problem. That is why a machine learning approach has been carried out: A database of observed signals for reference flaws is created (by using FEM tool) and used to calibrate a relationship giving the estimated area and the estimated dimension ratio from the observed signal. As the number of flaws in the database cannot be very large, the proposed approach overcomes the overfitting risk by performing a reduction of the data dimension.

PACS numbers:

Keywords: Nondestructive Evaluation, Inverse Problem, Hidden Corrosion, Eddy Currents

---

\*Department of Signal Processing and Electronic Systems , Supélec, 3 rue Joliot Curie, plateau de Moulon 91192, Gif-sur-Yvette Cedex, France.; Electronic address: `firstname.name@supelec.fr`

## I. INTRODUCTION

Application fields of eddy currents are numerous, for example nuclear electricity production (e.g. pipe testing), aircraft health monitoring (e.g. hidden corrosion), metallurgy (e.g. material classification, layer measurement). Eddy current techniques are particularly useful in Nondestructive Evaluation (NDE) to evaluate the dimensions of a flaw existing in a conductive material. Flaws can be grooves in pipe, cracks in a bulk material, interface flaws between two layers of aircraft fuselage [1], and so on.

In some NDE problems, the encountered flaw shapes can be represented by simple geometrical forms (e.g. recess, crack, and groove). This paper addresses the hidden corrosion problem for which a simple form is no longer suitable. That is why we have chosen to model the corrosion flaws as realizations of a 2D-random process. This representation well matches the random character and the erratic character of hidden corrosion.

This 2-D flaw is located at the surface of a bulk aluminum material which can be covered by a non-conductive material (e.g. coat of paint). It is thus non-visible. The eddy current technique consists in applying an alternating magnetic field above the surface to be inspected. This magnetic field creates eddy currents in aluminum, and so, the total magnetic field depends on the unknown flaw. Direct magnetic field measurements using giant magnetoresistive sensors (GMR) or Hall devices can be used [2, 3] and exploited to characterize the flaw.

When one is faced with hidden corrosion, estimating global geometrical parameters of the flaw seems easier than estimating its exact shape. It is nevertheless very useful to evaluate the damage. The proposed approach focuses on the estimation of the surface area and the dimension ratio of the flaw. This estimation is a classical inverse problem where the observed signals are not directly the parameters of interest (surface area and dimension ratio). Three difficulties arise:

1. The physical phenomena are too complex to be analytically described. This difficulty is overcome by adopting a machine learning approach: The searched relationship between observed signal and desired parameters is calibrated with a database of reference flaws.
2. The overfitting risk may lead to bad results. This well-known problem [4] occurs when the number of objects in the database (here the number of reference flaws) is not very

large compared to the length of the observed signal. A reduction of the dimension of the observed signal is used in the proposed approach. It has already been successfully used for another NDE problem [5, 6].

3. There are an infinite number of possible flaw shapes. This forces us to consider the variability of these flaws from the design of the inverse problem up to the test of the proposed estimation procedures.

The use of a random process to generate various flaw shapes is a key-point of the proposed approach to overcome these difficulties:

- It allows the construction of a database which includes flaw with erratic shapes, and in this way, flaws similar to the flaws which may be encountered in realistic situations.
- Realizations of the random process will be generated to define flaws used for testing. The independence of these realizations with the realizations used to calibrate the searched relationship will be a warrant for robustness.

The current study aims at showing the feasibility and the interest of such an approach. That is why a Finite Element Method (FEM) is used to create the database and why the study is restricted to 2D-flaws. The advantage of using FEM in association with the random process is the automatic generation of magnetic field signals for reference flaws and thus the possibility to test the method in numerous configurations.

Section II describes the creation of the database. Both the 2D-random process and FEM are explained. The learning approach carried out on the database is detailed in section III. It includes the reduction of the data dimension and the validation of the approach. Indeed, the proposed estimation of the flaw parameters has to remain accurate when it is applied for any flaw and with noisy signals. Thus, the robustness has been quantified.

At last, some conclusions are given concerning the feasibility of this study including random contours for flaw modeling.

## II. DATABASE CREATION

The aim is to characterize corrosion flaws hidden under an eventual layer of non-conductive material such as paint (see figure 8 for a geometrical description).

The principle of the presented eddy current technique is classic: An alternating magnetic field is created far away from the interface where a flaw may exist and parallel to this interface. Measurements of magnetic field are carried out in air near the interface in order to characterize a flaw located at the interface [7].

This section includes the definition of the simulated geometric forms, the FEM procedure, and the extraction of the observed magnetic field signals. The simulated geometric forms of flaws are generated with MATLAB while the FEM simulations are achieved with ANSYS.

### A. Parameters of interest for the flaw description

The database will be used to calibrate an inverse procedure whose aim will be to characterize the flaw shape from magnetic field measurements. The reconstruction of the exact shape is illusive in a context where the observed field depends more on the global shape than the details of the shape. In fact, we focus on 2 global parameters to describe a flaw:

- the surface area  $a$  of the defect. This parameter represents the importance of the flaw.
- a form parameter indicating if the flaw is flat or not. We define the dimension ratio parameter  $r$  as the ratio between the  $x$ -extension and the  $y$ -extension of the flaws (see figure 7 for a description of the  $x$  and  $y$  axis). In order to obtain a parameter robust against small variations of the shape, the  $x$ -extension and the  $y$ -extension are mean values defined from standard deviation expression:

$$r = \frac{\sigma_x}{\sigma_y} \text{ with } \sigma_\alpha = \frac{\iint_{\text{flaw}} (\alpha - \alpha_0)^2 dS}{\iint_{\text{flaw}} dS}, \alpha = x \text{ or } y \quad (1)$$

where  $x_0$  and  $y_0$  are the position of the center of mass of the flaw.

The reasons of choosing  $a$  and  $r$  are:

- $a$  and  $r$  are 2 meaningful parameters in a NDE context.
- It will be seen in section III that  $a$  and  $r$  are good candidates for the inverse problem procedure.

The database is populated by simulating contours for various areas and dimension ratio values. Ten values logarithmically and equally spaced between 0.05 and 1 for the  $r$  parameter, and between  $0.1\text{mm}^2$  and  $1\text{mm}^2$  for the  $a$  parameter have been chosen. The logarithmical

sampling is justified by the criterion used in the future calibration step (section III B) which favors small relative errors rather than absolute errors. Thus, the database includes 100 flaw contours. Figure 1 shows the chosen  $a$  and  $r$  values.

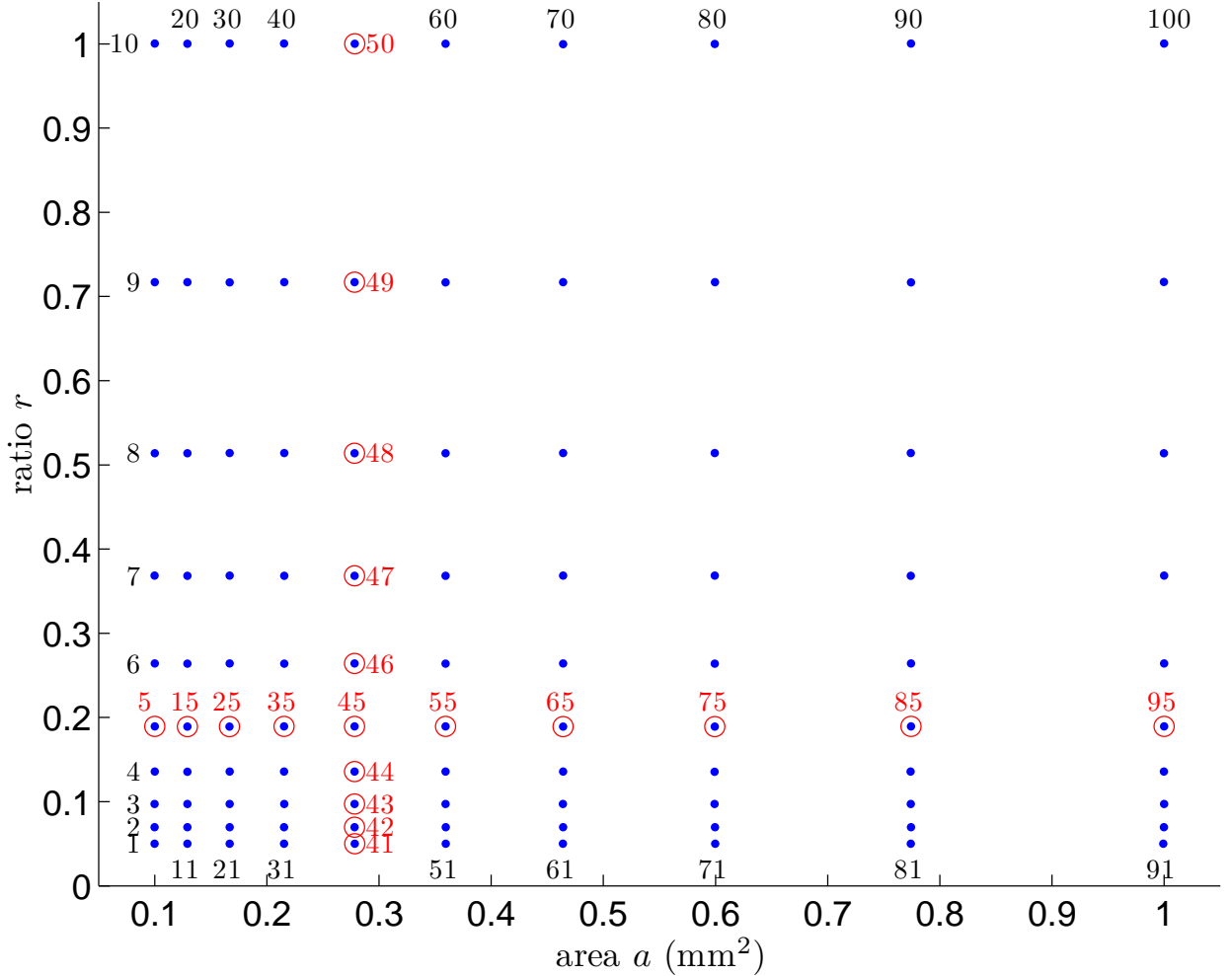


FIG. 1: Geometrical parameters ( $r$  and  $a$ ) of the simulated flaw contours. The points with a red circle represent the flaws that will be used for testing in section III C.

## B. Geometric forms of the simulated flaws

The literature about hidden corrosion flaws attention to the complex and erratic character of the corrosion phenomenon: Even for similar materials used in the same environment, corrosion flaws are irregular and are not repeatable [8]. That is why we have considered the geometrical descriptions of the flaws as realizations of a 2D-random process.

The literature also states that corrosion surfaces can be well described by fractals or multifractals [9][10]. The random process we choose is created in order to respect the fractal character of the flaw contours, and thus strives to create realistic flaws. Indeed, a simple process which permits the generation of a regular shape at an initial scale  $L_0$  is repeated at scales  $h_i = \frac{L_0}{2^i}$ ,  $i = 0 \dots I$  to obtain random variations from large scales to small scales.

In fact, the random process we use generates a 2D-function  $f(M)$  ( $M(x, y)$  is a point of the  $x - y$  plane) which is compared to a threshold  $f_{\text{thres}}$  to obtain the flaw:  $M$  belongs to the flaw if  $f(M) \geq f_{\text{thres}}$ . Suppose a contour  $\Gamma$  obtained from a 2D-function  $f(M)$ , is already available:

$$\Gamma = \{M/f(M) = f_{\text{thres}}\}$$

Figure 2 shows an example of a smooth function  $f(M)$  and the associated contour  $\Gamma$ . For  $y > 0$ ,  $f(M)$  is forced to be 0: it corresponds to the space over the material surface where the corrosion phenomenon does not exist.

Let us give details about the elementary process which adds details to a regular flaw contour. Suppose  $\Gamma$  must to be completed by details at scale  $h$ . A new 2D-function is obtained by adding local perturbations around points of the contour  $\Gamma$  in order to slightly change the contour shape.:

$$f(M) \leftarrow f(M) + \sum_{\{M_k \in \Gamma\}_{k=1 \dots K}} \alpha_k N \left( \frac{\|M - M_k\|}{h} \right) \quad (2)$$

where:

- $N$  is a centered function.
- $\alpha_k$  are coefficients randomly chosen with a zero mean.
- $K$  is the number of added local perturbations.

Many functions  $N(\|M\|)$  and many probability density functions (for  $\alpha$  and for the points  $M_k$ ) may be used since the aim of these functions is to create a local perturbation and not really to model the flaw shape. Indeed, the resulted flaw shapes depend more on the global approach (macroscopic aspect) than the parameter of the local perturbation (microscopic aspect). This point may be linked with the choice of a wavelet for the signal decomposition: Many choices exist. Nevertheless, some rules seem coherent:

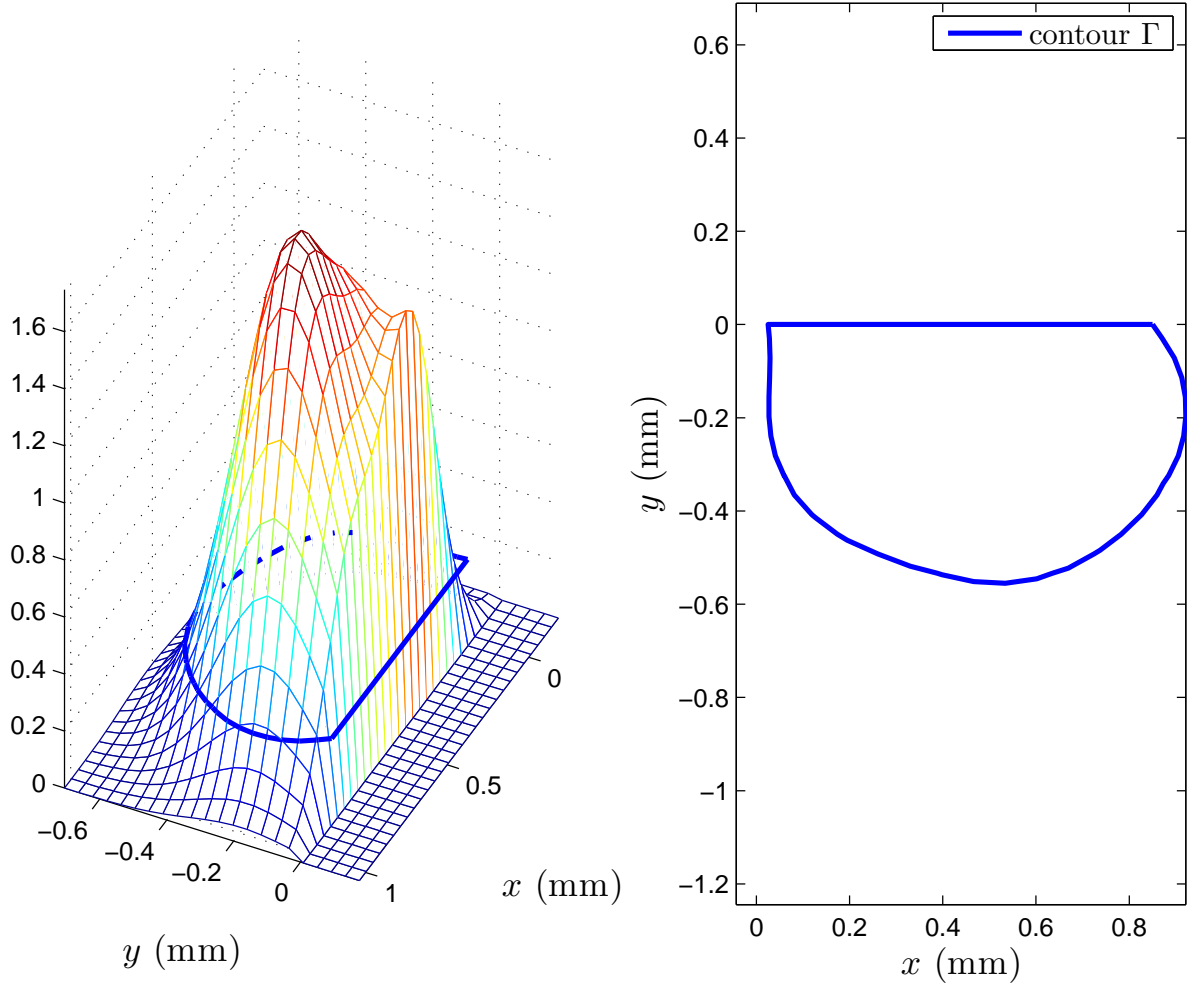


FIG. 2: Example of function  $f(M)$  (left) with the associated contour (right)

- $N$  has to be smooth since the high frequency variations will be created by the scale change. A commonly used type of RBF has been chosen:  $N(\|M\|) = \exp(-\|M\|^2)$ .
- $K$  depends on the influence radius of the RBF function, i.e.  $h_i$ . It is chosen approximately equal to  $\frac{l_i}{h_i}$  where  $l_i$  is the length of the contour at step  $i$ .
- No region of the contour is preferred over another and the  $K$  points  $M_k$  are randomly chosen along the contour with an uniform density in the curvilinear abscissa space.
- As  $N > 0$ , the sign of the local variation around  $M_k$  (see equation (2)) is the sign of



$\alpha_k$ . Thus  $\alpha_k < 0$  locally shrinks the flaw (and inversely  $\alpha_k > 0$  expands it). There is no reason to favor shrinkage or expansion and thus a symmetric probability function is chosen for  $\alpha_k$ . A Gaussian density has been chosen.

Figure 3 is an illustration of a more refined contour obtained from the smooth contour of figure 2. The process is repeated  $I$  times at scales  $h_i = \frac{L_0}{2^i}$ ,  $i = 1 \dots I$ . Figure 4 shows both final and intermediary contours. Examples of flaw contours are shown in figures 5 and 6.

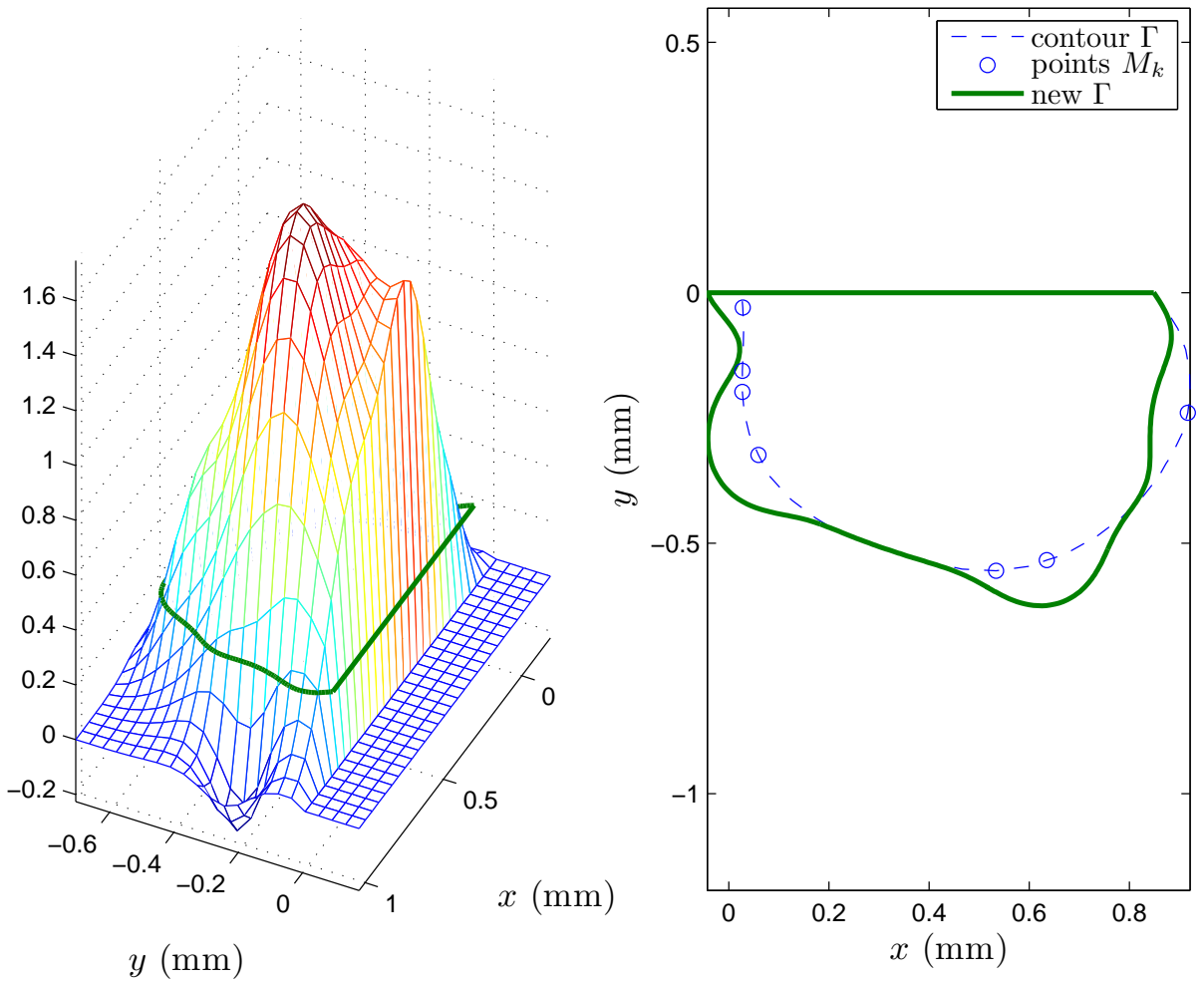


FIG. 3: Function  $f(M)$  of the figure 2 completed with smaller details (left) and the new associated contour (right).

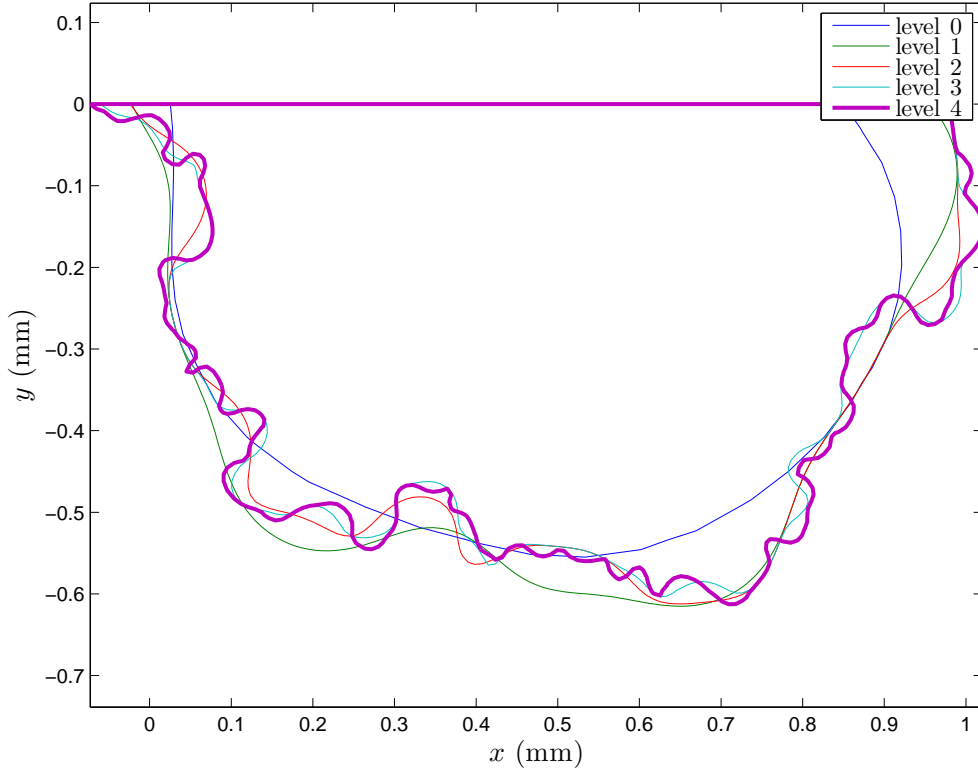


FIG. 4: Contours obtained after several applications of the elementary random process: level 0 stands for the initial smooth contour (figure 2 with details at scale  $h_0$ ) and level  $i$  stands for detailed contour with details at scales  $h_0 \rightarrow h_i$ .

Finally, a uniform sampling of the contour is performed in order to lead to a mesh adapted for the FEM, i.e. a regular mesh with a number of nodes as small as possible.

### C. Finite element model

Let us recall that a FEM tool is used because it is convenient for a feasibility study. The ANSYS software is used to solve the PDE for the magnetic vector potential.

The simulation of the experimental process carried out to characterize the desired flaws of section IIB needs 4 regions to be defined. Figures 7 and 8 show these different regions. Region B simulates aluminum, region D simulates the flaw whose contour points are imported from MATLAB calculations. Regions A and C represent non-conductive materials. The largest one (region A) represents air. The thinnest one (region C), above aluminum,

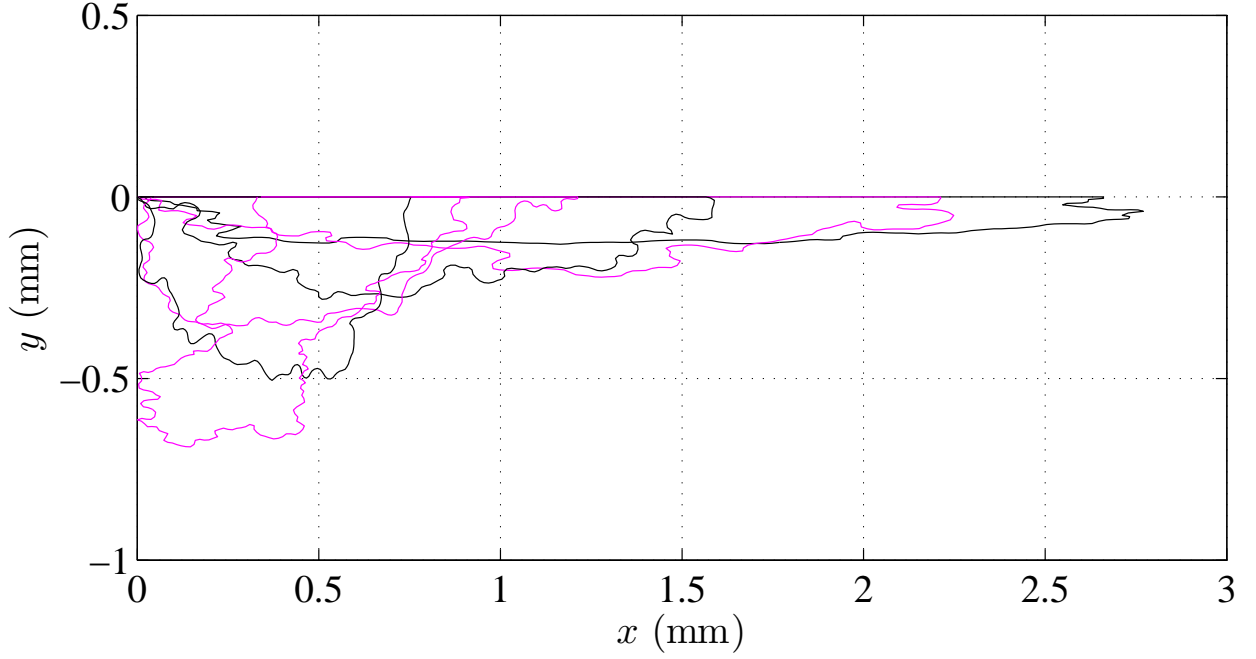


FIG. 5: Simulated flaw contours for  $a = 0.28 \text{ mm}^2$  and various dimension ratio values.

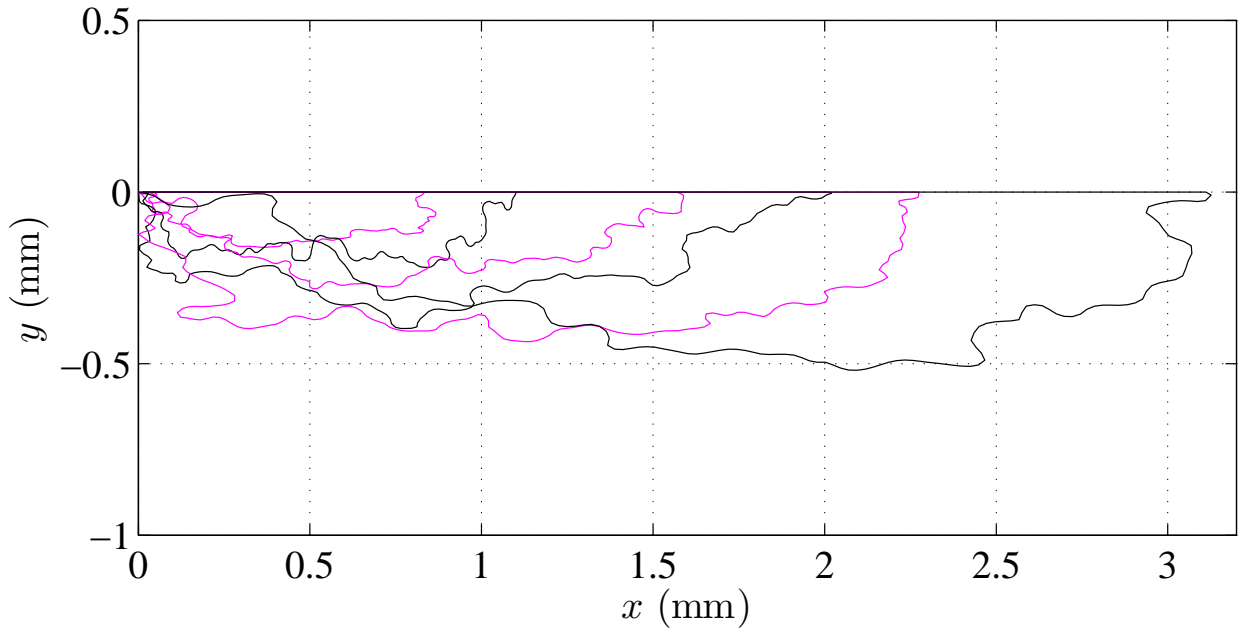


FIG. 6: Simulated flaw contours for  $r = 0.19$  and various area values.

simulates a layer of non-conductive material.

The physical properties of aluminum and aluminum oxide are considered to be homogeneous. The aluminum electrical conductivity is assumed to be  $3.7 \times 10^7 (\Omega\text{m})^{-1}$ . The

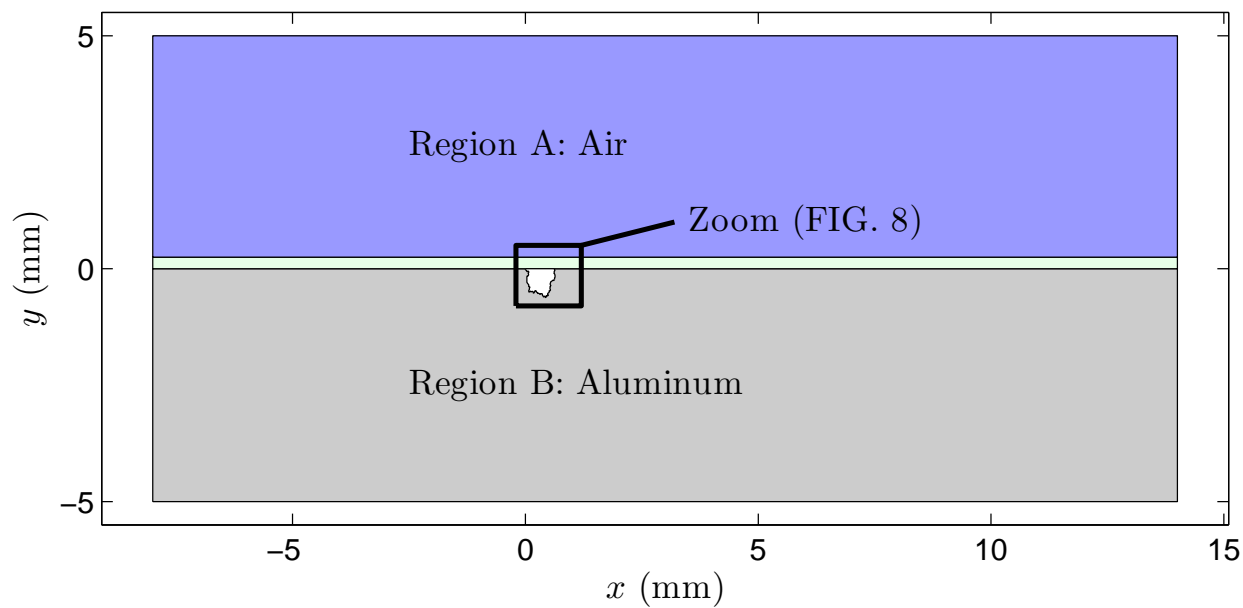


FIG. 7: Geometric description of the 4 regions simulated by the FEM.

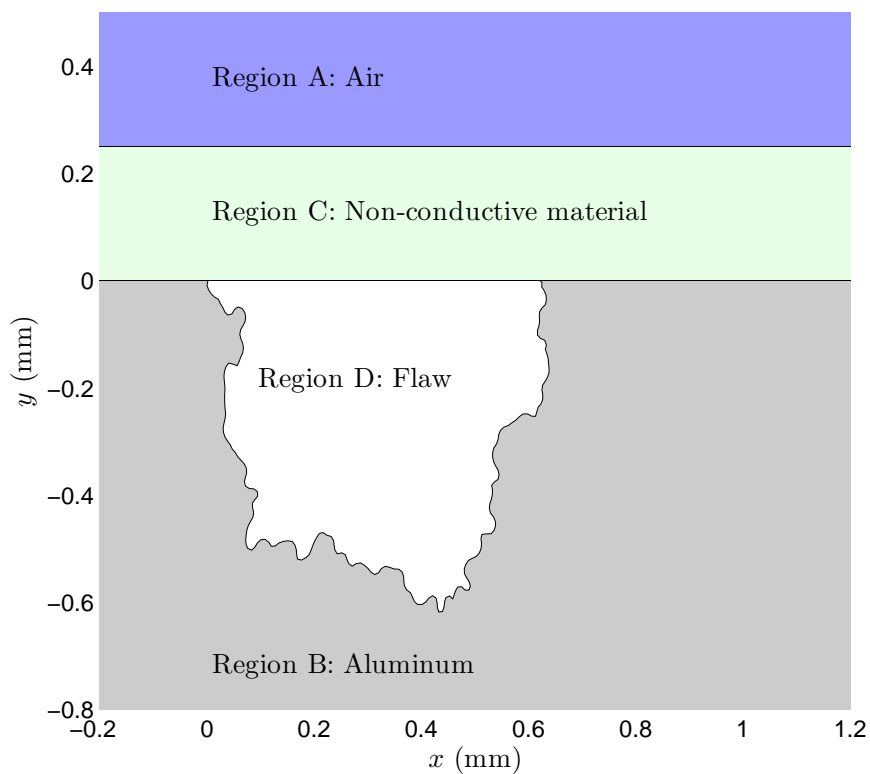


FIG. 8: Zoom on the flaw of the figure 7.

electrical conductivity of aluminum oxide is negligible in comparison to the electrical con-

ductivity of aluminum. So its exact value is of no significance.

It is supposed in this feasibility study that the sensor is small enough not to perturb the magnetic field. This is the reason why the FEM simulations do not include the sensor: The magnetic field values are picked up at a constant distance from the interface.

The alternating magnetic field is created by an alternating current applied at nodes located at the top of the air region in a direction perpendicular to the top edge of the air region. The current density of the alternating current is assumed to be  $4 \times 10^6$  A/m<sup>2</sup>. (It may be noted that the results of these calculations is linear with this excitation).

A last parameter is the alternating magnetic field frequency. Since the electrical conductivity of aluminum oxide is much more than that of aluminum, magnetic flux lines penetrate into the flaw region even if the excitation frequency is very high. For example, for 100 MHz, magnetic flux lines are concentrated in the air and some flux lines go down inside the flaw. In addition, the extension of the magnetic field variation is located near the flaw. Regarding low frequencies, the variation of magnetic field magnitude due to the flaw becomes less. A large range of frequencies can be used to detect the flaw. For a real experiment, some additional factors could reduce this range: the dimensions of the magnetic sensor, its resolution, the magnitude of measured signals, and so on. In short, the choice of frequency results in a trade-off between the sensitivity and the extension of the field variation due to the flaw.

For this feasibility study, the frequency has been assumed to be 10 kHz.

The whole dimension of the studied domain is: 22 mm  $\times$  10 mm.  $x$ -length and  $y$ -length of the domain were decided upon by assuming that far from the flaw, the magnetic field does not depend on the flaw. The chosen lengths are the smallest ones which satisfy this assumption at the domain boundary, even for the largest flaws. These lengths are coherent with the chosen frequency for the alternating magnetic field: At 10 kHz the skin depth in aluminium is about 0.8 mm and the  $x$ -length of 22 mm is appropriate to the region where variations of the  $x$ -component of the magnetic field exist.

The domain is automatically discretized with a triangular mesh by the ANSYS software. The mesh has been refined enough to obtain magnetic field values independent from the mesh. An example of mesh in the vicinity of a flaw is shown in figure 9.

Figure 10 shows the real part and the imaginary part of flux lines near the flaw. These illustrations show that the eddy current technique used can reveal the presence of a flaw. Without the flaw, these flux lines are parallel to the  $x$ -axis because the magnetic field created

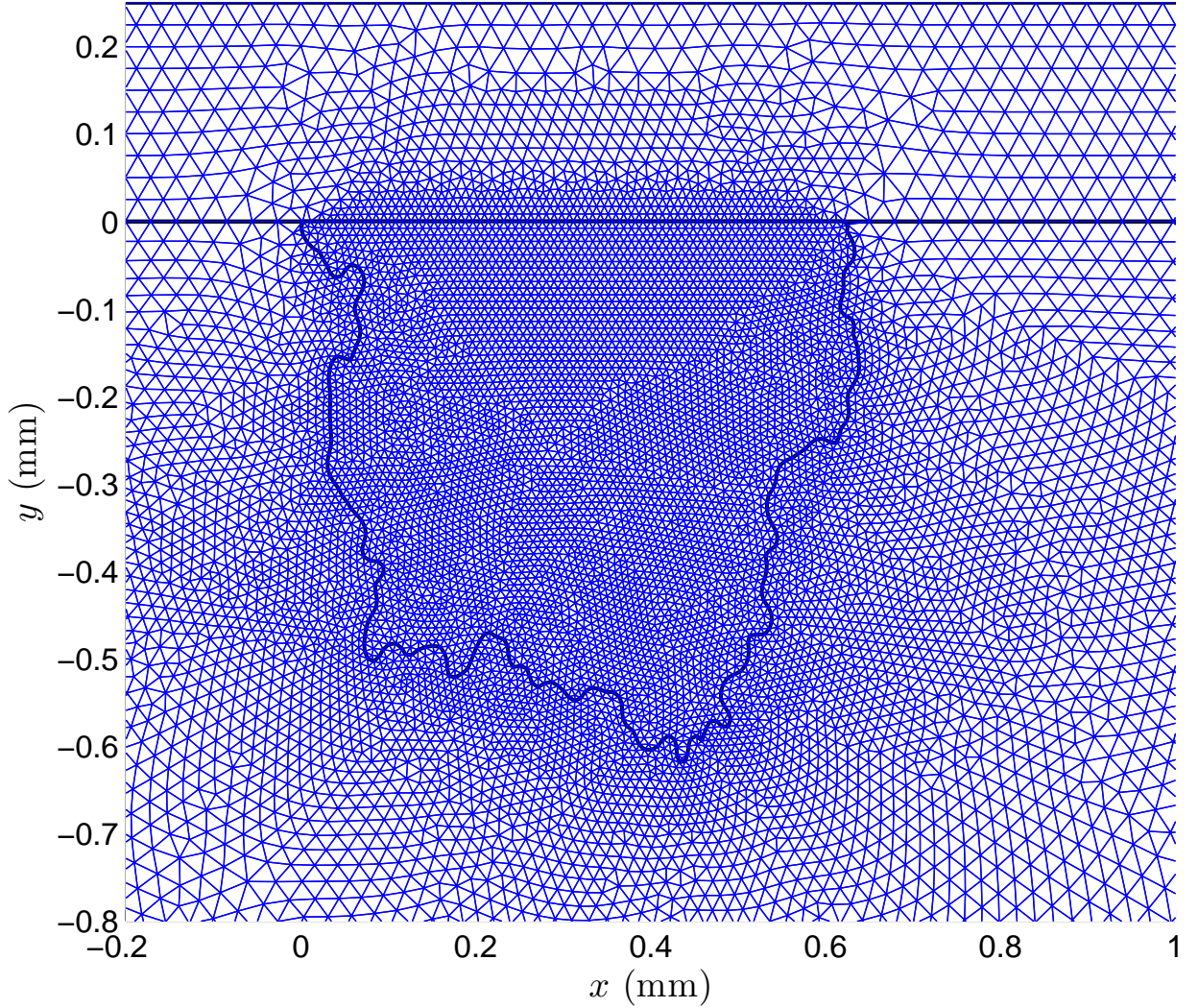


FIG. 9: Triangular mesh near flaw (flaw of the figure 7).

by the injected current is parallel to the  $x$ -axis.

#### D. Observed data

The FE calculations give the real part and the imaginary part of the two magnetic field components. As the  $y$ -component  $H_y$  is zero when there is no flaw in aluminum, this physical quantity is chosen to characterize the flaw. Magnetic field measurements are possible using an Hall device. It may be noted that the measurement of  $H_x$  is also possible by means of GMR.

Figures 11 and 12 show examples of real part and imaginary part (quantities calculated

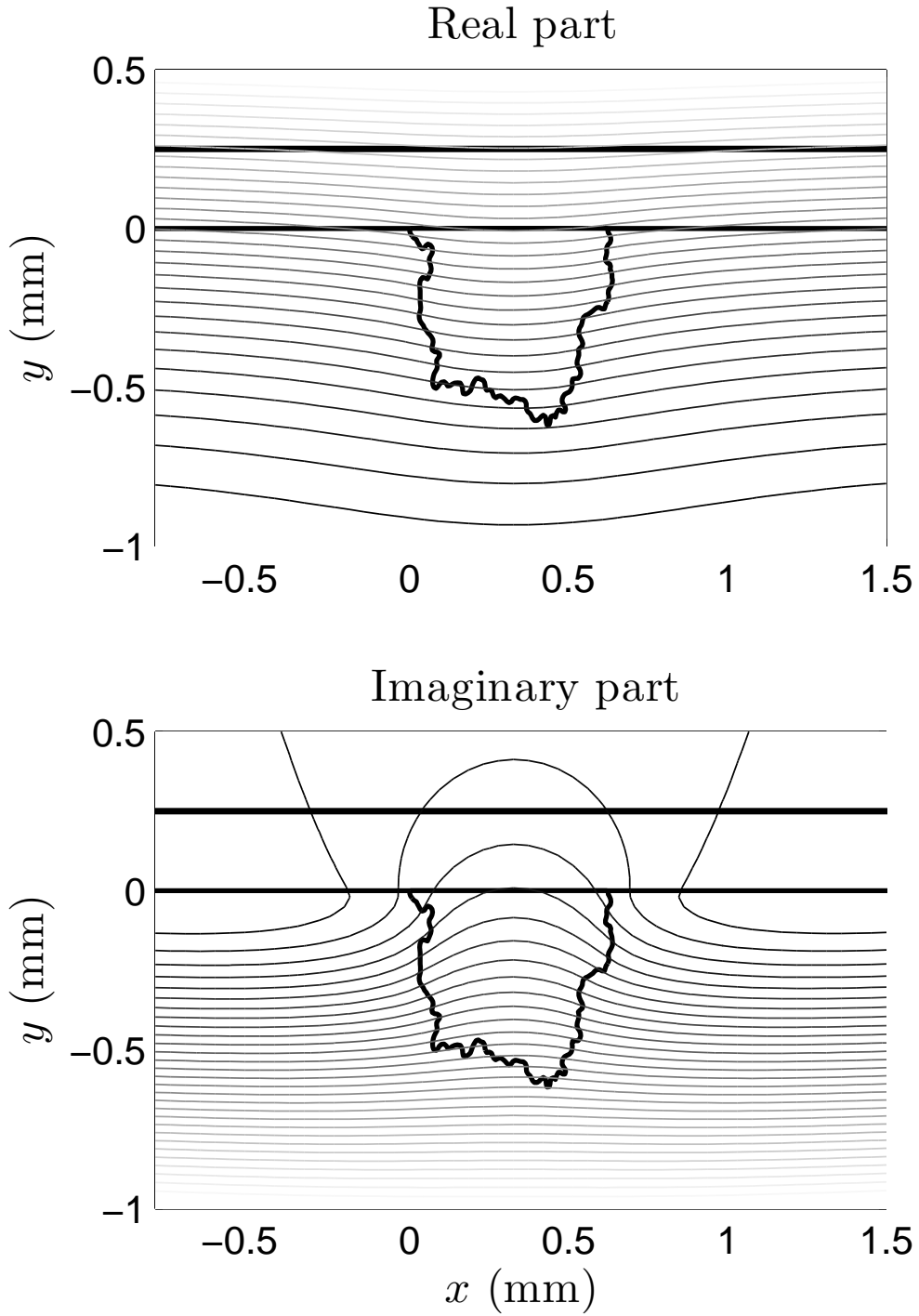


FIG. 10: Real part and imaginary part of the flux lines (flaw of the figure 7).

by the Ansys FEM software) of  $H_y$  for some flaws. The magnitude and phase of  $H_y$ , which are more significant data regarding magnetic field measurements, are calculated from these quantities.

In addition, when using FEM, the location of the flaw is known. This knowledge no longer exists with real flaws. The location of the flaw is estimated from the magnetic field values alone. The  $x$ -center of mass for the magnitude of  $H_y$  for each flaw is used as an estimator of location. A few examples of observed data (magnitude and phase) are shown in figure 13 and figure 14.

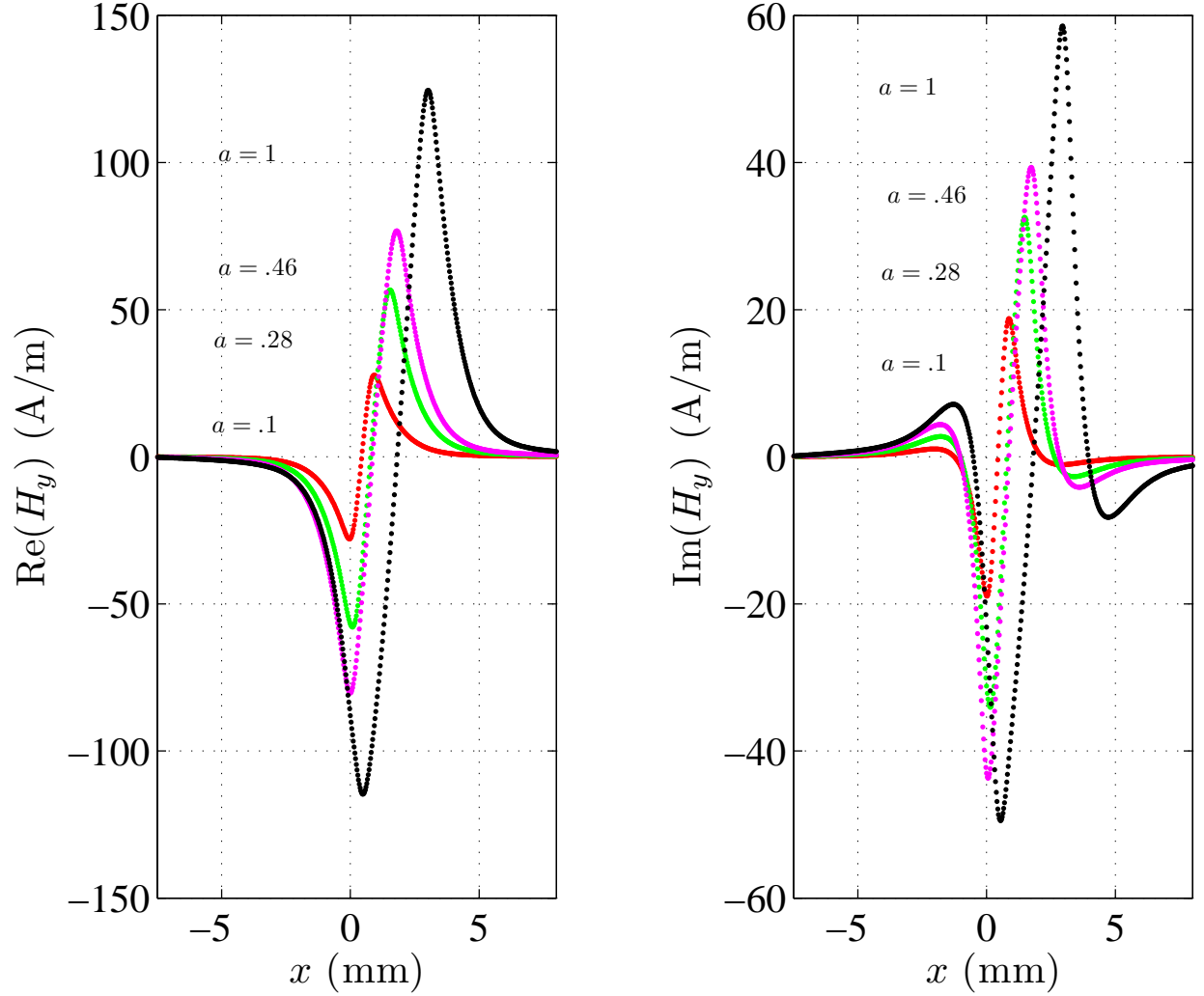


FIG. 11: Real part and imaginary part of  $H_y$  for 4 flaws with the same dimension ratio ( $r = .19$ ).

### E. Relevance of the surface area and dimension ratio

The relevance of the proposed sensor is related to its capacity to deliver observed signals which are more affected by the variations of the dimension parameters (area, dimension



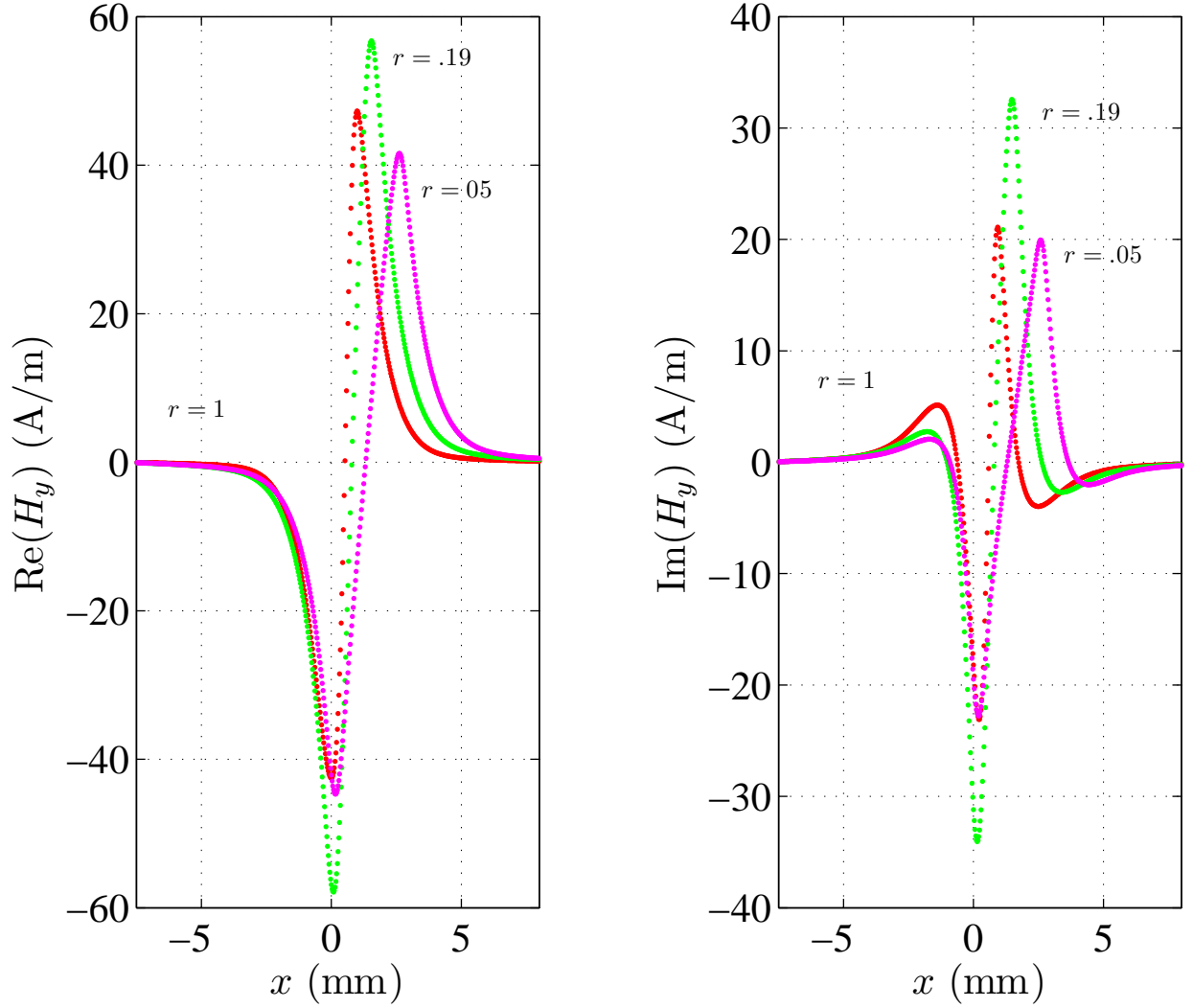


FIG. 12: Real part and imaginary part of  $H_y$  for 3 flaws with the same area ( $a = .28 \text{ mm}^2$ ).

ratio) than by the erratic variations of the contour flaw.

A criterion representing the variation level between the contours, whatever its origin is (parameter change or erratic character of the corrosion phenomenon) is required to quantify this point. Let  $\Gamma_1$  and  $\Gamma_2$  denote two contours to be compared. A natural criterion is using the ratio of the unshared area:

$$d_{\Gamma}(\Gamma_1, \Gamma_2) = 1 - \frac{S_{12}}{\sqrt{S_1 S_2}}$$

where  $S_1$  and  $S_2$  are the areas of the 2 flaws and  $S_{12}$  is the shared area. Figure 15 illustrates this criterion.

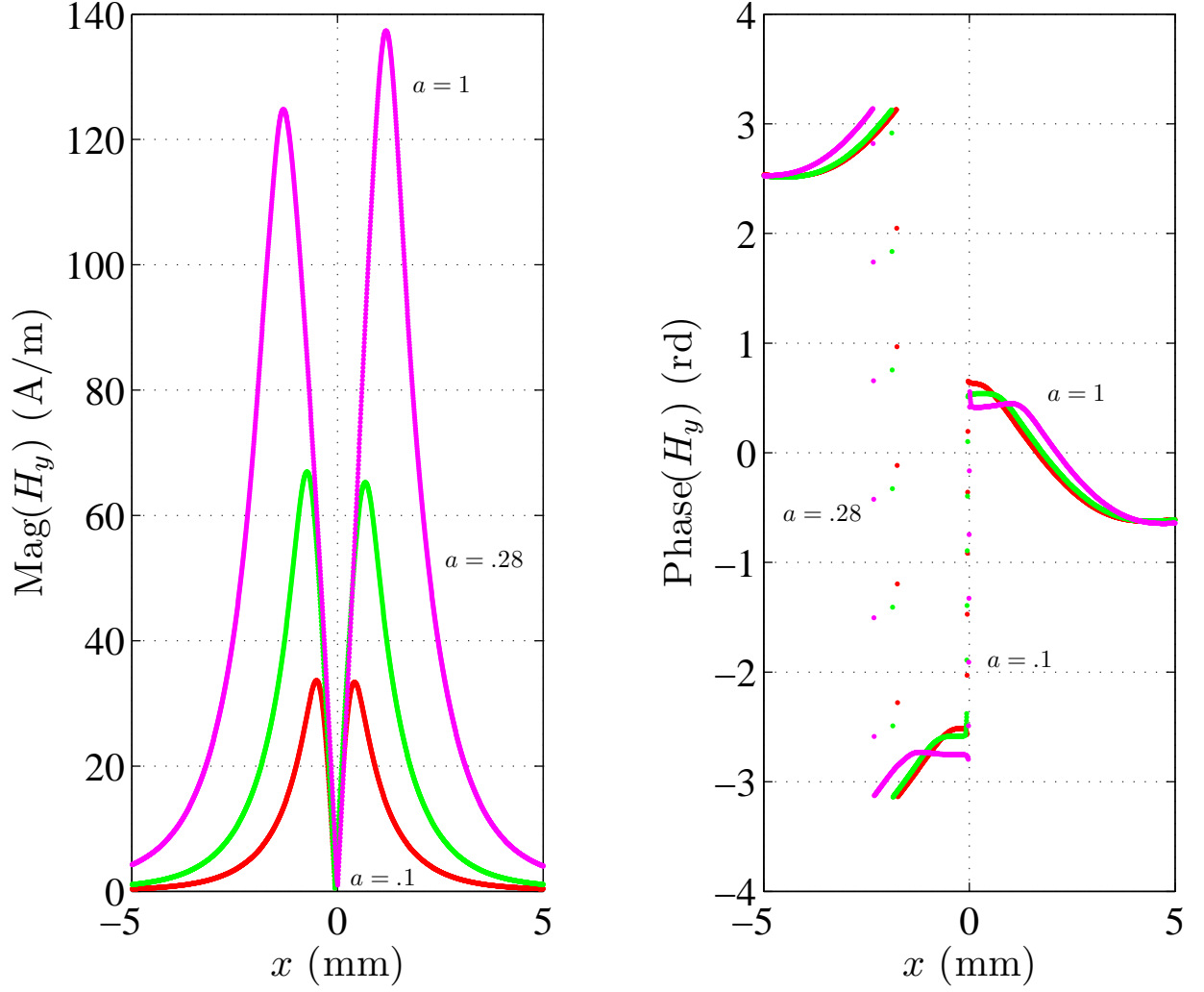


FIG. 13: Magnitude and phase of  $H_y$  (shifted with the barycenter of the magnitude of  $H_y$  for each flaw) for 3 flaws with the same dimension ratio ( $r = .19$ ).

For a fixed set of dimension parameters ( $a, r$ ), the dispersion level between flaw contours is evaluated by the following expected value:

$$\sigma = \mathbb{E}_{\Gamma \text{ of dimension } (a,r)} \{d_{\Gamma}(\Gamma, \bar{\Gamma}(a, r))\}$$

where  $\bar{\Gamma}(a, r)$  is defined as a mean contour over all the realizations of the random process with area  $a$  and dimension ratio  $r$ . More precisely, if  $(x(s), y(s))$  is the  $x - y$  description in the curvilinear coordinates of a contour  $\Gamma$ ,  $\bar{\Gamma}(a, r)$  is obtained by calculating the means of  $x(s)$  and  $y(s)$  over many realizations. Figure 16 shows the obtained  $\bar{\Gamma}$  for a unity area and an unity dimension ratio.

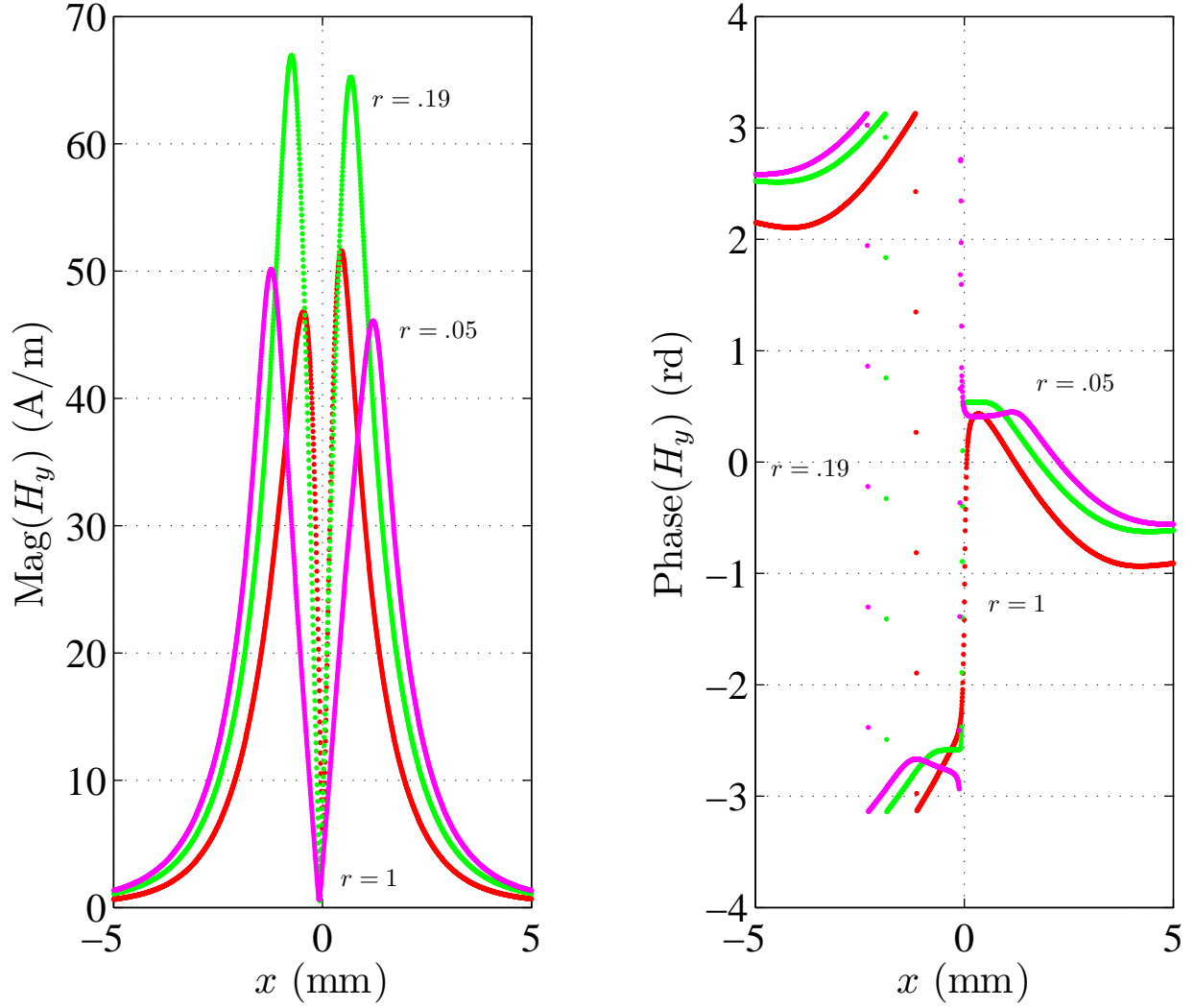


FIG. 14: Magnitude and phase of  $H_y$  (shifted with  $x$ -center of mass for the magnitude of  $H_y$  for each flaw) for 3 flaws with the same area ( $a = .28 \text{ mm}^2$ ).

As the random process for two sets of dimension parameters only differ from proportional transformations,  $\bar{\Gamma}(a, r)$  for any set of dimension parameters  $(a, r)$  is obtained from the previous  $\bar{\Gamma}(1, 1)$  with the same transformations. This also justified the independence of  $\sigma$  with  $a$  and  $r$ .

The first rows of Table I gives the dispersion value  $\sigma$ . It represents the effect of erratic variations of the contour shape. Similar dispersion values are reached by perturbing the dimension of the flaw (second and third rows of Table I).  $\Delta a = 29.2\%$  and  $\Delta r = 39.5\%$  are the increments used to build the database in section II B.

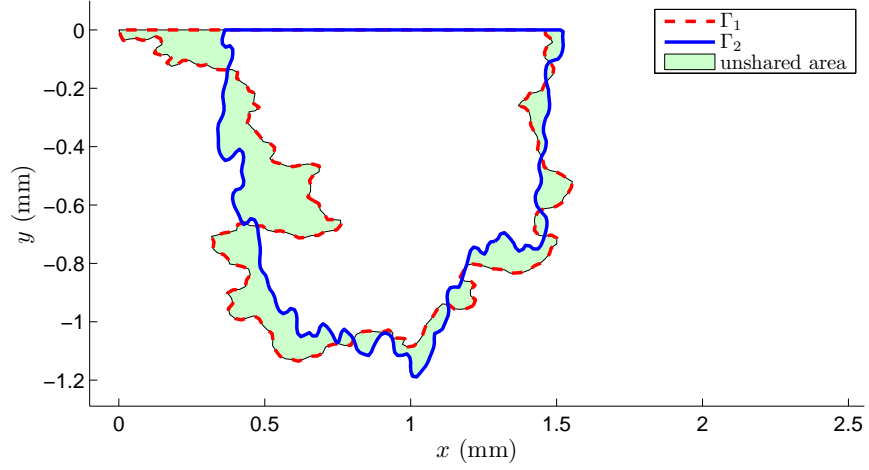


FIG. 15: Overlap between two flows:  $\Gamma_1$  and  $\Gamma_2$  are two random flows of unitary area ( $\text{mm}^2$ ). The fraction of unshared area is 27.8%.

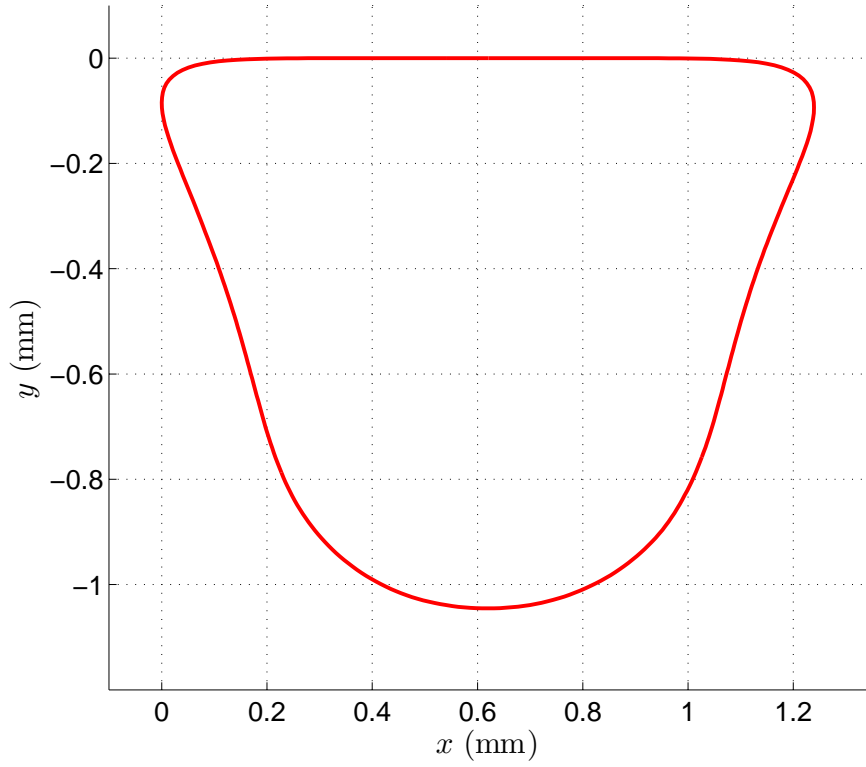


FIG. 16: Contour  $\Gamma_0(r, a)$ : Mean contour for a given set of dimension parameters ( $r = 1, a = 1$ ).

The consequences of the contour variations on the observed signals are quantified in the second column of Table I. The used criterion is the same as that used for the dispersion

TABLE I: Dispersion values, in the geometrical space (shape of the flaw) and in the signal space (magnetic field variations) for erratic changes of the flaw contour ( $\Gamma$  vs.  $\bar{\Gamma}(a, r)$ ) and for dimensional changes of the flaw ( $\bar{\Gamma}(a, r)$  vs.  $\bar{\Gamma}(a, r + \Delta r)$ ) and  $\bar{\Gamma}(a, r)$  vs.  $\bar{\Gamma}(a + \Delta a, r)$ ).  $\Delta a = 29.2\%$  and  $\Delta r = 39.5\%$  are the increments used to build the database in section II B.

	Geometry	Observed magnetic field
	$d_\Gamma$	$d_H$
$\sigma : \Gamma$ vs. $\bar{\Gamma}(a, r)$	17.0%	2.1%
$\bar{\Gamma}(a, r)$ vs. $\bar{\Gamma}(a, r + \Delta r)$	20.5%	21.9%
$\bar{\Gamma}(a, r)$ vs. $\bar{\Gamma}(a + \Delta a, r)$	21.6%	8.4%

level in the geometrical space:

$$d_H(H_1, H_2) = 1 - \frac{\int_{x_{\min}}^{x_{\max}} |H_1(x) - H_2(x)| dx}{\sqrt{\int_{x_{\min}}^{x_{\max}} |H_1(x)| dx \int_{x_{\min}}^{x_{\max}} |H_2(x)| dx}}$$

where  $H_1$  and  $H_2$  are the magnitudes of the magnetic field due to the two flaws ( $H_y$ ).

Table I shows that the observations are more influenced by changes in area  $a$  and in dimension ratio  $r$  than by the erratic shape of the contour. Using this point, we can feel confidence that we will be able to recover  $a$  and  $r$ , even for perturbed contours.

### III. INVERSE PROBLEM

The number of calibrated flaws (100) does not allow the direct use of the magnetic field observations ( $H_y$ ) in the inverse problem step: An extraction of features using the Principal Component Analysis is conducted in order to reduce the data dimension. After this step, the two unknown parameters  $r$  and  $a$  must be estimated from the sets of the components retained for each flaw. We have used in this second step, a bilinear relationship to link components to  $r$  and  $a$ . Previous studies [5, 6], concerning the robust estimation of groove dimensions existing in ferromagnetic pipes using Remote Field Eddy Current technique, have shown that the proposed method is accurate for estimating the two dimensions of grooves. Moreover these studies compare two approaches of dimension reduction as parameter modeling of

observed data and the PCA technique performed on the same observed data. This study shows that using the PCA approach gives better results in terms of accuracy and robustness.

So, we have chosen to solve the inverse problem by using PCA approach to reduce the data dimension, and the same bilinear relationship to estimate  $r$  and  $a$ .

### A. Principal Components Analysis

The advantage of such a reduction dimension method is that no precise prior knowledge is required. This well-known method [11] generates a new set of uncorrelated variables sorted by decreasing variances. In this way, it is possible to reduce the dimension of the observed data space while controlling the loss of information. In addition, since the new variables are uncorrelated, no redundant information is taken into account. This property will be of great interest in the regression step. Indeed, it allows a favorable trade-off between accuracy and robustness because it respects the parsimony principle (the information has to be represented by a small number of values in order to reduce the overfitting risk).

Otherwise, to keep all the information given by the database, a PCA is conducted on the magnitude of  $H_y$  concatenated with the phase of  $H_y$  for each flaw. The PCA technique is carried out after the magnitude data and the phase data are normalized. Figure 17 shows the variability of the first 10 components.

The dimension reduction is performed by keeping  $N_{pc}$  principal components. The choice of  $N_{pc}$  is a key-point of the approach since this parameter controls the trade-off between precision and robustness. This choice is detailed in the section III C.

Figure 18 shows an example of observed data and the resulting PCA modeled data with 6 kept components (which explain 97.2% of variability of the observed data.)

### B. Flaw parameter estimation

In order, to estimate the dimension ratio  $r$  and the area  $a$  of a flaw, a relationship between the PCA components and the flaw parameters have to be chosen. The simplest choice is a linear function of the PCA components but a previous study [12] proved that to improve

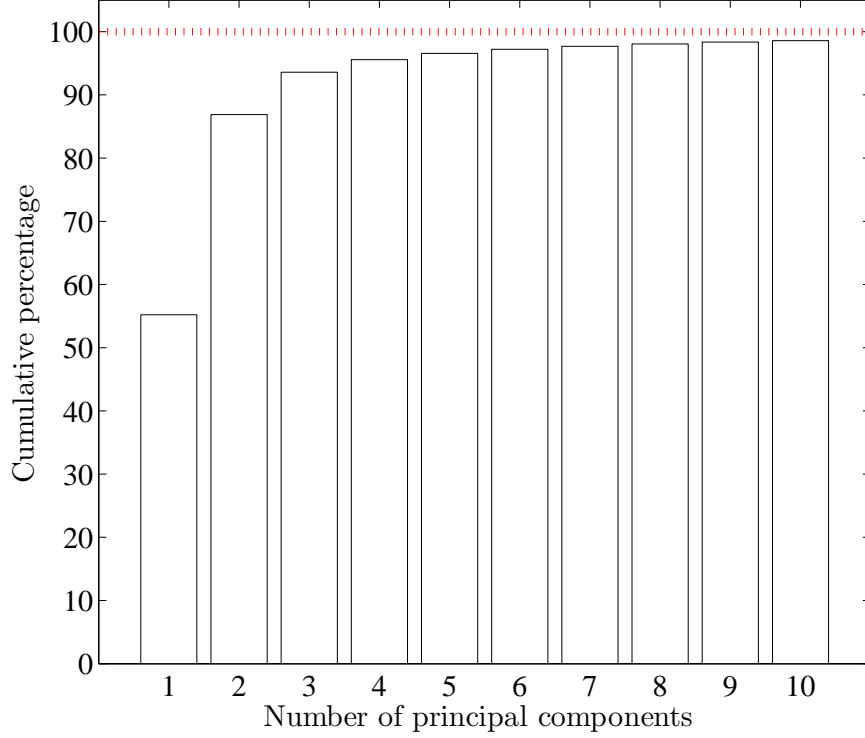


FIG. 17: Variability of the initial observed data explained by the first principal components.

the overall accuracy of the method, the bilinear one given by (3) is suitable.

$$\begin{aligned}
 r &= g_r(\boldsymbol{\theta}) = c_0^r + \sum_{i,j} c_{ij}^r \theta_i \theta_j + \sum_i c_i^r \theta_i, \\
 a &= g_s(\boldsymbol{\theta}) = c_0^a + \sum_{i,j} c_{ij}^a \theta_i \theta_j + \sum_i c_i^a \theta_i.
 \end{aligned} \tag{3}$$

The vector  $\boldsymbol{\theta}$  is the vector of the retained PCA components. By using matrix expressions, these equations may be written for all simulated defects as:  $\mathbf{r} = \boldsymbol{\Theta} \mathbf{c}^r$  and  $\mathbf{a} = \boldsymbol{\Theta} \mathbf{c}^a$ , where the matrix  $\boldsymbol{\Theta}$  contains PCA components for each flaw; each line of  $\boldsymbol{\Theta}$  is composed of the terms of the bilinear form for one flaw, as defined in (3). The  $c_{ij}$  coefficients are contained in vectors  $\mathbf{c}^r$  and  $\mathbf{c}^a$ , and vectors  $\mathbf{r}$  and  $\mathbf{a}$  contain the exact flaw parameters of the database. The  $c_{ij}$  coefficients may be calculated by solving the system of equations (3) in the least squares sense. As the error will be given in terms of relative errors, a relative criterion is used. The calibration phase (or learning phase) of the model building consists in calculating vectors  $\mathbf{c}_{\text{opt}}^r$  and  $\mathbf{c}_{\text{opt}}^a$  for a given number of flaws noted  $k$ . The vectors  $\mathbf{c}_{\text{opt}}^r$  and  $\mathbf{c}_{\text{opt}}^a$  are given

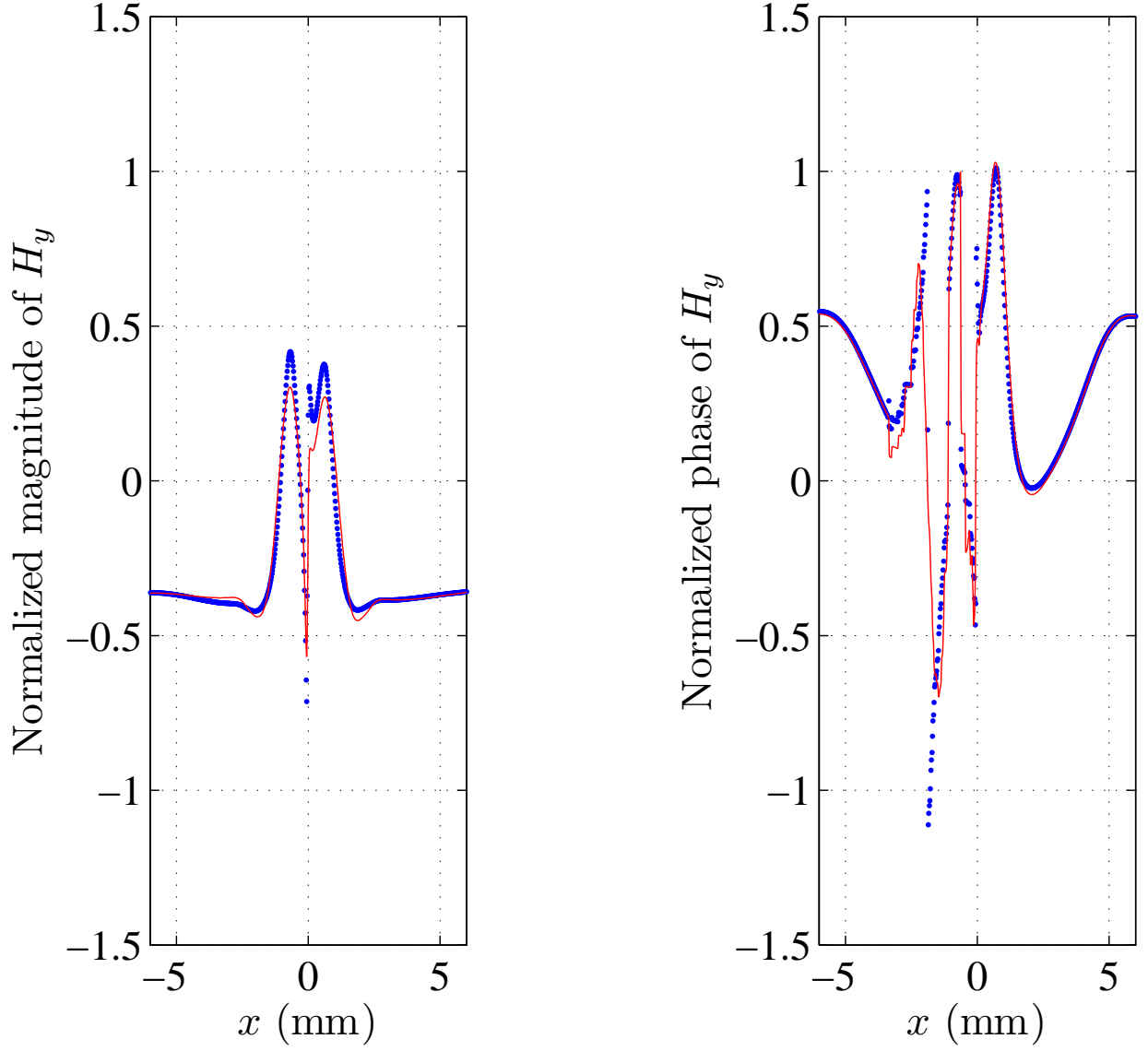


FIG. 18: Normalized observed data (.) and PCA normalized modeled data (-),  $a = .28 \text{ mm}^2$  and  $r = .19$ .

by:

$$\begin{aligned}
 \mathbf{c}_{\text{opt}}^r &= \arg \min_{\mathbf{c}^r} \sum_{i=1}^k \left( \frac{\boldsymbol{\theta}_i^\top \mathbf{c}^r - r_i}{r_i} \right)^2, \\
 \mathbf{c}_{\text{opt}}^a &= \arg \min_{\mathbf{c}^a} \sum_{i=1}^k \left( \frac{\boldsymbol{\theta}_i^\top \mathbf{c}^a - a_i}{a_i} \right)^2.
 \end{aligned} \tag{4}$$



Then, the estimated values of  $\mathbf{r}$  and  $\mathbf{a}$  can be calculated according to:

$$\begin{aligned}\hat{\mathbf{r}} &= \Theta \mathbf{c}_{\text{opt}}^r, \\ \hat{\mathbf{a}} &= \Theta \mathbf{c}_{\text{opt}}^a.\end{aligned}$$

The calibration error is defined as the minimum of the criteria in equation (4), i. e.  $\sqrt{\sum_{i=1}^k \left(\frac{\hat{r}_i - r_i}{r_i}\right)^2}$  for the dimension ratio and  $\sqrt{\sum_{i=1}^k \left(\frac{\hat{a}_i - a_i}{a_i}\right)^2}$  for the surface area.

Equations (3) show that if  $n$  stands for the number of coefficients  $c_{ij}$  for  $r$  and  $a$ ,  $n$  is given by:

$$n = 1 + N_{\text{pc}} + (N_{\text{pc}}(N_{\text{pc}} + 1))/2.$$

$n$ , which depends on  $N_{\text{pc}}$  cannot be greater than the number of equations of the rectangular systems (equation (3)) because the mean square criterion is no longer adapted (for 100 flaws,  $\max(N_{\text{pc}}) = 12$ ).

The overfitting phenomenon can appear when  $n$  is too large, so, to avoid the overfitting risk, smallest possible  $N_{\text{pc}}$  should be chosen.

### C. Choice of the number of components

The robustness of the obtained estimators is evaluated as a function of  $N_{\text{pc}}$ . The aim is to find estimators that are robust with:

1. noisy observations,
2. flaws which do not exist in the training database.

Both kinds of robustness are evaluated by using a cross-validation technique performed on noisy observations.

A leave-one-out cross-validation is performed: 99 calibrated flaws are used to estimate  $\mathbf{c}_{\text{opt}}^r$  and  $\mathbf{c}_{\text{opt}}^a$  and the left one is used to generate noisy observations that are used for testing.

The flaws whose surface area is  $a = .28 \text{ mm}^2$  or whose dimension ratio is  $r = .19$  have been tested. These  $n_{\text{test}} = 19$  selected flaws are marked with a red circle in figure 1. For each of these flaws,  $n_{\text{noise}} = 100$  noisy observations have been built by considering them as observations associated with several erratic flaw shapes with a constant surface area and a constant dimension ratio.

FEM calculations are then carried out for each realization. Figure 19 shows some realizations of the same flaw. Figure 20 shows the observations corresponding to these realizations. For each tested flaw, the coefficients  $\mathbf{c}_{\text{opt}}^r$  and  $\mathbf{c}_{\text{opt}}^a$  in equations (4) are used to calculate esti-

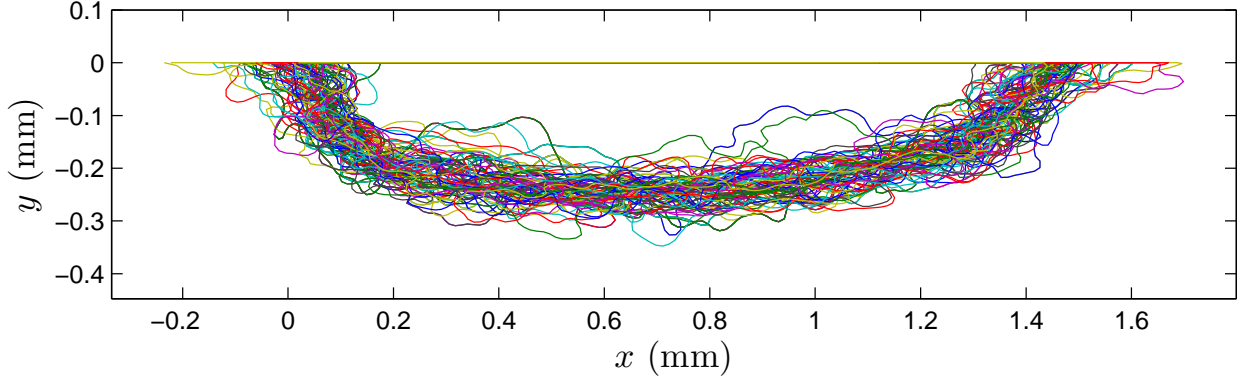


FIG. 19: Some realizations of contours for  $a = .28 \text{ mm}^2$  and  $r = .19$  (flaw #45).

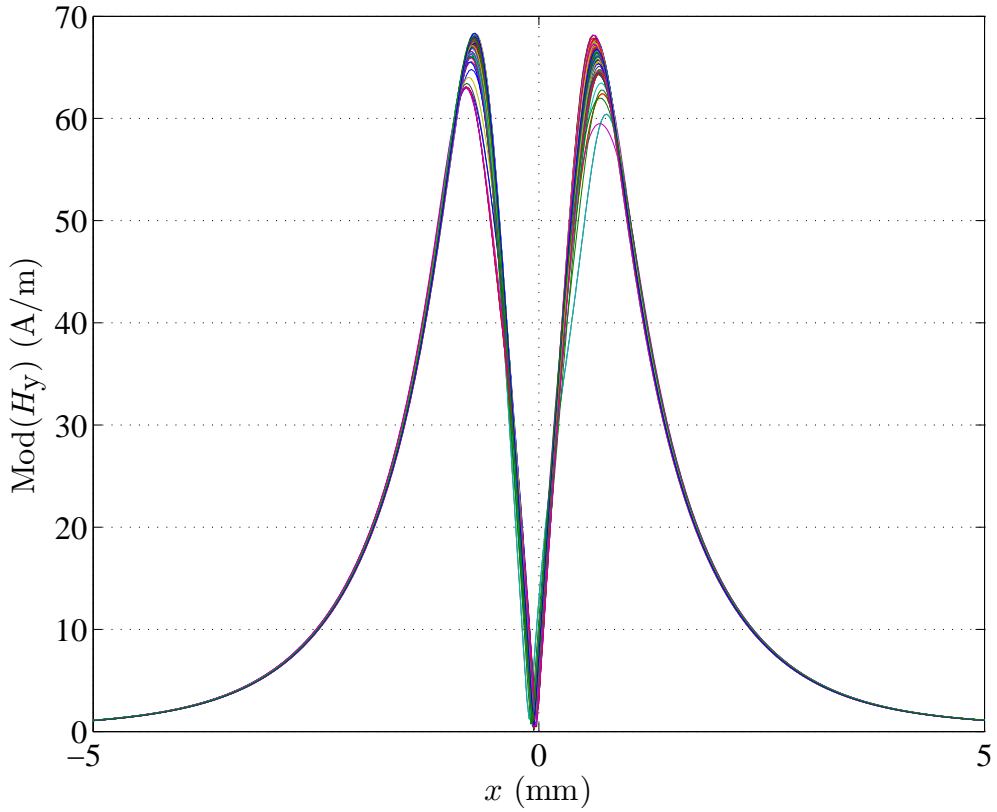


FIG. 20: Noisy centered observations (magnitude of  $H_y$ ) for  $a = .28 \text{ mm}^2$  and  $r = .19$  (flaw #45).

mates of  $r$  and  $a$  and so, the  $n_{\text{test}} \times n_{\text{noise}}$  relative estimation errors. The cross-validation error

$e_{CV}$  is defined as the mean error (in the least-squares sense) calculated over the  $n_{\text{test}} \times n_{\text{noise}}$  tested observations.

Figure 21 gives both calibration and cross-validation errors and shows the effect of over-fitting. For both parameters  $r$  and  $a$ , the optimal number of components  $N_{\text{pc}}$  is equal to 6 regarding the minimum of the mean cross-validation error. That is why this value was chosen for the evaluation of the method.

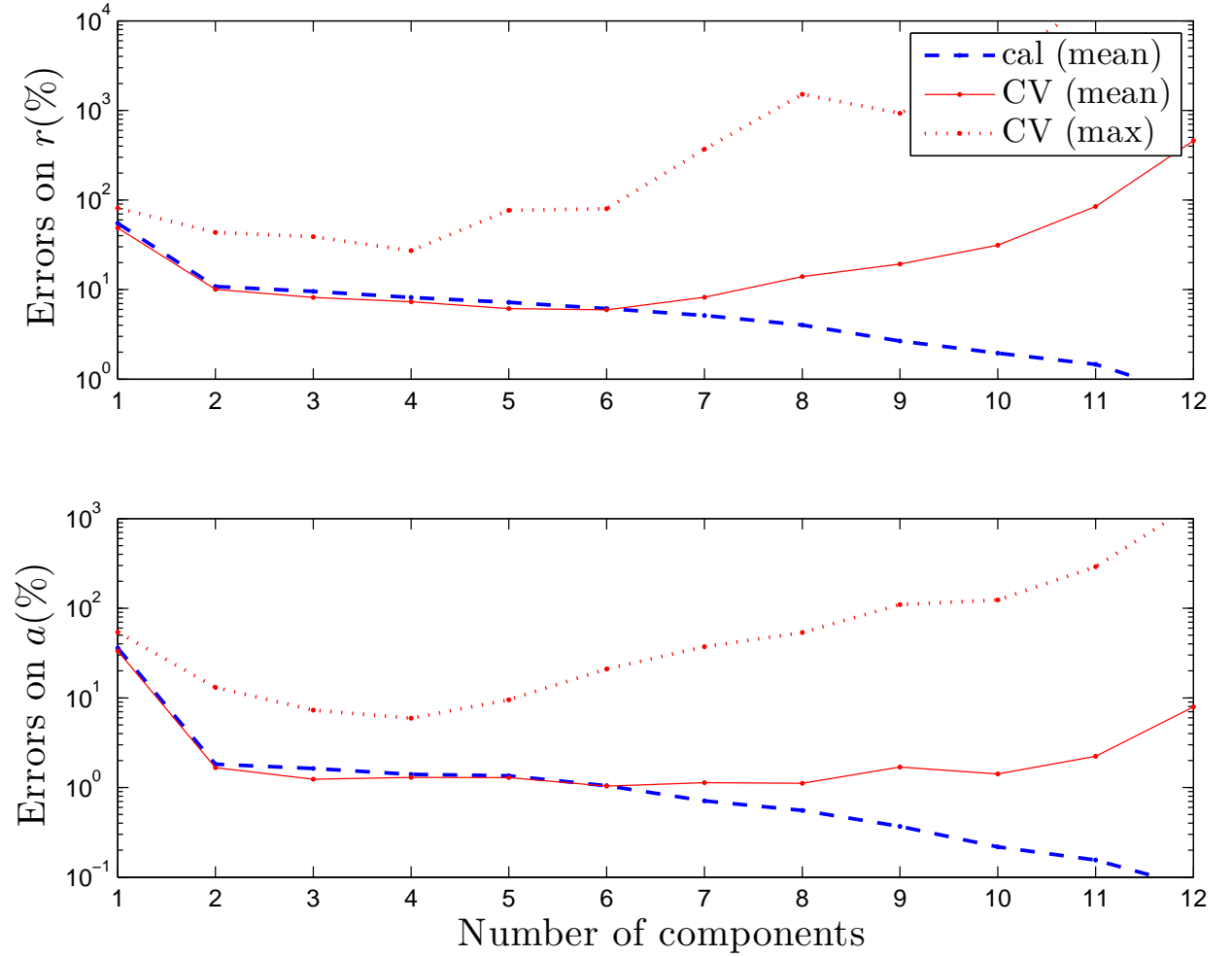


FIG. 21: Calibration errors (cal) and Cross-validation errors (CV) on  $r$  and  $a$  versus the number of components. It could be noted that the calibration error  $e_{\text{cal}}$  (dashed line) decreases with increasing the number of Principal Components.

## D. Results

Figures 22 and 23 gives for each tested flaw the histograms of the cross-validation errors for the dimension ratio and for the surface area. The largest errors are often reached for flaws at the limit of the discretized  $(a, r)$  domain. It is particularly true for flaw #41 and #95 (see the  $(a, r)$  domain in figure 1). This point may be linked with the fact that the overfitting risk increases when data used in the testing step are far from the data used in the calibration step.

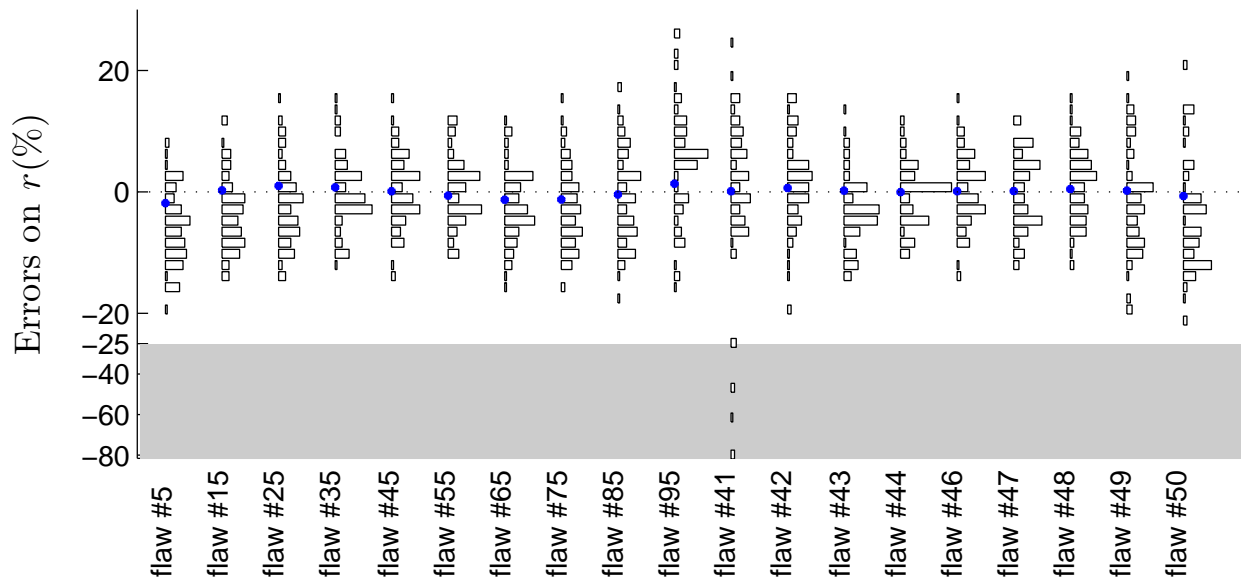


FIG. 22: Histograms of the cross-validation errors (%) on the dimension ratio for the tested flaws marked with a red circle in figure 1. The  $y$ -axis for errors  $\leq -25\%$  has been compressed in order to reach more readable histograms. The blue points are the mean errors.

Table II summarizes the results for the 19 tested flaws. The table shows that even if the maximum errors are much larger than mean errors, errors are less than 14% for  $r$  and than 3% for  $a$  in 95% ( $e_{CV_{95\%}}$ ) of all cases. It could also be noted that the cross-validation errors, even for tests on noisy observations obtained from perturbed flaw shapes, are barely greater than the calibration errors. This robustness is reached thanks to the use of appropriate global shape parameters and thanks to the control of the overfitting risk (achieved by the choice of the number of PCA components). The choice of the surface area  $a$  and the dimension ratio  $r$  as global shape parameters to be recovered is a key-point because these parameters are meaningful for the addressed problem (they well show the extensiveness of the corrosion

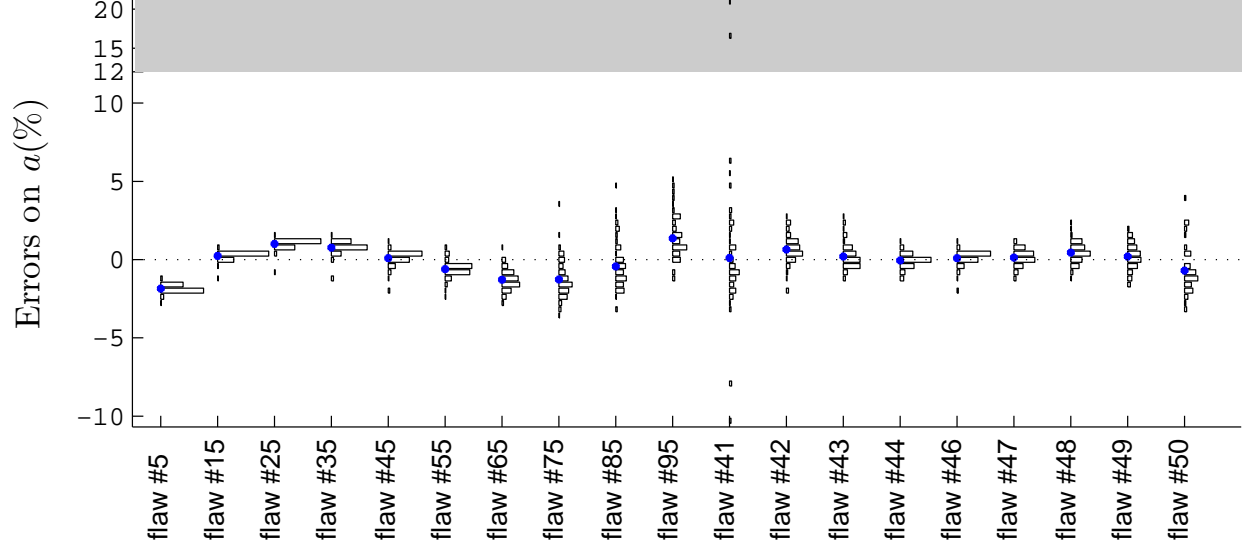


FIG. 23: Same caption than in figure 22 but for the surface area. The  $y$ -axis for errors  $\geq 12\%$  has been compressed.

TABLE II: Calibration and cross-validation errors (%).

Dimension ratio ( $r$ )	$\sqrt{\langle e_{\text{cal}}^2 \rangle}$	7.71
	$\sqrt{\langle e_{\text{CV}}^2 \rangle}$	7.90
	$e_{\text{CV}95\%}$	13.65
	$\max  e_{\text{CV}} $	79.80
	$N_{\text{pc}}$	6
Area ( $a$ )	$\sqrt{\langle e_{\text{cal}}^2 \rangle}$	1.39
	$\sqrt{\langle e_{\text{CV}}^2 \rangle}$	1.57
	$e_{\text{CV}95\%}$	2.58
	$\max  e_{\text{CV}} $	20.95
	$N_{\text{pc}}$	6

damage) and because the use of integral expressions in their definition (see equation (1)) leads to robust estimations.

## IV. CONCLUSION

The paper proposes a full approach, i.e. both the non-destructive measurement device and the associated signal processing, to estimate geometrical parameters of hidden corrosion flaws. The major part of the work is dedicated to robustness: The geometrical parameters have to be correctly estimated for a large range of flaw contours. However, the shapes of the corrosion flaws are erratic and may lead to poor estimates if the erratic character is not considered. Erratic contours have been generated and the proposed estimator is built by considering a cross-validation technique in order to address this problem. Results attest to the correct behavior of the estimator since a parameter regarding the corrosion level (the area has been chosen to consider this level) is estimated with mean errors less than 2% and a parameter regarding the flaw shape (the dimension ratio has been chosen) is estimated with mean errors less than 10%.

The proposed methodology has been tested on simulated flaws to show its feasibility and its benefits. Simulation is a convenient approach to generate a number of hidden corrosion flaws as large as required to test the approach. Combining a FEM tool with a procedure which automatically generates erratic flaw shapes is in this work the key to reach the required robustness.

- 
- [1] A. Y. M. Spies and C. Kropas-Hugues, "Simulation of ultrasonic and electromagnetic non-destructive evaluation of corrosion," *Review of Quantitative Nondestructive Evaluation*, vol. 22, pp. 367–374, 2003.
  - [2] T. Dogaru and S. Smith, "Giant Magnetoresistance-Based Eddy-Current Sensor," *IEEE Transactions on Magnetics*, vol. 37, no. 3, pp. 3831–3838, 2001.
  - [3] V. Katyal and J. Bowler, "Simulation of Eddy-Current Corrosion Detection using a sensor array," *AIP Conference Proceedings*, vol. 657A, pp. 359–366, 2003.
  - [4] C. Schaffer, "Overfitting Avoidance as Bias," *Machine Learning*, vol. 10, no. 2, pp. 153–78, 1993.
  - [5] L. L. M.-E. Davoust and G. Fleury, "Robust estimation of flaw dimensions using Remote Field Eddy Current Inspection," *Measurement Science and Technology*, vol. 17, pp. 3006–3014, 2006.

- [6] M.-E. D. L. Le Brusquet and G. Fleury, “A comparison between parametric and PCA methods of dimension reduction - Application to an Eddy current non destructive evaluation,” *International Journal of Applied Electromagnetics and Mechanics*, vol. 19, no. 1-4, pp. 119–124, 2004.
- [7] Z. W. J. C. P. Huang, G. Zhang and Z. Zhou, “Inspection of defects in conductive multi-layered structures by an eddy current scanning technique: Simulation and experiments,” *NDT & E international*, vol. 39, no. 7, pp. 578–84, 2006.
- [8] S. Xu and Y. Weng, “A new approach to estimate fractal dimensions of corrosion images,” *Pattern Recognition Letters*, vol. 27, pp. 1942–1947, 2006.
- [9] F. Jin and F. Chiang, “Nondestructive Evaluation of Corrosion by Fractal Geometry,” *Research in Nondestructive Evaluation*, vol. 7, no. 4, 1996.
- [10] F. S. J. M. Costa and M. Vilarrasa, “Fractal patterns from corrosion pitting,” *Corrosion Science*, vol. 32, no. 5-6, 1991.
- [11] J. K. A. Hyvärinen and R. Oja, *Independent Component Analysis*. Willey-Interscience, 2001.
- [12] G. F. M.-E. Davoust and J. Oksman, “A Parametric Estimation Approach for Groove Dimensioning Using Remote Field Eddy Current Inspection,” *Research in Nondestructive Evaluation*, vol. 11, pp. 39–57, 1999.