Robustness of the Data-Driven Identification algorithm with incomplete input data

Marie Dalémat, Michel Coret, Adrien Leygue, Erwan Verron

OVERLAY

Journal of Theoretical, Computational and Applied Mechanics

JTCAM

DIAMOND OPEN ACCESS

Identifiers
DOI 10.46298/jtcam.12590
OAI hal-04301210v2

History
Received Feb 04, 2021
Accepted Oct 15, 2021
Published Dec 11, 2023

Associate Editor
Anna Pandolfi

Reviewer
Anonymous

Open Review
OAI hal-03647968

Supplementary Material
Data
DOI
10.5281/zenodo.10090468

Robustness of the Data-Driven Identification algorithm with incomplete input data

Marie Dalémat, Michel Coret, Adrien Leygue, and Erwan Verron

Institut de Recherche en Génie Civil et Mécanique (GeM), UMR CNRS 6183, École Centrale de Nantes, France

Identifying the mechanical response of a material without presupposing any constitutive equation is possible thanks to the Data-Driven Identification algorithm developed by the authors. It allows to measure stresses from displacement fields and forces applied to a given structure; the peculiarity of the technique is the absence of underlying constitutive equation. In the case of real experiments, the algorithm has been successfully applied on a perforated elastomer sheet deformed under large strain. Displacements are gathered with Digital Image Correlation and net forces with a load cell. However, those real data are incomplete for two reasons: some displacement values, close to the edges or in a noise-affected area, are missing and the force information is incomplete with respect to the original DDI algorithm requirements. The present study proves that with appropriate data handling, stress fields can be identified in a robust manner. The solution relies on recovering those missing data in a way that no assumption, except the balance of linear momentum, has to be made. The influence of input parameters of the method is also discussed. The overall study is conducted on synthetic data: perfect and incomplete data are used to prove robustness of the proposed solutions. Therefore, the paper can be considered as a practical guide for implementing the DDI method.

Keywords: Data Driven Identification, Digital Image Correlation, incomplete data, stress measurement

1 Introduction

Constitutive equations are historically essential in Mechanics of Materials to perform analytical or numerical calculations: they close the problem when combined with mechanical equilibrium. In practice, the identification procedure consists in choosing or deriving a constitutive model that describes well the material response. Then, the calibration of the model parameters has to be done ideally considering multiple deformation states (uniaxial tension, pure shear, biaxial tension...), which renders difficult the experimental process. These steps are often done iteratively to end up with a robust and well-calibrated constitutive model. With the evolution of full-field measurement techniques such as Digital Image Correlation (Sutton et al. 2009), identification methods are constantly being improved, specifically with non-standards tests. For example, Avril et al. (2008) and Roux and Hild (2020) proposed an overview of identification techniques such as the Virtual Fields Method or the Finite Element Model Updating Method. Concomitantly, with the emergence of Data Sciences, other methods are proposed: for example, Furukawa and Yagawa (1998) and Yang et al. (2019) trained a neural network for identification purposes.

Here, a new path is chosen: identifying the material response with no underlying constitutive equation. Indeed, it is possible to use the previous techniques (full-field methods and Data Sciences) to create rich databases that can be used for identification but also for simulation. It overcomes the difficulties in getting a robust identification of the model parameters. This has been introduced in (Kirchdoerfer and Ortiz 2016) where the constitutive equation is replaced by a discrete database of strain-stress couples. The corresponding approach is referred to as Data-Driven Computational Mechanics (DDCM). Slightly different formulations of this solver are proposed in (Ayensa-Jiménez et al. 2018; Kanno 2018; Kirchdoerfer and Ortiz 2017; Nguyen and

Keip 2018) and several extensions are discussed in (Kirchdoerfer and Ortiz 2018; Conti et al. 2018; Eggersmann et al. 2019).

Concerning material characterization, non-parametric approaches are proposed in (Latorre and Montáns 2020; Crespo and Montáns 2019) in which the strain energy function of a hyperelastic material is not presupposed but expressed with splines. In (Amores et al. 2020) splines are further used to build a structure-based non-parametric constitutive manifold of the material, using simple experimental tests which explicitly provide the stress values. Furthermore, it is possible to account for the thermodynamic consistency of the data-driven procedure through well established formalism as proposed by González et al. (2019).

For more complex testing conditions, the stress field is heterogeneous and cannot be obtained in a straightforward manner. In (Réthoré et al. 2018), a decomposition of the strain field obtained with Digital Image Correlation is made in order to compute the stress field without constitutive equation. In (Seghir and Pierron 2018), experimental dynamic measurements are used in the balance equations so that stress fields can be directly computed. Additionally, several manifold learning approaches have been proposed and validated on synthetic data to identify a material constitutive manifold, see for example (Ibañez et al. 2017; Kanno 2021).

In the present paper, a specific algorithm called Data Driven Identification (DDI) is considered; it has been recently proposed in (Leygue et al. 2018). It allows to identify heterogeneous stress fields from measured displacement fields and external forces, without constitutive equation. It relies on the availability of heterogeneous and rich data which can be smartly clustered so that a strain-stress database is built without constitutive equations. It has been validated with synthetic data (Leygue et al. 2018) and its application to real data has been recently assessed (Dalémat et al. 2019). It is an innovative tool to measure stress fields, from DIC gathered displacement fields and net forces measured by load cells.

The difficulty in applying the DDI algorithm to real data lies mainly in the incompleteness and noisiness of data in some areas of the samples. Indeed, unlike a synthetic problem where everything is perfectly known, neither all forces nor all displacements can be perfectly measured; these difficulties are overcome by making *preprocessing choices* (both on the two intrinsic parameters of the algorithm and on the experimental input data). The so-called preprocessing step transforms raw input data to well-conditioned input data with consistent parameters for the DDI algorithm. The present work demonstrates the robustness of the DDI algorithm when applied to incomplete data: several possible preprocessing choices are compared so that the proper one can be applied with confidence. It is to note that although the discussion is here illustrated on a single non-linear hyperelastic case study, it stems from the experience accumulated in applying DDI to many cases involving synthetic and real data, and linear and non-linear material behaviors (Leygue et al. 2018; Dalémat et al. 2019; Stainier et al. 2019).

The paper is organized as follows. First, a brief recall of the algorithm is proposed in order to highlight its optimal parameters and input data. Then, a case study is built to study several preprocessing choices. Synthetic data are considered for which the reference stress response is known. These synthetic data are modified to simulate incomplete data representative of reality. Then, a parametric study is conducted to find the proper preprocessing choice. It focuses on:

(i) the *intrinsic parameters* (of the algorithm) when using the DDI with perfect data;
(ii) the preprocessing step for *incomplete input data* (missing displacements and forces);

Finally, the proper preprocessing choice is summarized so that the DDI method can be applied with confidence on real (*i.e.* partial and noisy) data. It gives the reader the possibility to implement him/herself the DDI method for real data.

2 Recall of the Data Driven Identification algorithm

This section is recalling the DDI algorithm so that its optimal parameters and input data are highlighted. The Data Driven Identification (DDI) (Leygue et al. 2018) corresponds to the inverse method of Data Driven Computational Mechanics (DDCM) derived in (Kirchdoerfer and Ortiz 2016). This method identifies the complete response of a structure without using constitutive equation, from a large database.

2.1 Input data

We consider a 2D-meshed geometry, deformed over N_X increments indexed by X. For this geometry, the following data are the inputs of the algorithm and are considered to be available:

(I-1) the nodal displacements $\mathbf{u}_j^X$, j being the node number. The strain derived from the displacements is the Hencky true strain tensor $\ln \mathbf{v}$. It is defined from $\mathbf{b}$, the left Cauchy-Green strain tensor, by $2\ln \mathbf{v} = \ln \mathbf{b}$ with $\mathbf{b} = \mathbf{F}\mathbf{F}^\top$, $\mathbf{F}$ being the deformation gradient tensor. In practice, a Digital Image Correlation software provides displacement fields on a grid. A mesh with associated connectivity is built from it to compute $\mathbf{F}$,

(I-2) the matrix $\mathbf{B}_{ej}^X$ which encodes both geometry and connectivity, e being the quadrature point number. In particular, the mechanical balance can be evaluated at all nodes by:

$$\sum_e w_e^X \mathbf{B}_{ej}^X \cdot \boldsymbol{\sigma}_e^X = \mathbf{f}_j^X \quad \forall X, j, \tag{1}$$

where w_e^X is the integration weight of point e at loading step X,

(I-3) the nodal forces $\mathbf{f}_j^X$. These are zero in the absence of body forces, excepted for boundary nodes.

Additionally, the method has two intrinsic parameters:

(Inp-1) the size N^* of the (stress-strain) database that samples the material response,

(Inp-2) the positive definite tensor $\mathbb{C}$ that defines the distance between two points in the phase space (here, the stress-strain space).

2.2 Output of the method

After convergence is achieved, the N_X mechanical problems are solved and the method provides:

(O-1) the stress fields $\boldsymbol{\sigma}_e^X$ that satisfy the mechanical balance in each node j according to Equation (1). The stress $\boldsymbol{\sigma}_e^X$ (calculated) and the strain $\ln \mathbf{v}_e^X$ (measured) are referred to as a *mechanical state*, as they are mechanically admissible (balanced and compatible),

(O-2) the N^* material states $(\ln \mathbf{v}_i^*, \boldsymbol{\sigma}_i^*)$, N^* being chosen by the user (Inp-1). These material states can be interpreted as a sampling of the material strain-stress response surface. Their distance from mechanical states is defined by a norm $||\cdot||_{\mathbb{C}}^2$ defined in Equation (2) where $\mathbb{C}$ is a fourth order positive definite tensor also chosen by the user (Inp-2).

2.3 Solver

The algorithm aims at finding material states that are as close as possible to statically and kinematically admissible mechanical states (the latter being half known: the strain field is known, the stress field not), according to the norm $||\cdot||_{\mathbb{C}}^2$ defined by:

$$||(\ln \mathbf{v}, \boldsymbol{\sigma})||_{\mathbb{C}}^2 = \frac{1}{2}(\ln \mathbf{v} : \mathbb{C} : \ln \mathbf{v} + \boldsymbol{\sigma} : \mathbb{C}^{-1} : \boldsymbol{\sigma}). \tag{2}$$

Although this norm has the form and the units of an energy density through $\mathbb{C}$ [Pa] it is not related to any actual energy in the system: the magnitude of $\mathbb{C}$ simply allows to weight the respective contributions of strain and stress.

The problem is formulated as follows:

$$\text{solution} = \underset{\boldsymbol{\sigma}_e^X, (\ln \mathbf{v}_i^*, \boldsymbol{\sigma}_i^*)}{\arg\min} \; \mathcal{E}(\boldsymbol{\sigma}_e^X, \ln \mathbf{v}_e^{*X}, \boldsymbol{\sigma}_e^{*X}), \tag{3}$$

with

$$\mathcal{E}(\boldsymbol{\sigma}_e^X, \ln \mathbf{v}_{e^*}^X, \boldsymbol{\sigma}_{e^*}^X) = \sum_{X'} \sum_{e'} w_{e'}^{X'} ||(\ln \mathbf{v}_{e'}^{X'} - \ln \mathbf{v}_{e'^*}^{X'}, \boldsymbol{\sigma}_{e'}^{X'} - \boldsymbol{\sigma}_{e'^*}^{X'})||_{\mathbb{C}}^2, \tag{4}$$

and subject to the constraints:

- satisfy Equation (1),
- material state $(\ln \mathbf{v}_e^{*X}, \boldsymbol{\sigma}_e^{*X})$ associated to the element e of increment X belongs to the database $(\ln \mathbf{v}_i^*, \boldsymbol{\sigma}_i^*)_{i=1}^{N^*}$.

Therefore, the DDI outputs are:

- the mechanical states,
- the database of material states, and
- the mapping between mechanical and material states.

In (Leygue et al. 2018), the validity of the method has been demonstrated with perfect synthetic data, from (I-1) to (I-3).

In the experimental validation (Dalémat et al. 2019), the algorithm has been applied with incomplete data that are well-preprocessed. The purpose of this paper is to carefully study the preprocessing choices and their influence on the robustness of the algorithm.

3 Building the case study

In this section, a case study is developed to study several preprocessing choices. First, features of usual real data are presented then several preprocessing options are proposed, with a focus on missing data. Finally, the methodology for the next section is summarized.

3.1 From idealized to realistic input data

Experimental data might have missing information and can be noisy. Here, the construction of the actual realistic problem from perfect synthetic data is explained thanks to a 2D example. Noise on the displacement field measurements has to be taken into account. The typical noise in DIC is considered to have an amplitude of the order of 1 pixel, independently of the measured displacement. The discussion of the effect of noisy displacement values is beyond the scope of this paper and has already been partially addressed in the original DDI publication (Leygue et al. 2018).

3.1.1 Why are some data missing?

The DDI method is applied to a perforated hyperelastic membrane subjected to uniaxial tension. The mechanical problem and the different notations are provided in Figure 1(a).

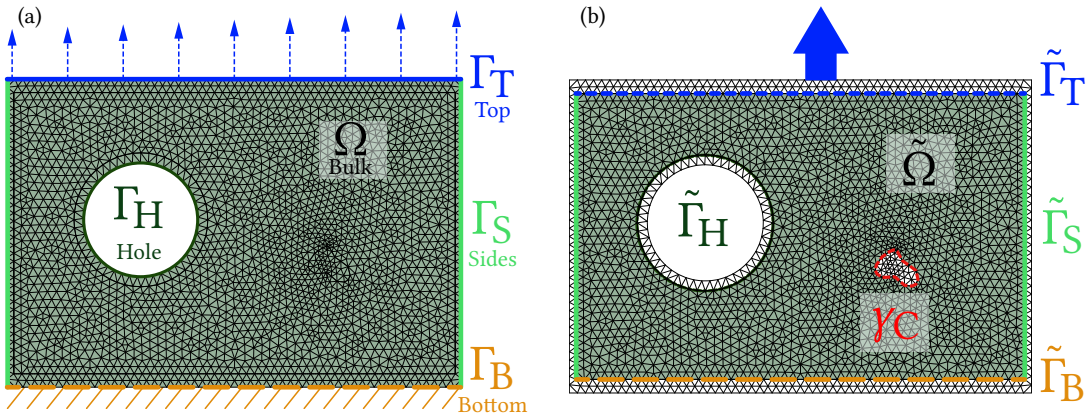

Figure 1 Problem formulation: (a) theoretical and (b) real problems with three particular modified boundaries: top boundary with grip, *cluster* of missing data and imperfectly defined edges close to hole.

Synthetic data are rendered incomplete according to usual experimental constraints:

- we cannot measure the nodal forces but a net force,
- displacements are sometimes missing in areas called *clusters* (which are larger than just a few pixels): the latter are the DIC results when using a software that does not provide the considered unreliable displacements (due to large strain, noise or loss of speckles for example),
- displacements are also missing close to edges: both the camera and the DIC software which works on a manually preselected region cannot resolve the edges of the part. In addition, most correlation software use rectangular patterns that cannot account for curved edges.

The mechanical problem with real boundaries is thus depicted in Figure 1(b). Also, in Figure 1, the theoretical boundaries are the top boundary (Γ_T) where the force is applied, the sides boundaries (Γ_S) that are free edges, the displacement-free bottom boundary (Γ_B) which is clamped

and the hole boundary which is stress-free (Γ_H). In the real problem, all the boundaries are close but not exactly identical to the actual ones. They are noted $\tilde{\Gamma}_T$ for the top, $\tilde{\Gamma}_S$ for the sides, $\tilde{\Gamma}_B$ for the bottom and $\tilde{\Gamma}_H$ for the hole boundaries. Plus, the cluster of missing data is defined by its boundary denoted γ_C.

3.2 Possible preprocessing options for missing data

The preprocessing choices concern both the intrinsic parameters of the DDI method and the way of dealing with raw data. First, the preprocessing choices regarding the missing data are detailed.

With such experimental data, it is necessary to rewrite some equations of the initial DDI algorithm given in Section 2. Indeed, handling properly the areas where data are missing is fundamental to insure robustness. Several possibilities are proposed to deal with the missing data:

- In the area near the grip, where the force is measured (in the following, the "t-" stands for the top boundary):

(t-0) In the synthetic case, we know each nodal force $\mathbf{f}_j^X$ at the top boundary Γ_T;

(t-1) Using a load cell, only the sum of the forces $\mathbf{f}_j^X$ on the top boundary Γ_T in the loading direction $\mathbf{n}_{sol}$ is known:

$$\sum_{j\in\Gamma_T} \mathbf{f}_j^X \cdot \mathbf{n}_{sol} = f_{cell}^X \quad \forall X. \tag{5}$$

It is thus possible to define a global equilibrium condition on the boundary, by combining Equation (1) and Equation (5):

$$\sum_{j\in\Gamma_T} \sum_e w_e^X \mathbf{B}_{ej}^X \cdot \boldsymbol{\sigma}_e^X \cdot \mathbf{n}_{sol} = f_{cell}^X \quad \forall X. \tag{6}$$

(t-2) In the real case, displacements close to the grips are not measured and the true boundary cannot be considered in the algorithm. Thus, the boundary Γ_T cannot be considered and is replaced by $\tilde{\Gamma}_T$. To deal with the force information, the simplest solution is to assume that Equation (6) applies also on $\tilde{\Gamma}_T$ as follow:

$$\sum_{j\in\tilde{\Gamma}_T} \sum_e w_e^X \mathbf{B}_{ej}^X \cdot \boldsymbol{\sigma}_e^X \cdot \mathbf{n}_{sol} = f_{cell}^X \quad \forall X. \tag{7}$$

- For *clusters* of missing displacement values, the objective function Equation (4) cannot be evaluated in some elements which should be removed from the problem along with associated nodes (in the following, the "c-" stands for clusters).

(c-1) A simple and naive solution is to simply discard the equilibrium constraint for these nodes.

(c-2) Another solution is to consider that the boundary of *clusters* is the boundary of a mechanically balanced subset. Indeed, a global balance condition is prescribed on the boundary γ_C. This is equivalent to consider a zero net force on this boundary. This can be easily explained by the Ostrogradsky-Gauss theorem in the continuous formulation:

$$\int_{\Omega_C} \mathrm{div}\boldsymbol{\sigma}\, \mathrm{dV} = \int_{\gamma_C} \boldsymbol{\sigma}\cdot\mathbf{n}\, \mathrm{dS} = \int_{\gamma_C} \mathbf{f}\, \mathrm{dS}, \tag{8}$$

which gives, for the discrete formulation:

$$\sum_{j\in\gamma_C} \sum_e w_e^X \mathbf{B}_{ej}^X \cdot \boldsymbol{\sigma}_e^X = \mathbf{0}. \tag{9}$$

- For edges close to holes, the perfect case is the one where the mesh boundary coincides with the real edge of the hole and the free edge condition applies. It is denoted (h-o) and will be the reference case (in the following, the "h-" stands for the hole boundary). In the real case, due to the imperfect edge definition, the displacement values in the vicinity of holes edges are not known. Therefore, the data on the real boundary Γ_H are not known and $\tilde{\Gamma}_H$ must be considered instead. On this boundary, several assumptions can be made:

(h-1) The free edge assumption can be adopted if we consider that $\tilde{\Gamma}_H$ is really close to Γ_H so the edge is free. This incorrect assumption is likely to introduce a bias in the predictions.

(h-2) A weaker assumption consists in applying a zero net force on this boundary. It is verified as the missing matter should be mechanically balanced (like in (c-2)):

$$\sum_{j\in\tilde{\Gamma}_H}\sum_{e} w_e^X \mathbf{B}_{ej}^X \cdot \boldsymbol{\sigma}_e^X = \mathbf{0}. \tag{10}$$

These strategies to deal with missing data are summarized in Figure 2.

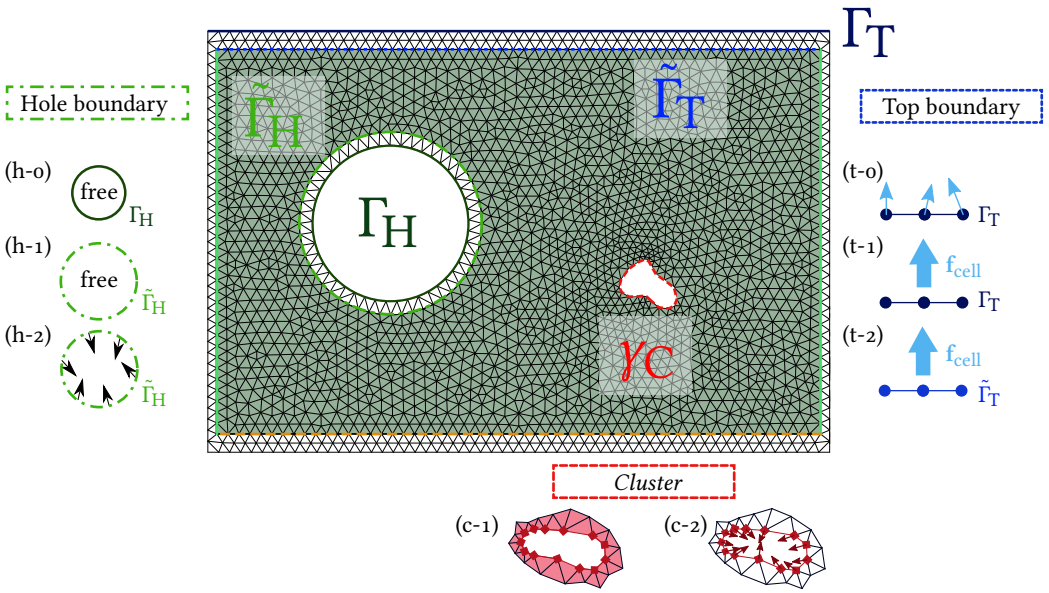

Figure 2 Summary of the possible preprocessing choices. They concern missing data for three particular boundaries: top boundary with grip, *cluster* of missing data and imperfectly defined edges close to hole.

3.3 Methodology in practice

3.3.1 Inputs and parameters

The methodology to investigate the robustness of the DDI is to compare several cases of synthetic input data that are deteriorated on purpose. The cases are studied with respect to the intrinsic parameters of the DDI algorithm. As a recall, the inputs of the algorithm are:

- the algorithm parameters (intrinsic to the resolution method): N^* and $\mathbb{C}$,
- the measured data, especially displacements and forces, which can be incomplete.

Therefore, the discussion is organized as follows:

1. First, the effects of intrinsic parameters on a case where the input data are prefect are analyzed;
2. Second, the influence of the incomplete measured data is analyzed: the cases of (t-o), (t-1) and (t-2) related to the top grip are compared, the cases of (c-1) and (c-2) related to the *clusters* of missing data are discussed, and finally the cases of (h-o), (h-1) and (h-2) related to the edges close to the holes are considered.

3.3.2 Reference model

It is necessary to build synthetic data for which the reference response is known. Thus, a standard Finite Element model (made with the software Abaqus™) is used. The geometry is indicated in Figure 3 where both the initial and deformed meshes of the problem are presented. The initial height is denoted h_0.

The Ogden model (Ogden 1972) is chosen with the corresponding strain energy density

$$W(\lambda_1,\lambda_2,\lambda_3) = \sum_{i=1}^{n} \frac{\mu_i}{\alpha_i}\left(\lambda_1^{\alpha_i}+\lambda_2^{\alpha_i}+\lambda_3^{\alpha_i}-3\right). \tag{11}$$

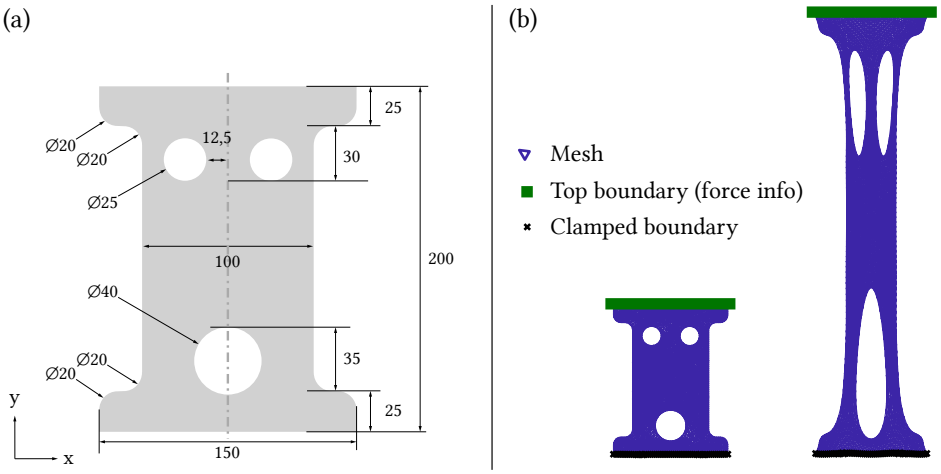

Figure 3 Case study of a perforated hyperelastic membrane under uniaxial tension: (a) initial geometry, (b) mesh after 200 % of total macroscopic strain. On the top nodes, information on force is available while the bottom ones are clamped.

The parameters are listed in Table 1. They are identified in (Ogden 1972) to fit experimental data of Treloar 1944. $N_e = 6108$ linear triangular finite elements under plane stress condition are chosen. The displacements are prescribed using a (x, y) coordinate system corresponding to the horizontal and vertical directions, respectively. They are given for the top and bottom boundaries (denoted Γ_T and Γ_B) by

$$\begin{cases} u_x = 0 \text{ and } u_y = 2h_0 & \text{on } \Gamma_T \\ u_x = 0 \text{ and } u_y = 0 & \text{on } \Gamma_B. \end{cases} \tag{12}$$

The finite element computation is decomposed into $N_X = 21$ increments under quasi-static loading conditions. It gives the reference stresses in each element denoted $\boldsymbol{\sigma}_{\text{FE}}$. The strain fields, meshes and loading conditions are used as inputs in the DDI algorithm with the preprocessing choices introduced in the previous section, resulting in an identified stress field, denoted $\boldsymbol{\sigma}_{\text{DDI}}$.

Table 1 Ogden parameters to build the reference solution (Ogden 1972).

Coefficient	Value	Units
μ_1	$6.18 \cdot 10^5$	Pa
μ_2	$1.18 \cdot 10^3$	Pa
μ_3	$-9.81 \cdot 10^3$	Pa
α_1	1.3	–
α_2	5.0	–
α_3	−2.0	–

3.3.3 Error in stress identification

As the purpose of the DDI is to measure stress field without constitutive equation, the global error between the stress field identified by the DDI $\boldsymbol{\sigma}_{\text{DDI}}$ and the reference one $\boldsymbol{\sigma}_{\text{FE}}$ is computed for all loading increments X and all elements e by

$$\text{e} = \frac{\sum\limits_{X,e} ||\boldsymbol{\sigma}_{\text{FE},e^X} - \boldsymbol{\sigma}_{\text{DDI},e^X}||_2}{\sum\limits_{X,e} ||\boldsymbol{\sigma}_{\text{FE},e^X}||_2}. \tag{13}$$

4 Results and discussion

This section presents the results obtained by comparing several cases of incomplete data. The aim is to determine the proper preprocessing choices that ensure robustness and reliability for stress identification. The influence of the intrinsic parameters of the DDI algorithm is first discussed with perfect input data. Then, the incompleteness of input data and the preprocessing choices associated are discussed.

4.1 Influence of intrinsic parameters

The number of material states N^* is the parameter that allows to sample more or less finely the response of the material. It is to be compared to the total number of degrees of freedom of the problem: $N_e \times N_X = 128\,268$. We define the sampling ratio $r^* = (N_e \times N_X)/N^*$ and consider that it varies between 2 and 10^4.

The distance to mechanical states $(\ln \mathbf{v}_e^X, \boldsymbol{\sigma}_e^X)$ is defined by the norm $||\cdot||^2_{\mathbb{C}}$ of Equation (2). The simplest form for the tensor $\mathbb{C}$ is spherical with an amplitude C, that is $\mathbb{C} = C\mathbb{I}$ where $\mathbb{I}$ is the fourth-order identity tensor. This form aims to equally weight all components of the strain and stress fields. The tensor C is defined accordingly to a pseudo-tangent elasticity modulus of the behavior model used in finite element analyses: $C_0 = 2.3 \times 10^6$ Pa. It is computed by the slope of the straight line found by the least mean square method in the $(||\ln \mathbf{v}||_{\text{VM}}, ||\boldsymbol{\sigma}||_{\text{VM}})$ space (von Mises norm). Practically, we choose values of C ranging from $10^{-6}C_0$ to $10^6 C_0$.

Figure 4 shows the identification error after convergence. Figure 4(a) presents the error as a function of the sampling ratio r^* for different values of C. For each C value, the minimum error (with respect to r^*) is reported. Then, the minimum error (for the optimal value of r^*) in relation to C/C_0 is shown in Figure 4(b). The error is minimal for $r^* \approx 20$ (10 to 50 depending on the C value). A large ratio (not enough material states) implies a sub-sampling of the response and therefore a significant error. Conversely, a too small ratio (too many material states) does not provide enough regularization to the stress estimation problem as the behavior is no longer averaged sufficiently, which also leads to a significant error. It is therefore necessary to choose a value between these two extrema; similar results are reported in (Leygue et al. 2018).

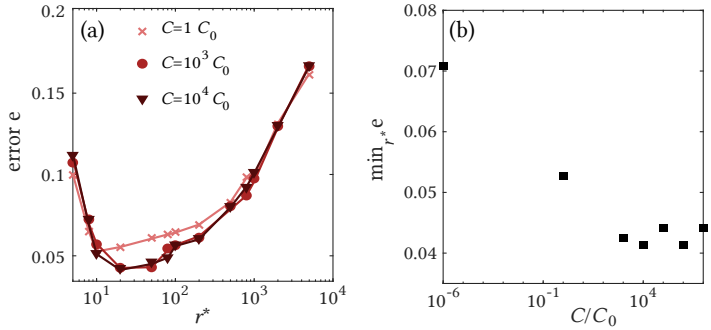

Figure 4 Influence of the intrinsic parameters N^* (related to r^*) and C without missing data: (a) error compared to r^* and (b) minimum error compared to C.

In addition it is shown that C significantly contributes to the convergence of the method: the higher it is (to a certain extent), the lower the error is. Indeed, the distance defined by the norm influences the mapping between material and mechanical states. By choosing a large value of C, the mapping based on strain values is favored, which is relevant since they are measured (and so reliable) unlike stresses which evolve during the convergence of the algorithm. Finally, N^* is more influential than C: without missing data, a bad choice of N^* will never be compensated by a good choice of C.

4.2 Influence of the incompleteness of input data

4.2.1 Force input

First, we consider preprocessing choices related to force information: either with all nodal forces (t-0), or represented by their net value on the true boundary (t-1) or the net force on the approximate boundary (t-2). The influence of N^* on the error is reported in Figure 5 for $C = 10^3 C_0$. Global errors are similar: a sampling ratio r^* from 20 to 100 is preferable. It shows that the DDI results are only slightly influenced by the way these equilibrium conditions are prescribed on the top boundary.

For a local insight, Figure 6 presents the nodal forces computed with the stress identified with the DDI in case (t-1). They are compared to the reference case (t-0). They are really similar which means that stresses computed with the DDI are almost as perfect as the reference ones, even if the input in force is the net force only.

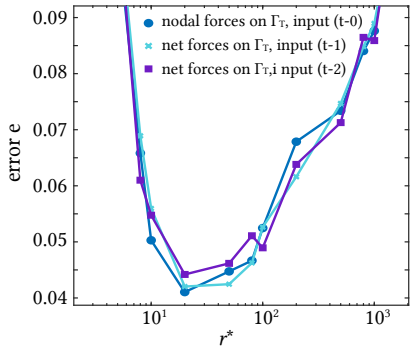

Figure 5 Influence of force inputs on the error as a function of r^* for (t-0) the given nodal forces, and (t-1) and (t-2) the given net force on respectively the true boundary and the approximate boundary.

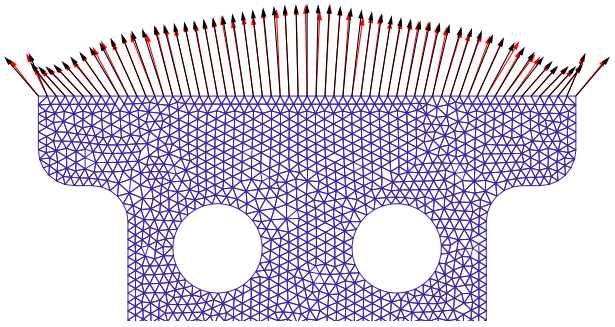

Figure 6 Nodal forces computed with the identified stress for the case (t-1) compared to the reference nodal forces (t-0). Red arrows: nodal forces recomputed from stress identified by DDI, case (t-1); black arrows: reference nodal forces.

4.2.2 *Cluster* of missing data

We consider the influence of the preprocessing choices related to a *cluster*: by handling it naively (c-1) and in a mechanically optimal way (c-2). The influences of r^* and C on the error are shown in Figure 7.

In the case of a naively handled *cluster* (c-1), it is difficult to achieve a small error. Too many or too few material states lead to more important errors. Here, the choice of C is crucial: the larger it is (within a certain limit), the closer we get to a mapping based on strains (which are known). By simply adding the zero net force condition (c-2) on the boundary, as proposed in Equation (9), a robustness similar to results without missing data is recovered. In this case, the choice of C is much less critical than that of N^*.

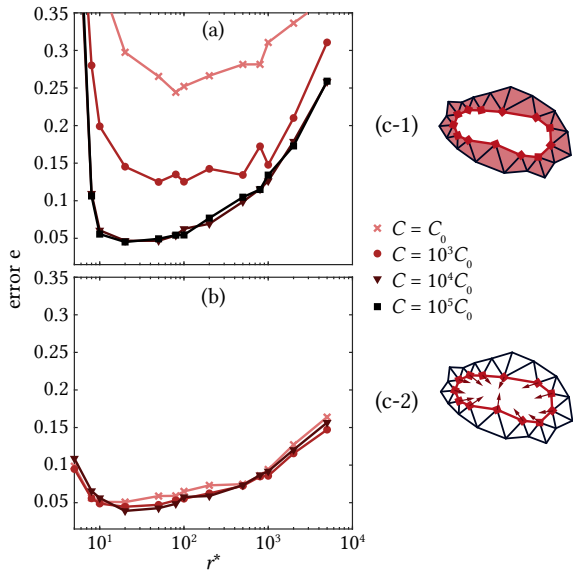

Figure 7 Influence of the preprocessing choice with a *cluster* on the error as a function of r^* and C for (a) case (c-1) naive and (b) case (c-2) mechanically optimal.

4.2.3 Imperfect resolution close to holes

Finally, we consider the influence of the preprocessing choices in the case of an imperfect resolution close to the edges, on the boundary $\tilde{\Gamma}_H$. The case of (h-1) the free edge assumption on this boundary and the case of (h-2) zero net force over the boundary $\tilde{\Gamma}_H$ are compared to the perfect case with no missing data close to the edge (h-0). Errors are plotted for a given C with

respect to the sampling ratio r^* in Figure 8. Considering the free edge assumption leads to a large

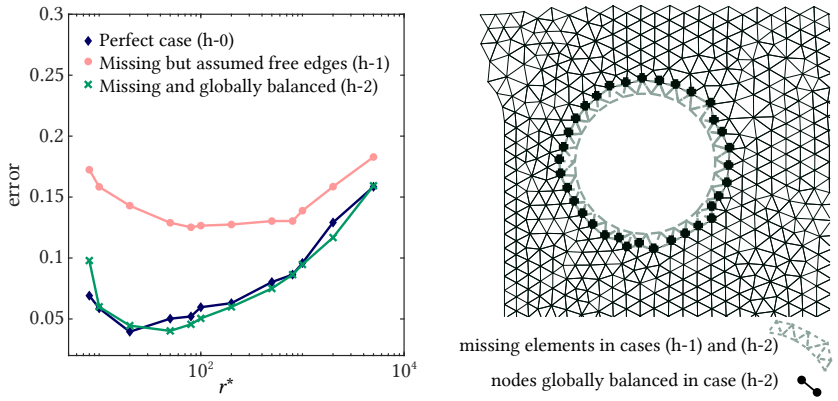

Figure 8 Influence of preprocessing choice for the hole edge definition on the error as a function of r^* for a given C, for the cases of an imperfectly defined edge close to the hole (with (h-1) the assumption of free edge, (h-2) the global balance condition) and of a perfectly defined edge (h-0) (left subfigure). Nodes/elements used in the calculations (right subfigure).

error, whereas the globally balanced assumption again induces a small error, close to the ideal case. Then, it is interesting to study the stress distribution as one approaches the hole: the von Mises stress is plotted along a line of the sample for the three cases, as depicted in Figure 9. For the free edge assumption (h-0), the algorithm predicts a misplaced stress increase close to the wrongly presumed free edge. Stresses are overestimated around the hole and this overestimation propagates to the bulk by equilibrium relations which are global. Therefore, the best manner to handle an imperfect edge consists in adopting a mechanically correct assumption: only a zero net force condition must be enforced. In this case (h-2), the error is similar to the one with no missing data. The optimal ratio r^* is again between 20 and 100.

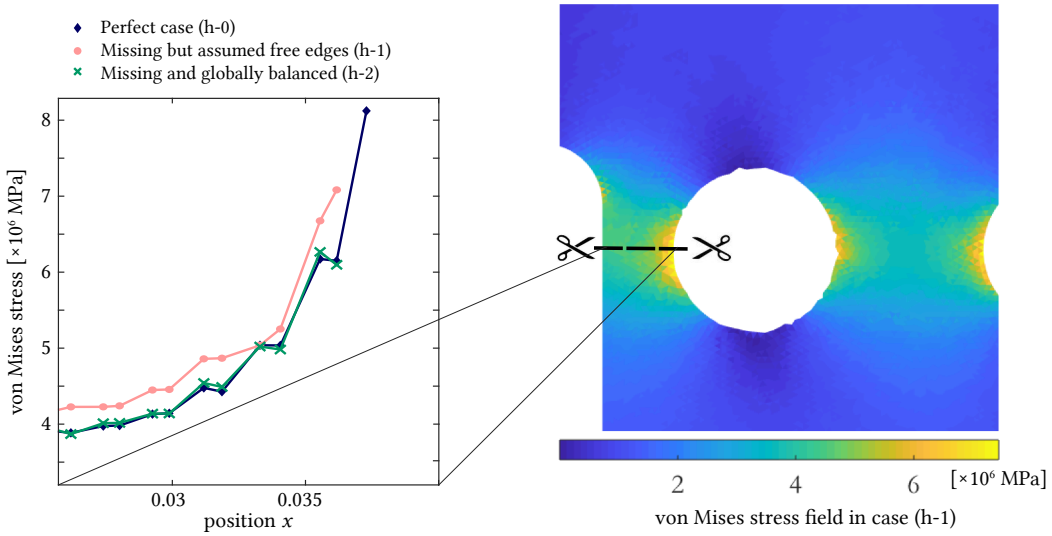

Figure 9 von Mises stress field (right) and reported values along a horizontal line going through the sample (left), for the three cases between reference and identified stress fields

4.2.4 Summary

To close this discussion on how to handle properly incomplete input data, Figure 10 presents the three stress components identified considering the corrections depicted in the depicted on the mesh in the top left corner of the figure. The incomplete input data have to be completed smartly with only the mechanical balance equation: (t-2), (c-2), (h-2). It leads to results similar to the perfect case: the global error is less than 0.05.

The preprocessing choices with best robustness for missing data are summarized in Figure 11. They are selected in the experimental validation of the DDI algorithm (Dalémat et al. 2019).

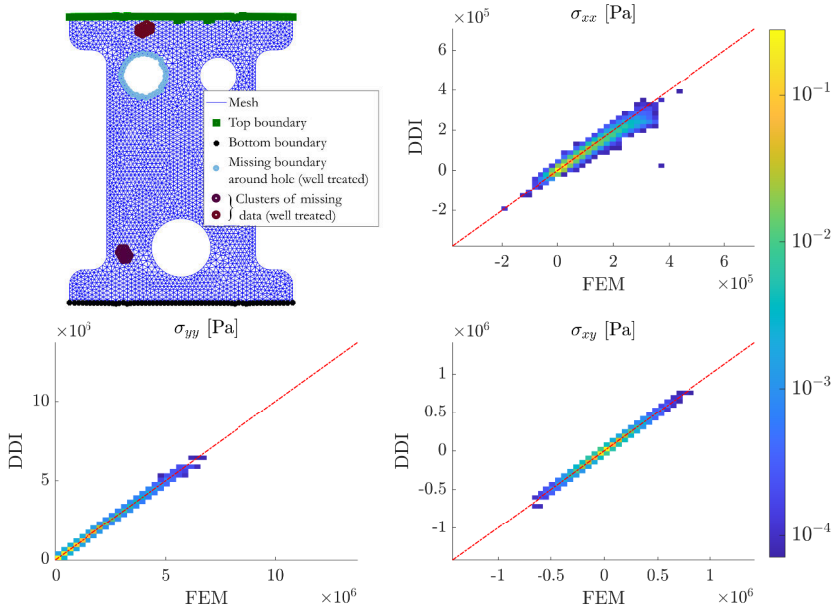

Figure 10 Comparison between reference stress fields and identified stress fields with DDI for a geometry with the proper preprocessing choices for missing data. Colors in 2D histograms represent the histogram bin probability for each stress component.

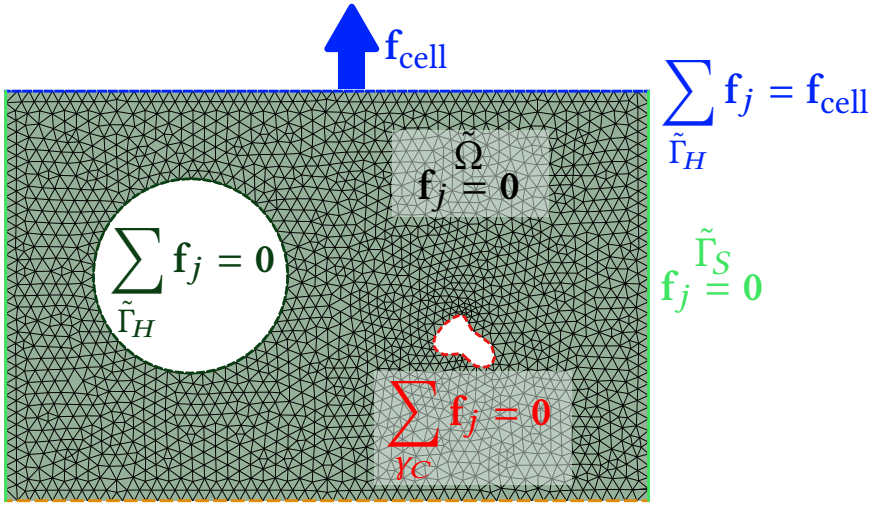

Figure 11 Summary of the proper preprocessing choice to deal with three particular boundaries (top boundary with a global force information, *cluster* of missing data and imperfectly defined edges close to a hole).

5 Closure: implementation of the DDI with real data

In this work, the input parameters of the DDI algorithm have been examined, with the objective of identifying correctly the stress field without constitutive equation. A study of its intrinsic parameters confirms our previous work. In particular, the consequences of incomplete data (inherent to experimental data) is analyzed through two aspects: the availability of net forces instead of nodal forces on the computational mesh, and the difference between the actual part geometry and the computational mesh. This last aspect appears either through *clusters* of missing data (areas of a few pixels/elements) and the imperfect edge definitions close to holes and boundaries. We show that the robustness of the method is ensured when incomplete data are managed under a strict mechanical point of view. Although these recommendations are here illustrated on a single example, they are drawn from our experience in applying DDI to many cases involving synthetic and real data, linear and non-linear material behaviors (Leygue et al. 2018; Dalémat et al. 2019; Stainier et al. 2019).

To conclude, we propose to adapt the original DDI algorithm to real experimental data. The

boundaries defined in Figure 1 are considered. The solution of the problem is still defined as:

$$\text{solution} = \underset{\boldsymbol{\sigma}_e^X,(\ln \mathbf{v}_i^*,\boldsymbol{\sigma}_i^*)}{\arg\min} \; \mathcal{E}(\boldsymbol{\sigma}_e^X, \ln \mathbf{v}_e^*, \boldsymbol{\sigma}_e^*), \tag{14}$$

with $\mathcal{E}$ defined in Equation (4), and subject to the (new) constraints:

- respecting the mechanical balance equations:
  - locally:

$$\sum_e w_e^X \mathbf{B}_{ej}^X \cdot \boldsymbol{\sigma}_e^X = \mathbf{0} \quad \forall X, \forall j \in \tilde{\Omega} \cup \tilde{\Gamma}_{\mathrm{B}}, \tag{15}$$

  - globally:

$$\sum_{j \in \Delta^X} \sum_e w_e^X \mathbf{B}_{ej}^X \cdot \boldsymbol{\sigma}_e^X = \mathbf{F}_{\Delta^X} \quad \forall X, \tag{16}$$

    with Δ^X representing each boundary on which is applied the net force $\mathbf{F}_{\Delta^X}$, at the increment X *i.e.*:
    - the top boundary $\tilde{\Gamma}_{\mathrm{T}}^X$ on which the net force is the one measured by the load cell $f_{\mathrm{cell}}^X \mathbf{n}_{\mathrm{sol}}$;
    - the boundary around *clusters* of missing data γ_{C}^X on which a zero net force is applied;
    - the boundary around a hole $\tilde{\Gamma}_{\mathrm{H}}^X$ on which a zero net force is applied;
- ensuring that the material state $(\ln \mathbf{v}_{e^X}^*, \boldsymbol{\sigma}_{e^X}^*)$ associated to the element e^X belongs to the database $(\ln \mathbf{v}_i^*, \boldsymbol{\sigma}_i^*)_{i=1}^{N^*}$.

The implementation details can be found in (Leygue et al. 2018). For the intrinsic parameters, it is advised to choose N^* so that the sampling ratio r^* is 100 and to choose $\mathbb{C} = C\mathbb{I}$ with $C \geqslant 10^3 C_0$, C_0 being the average stiffness of the material and $\mathbb{I}$, the fourth-order identity tensor.

References

Amores, V. J., J. M. Benítez, and F. J. Montáns (2020). Data-driven, structure-based hyperelastic manifolds: A macro-micro-macro approach to reverse-engineer the chain behavior and perform efficient simulations of polymers. *Computers & Structures* 231:106209. [DOI], [ARXIV].

Avril, S., M. Bonnet, A.-S. Bretelle, M. Grédiac, F. Hild, P. Ienny, F. Latourte, D. Lemosse, S. Pagano, E. Pagnacco, and F. Pierron (2008). Overview of identification methods of mechanical parameters based on full-field measurements. *Experimental Mechanics* 48(4):381–402. [DOI], [HAL].

Ayensa-Jiménez, J., M. H. Doweidar, J. A. Sanz-Herrera, and M. Doblaré (2018). A new reliability-based data-driven approach for noisy experimental data with physical constraints. *Computer Methods in Applied Mechanics and Engineering* 328:752–774. [DOI], [OA].

Conti, S., S. Müller, and M. Ortiz (2018). Data-Driven Problems in Elasticity. *Archive for Rational Mechanics and Analysis* 229(1):79–123. [DOI], [OA].

Crespo, J. and F. J. Montáns (2019). General solution procedures to compute the stored energy density of conservative solids directly from experimental data. *International Journal of Engineering Science* 141:16–34. [DOI].

Dalémat, M., M. Coret, A. Leygue, and E. Verron (2019). Measuring stress field without constitutive equation. *Mechanics of Materials* 136:103087. [DOI], [OA].

Eggersmann, R., T. Kirchdoerfer, S. Reese, L. Stainier, and M. Ortiz (2019). Model-Free Data-Driven inelasticity. *Computer Methods in Applied Mechanics and Engineering* 350:81–99. [DOI], [OA].

Furukawa, T. and G. Yagawa (1998). Implicit constitutive modelling for viscoplasticity using neural networks. *International Journal for Numerical Methods in Engineering* 43(2):195–219. [DOI].

González, D., F. Chinesta, and E. Cueto (2019). Thermodynamically consistent data-driven computational mechanics. *Continuum Mechanics and Thermodynamics* 31(1):239–253. [DOI], [HAL].

Ibañez, R., D. Borzacchiello, J. V. Aguado, E. Abisset-Chavanne, E. Cueto, P. Ladeveze, and F. Chinesta (2017). Data-driven non-linear elasticity: constitutive manifold construction and problem discretization. *Computational Mechanics* 60(5):813–826. [DOI], [HAL].

Kanno, Y. (2018). Simple Heuristic for Data-Driven Computational Elasticity with Material Data Involving Noise and Outliers: A Local Robust Regression Approach. *Japan Journal of Industrial and Applied Mathematics* 35(3):1085–1101. [DOI], [ARXIV].

Kanno, Y. (2021). A kernel method for learning constitutive relation in data-driven computational elasticity. *Japan Journal of Industrial and Applied Mathematics* 38(1):39–77. [DOI].

Kirchdoerfer, T. and M. Ortiz (2016). Data-driven computational mechanics. *Computer Methods in Applied Mechanics and Engineering* 304:81–101. [DOI], [OA].

Kirchdoerfer, T. and M. Ortiz (2017). Data Driven Computing with noisy material data sets. *Computer Methods in Applied Mechanics and Engineering* 326:622–641. [DOI], [OA].

Kirchdoerfer, T. and M. Ortiz (2018). Data-driven computing in dynamics. *International Journal for Numerical Methods in Engineering* 113(11):1697–1710. [DOI], [OA].

Latorre, M. and F. J. Montáns (2020). Experimental data reduction for hyperelasticity. *Computers & Structures* 232:105919. [DOI], [OA].

Leygue, A., M. Coret, J. Réthoré, L. Stainier, and E. Verron (2018). Data-based derivation of material response. *Computer Methods in Applied Mechanics and Engineering* 331:184–196. [DOI], [HAL].

Nguyen, L. T. K. and M.-A. Keip (2018). A data-driven approach to nonlinear elasticity. *Computers & Structures* 194:97–115. [DOI].

Ogden, R. W. (1972). Large deformation isotropic elasticity: on the correlation of theory and experiment for compressible rubberlike solids. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* 328(1575):567–583. [DOI].

Roux, S. and F. Hild (2020). Optimal procedure for the identification of constitutive parameters from experimentally measured displacement fields. *International Journal of Solids and Structures* 184:14–23. [DOI], [OA].

Réthoré, J., A. Leygue, M. Coret, L. Stainier, and E. Verron (2018). Computational measurements of stress fields from digital images. *International Journal for Numerical Methods in Engineering* 113(12):1810–1826. [DOI], [HAL].

Seghir, R. and F. Pierron (2018). A Novel Image-based Ultrasonic Test to Map Material Mechanical Properties at High Strain-rates. *Experimental Mechanics* 58(2):183–206. [DOI], [OA].

Stainier, L., A. Leygue, and M. Ortiz (2019). Model-Free Data-Driven Methods in Mechanics: Material Data Identification and Solvers. *Computational Mechanics* 64(2):381–393. [DOI], [OA].

Sutton, M. A., J. J. Orteu, and H. Schreier (2009). *Image Correlation for Shape, Motion and Deformation Measurements: Basic Concepts, Theory and Applications*. Springer US. ISBN: 978-0-387-78747-3. [DOI], [OA].

Treloar, L. R. G. (1944). Stress-strain data for vulcanised rubber under various types of deformation. *Transactions of the Faraday Society* 40:59. [DOI].

Yang, H., X. Guo, S. Tang, and W. K. Liu (2019). Derivation of heterogeneous material laws via data-driven principal component expansions. *Computational Mechanics* 64(2):365–379. [DOI].

Authors' contributions Dalémat carried out most of the study, performed numerical simulations, and drafted the original manuscript. Leygue originaly developed DDI matlab libraries, he largely helped with implementation and numerical issues. All authors conceived the study and developed the methodology. All authors participated in the critical review of the manuscript. All authors read and approved the final manuscript.

Supplementary Material Synthetic Hyperelastic data for DDI have been shared in Zenodo website (see permalink on the front page).

Acknowledgements This work was performed by using HPC resources of Centrale Nantes Supercomputing Center on the cluster Liger, granted and identified D1705030 by the High Performance Computing Institute(ICI).

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare that they have no competing interests.

Journal's Note JTCAM remains neutral with regard to the content of the publication and institutional affiliations.