

# Sensitivity Analysis Using a Sparse Grid Surrogate Model in Electromagnetic NDE

Arnold BINGLER<sup>1</sup> and Sándor BILICZ<sup>2</sup>

*Budapest University of Technology and Economics*

**Abstract.** The global sensitivity analysis of electromagnetic nondestructive evaluation (NDE) by means of Sobol' indices are considered in this work. To reduce the computational burden, a sparse grid surrogate model is used. The latter can simply replace the true simulator to some extent, but it can also be used to numerically evaluate the integrals defining the Sobol' indices. In most of the NDE setups, the output is not a scalar quantity but functional data (e.g., a surface scan); a method is presented to take this into account. The sparse grid based sensitivity analysis is compared to classical techniques via examples drawn from electromagnetic NDE.

**Keywords.** sensitivity analysis, Sobol' index, sparse grid

## 1. Introduction

To characterize the uncertainty of the output of a simulation model due to its uncertain input parameters, Sobol' indices are commonly used. However, the calculation of the Sobol' indices can be computationally demanding, especially in the case when a heavy simulator is considered and/or the number of input parameters is high. The application of surrogate models provides a way to reduce this computational burden. Polynomial Chaos Expansion (PCE) supplies an efficient, low-cost technique to compute the Sobol' indices. Recently, the use of sparse grids (SGs) in electromagnetic nondestructive evaluation (ENDE) has been proposed [1]. SG can be used both as an approximation of the model (e.g., for Monte Carlo method) and as a numerical quadrature to evaluate integrals involved in the calculation of Sobol' indices. The surrogate model-based sensitivity analysis in the context of ENDE is considered, e.g., in [2]. The present work aims at (i) comparing different surrogate model based approaches of Sobol' index calculation for ENDE problems and (ii) extending the formulations to the case when the output consists of functional data, being typical in ENDE (e.g., surface scan of an eddy-current probe).

---

<sup>1</sup>E-mail: arni38@sch.bme.hu

<sup>2</sup>Corresponding Author: Sándor Bilicz; E-mail: bilicz@evt.bme.hu

## 2. Sensitivity Analysis

In this section a brief overview is given on the purpose of the sensitivity analysis, particularly with respect to the method of Sobol’ indices as well as the possibilities of its numerical calculation.

Suppose that we have a mathematical model  $y = f(\mathbf{x})$  as a multivariate, square integrable function over a domain defined by a set of independent variables,  $\mathbf{x} = (x_1, \dots, x_n)$ . In general the aim of sensitivity analysis is the quantification of how the uncertainty of the model output is affected by the uncertain input parameters. The *black-box* approach is commonly used, i.e., one relies only on certain number of input samples and the corresponding output ones. This approach provides the flexibility of applications in wide range of industrial and scientific fields.

### 2.1. Sobol’ indices

The technique of Sobol’ indices is a variance-based sensitivity analysis method. After its introduction in [3], by now it has become a widely used approach in several fields. A detailed presentation is given in, e.g., [4]. A brief summary is given below, such that the novelty of the present work can be clearly pointed out.

Without the loss of generality we can assume that all input variables are uniformly and independently distributed in the unit hypercube  $\mathcal{D}_{\mathbf{x}} = [0, 1]^n$ . The key idea is to decompose  $f(\mathbf{x})$  into the sum of subfunctions of increasing dimension, such that

$$f(\mathbf{x}) = f_0 + \sum_{i=1}^n f_i(x_i) + \sum_{1 \leq i < j \leq n} f_{ij}(x_i, x_j) + \dots + f_{1,2,\dots,n}(x_1, x_2, \dots, x_n) \quad (1)$$

with  $f_0$  being the expected value of  $f(\mathbf{x})$  and the integral of the subfunctions with respect to any of their arguments is zero, i.e., the subfunctions are pairwise orthogonal. This Sobol-decomposition is shown to be uniquely exist in [3], and the subfunctions are defined in a recursive manner:

$$f_0 = \int_{\mathcal{D}_{\mathbf{x}}} f(\mathbf{x}) d\mathbf{x} \quad (2a)$$

$$f_i(x_i) = \int_0^1 \dots \int_0^1 f(\mathbf{x}) d\mathbf{x}_{\sim i} - f_0 \quad (2b)$$

$$f_{ij}(x_i, x_j) = \int_0^1 \dots \int_0^1 f(\mathbf{x}) d\mathbf{x}_{\sim i,j} - f_0 - f_i(x_i) - f_j(x_j) \quad (2c)$$

where  $x_{\sim j}$  is a set without  $x_j$ . In addition the variance of  $f(\mathbf{x})$  can be partitioned into the sum of sub-variances:

$$D = \sum_{\mathbf{v} \subseteq \mathbf{x} \setminus \{0\}} D_{\mathbf{v}}, \quad (3)$$

with  $D_{\mathbf{v}}$  denoting the variance of the subfunction described by a group of variables  $\mathbf{v} = (x_{i_1}, x_{i_2}, \dots, x_{i_s})$ . Each sub-variance can be considered as the contribution of a group

of variables to the total variance. The measure of their importance, namely the Sobol' indices thus can be defined as

$$S_{\mathbf{v}} = D_{\mathbf{v}} / D. \quad (4)$$

The sum of all possible indices is 1, providing us a simple interpretation of their meaning. Groups consist of only one variable yield the 1st order indices. The 2nd order ones mean the effects caused by the interaction of two parameters without their 1st order effects; higher order indices can be analogously defined. Traditionally, the 1st order indices are estimated by Monte Carlo formulas:

$$\hat{f}_0 = \sum_{k=1}^M f(\mathbf{x}^{(k)}) \quad (5a)$$

$$\hat{D} = \sum_{k=1}^M f(\mathbf{x}^{(k)})^2 - \hat{f}_0^2 \quad (5b)$$

$$\hat{D}_i = \sum_{k=1}^M f(x_i^{(k)}, \mathbf{x}_{\sim i}^{(k)}) f(x_i^{(k)}, \mathbf{x}'_{\sim i}^{(k)}) - \hat{f}_0^2 \quad (5c)$$

where  $\mathbf{x}^{(k)}$  is the  $k$ -th representation of the  $M$  samples and  $\mathbf{x}'_{\sim i}^{(k)}$  denotes a sample independent from  $\mathbf{x}_{\sim i}^{(k)}$

## 2.2. Extension of Sobol' indices

In the original framework of Sobol' indices, the output is a scalar function and the variables have to be independent. In certain cases this can be sufficient (e.g., scalar-output indices for POD studies), however, several NDE related applications cannot be directly treated due to these limitations. For example, vector output functions occur when a surface scan of impedance variation is considered or geometric constraints apply for the defect parameters, making them dependent on each other. Therefore, the original definition needs to be extended and generalized regarding the above cases, which is the main contribution of this work.

### 2.2.1. Multiple output functions

Supposing we have a multiple output function  $\mathbf{F}(\mathbf{x}) = (F_1(\mathbf{x}), F_2(\mathbf{x}), \dots, F_P(\mathbf{x}))$ , a natural extension of Sobol' indices is taking the average of the previously calculated indices of the component functions, i.e.,

$$S_i^{avg} = \frac{1}{P} \sum_{j=1}^P S_{F_j}^{(i)}. \quad (6)$$

This method has its drawback by not taking into account the vectorial nature of the output and the correlation between the component functions. A more suitable solution requires the definition of global indices to the entire functional output. A method with

this consideration has recently been introduced in [5]. Hereby we give another, simpler approach to extend the 1st order indices, relying on an equivalent stochastic definition:

$$S_i = \text{Var}_{x_i} [\mathbb{E}_{\mathbf{x}_{\sim i}} [f(\mathbf{x})|x_i]] / \text{Var} [f(\mathbf{x})]. \tag{7}$$

In 1 dimension (1-D), variance can be considered as the expected value of a squared error function, defined by the deviation of  $f(\mathbf{x})$  from its own expected value. In  $N$ -D, variance can be defined analogously with the error being a vector and the 2-norm is used as a metric of distance instead of absolute value:

$$D^{vec} = \mathbb{E} [\| \mathbf{h} \|^2] = \mathbb{E} [\| \mathbf{F}(\mathbf{x}) - \mathbf{F}_0 \|^2], \tag{8}$$

where  $\mathbf{F}_0$  denotes the vectorial expected value of  $\mathbf{F}(\mathbf{x})$ . This leads us the definition of vectorial Sobol’ indices:

$$S_i^{vec} = \text{Var}_{x_i} [\mathbb{E}_{\mathbf{x}_{\sim i}} [\mathbf{F}(\mathbf{x})|x_i]] / \text{Var} [\mathbf{F}(\mathbf{x})] = D_i^{vec} / D^{vec}. \tag{9}$$

Monte Carlo estimators are also extended to vectorial outputs as

$$\hat{\mathbf{F}}_0 = \frac{1}{M} \sum_{k=1}^M \mathbf{F}(\mathbf{x}^{(k)}) \tag{10a}$$

$$\hat{D}^{vec} = \frac{1}{M} \sum_{k=1}^M \| \mathbf{F}(\mathbf{x}^{(k)}) \|^2 - \| \hat{\mathbf{F}}_0 \|^2 \tag{10b}$$

$$\hat{D}_i^{vec} = \frac{1}{M} \sum_{k=1}^M \mathbf{F}(x_i^{(k)}, \mathbf{x}_{\sim i}^{(k)}) \bullet \mathbf{F}(x_i^{(k)}, \mathbf{x}'_{\sim i}{}^{(k)}) - \| \hat{\mathbf{F}}_0 \|^2 \tag{10c}$$

with  $\bullet$  denoting the scalar multiplication. Though the squared Euclidean norm fits well for our purpose, we note that this choice is not obvious. In the case of a sparse output signal—which is typical, e.g., when a temporal echo is recorded in an ultrasonic testing method—other norms (such as the maximum norm) might be preferred. One may also consider an appropriate pre-processing of the signal before using the Euclidean norm.

### 2.2.2. Dependent variables

A method of sensitivity analysis over non-rectangular domain has recently been studied in [6]. Herein we introduce a method based on the *Rosenblatt transform* (RT) proposed in [7]. Let us denote with  $\mathbf{x}^*$  a permutation of  $\mathbf{x}$  and with  $F_i(x_i^* | \mathbf{v})$  being the conditional cumulative distribution function of  $x_i^*$  with respect to a subset  $\mathbf{v} \subseteq \mathbf{x}_{\sim i}^*$ . It is known that the joint probability density function of  $\mathbf{x}^*$  can be decomposed as follows:

$$p(\mathbf{x}^*) = p(x_1^*)p(x_2^*|x_1^*)p(x_3^*|x_1^*,x_2^*) \dots p(x_n^*|\mathbf{x}_{\sim n}^*). \tag{11}$$

Based on (11), Rosenblatt transform provides a bijective mapping to the unit hypercube  $U \sim [0, 1]^n$ :

$$F_1(x_1^*) = u_1, \quad F_2(x_2^*|x_1^*) = u_2, \quad \dots, \quad F_n(x_n^*|\mathbf{x}_{\sim n}^*) = u_n. \tag{12}$$

There are  $n!$  different transforms due to the permutations of  $\mathbf{x}$ , however, the results of interest narrow into two special cases concerning  $x_i$ . In the case of  $x_i = x_1^*$ , the transform is performed via the marginal distribution function of  $x_i$ , hence contains the effect from the constraints as well, while in the case of  $x_i = x_n^*$ , these effects are excluded due to the fact that  $F_i$  is conditional to all other variables. Therefore the authors in [7] call the 1st order Sobol' indices of  $x_i$  gained from these special transforms as full and individual indices, respectively. These indices characterize the uncertainty contribution of dependent variables.

### 3. Surrogate Models

Performing sensitivity analysis usually requires the modelling of the examined configuration in many settings of the input parameters in order to get the proper number of samples to Monte Carlo simulation. At each sample the electromagnetic model is evaluated (e.g., finite-element model, integral equation model), which might lead to a very long simulation. To reduce the computational burden due to the "curse of dimensionality", surrogate models are used as low-cost approximations of the true simulator, usually constructed as a linear combination of a set of orthonormal multivariate basis functions. In general a set of orthonormal univariate basis functions  $\Psi(\mathbf{x}) = \{\Psi_1(x), \Psi_2(x), \dots, \Psi_l(x)\}$  is created in the first step, then the multivariate ones are built as their tensor product:

$$\Phi(\mathbf{x}) = \Psi(x_1) \otimes \Psi(x_2) \otimes \dots \otimes \Psi(x_n). \quad (13)$$

This set can be truncated to a set of lower cardinality, yielding the surrogate model as

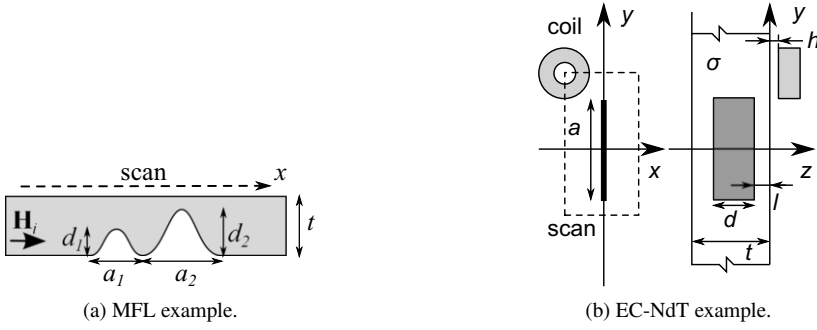
$$f(\mathbf{x}) \approx \hat{f}(\mathbf{x}) = \sum_{i \in \mathcal{A}} \alpha_i \Phi_i(\mathbf{x}) \quad (14)$$

with  $\mathcal{A}$  being an index set. Basically, the differences between the numerous data-fit models are the strategy behind the construction of the univariate basis, the truncation scheme and the estimation of the coefficients. Herein two commonly used methods are briefly summarized: the Polynomial Chaos Expansion (PCE) and the Sparse Grid (SG) interpolation.

*Polynomial Chaos Expansion.* The PCE provides approximation in a stochastic framework of  $f(\mathbf{x})$  by choosing the basis functions to be orthonormal with respect to the joint probability density function of  $\mathbf{x}$ , e.g., Legendre-polynomials for variables uniformly distributed in  $] - 1, 1[$ , Hermite-polynomials for Gaussian distributions, etc. [4] The truncation might be performed by giving a limit on the highest occurring polynomial degree. The coefficients are traditionally calculated from random input and output samples by the ordinary least-squares method:

$$\hat{\alpha} \approx \arg \min \mathbb{E} \left[ \left( f(\mathbf{x}) - \sum_{i \in \mathcal{A}} \alpha_i \Phi_i(\mathbf{x}) \right)^2 \right]. \quad (15)$$

There is strong link between Sobol' decomposition and PCE due to the uniqueness of the former one and the orthogonality of the basis functions. The variance of  $f(\mathbf{x})$  is



**Figure 1.** Sketch of the configurations. MFL parameters:  $a_{1,2} \in [0.6, 1.2]$ mm and  $d_{1,2} \in [0.2, 1.6]$ mm. EC-NdT parameters:  $t = (1.25 \pm 0.01)$ mm,  $\sigma = (1 \pm 0.01)$ MS/m,  $h = (0.5 \pm 0.05)$ mm,  $a \in [2, 10]$ mm,  $d \in [0.125, 1]$ mm and  $l \in [0.125, 1.125]$ mm.

partitioned into the sum of the square of the coefficients, providing a convenient way to evaluate the sub-variances and the Sobol’ indices.

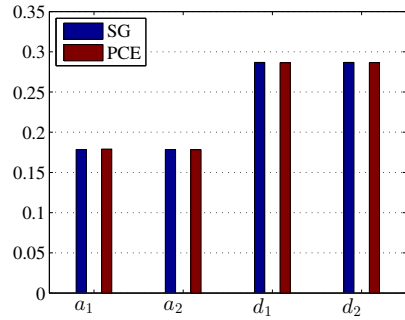
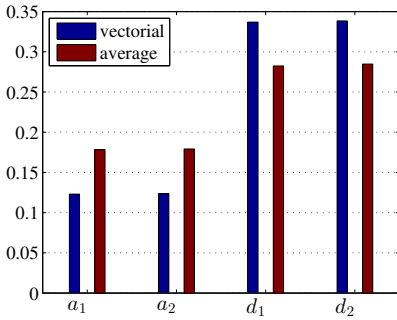
*Sparse grid interpolation.* SG interpolation (detailed in [1], [8]) is based on the evaluation of the original model at specific points, calling them the supporting nodes. The basis functions have a hierarchical structure of level-by-level, each of them belongs to a supporting node  $x_i^{(i)}$ . Their tensor product results in a multivariate basis that can be truncated by, e.g., a linear constraint on the sum of the levels:  $\sum_{i=1}^n l_i = l$ . In the case of linear (“hat”) basis functions, the interpolant at depth  $d$  equals to the sum of interpolant from the previous depth and the linear combination of the multivariate functions at level  $d$ :

$$\hat{f}(\mathbf{x}) \approx \hat{f}_d(\mathbf{x}) = \hat{f}_{d-1}(\mathbf{x}) + \underbrace{\sum_{i=1}^{m_d} \Phi_d^{(i)}(\mathbf{x}) \left[ f(\mathbf{x}_d^{(i)}) - \hat{f}_{d-1}(\mathbf{x}_d^{(i)}) \right]}_{v_d^{(i)}} \quad (16)$$

with  $\hat{f}_0(\mathbf{x}) = \Phi_0^{(1)}(\mathbf{x})f(\mathbf{x}_0^{(1)})$  and  $\mathbf{x}_d^{(i)}$  denoting the vector of 1-D nodes belonging to the 1-D functions of  $\Phi_d^{(i)}$ . The coefficients  $v_d^{(i)}$  are equal to the difference between  $f(\mathbf{x})$  and  $\hat{f}_{d-1}(\mathbf{x})$  at the nodes  $\mathbf{x}_d^{(i)}$ . The SG provides a surrogate model to generate Monte Carlo samples, however, it can be used as numerical quadrature to directly evaluate the integrals in (2), similarly to PCE [9].

#### 4. Numerical examples

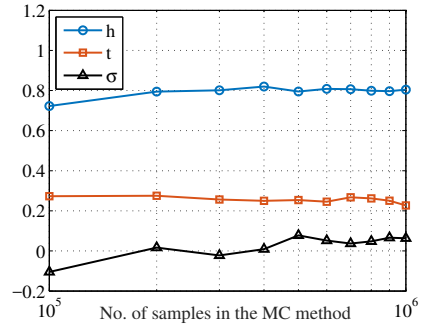
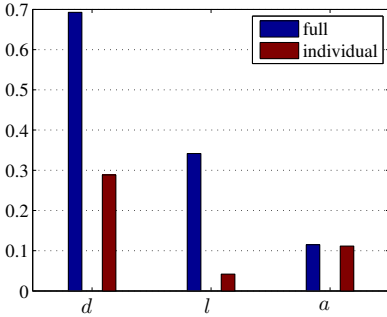
*Magnetic Flux Leakage NdT (4-parameter model).* A ferromagnetic plate of thickness  $t = 2$  mm and  $\mu_r = 100$  is corrupted by grooves infinitely perpendicular to the  $x$ -axis. The grooves are described by 4 parameter-cubic splines, with parameters uniformly distributed as given in the caption of Fig. 1. A homogeneous  $H_{x0} = 1$  A/m magnetic field is imposed in the  $x$  direction. The distortion field ( $\Delta H_x$ ) is measured at 51 equidistant points on the top of the surface. The depths of the grooves were found to be more important parameters compared to their widths (Fig. 2a). Good correlation between the vectorial and



(a) 1st order vectorial and average Sobol' indices.

(b) Comparison of the SG and PCE based direct approaches by the average indices.

Figure 2. Results of the MFL example.



(a) Full and individual 1st order indices of the defect parameters while  $h, t, \sigma$  are kept at their mean value.

(b) Convergence of the setup parameters while  $d, l, a$  are kept at their mean value.

Figure 3. Results of the EC-NdT example.

averaging method can be observed. Direct methods equally resulted in the same outcome as MC-based calculation (Fig. 2b).

*Eddy-Current NdT (3+3 parameter model).* An infinite, non-ferromagnetic, conductive plate with a thickness  $t$  and conductivity  $\sigma$  is investigated as shown in Fig. 1b. The plate includes an ideal crack described by 3 parameters from which  $d$  and  $l$  are dependent ones due to the  $d + l \leq t$  constraint. A coil with time-harmonic excitation of 150 kHz is scanning over the surface at a lift-off  $h$ . The change of its impedance is measured at 297 test points of the grid  $(x, y) \in \{-2 : 0.5 : 2\} \times \{-8 : 0.5 : 8\}$  mm. A single model evaluation needs 15 seconds by the integral equation simulation [10], thus a SG model was built to reduce computation time. A cross-validation was also performed to ensure its accuracy; RMS-error of 4 % with depth=5, below 0.1 % with depth=7 was achieved. Due to the complex nature of the output vector, it had to be transformed to a 594-element real-valued vector as  $\mathbf{F}(\mathbf{x}) \Rightarrow [\text{Re}\{\mathbf{F}(\mathbf{x})\}; \text{Im}\{\mathbf{F}(\mathbf{x})\}]$ . The sensitivity analysis was performed by dividing the parameters into two groups: defect parameters ( $a, d, l$ ) and parameters of the measurement setup ( $h, t, \sigma$ ).

The depth of the crack was repeatedly found to be the most important defect parameter as both its individual and full index are the highest ones (Fig. 3a). The lift-off  $h$  exceeds out of the setup parameters, while the conductivity has almost no effect. The convergence study confirmed that the number of required samples need to be close to the order of millions (Fig. 3b).

## 5. Summary

The proposed extension of the Sobol' indices is found to be an appropriate technique to characterize the effect of uncertain parameters on a complete line/surface scan. To some extent, dependent uncertain input parameters can also be treated. The presented tools are shown to apply well to various NdT examples, the conclusions coincide with the physical expectations. Future work will include further reduction of the required sample number by means of PCE coefficients.

## Acknowledgment

The work was created in commission of the National University of Public Service under the priority project KÖFOP-2.1.2-VEKOP-15-2016-00001 titled "Public Service Development Establishing Good Governance" in the Bay Zoltán Ludovika Workshop. Further support was provided by the Hungarian Scientific Research Fund under grant K-111987 and by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

## References

- [1] S. Bilicz, "Sparse grid surrogate models for electromagnetic problems with many parameters," *IEEE Transactions on Magnetics*, vol. 52, no. 3, pp. 1–4, 2016.
- [2] R. Miorelli, X. Artusi, A. B. Abdesslem, and C. Reboud, "Database generation and exploitation for efficient and intensive simulation studies," in *AIP Conference Proceedings 1706*, 2016.
- [3] I. M. Sobol', "Sensitivity estimates for nonlinear mathematical models," *Mathematical Modelling and Computational Experiments*, vol. 1, no. 4, pp. 407–414, 1993.
- [4] B. Sudret, "Global sensitivity analysis using polynomial chaos expansions," *Reliability Engineering & System Safety*, vol. 93, no. 7, pp. 964–979, 2008.
- [5] F. Gamboa and A. L. Alexandre Janon, Thierry Klein, "Sensitivity indices for multivariate outputs," *Comptes Rendus Mathématique*, vol. 351, no. 7-8, pp. 307–310, 2013.
- [6] S. Kucherenko, O. Klymenko, and N. Shah, "Sobol indices for problems defined in non-rectangular domains," *Reliability Engineering & System Safety*, vol. 167, pp. 218–231, 2017.
- [7] T. A. Mara, S. Tarantola, and P. Annoni, "Non-parametric methods for global sensitivity analysis of model output with dependent inputs," *Environmental Modelling and Software, Elsevier*, vol. 72, pp. 173–183, 2015.
- [8] H.-J. Bungartz and M. Griebel, "Sparse grids," *Acta Numerica*, vol. 13, pp. 147–269, 2004.
- [9] G. T. Buzzard, "Global sensitivity analysis using sparse grid interpolation and polynomial chaos," *Reliability Engineering & System Safety*, vol. 107, pp. 82–89, 2012.
- [10] J. Pávó and D. Lesselier, "Calculation of eddy current testing probe signal with global approximation," *IEEE Transactions on Magnetics*, vol. 42, no. 4, pp. 1419–1422, 2006.