Available online at www.eccomasproceedia.org
Eccomas Proceedia UNCECOMP (2017) 372-380

ECCOMAS
Proceedia

UNCECOMP 2017
2nd ECCOMAS Thematic Conference on
Uncertainty Quantification in Computational Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou (eds.)
Rhodes Island, Greece, 15–17 June 2017

# PARAMETER IDENTIFIABILITY THROUGH INFORMATION THEORY

**G. Capellari**[1], **E. Chatzi**[2], **and S. Mariani**[1]

[1] Politecnico di Milano, Dipartimento di Ingegneria Civile e Ambientale
Piazza Leonardo da Vinci, 32, 20133 Milano (IT)
e-mail: {giovanni.capellari, stefano.mariani}@polimi.it

[2] ETH Zürich, Institut für Baustatik und Konstruktion
Stefano-Franscini-Platz, 5, 8093 Zürich (CH)
e-mail: chatzi@ibk.baug.ethz.ch

**Keywords:** Identifiability, Information Theory, Parameter Estimation, Uncertainty Quantification.

**Abstract.** *In this paper, we address the problem of assessing the identifiability of model parameters in a mechanical system, i.e., whether unknown parameters can be estimated given a set of measurements collected through sensor networks. Practical identifiability can arise due to either a lack of sensitivity or a joint effect of the parameters on the measurements. Information theory can be used to detect the sources of non-identifiability, with the purpose of establishing an efficient sensor network design. Mutual Information between the parameter and the measured outputs, and Conditional Mutual Information of each parameter couple, conditioned on the measurements, are considered. Adoption of these indices is overviewed for practically assessing the identifiability of the mechanical properties of a non-linear structural model.*


Peer-review under responsibility of the organizing committee of UNCECOMP 2017.
doi: 10.7712/120217.5376.17179

# 1 INTRODUCTION

In designing structural health monitoring (SHM) systems, the identifiability and observability of the quantities to be estimated ought to be taken into account [1]. Any set of model parameters is deemed as identifiable if it is possible to uniquely estimate this on the basis of the measured data. The concept of identifiability in mechanical systems has been defined in [2] and [3]. While model identifiability is related to the model mathematical structure only, practical identifiability is referred to the relation between measurements and estimated parameters. Therefore, unlike its purely theoretical counterpart, practical identifiability does take into account uncertainties in the considered variables. In [4], a thorough review of existing methods to study structural identifiability and observability for nonlinear structural models is provided. Practical identifiability is usually addressed through the Fisher information matrix, i.e., the sensitivity of the measured quantities with respect to the parameters [5]. These methods allow to study the identifiability of a certain set of parameters, but they do not highlight the relations among the parameters and the link between each of them and the measured data. By means of such methods, non-identifiability may basically be attributed to either the lack of sensitivity of measured quantities or the compensation of the effects of the parameters on the model response. Information theory has been employed in [6] and [7] to study the relations among parameters; in [8], the concepts of stability and observability have been highlighted within an information theory approach. In this paper, we propose to employ mutual information and conditional mutual information to address the two aforementioned causes of non-identifiability.
In the remainder of this paper, a brief review of information measures and associated definitions is first provided (Section 2.1). Then, the identifiability problem is discussed (Section 2.2) and the employed numerical methods are described (Section 2.3). Finally, the application of the mentioned methods to a non-linear system is presented (Section 3).

# 2 IDENTIFIABILITY AND INFORMATION THEORY

## 2.1 Preliminary definitions

In the present section, only the main definitions of the information measures that will be used hereinafter are briefly revisited.
The Mutual Information (MI) between two random variables $X$ and $Y$ is defined as:

$$I(X;Y) = \int_X \int_Y p(x,y) \log \left[ \frac{p(x,y)}{p(x)p(y)} \right] dxdy \tag{1}$$

where $p(x, y)$ is the joint probability distribution function, while $p(x)$ and $p(y)$ denote marginals. The mutual information may be also interpreted as the Kullback-Leibler divergence (expressed as $D_{KL}[\cdot||\cdot]$) of the product of the marginal distributions from the joint distributions:

$$I(X;Y) = D_{KL}[p(x,y)||p(x)p(y)] \tag{2}$$

In this sense, the MI gives a measure of the difference (or similarity) in information between the joint probability distribution and the marginals, i.e., the degree of correlation between $X$ and $Y$. If $X$ and $Y$ are independent, then $p(x, y) = p(x)p(y)$ and therefore $I(X;Y) = 0$.
A slightly different interpretation relies on the definition of Shannon entropy, according to:

$$I(X;Y) = H(X) - H(X|Y) \tag{3}$$

where $H(X)$ is the Shannon entropy of $p(x)$ and $H(X|Y)$ is the conditional Shannon entropy of $X$, conditioned on $Y$. Within the Bayesian inference framework, by considering $X$ as the

parameter variable and $Y$ as the measurement variable, $I(X;Y)$ may be interpreted as the increase in information between the prior distribution $p(x)$ and the posterior distribution $p(x|y)$; in other words, the MI quantifies how informative the measured data is with respect to the parameters to be estimated.
Let us consider now three random variables $X$, $Y$ and $Z$. The Conditional Mutual Information (CMI) is defined as follows:

$$I(X;Y|Z) = \mathbb{E}_Z[I(X;Y)|Z] = \int_Z p(z) \int_X \int_Y p(x,y|z) \log\left[\frac{p(x,y|z)}{p(x|z)p(y|z)}\right] dxdydz \tag{4}$$

where $\mathbb{E}_Z[\cdot]$ is the conditional expectation with respect to the variable $Z$. The CMI is the variant of the MI when the aforementioned probability distributions are conditioned with respect to the additional random variable $Z$. Analogously to Eq. (2), the CMI can be written as:

$$I(X;Y|Z) = \mathbb{E}_Z[D_{KL}[p(x,y|z)||p(x|z)p(y|z)]] \tag{5}$$

The interaction information is instead defined as:

$$I(X;Y;Z) = I(X;Y) - I(X;Y|Z) \tag{6}$$

Recalling the previous definitions of MI (Eq. (1)) and CMI (Eq. (4)), the interaction information represents the difference between the information shared by $X$ and $Y$ when $Z$ is given, and the same quantity when $Z$ is not given. It can be proved that while $I(X;Y)$ and $I(X;Y|Z)$ are both strictly non-negative quantities, $I(X;Y;Z)$ may assume any value. If $I(X;Y;Z) > 0$, then the random variable $Z$ enhances the correlation between $X$ and $Y$ (redundancy), otherwise it reduces it (synergy). For further details on the interpretation of interaction information, the reader may refer to [9].

### 2.2 Identifiability

Let us consider a structural system, and assume we want to assess its relative mechanical properties through measured data, as obtained from a set of $n_{sens}$ sensors. The vectorial random variable $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{n_{sens}}$ is referred to the measurements, while $\boldsymbol{\Theta} \in \mathbb{R}^{n_{\Theta}}$ represents the $n_{\boldsymbol{\Theta}}$ parameters to be estimated. The relation between parameters and measurements is assumed to be described by the model $\mathcal{M} : \mathbb{R}^{n_{\Theta}} \times \mathbb{R}^{n_{dof}} \longrightarrow \mathbb{R}^{n_{sens}}$ as follows:

$$\mathbf{y} = \mathcal{M}(\boldsymbol{\theta}, \mathbf{f}) + \boldsymbol{\epsilon} \tag{7}$$

where $\mathbf{f} \in \mathbb{R}^{n_{\mathbf{f}}}$ are the model inputs, namely the loads applied on the structure; $\boldsymbol{\epsilon} \in \mathbb{R}^{n_{sens}}$ represents the measurement error and is relevant to the sensors accuracy. We assume $\boldsymbol{\epsilon}$ to be a zero-mean Gaussian noise, sampled from the probability density function $p(\boldsymbol{\epsilon}) = \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the relative covariance matrix.
Model identifiability for structural model updating has been defined in [3]: a set of parameters are said to be identifiable if they can be uniquely determined by the input-output data. In the Bayesian inference framework, this statement holds if the posterior probability distribution $p(\boldsymbol{\theta}|\mathbf{y})$ presents a finite number of maxima in the $\boldsymbol{\Theta}$ space [2]. The parameters are: globally identifiable if there is a unique maximum point $\boldsymbol{\theta}^*$, locally identifiable if there is more than one maximum, and non-identifiable if there the are infinite maxima.
Most of the methods to study practical identifiability are based on measures of the local curvature of the likelihood function in the neighbourhood of $\boldsymbol{\theta}^*$ using , e.g., Hessian or Fisher information

matrix [10, 5] or confidence intervals [11]. These methods can only provide information on the identifiability of all the set of parameters, but not on the relations between them.
Non-identifiability can basically arise from two causes:

(a) Compensation of a parameter by others (also known as collinearity): this happens whenever some parameters have the same effect on the model response. In this case, the parameters are somehow redundant and therefore they may not be separately estimated.

(b) Lack of sensitivity of the measurements to the parameters. For instance, this can occur if the model input $\mathbf{f}$ is such that the measurements do not depend on the parameters, i.e., the parameters are not "activated".

In order to address (a), we can point out that any couple of parameters $\{\Theta^i, \Theta^j\}$ can be considered as a common influence of the measurement $\mathcal{Y}$ if they are highly correlated, having fixed $\mathbf{y}$. Therefore, the higher the difference in information between the conditional joint probability distribution $p(\theta^i, \theta^j|\mathbf{y})$ and the product of the conditional marginal distributions $p(\theta^i|\mathbf{y})p(\theta^j|\mathbf{y})$, the harder it is to estimate these parameters separately. As suggested in [7], a natural measure of the aforementioned occurrence is the CMI $I(\Theta^i; \Theta^j|\mathcal{Y})$ between any parameter couple $\{\Theta^i, \Theta^j\}$: the higher this is, the more correlated the parameters are, given the measurements, and therefore the less they are identifiable together.
On the other hand, the interaction information $I(\Theta^i; \Theta^j; \mathcal{Y})$ can be employed to assess the extent of correlation between the parameters which may be attributed to the measurements or, in other words, if the correlation increases or decreases as the model response is measured. In our applications, we assume that the parameters are not correlated prior to acquiring the measurements. Therefore, the prior joint probability distribution is $p(\theta^i, \theta^j) = p(\theta^i)p(\theta^j)$ and the MI is $I(\Theta^i; \Theta^j) = 0$. From Eq. (6), it turns out that $I(\Theta^i; \Theta^j; \mathcal{Y}) = -I(\Theta^i; \Theta^j|\mathcal{Y})$ and we can conclude that, in this special case, the interaction information yields the same results as the MI. Nevertheless, for a general case, it cannot be used to assess identifiability.
Problem (b) is addressed by exploiting $I(\Theta^i; \mathcal{Y})$, i.e., the MI between each parameter and the measurements. If $I(\Theta^i; \mathcal{Y}) = 0$, then $\mathcal{Y}$ does not depend on $\Theta^i$, and therefore the latter cannot be identified. The dependency of the measurements on the parameters to be estimated can be affected by several factors, such as the model input, the sensor placement and the type of physical quantities to be measured. Regarding the sensor placement, in [5, 12] a method to optimally place them is presented, by maximizing the relative information content with respect to the parameters to be estimated and, hence, reduce problems of non-identifiability.
It is important to underline that the method places no assumptions on the model linearity or on the types of prior distributions.

### 2.3 Numerical solution

If the probability distribution functions defined in the previous sections are not available analytically, the CMI $I(\Theta^i; \Theta^j|\mathcal{Y})$ and the MI $I(\Theta^i; \mathcal{Y})$ cannot be computed analytically from Eqs. (4) and (1), thereby creating the need for an numerical approach. There are essentially three classes of methods for evaluating the MI: methods based on Monte Carlo approximations [13]; Kernel Density Estimation (KDE) based methods [14]; and k-Nearest Neighbors (kNN) based approaches [15].
Here, due to its ease of implementation, we employ the Gaussian KDE method proposed in [16].

The kernel density estimator $\hat{p}(x)$ of the probability density function $p(x)$ is defined as:

$$\hat{p}(x) = \frac{1}{Nh^d} \sum_{i=1}^{N} \mathcal{K}\left(\frac{||x - x_i||}{h}\right) \tag{8}$$

where $\mathcal{K}(\cdot)$ is the Gaussian kernel, $d$ is the dimension of the random variable $X$, $h$ is the kernel bandwidth and $N$ is the sample size. The estimated MI reads:

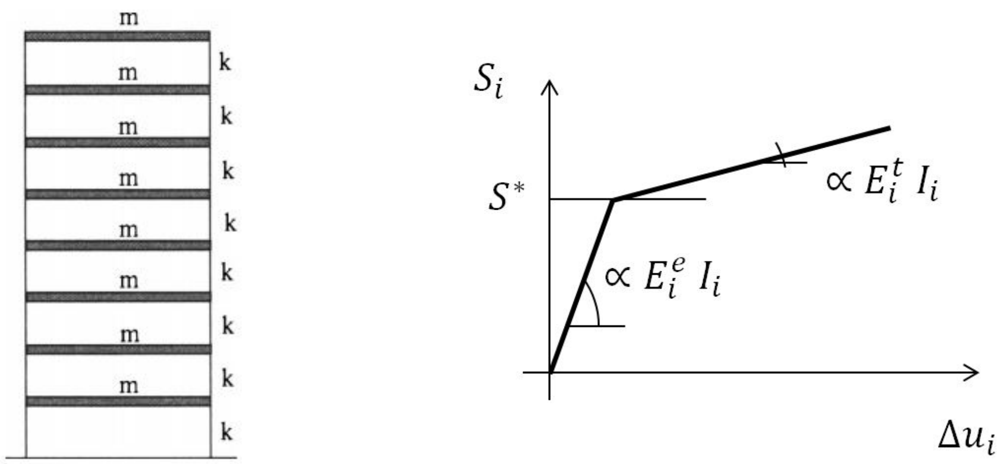
$$\hat{I}(X;Y) = \frac{1}{N} \sum_{i=1}^{N} \log \frac{\hat{p}(x_i, y_i)}{\hat{p}(x_i)\hat{p}(y_i)} \tag{9}$$

where $\hat{p}(x_i, y_i)$, $\hat{p}(x_i)$ and $\hat{p}(y_i)$ are computed according to Eq. (8). In order to reduce the allocated memory required for the computation, an ensemble estimator has been used, as suggested in [17]. The $N$ samples are thus divided into $M$ groups and the MI is simply computed as:

$$\hat{I}(X;Y) = \frac{1}{M} \sum_{i=1}^{M} \hat{I}^i(X;Y) \tag{10}$$

where $\hat{I}^i(X;Y)$ is the MI estimation related to the $i-$th group of samples.
In [13, 12], MI has been computed through a Monte Carlo approximation of Eq. (1). Despite a faster convergence rate, the latter approach is practically unsuitable for the computation of the CMI, because of the high computational cost of the multi-dimensional numerical integration. For this reason, the computation of the CMI is performed through the same KDE approach adopted for the MI.

## 3 APPLICATION TO A STRUCTURAL MODEL



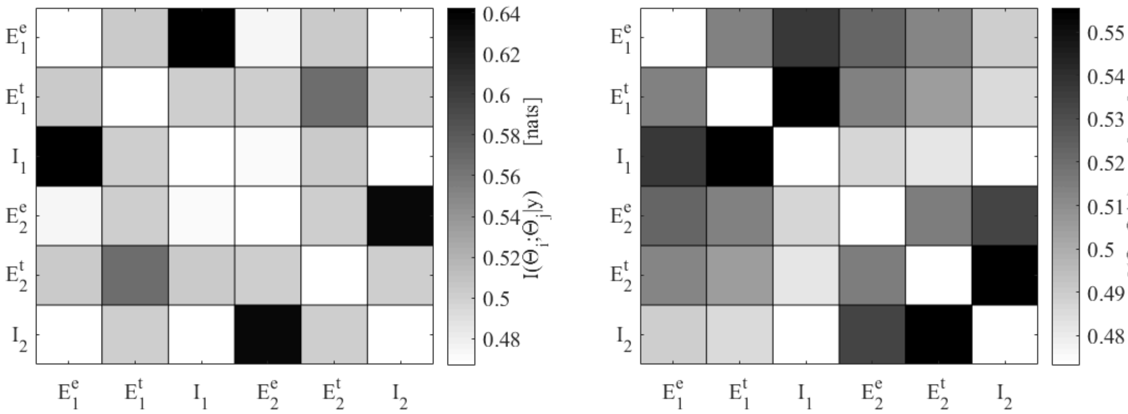

(a) (b)

Figure 1: (a) Shear-type 8-stories building [18]; (b) Relation between non-linear inter-storey drift and shear force, as defined in Eq. (11).

In this section a simple example is chosen so that the practical non-identifiability issues described in Section 2.2 are clearly manifested from the model and, hence, the working of the

| | $I(E_1^e;\mathcal{Y})$ | $I(E_1^t;\mathcal{Y})$ | $I(I_1;\mathcal{Y})$ | $I(E_2^e;\mathcal{Y})$ | $I(E_2^t;\mathcal{Y})$ | $I(I_2;\mathcal{Y})$ |
|---|---|---|---|---|---|---|
| $S_i < S_i^*$ | 0.2280 | 0.0419 | 0.2256 | 0.2240 | 0.0427 | 0.2233 |
| $S_i > S_i^*$ | 0.1040 | 0.1336 | 0.2238 | 0.1052 | 0.1353 | 0.2239 |

Table 1: Mutual Information $I(\Theta_i;\mathcal{Y})$ of each parameter in $\mathbf{\Theta} = [E_1^e; E_1^t; I_1; E_2^e; E_2^t; I_2]$ and the measured top-floor displacement $\mathcal{Y}$, considering the cases $S_i < S_i^*$ and $S_i > S_i^*$.




(a) (b)

Figure 2: Conditional Mutual Information $I(\Theta_i;\Theta_j|\mathcal{Y})$ of each couple of parameters in $\mathbf{\Theta} = [E_1^e; E_1^t; I_1; E_2^e; E_2^t; I_2]$ and the measured top-floor displacement $\mathcal{Y}$, considering the cases (a) $S_i < S_i^*$ (b) and $S_i > S_i^*$ .

method may be checked and validated.

The approach discussed in the previous section is now applied to a shear-type 8-storey building model (Figure 1a). We assume the flexural rigidity of all the horizontal members to be much higher than that of the column elements, so that the only relevant degrees of freedom are the horizontal displacements of each floor. According to this assumption, the floor stiffness at the $i$−th storey is $k = c\frac{12E_iI_i}{h^3}$, where $c$ is the number of columns per floor, $E_i$ is the material elastic modulus, $I_i$ is the moment of inertia in the storey columns and $h$ the floor height.

The inter-story drifts $\Delta u_i$ depend on the shear force $S_i$ according to the following relation (Figure 1b):

$$\Delta u_i = \begin{cases} \frac{h^3}{c12E_i^eI_i}S_i & \text{if } S_i < S_i^* \\ \frac{h^3}{c12E_i^eI_i}S_i^* + \frac{h^3}{c12E_i^tI_i}(S_i - S_i^*) & \text{if } S_i \geq S_i^* \end{cases} \tag{11}$$

where we have assumed a simple bi-linear rule for the modulus $E_i$:

$$E_i = \begin{cases} E_i^e & \text{if } S_i < S_i^* \\ E_i^t & \text{if } S_i \geq S_i^* \end{cases} \tag{12}$$

where $E_i^e$ is the elastic modulus, $E_i^t < E_i^e$ is the tangent modulus and $S_i^*$ is the shear yield capacity.

We assume availability of displacement measurements $u_8$ at the top floor (the corresponding random variable is named as $\mathcal{Y}$) and the aim is to study the identifiability of the parameters relevant to the first two floors, i.e. $\mathbf{\Theta} = [E_1^e; E_1^t; I_1; E_2^e; E_2^t; I_2]$. The parameters are assumed as

uniformly distributed.
First, we consider the case in which $S_i < S_i^* \; \forall i$: from Eq. 11, we can point out that the displacements and, hence, the measurement, do not depend on $E_i^t$. The non-identifiability of $E_i^t$ shows up in the values of the MI $I(\Theta_i;\mathcal{Y})$ in Table 1: $I(E_1^t;\mathcal{Y})$ and $I(E_2^t;\mathcal{Y})$ are one order of magnitude lower than the other MI value. The MI is not exactly zero as one may expect from the definition in Eq. (1), because of the round-off error of the KDE method. The measured displacement depends only on the flexural stiffness $E_i^e I_i$: these two parameters cannot be estimated separately, since they offer a joint influence to the model response. This is highlighted by the CMI values $I(\Theta_i;\Theta_j|\mathcal{Y})$ reported in Figure 2a: as expected, the maximum values are reached for the couples $\{E_1^e, I_1\}$ and $\{E_2^e, I_2\}$.
On the other hand, if $S_i > S_i^* \; \forall i$, the non-linear mechanical behaviour described in Eq. (11) affects the solution. As noted from Table 1, there are no parameters for which $I(\Theta_i;\mathcal{Y}) \approx 0$. However, from Eq. (11), we can point out that the model responses $\Delta u_i$ depends with the same relationship on $E_i^e/E_i^t$ and $I_i$, and therefore this prevents identifiability. The related CMI values stem from the latter fact: in Figure 2b, $I(\Theta_i;\Theta_j|\mathcal{Y})$ is maximum for the couples $\{E_1^e, I_1\}$ and $\{E_2^e, I_2\}$. Moreover, $I(E_{1,2}^t; I_{1,2}|\mathcal{Y}) > I(E_{1,2}^e; I_{1,2}|\mathcal{Y})$ as, since $E_i^t I_i < E_i^e I_i$, the resulting displacement is heavily dependent on $E_i^t$. The same fact can be underlined in Table 1, as $I(E_i^e;\mathcal{Y}) < I(E_i^t;\mathcal{Y})$.

## 4 CONCLUSIONS

In the present paper, we have addressed the problem of parameter-identifiability in mechanical systems. The non-identifiability of parameters can arise due to either lack of sensitivity or compensations in the dependency of measurements on the parameters. Within an information theoretic approach, identifiability is here detected in terms of these two occurrences using the mutual information between each parameter and the measurements, and the conditional mutual information between each couple of parameters, conditioned on the measurements.
The methodology has been applied to a non-linear mechanical model, namely a shear-type 8-storeys building model, where the former causes of non-identifiability are easily recognizable. The MI and the CMI proved able to detect and quantify identifiability, and to reveal the relations between parameters. Moreover, no unnecessary assumptions on the model linearity or on distributions gaussianity were placed.

### Acknowledgements

G. C. wishes to acknowledge the partial financial support by IDEA League through a Student Grant.

## REFERENCES

[1] Manolis Chatzis, Eleni Chatzi, and Savvas Triantafyllou. A discontinuous class of filtering methods for the identification of non-smooth dynamical systems. *Proceedings of the 5th International Symposium on Life-Cycle Civil Engineering (IALCCE 2016), Delft (NL)*, 2016.

[2] K.-V. Yuen. *Bayesian Methods for Structural Dynamics and Civil Engineering*. John Wiley & Sons, Ltd, 2010.

[3] L. S. Katafygiotis and J. L. Beck. Updating models and their uncertainties. ii: Model identifiability. *Journal of Engineering Mechanics*, 124(4):463–467, 1998.

[4] M. N. Chatzis, E. N. Chatzi, and A. W. Smyth. On the observability and identifiability of nonlinear structural and mechanical systems. *Structural Control and Health Monitoring*, 22(3):574–593, 2014.

[5] C. Papadimitriou. Optimal sensor placement methodology for parametric identification of structural systems. *Journal of Sound and Vibration*, 278(4-5):923–947, 2004.

[6] K. Nienaltowski, M. Wlodarczyk, T. Lipniacki, and M. Komorowski. Clustering reveals limits of parameter identifiability in multi-parameter models of biochemical dynamics. *BMC Systems Biology*, 9(65), 2015.

[7] S. Pant and D. Lombardi. An information-theoretic approach to assess practical identifiability of parametric dynamical systems. *Mathematical Biosciences*, 268:66–79, 2015.

[8] M. Kam, R. Cheng, and P. Kalata. An information-theoretic interpretation of stability and observability. *American Control Conference, Minneapolis (US)*, pages 1957–1962, 1987.

[9] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann Publishers Inc., 1988.

[10] J. Jacquez and P. Greif. Numerical parameter identifiability and estimability: integrating identifiability, estimability, and optimal sampling design. *Mathematical Biosciences*, 77:201–228, 1985.

[11] A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, and J. Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929, 2009.

[12] G. Capellari, E. Chatzi, and S. Mariani. An optimal sensor placement method for shm based on bayesian experimental design and polynomial chaos expansion. *Proceedings of the VII European Congress on Computational Methods in Applied Sciences and Engineering, Crete Island (GR)*, 3:6272–6282, 2016.

[13] X. Huan and Y. M. Marzouk. Simulation-based optimal bayesian experimental design for nonlinear systems. *Journal of Computational Physics*, 232(1):288–317, 2013.

[14] Y-I Moon, B. Rajagopalan, and U. Lall. Estimation of mutual information using kernel density estimators. *Physical review. E*, 52(3):2318–2321, 1995.

[15] A. Kraskov, H. Stögbauer, and P. Grassberger. Estimating mutual information. *Physical Review E*, 69(6):066138, 2004.

[16] R. Steuer, J. Kurths, C. O. Daub, J. Weise, and J. Selbig. The mutual information: Detecting and evaluating dependencies between variables. *Bioinformatics*, 18(2):S231–S240, 2002.

[17] J. Kybic. High-dimensional mutual information estimation for image registration. *International Conference on Image Processing ICIP '04, Singapore (SGP)*, pages 1779–1782, 2004.

[18] R. A. De Callafon, B. Moaveni, J. P. Conte, X. He, and E. Udd. General realization algorithm for modal identification of linear dynamic systems. *Journal of Engineering Mechanics*, 134(9):712–722, 2008.