



HAL
open science

Multimodal data augmentation for digital twinning assisted by artificial intelligence in mechanics of materials

Axel Aublet, Franck N'Guyen, Henry Proudhon, David Ryckelynck

► **To cite this version:**

Axel Aublet, Franck N'Guyen, Henry Proudhon, David Ryckelynck. Multimodal data augmentation for digital twinning assisted by artificial intelligence in mechanics of materials. *Frontiers in Materials*, 2022, 9, pp.971816. 10.3389/fmats.2022.971816 . hal-03815432

HAL Id: hal-03815432

<https://hal.science/hal-03815432>

Submitted on 14 Oct 2022

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.

Multimodal data augmentation for digital twinning assisted by artificial intelligence in mechanics of materials

Axel Aublet^{1,2*}, Franck N’Guyen², Henry Proudhon², David Ryckelynck^{2*}

¹ Mines Paris PSL, CNRS UMR 7633, Centre des Matériaux, 63 - 65 rue Henry Auguste Desbrières BP 87 F-91003 Evry cedex FRANCE,

² Safran Tech, Rue des Jeunes Bois, 78117 Châteaufort, FRANCE

Correspondence*:

Corresponding Author

axel.aublet@minesparis.psl.eu, david.ryckelynck@minesparis.psl.eu

2 ABSTRACT

3 Digital twins in mechanics of materials usually involve multimodal data in the sense that
4 an instance of a mechanical component has both experimental and simulated data. These
5 simulations aim not only to replicate experimental observations but also to extend the data.
6 Whether spatially, temporally or functionally for various possible uses of the components. Related
7 multimodal data are scarce, high dimensional and a physics-based causality relation exists
8 between observational and simulated data. We propose a data augmentation scheme coupled to
9 data pruning, in order to limit memory requirements for high-dimensional augmented data. This
10 augmentation is desirable for digital twinning assisted by artificial intelligence when performing
11 nonlinear model reduction. Here, data augmentation aims at preserving similarities in terms
12 of validity domain of reduced digital twins. In this paper, we consider a specimen subjected
13 to a mechanical test at high temperature, where the as-manufacturing geometry may impact
14 the lifetime of the component. Hence, an instance is represented by a digital twin that includes
15 a 3D X-Ray tomography data of the specimen, the related finite element mesh and the finite
16 element predictions of thermo-mechanical variables at several time steps. There is thus, for each
17 specimen, geometrical and mechanical information. Multimodal data, which couples different
18 representation modalities together, are hard to collect and annotate them requires much more
19 effort. Thus, the analysis of multimodal data generally suffers from the problem of data scarcity.
20 This problem is particularly pronounced when considering digital twins of thermo-mechanical tests,
21 which are very complex to implement in large amounts. The proposed data augmentation scheme
22 aims to train a recommending system that recognizes a category of data available in a training set
23 that has already been fully analysed by using high fidelity models. Such a recommending system
24 enables the use of a ROM-net for fast lifetime assessment via local reduced order models.

25 **Keywords:** High dimensional data, Model Order Reduction, Data Pruning, Oversampling, Computational Mechanics, Lifetime
26 Prediction

1 BACKGROUND

27 A Digital Twin (DT) is an ultra-realistic model in geometric details, including manufacturing anomalies, or
28 in material Bellinger et al. (2011), specific to an instance of an engineering system or of a component in
29 a system. More precisely, when considering model-based engineering systems, DT integrate ultra-high
30 fidelity simulations, maintenance history and all available historical among real data. Those information
31 are used to mirror the life of their respective "flying twin", as proposed in Glaessgen and Stargel (2012) for
32 aeronautical applications. In this paper, we consider Real Instances of Mechanical Components (RIMC)
33 that undergo thermo-mechanical loadings in an experiment that reflects what an aircraft engine imposes
34 to a high pressure turbine blade. The as-manufactured geometries are observed by X-Ray Computed
35 Tomography (CT) that generates 3D accurate digital images of RIMC. A numerical model is a DT only if

36 its corresponding RIMC exists. How to make the best use of existing data while exploiting the information
 37 with classical mathematical models is an important issue. Data-driven modelling is emerging as a powerful
 38 paradigm for physical and biological systems Zhang et al. (2019). In various industrial fields related to
 39 structural mechanics, most of numerical models incorporate scientific knowledge about mechanics of
 40 materials. Hence, the numerical instance of a DT incorporates multimodal data. Each modality of data has
 41 its own ambient space related either to material parameters, geometrical and morphological representation
 42 or boundary and initial conditions for partial differential equations. One important modality is related to
 43 the synthetic data forecast by model-based numerical simulations, predicting the lifetime of the RIMC for
 44 instance, upon observational data. Such synthetic data support decision in engineering for maintenance,
 45 operating optimization or decommission of RIMC. With the development of Non Destructive Testing
 46 (NDT) in the manufacturing industry, we can expect a growing activity on image-based DT Seon et al.
 47 (2020); Launay et al. (2021a) for accurate descriptions of as-manufactured geometries and microstructural
 48 properties of structural components. Such data are crucial for lifetime predictions Aublet et al. (2022).
 49 Unfortunately, image-based DT in material science is so complex and time-consuming, so that in practice,
 50 this task does not scale with the frequency of quality inspection in manufacturing. This is a major scaling
 51 issue for online DT of as-manufactured structural components, where lifetime prediction is strongly
 52 affected by geometrical defects. Another issue is the curse of dimensionality related to the dimension of the
 53 geometrical ambient space, so that a dense sampling of this space is unaffordable. So, learning regression
 54 for lifetime prediction from observational data, as in cross-modal predictions (Launay et al., 2021a), may be
 55 not relevant for such image-based modelling. In this work, we augment multimodal data in order to train a
 56 ROM-net Daniel et al. (2020). A ROM-net learns a piecewise linear latent-space for model order-reduction.
 57 It incorporates a classifier, in its first layers, that recommend the selection of a local reduced-order model
 58 (LROM), from a digital image of the RIMC, where each LROM has its own label. In this paper, we assume
 59 that the latent space has a structure of vector bundle that is locally trivial. This means that LROM's have
 60 a validity domain larger than the support of training data Ryckelynck et al. (2015). In the sequel, DTs
 61 that have similar validity domains for their specific LROM are termed ROM-similar. Therefore, ROM-net
 62 predictions should be accurate when the input data have ROM-similar data in the dictionary of LROM.
 63 Unfortunately, the multimodal data related to image-based DT are very scarce, compare to the dimension
 64 of the ambient spaces of these data. Furthermore, we need to prove the efficiency of digital twinning assisted
 65 by artificial intelligence prior getting more data related to manufactured components. Hence ROM-net,
 66 or more precisely the classifier of the ROM-net, needs augmented data (Lecun et al., 1998). A review on
 67 image-data augmentation schemes is available in Shorten and Khoshgoftaar (2019). Both image warping
 68 and oversampling augmentations are proposed in the literature, so that labels are preserved. Here, the
 69 multimodal data involved in digital twinning require to account for the causality relation that exists between
 70 the model-based predictions and observational image data. In data warping augmentation, alone, the
 71 causality relation between observational data and simulated data is not preserved. Hence, labels that refer
 72 to LROM can't be preserved by data warping. So, we propose a dedicated oversampling procedure based
 73 on ROM-similarity. A schematic view of the assumed piecewise linear latent space is shown in Figure
 74 ???. In this figure, ROM-similar data are aligned on the same slope. We consider both experimental data
 75 and simulated data related to the lifetime of a realistic structural component, composed of CMSX4-PLUS
 76 super alloy, undergoing cyclic loading and plastic strains at high temperature. The next section details the
 77 proposed scheme for multimodal data oversampling. In Section 3 two examples are provided. The first
 78 one is an augmentation around a defect in a 2D picture which affect the local strain fields. In the second
 79 example, we give a detailed description of the considered multimodal data for an academic example and
 80 for a mechanical component undergoing cyclic loading at high temperature for its lifetime characterization.
 81 The ambient spaces related to the multimodal data in this paper are realistic regarding manufactured parts
 82 of the aeronautics industry Aublet et al. (2022).

2 METHODS

83 The multimodal data involved in a DT are denoted by $\boldsymbol{\mu}$ for material or metallurgical parameters, $\boldsymbol{\mu}_{\square}$ for
 84 digital images, p for the lifetime prediction and a matrix $\mathbf{u} \in \mathbb{R}^{\mathcal{N} \times n_t}$ of model-based simulated data over
 85 n_t time steps. The residuals of the physics-based equations read:

$$\mathbf{r}(\mathbf{u}; k, \boldsymbol{\mu}, \boldsymbol{\mu}_{\square}) = 0, \quad k = 1, \dots, n_t, \quad p - \ell(\mathbf{u}) = 0 \quad (1)$$

86 where $\mathbf{r}(\cdot; k, \cdot, \cdot) \in \mathbb{R}^{\mathcal{N}}$, $k = 1 \dots n_t$ is a given series of complex physics-based equations that forecast
 87 primal variables $\mathbf{u}[:, k]$ (i.e. displacements in mechanical models) and ℓ is a much less complex function for
 88 lifetime prediction knowing \mathbf{u} . n_t is the number of residual vectors required to predict the n_t columns of \mathbf{u} .
 89 For instance, n_t is the number of loading steps in a mechanical problem. Residuals \mathbf{r} setup the causality
 90 relation between the multimodal data.

91 The solution manifold of all \mathbf{u} fulfilling equation 1, for given $\boldsymbol{\mu}$ and $\boldsymbol{\mu}_\square$, is denoted by E . For reduced
 92 order modelling purpose, we assume that E has a structure that the data augmentation scheme has to
 93 preserve. This structure is a vector bundle of rank N . It is a family of vector spaces of dimension N
 94 parametrized by a base space denoted by M . In the proposed framework, M is the manifold of primal
 95 variables \mathbf{u} for digital twins. The vector bundle is assumed to be locally trivial, in the sense that the
 96 local vector space is constant on subsets of M . In the sequel, these subsets of M have a dedicated label
 97 y and this label refers to a local reduced basis $\mathbf{V}(y) \in \mathbb{R}^{\mathcal{N} \times N}$. An instance of a DT is denoted by
 98 $\mathbf{X}^{(i)} = \{\boldsymbol{\mu}^{(i)}, \boldsymbol{\mu}_\square^{(i)}, \mathbf{u}^{(i)}, p^{(i)}\}$, with $\mathbf{u}^{(i)} \in M$. Its label is denoted by $y^{(i)}$ and $\mathbf{V}(y^{(i)})$ is the local reduced
 99 basis attached to the i^{th} DT. The reduced basis $\mathbf{V}(y^{(i)})$ contains the left eigenvectors of the truncated
 100 singular value decomposition of $\mathbf{u}^{(i)}$. It is an orthogonal reduced basis. The original set of DTs indices is
 101 denoted by \mathcal{D}_{t0} . The set of indices related to augmented data is denoted by \mathcal{D}_{aug} , such that $\mathcal{D}_{t0} \cap \mathcal{D}_{aug} = \emptyset$.
 102 Augmented data $\mathbf{X}^{(a)}$, $a \in \mathcal{D}_{aug}$, are admissible, if they preserve the structure of E in the following sense:
 103 it exists $i \in \mathcal{D}_{t0}$ such that:

$$\mathbf{u}^{(a)} \in \text{span}(\mathbf{V}(y^{(i)})) \tag{2}$$

$$y^{(a)} = y^{(i)} \tag{3}$$

$$\mathbf{r}(\mathbf{u}^{(a)}; k, \boldsymbol{\mu}^{(a)}, \boldsymbol{\mu}_\square^{(a)}) = 0, k = 1, \dots, n_t, \quad p^{(a)} - \ell(\mathbf{u}^{(a)}) = 0 \tag{4}$$

104 This admissibility criteria involves, in Equation 3, the usual constrain that preserve labels of original
 105 data. We acknowledge that the proposed data augmentation scheme does not improve the sampling of the
 106 base space M . Its purpose is to augment the number of instances that share the same labels. Figure ??
 107 is illustrating the structure of manifold E . In this figure, the sampling points on manifold M are the red
 108 points. The local vector spaces are represented as green lines. The desired augmented data are represented
 109 by blue crosses. The labels have also been represented on each axis to this plot.

110 Usually, \mathcal{N} is large and can range from 10^5 to 10^7 . The higher the resolution of digital images, the
 111 higher the dimensions of $\boldsymbol{\mu}_\square$ and \mathbf{u} . In practice, \mathbf{r} and \mathbf{u} are related to a finite element mesh of the digital
 112 image $\boldsymbol{\mu}_\square$. For such high dimensional multimodal data, preserving memory storage capabilities while
 113 performing data augmentation is the main issue to achieve feasible augmentations. The storage limits of
 114 high dimensional multimodal data was not in the scope of previous papers on simulated data augmentation
 115 Daniel et al. (2021); Launay et al. (2021b). Without solving this issue, no data oversampling is possible
 116 here. Therefore, to limit the memory requirement and computational time, oversampling is coupled to a
 117 data pruning technique Hilth et al. (2019) and numerical approximations. The method intends to restrict
 118 Equation 2 and Equation 4 to sets of selected rows among the complete set of rows. These sets are denoted
 119 respectively $\overline{\mathcal{F}}$ and \mathcal{F} . In order to have a consistent selection of these sets of rows, the data pruning
 120 procedure accounts for the finite element description of the residual \mathbf{r} and the spatial interpolation of the
 121 primal variable \mathbf{u} . This interpolation uses finite element shape functions denoted by $(\varphi_j)_{j=1, \dots, \mathcal{N}}$. The
 122 interpolated primal variable reads:

$$\tilde{\mathbf{u}}(\boldsymbol{\xi}; k, \boldsymbol{\mu}, \boldsymbol{\mu}_\square) = \sum_{j=1}^{\mathcal{N}} \varphi_j(\boldsymbol{\xi}) u[j, k](\boldsymbol{\mu}, \boldsymbol{\mu}_\square) \quad \forall \boldsymbol{\xi} \in \Omega \tag{5}$$

123 where Ω is the geometrical domain occupied by the DT and $\boldsymbol{\xi}$ is the position vector in this domain. In
 124 image-based DTs, the finite element shape functions and the domain Ω are specific to each instance of
 125 $\boldsymbol{\mu}_\square^{(i)}$, for i in \mathcal{D}_{t0} . But this dependence is not included in notations here, for the sake of simplicity. The
 126 data pruning procedure proposed in Hilth et al. (2019) introduces a reduced domain denoted by Ω_R . This

127 reduced domain is the support of finite element shape functions related to \mathcal{F} :

$$\Omega_R = \cup_{j \in \mathcal{F}} \text{supp}(\varphi_j) \quad (6)$$

128 and $\overline{\mathcal{F}}$ is the set of indices of the variable \mathbf{u} restricted to Ω_R , so that the residual restricted to \mathcal{F} can be fully
129 computed using solely $\mathbf{u}[\overline{\mathcal{F}}]$. For the sake of simplicity this restricted residual is denoted by $\mathbf{r}(\mathbf{u}[\overline{\mathcal{F}}, :]; \cdot)[\mathcal{F}]$.
130 It is important to note that Ω_R must cover the region of interest for geometrical variations.

131 For pruned data, approximation errors in Equation 2 must be tolerated. We propose a two steps
132 augmentation technique, that contains: 1) a random sampling step of parameters $\boldsymbol{\mu}, \boldsymbol{\mu}_{\square}$ in the vicinity of
133 $(\boldsymbol{\mu}^{(i)}, \boldsymbol{\mu}_{\square}^{(i)})_{i \in \mathcal{D}_{t0}}$; 2) a screening step in accordance with the admissibility criteria for pruned data. The
134 admissibility of pruned augmented data fulfill the following equations:

$$\boldsymbol{\gamma}^{(a)}[:, k] = \underset{\mathbf{g} \in \mathbb{R}^N}{\text{argmin}} \|\mathbf{u}^{(a)}[\overline{\mathcal{F}}, k] - \mathbf{V}^{(i)}[\overline{\mathcal{F}}, :] \mathbf{g}\|, \quad k = 1, \dots, n_t \quad (7)$$

$$\eta^{(a)} = \frac{\|\mathbf{u}^{(a)}[\overline{\mathcal{F}}, :] - \mathbf{V}^{(i)}[\overline{\mathcal{F}}, :] \boldsymbol{\gamma}^{(a)}\|_F}{\|\mathbf{u}[\overline{\mathcal{F}}, :]\|_F} \quad (8)$$

$$s = 1 - \eta^{(a)} > s_{tol} \quad (9)$$

$$y^{(a)} = y^{(i)} \quad (10)$$

$$\mathbf{r}(\mathbf{u}^{(a)}[\overline{\mathcal{F}}, :]; k, \boldsymbol{\mu}^{(a)}, \boldsymbol{\mu}_{\square}^{(a)})[\mathcal{F}] = 0, \quad k = 1, \dots, n_t \quad (11)$$

$$\mathbf{u}^{(a)}[j, :] = \mathbf{u}^{(i)}[j, :] \quad \forall j \in \overline{\mathcal{F}} \setminus \mathcal{F} \quad (12)$$

$$p = \ell(\mathbf{V}^{(i)} \boldsymbol{\gamma}^{(a)}) \quad (13)$$

135 where $\|\cdot\|_F$ is the Frobenius norm. Here we assume that $\mathbf{V}^{(i)}[\overline{\mathcal{F}}, :]$ is full column rank. The above
136 computation of the reduced coordinates $\boldsymbol{\gamma}^{(a)}$ is known as the Gappy POD Everson and Sirovich (1995).
137 Again, we assume that \mathbf{u} is supported by the mesh of the i^{th} DT, so that the evaluation of $\eta^{(a)}$ makes
138 sense. Equation (9) is checking the ROM-similarity of the augmented data. Equation (12) in finite element
139 modelling is termed a Dirichlet boundary condition. Other interesting boundary conditions are proposed in
140 Hilth et al. (2019) for hyper-reduction. s_{tol} is a hyper-parameter of the augmentation scheme.

141 Remark: It is possible to update the vector bundle after the data augmentation. The accuracy of the
142 LROM that we can compute after data augmentation is not directly related to the threshold used here for
143 the ROM-similarity. For instance, after data augmentation, we can aggregate in a matrix $\widehat{\mathbf{u}}^{(i)}$ the primal
144 variable having the same label, for both original and augmented data. Therefore, a truncated singular value
145 decomposition of $\widehat{\mathbf{u}}^{(i)}$ can give a very accurate reduced basis for the category of data $y = y^{(i)}$.

3 RESULTS AND DISCUSSION

146 3.1 Modelling the local mechanical response of voids in an elastic material

147 In a previous work Ryckelynck et al. (2020), we published 2D DTs of voids in elastic bodies. In this
148 example, we augment the data around two DTs from this previous work. The augmentation ratio is 100
149 for each DT, prior fulfilling the constrains for the construction of \mathcal{D}_{aug} . Here, mechanical variables are
150 structured data similar to digital images Launay et al. (2021a). But not all data augmentation techniques
151 proposed in the literature for digital images are relevant for reduced digital twining. For instance rotation
152 and shear transformations must be consistent for both geometrical and mechanical data. In this example, the
153 domain Ω is a square (on the left in Figure 1). The manufacturing process creates voids and each instance
154 of void is placed in the centre of Ω . The mesh is specific to each DT in order to capture the geometry of the
155 void (at the centre of Figure 1). In this mesh a phase field represents the void: if $\phi(\boldsymbol{\xi}) > 0$ then $\boldsymbol{\xi}$ is in the
156 void, else $\boldsymbol{\xi}$ is outside the void ($1 \geq \phi \geq 0$). In mechanical equations of elasticity for digital twining, the
157 Young modulus reads $E(\boldsymbol{\xi}) = E_o (1 - \phi(\boldsymbol{\xi}))$, where E_o is the Young modulus in the bulk. In this simple
158 example, we restrict our attention to the prediction of shear strains around each void, under a macroscopic

159 shear strain. The reduced domain is shown on the right in Figure 1. Here, the cardinal of \mathcal{F} is only 30%
 160 smaller than \mathcal{N} . Simulation errors of shear strains is about 18% when solving equations (11)-(12) instead
 161 of equation (1). The mechanical problem has 3 time steps t_0, t_1, t_2 . In the first time step a longitudinal
 162 traction is imposed to the boundaries of Ω . In the second time step it is transverse traction. In the last time
 163 step, we impose a shear strain at the boundaries of Ω .

164 Multimodal data are the 2D image of the void (i.e. the image of ϕ) and the mechanical shear strain
 165 in Ω_R . These data are shown in Figure 2 and Figure 3 for Digital twins $i = 1$ and $i = 2$ respectively.
 166 Augmented image-data are obtained by usual shear and rotation of the void image. The resulting phase
 167 field is denoted by ϕ' . The mechanical expression of the strain tensor is the symmetric part of the gradient
 168 of the displacement field \mathbf{u} . This strain tensor is denoted by ε , $\varepsilon(\mathbf{u}) = (\nabla \mathbf{u} + \nabla \mathbf{u}^T)/2$. The solution of
 169 elastic equations (11)-(12) forecasts the augmented value for the shear strain $\varepsilon(\mathbf{u}')$. We can see on the full
 170 order simulations, on Figures 2 and Figure 3, that the mechanical strains are highly dependent on the void
 171 morphology.

172 Two examples of augmented data are shown in Figure 4 and Figure 5. In these examples, it is clear that
 173 the mechanical variables do not follow the warping imposed to μ_{\square} , especially far from the void. Local
 174 reduced bases have 3 modes. The reduced basis of the first DT is shown in Figure 6. The ROM-Similarity
 175 of augmented data is shown in Figure 7. Only one augmented data in $\tilde{\mathcal{D}}^{(1)}$ fails to be ROM-Similar. They
 176 are 28 for the second DT. Blue and red curves are well separated. None of augmented data for one DT is
 177 ROM-Similar to the other DT.

178 3.2 Application on fatigue specimen observed via X-ray computer tomography

179 In this subsection, we consider a mechanical specimen subjected to a mechanical test at high temperature,
 180 where as-manufacturing geometry may impact the lifetime of the specimen. The DT includes a 3D X-Ray
 181 tomography of the specimen (Figure 8), the related finite element mesh (Figure 9) and the finite element
 182 predictions of thermo-mechanical variables that are modified at several time steps during the test (on the
 183 pruned mesh in Figure 12). The problem of data scarcity is particularly pronounced for this kind of DTs. In
 184 this example, $\mathcal{N} > 3 \cdot 10^5$, $m = 20$, and the target reduced dimension of LROM is setup to $N = 6$ in order
 185 to have $s_{tol} = 0.99$. In this example, μ is a vector of 3 angles related to single crystal CMSX4-PLUS of the
 186 specimen as presented in Aublet et al. (2022). This parameter is augmented randomly using angle variations
 187 below 12° . Usual image warping techniques are not relevant on the 3D tomographic image of DT. Hence, a
 188 random mesh morphing have been applied to the mesh of the DT in order to augment observational data
 189 μ_{\square} . The geometrical variations were generated around a DT^i using a Gaussian perturbation on geometrical
 190 parameters with the following model. For a parameter q , the new value p' was randomly chosen in the
 191 vicinity of q with a common log-normal \mathcal{X} law centred on q with a standard deviation of $0.2 * q$ such as
 192 $p' \in \mathcal{X}(q, 0.2 * q)$. We generated 10 meshes around each DT and 10 cycles of triangular thermo-mechanical
 193 loading were applied. An example of node displacement for this morphing step is shown in Figure 10. Most
 194 finite elements removed for the data pruning are located around a cooling hole in the specimen (Figure 11).
 195 To ensure the consistency of HROM equations 11 under Neumann boundary conditions, some sets have to
 196 be preserved during the pruning. $\text{Card}(\bar{F})$ is twice smaller than \mathcal{N} , so the memory saving factor is 2. The
 197 prediction error of finite element simulation using the pruned mesh is 10% in Ω_R . Data augmentation has
 198 been performed for two DTs $X^{(1)}$ and $X^{(2)}$. After augmentation, the ROM-similarity is computed twice
 199 for all augmented data. This enables to check the labels of augmented data. The original DTs are quite
 200 ROM-similar, so values of s are close to 0.99. They are reported in Figure 13. Labels of augmented data
 201 are consistent with the ROM-Similarities. However, all ROM-Similarities are very high in this example.

202 Here, we consider a time-dependent problem due to plasticity computed with a Crystal Plasticity model
 203 developed by Méric and Cailletaud in Méric et al. (1991). This kind of problem implies a temporal analysis
 204 with initial important spread on the domain (low values with importation gradients) at the first times and
 205 then, a localization due to a mechanical defect such as a hole as shown in Figure 9. This localization
 206 implies local variations of displacement fields around the defect which needs a higher value of N to catch
 207 the information in LROM. As we performed a truncature with $N = 6$, we can't gather the local information
 208 on plasticity. The development of ROM-Similarity criteria is still a work in progress in plasticity. Here,
 209 $\mathbf{V}(y^{(1)})$ and $\mathbf{V}(y^{(2)})$ seem pretty similar hence the close ROM-Similarities reported in Figure 13.

4 CONCLUSION

210 The proposed data augmentation method is very versatile for multimodal data that includes simulated data
211 in the framework of image-based digital twinning. It requires the access to the finite element model used
212 for digital twinning. The method assumes that the latent space hidden in the multimodal data is piecewise
213 linear, so that nonlinear model reduction uses local reduced bases. The validity domains of this reduced
214 bases enable the definition of similarities, termed ROM-Similarity, between multimodal data. The pruning
215 technique, used to limit memory requirement for high-dimensional augmented-data, has an hyper-parameter
216 Ω_R . The larger the domain Ω_R the more accurate the evaluation of ROM-Similarities. The smaller Ω_R
217 the higher the memory savings. The development of ROM-Similarity criteria is still a work in progress in
218 plasticity.

CONFLICT OF INTEREST STATEMENT

219 The authors declare that the research was conducted in the absence of any commercial or financial
220 relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

221 Axel Aublet performed the Digital Twinning for the thermo-mechanical specimens and made the X-Ray
222 computed tomographies. Franck N'Gyuen did all image-based finite element meshes. Henry Proudhon
223 brought his help for the elasto-viscoplastic simulations, edited and commented the initial manuscript. David
224 Ryckelynck proposed the augmentation scheme and he applied it the numerical results on the 2D example.
225 Axel Aublet made the augmentation scheme onto the thermo-mechanical specimens.

FUNDING

226 Axel Aublet's PhD thesis is financially supported by SAFRAN.

REFERENCES

- 227 Aublet, A., Rambaudon, M., N'Guyen, F., Ryckelynck, D., Remacha, C., Cariou, R., et al. (2022).
228 Mechanical fatigue testing under thermal gradient and manufacturing variabilities in nickel-based
229 superalloy parts with air-cooling holes. *Experimental Mechanics* doi:10.1007/s11340-022-00868-0
- 230 Bellinger, N., Tuegel, E. J., Ingraffea, A. R., Eason, T. G., and Spottswood, S. M. (2011). Reengineering
231 aircraft structural life prediction using a digital twin. *International Journal of Aerospace Engineering*
232 doi:10.1155/2011/154798
- 233 Daniel, T., Casenave, F., Akkari, N., and Ryckelynck, D. (2020). Model order reduction assisted by
234 deep neural networks (ROM-net). *Advanced Modeling and Simulation in Engineering Sciences* 7.
235 doi:https://doi.org/10.1186/s40323-020-00153-6
- 236 Daniel, T., Casenave, F., Akkari, N., and Ryckelynck, D. (2021). Data augmentation and feature selection
237 for automatic model recommendation in computational physics. *Mathematical and Computational*
238 *Applications* 26. doi:10.3390/mca26010017
- 239 Everson, R. and Sirovich, L. (1995). Karhunen-Loeve procedure for gappy data. *JOSA A* 12. doi:10.1364/
240 JOSAA.12.001657
- 241 Glaessgen, E. and Stargel, D. (2012). The digital twin paradigm for future nasa and u.s. air force
242 vehicles. In *Structures, Structural Dynamics, and Materials Conference* (Aerospace Research Central).
243 doi:10.2514/6.2012-1818
- 244 Hilth, W., Ryckelynck, D., and Menet, C. (2019). Data pruning of tomographic data for the calibration
245 of strain localization models. *Mathematical and Computational Applications* 24. doi:10.3390/
246 mca24010018
- 247 Launay, H., Ryckelynck, D., Lacourt, L., Besson, J., Mondon, A., and Willot (2021a). F. deep multimodal
248 autoencoder for crack criticality assessment. *Int. J. Numer. Methods Eng.* , 1–25doi:10.1002/nme.6905
- 249 Launay, H., Willot, F., Ryckelynck, D., and Besson, J. (2021b). Mechanical assessment of defects in
250 welded joints: morphological classification and data augmentation 11. doi:10.1186/s13362-021-00114-7

- 251 Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document
252 recognition. *Proceedings of the IEEE* 86, 2278–2324. doi:10.1109/5.726791
- 253 Méric, L., Poubanne, P., and Cailletaud, G. (1991). Single Crystal Modeling for Structural Calculations:
254 Part 1 — Model Presentation. *Journal of Engineering Materials and Technology* 113, 162–170.
255 doi:10.1115/1.2903374
- 256 Ryckelynck, D., Gallimard, L., and Jules, S. (2015). Estimation of the validity domain of hyper-reduction
257 approximations in generalized standard elastoviscoplasticity. *Advanced Modeling and Simulation in*
258 *Engineering Sciences* 2. doi:10.1186/s40323-015-0027-7
- 259 Ryckelynck, D., Goessel, T., and Nguyen, F. (2020). Mechanical dissimilarity of defects in welded joints
260 via Grassmann manifold and machine learning. Preprint
- 261 Seon, G., Nikishkov, Y., Makeev, A., and Ferguson, L. (2020). Towards a digital twin for mitigating void
262 formation during debulking of autoclave composite parts. *Engineering Fracture Mechanics* 225, 106792.
263 doi:https://doi.org/10.1016/j.engfracmech.2019.106792
- 264 Shorten, C. and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning.
265 *Journal of Big Data* 6. doi:10.1186/s40537-019-0197-0
- 266 Zhang, D., Lu, L., Guo, L., and Karniadakis, G. E. (2019). Quantifying total uncertainty in physics-
267 informed neural networks for solving forward and inverse stochastic problems. *Journal of Computational*
268 *Physics* 397, 108850. doi:https://doi.org/10.1016/j.jcp.2019.07.048

FIGURE CAPTIONS

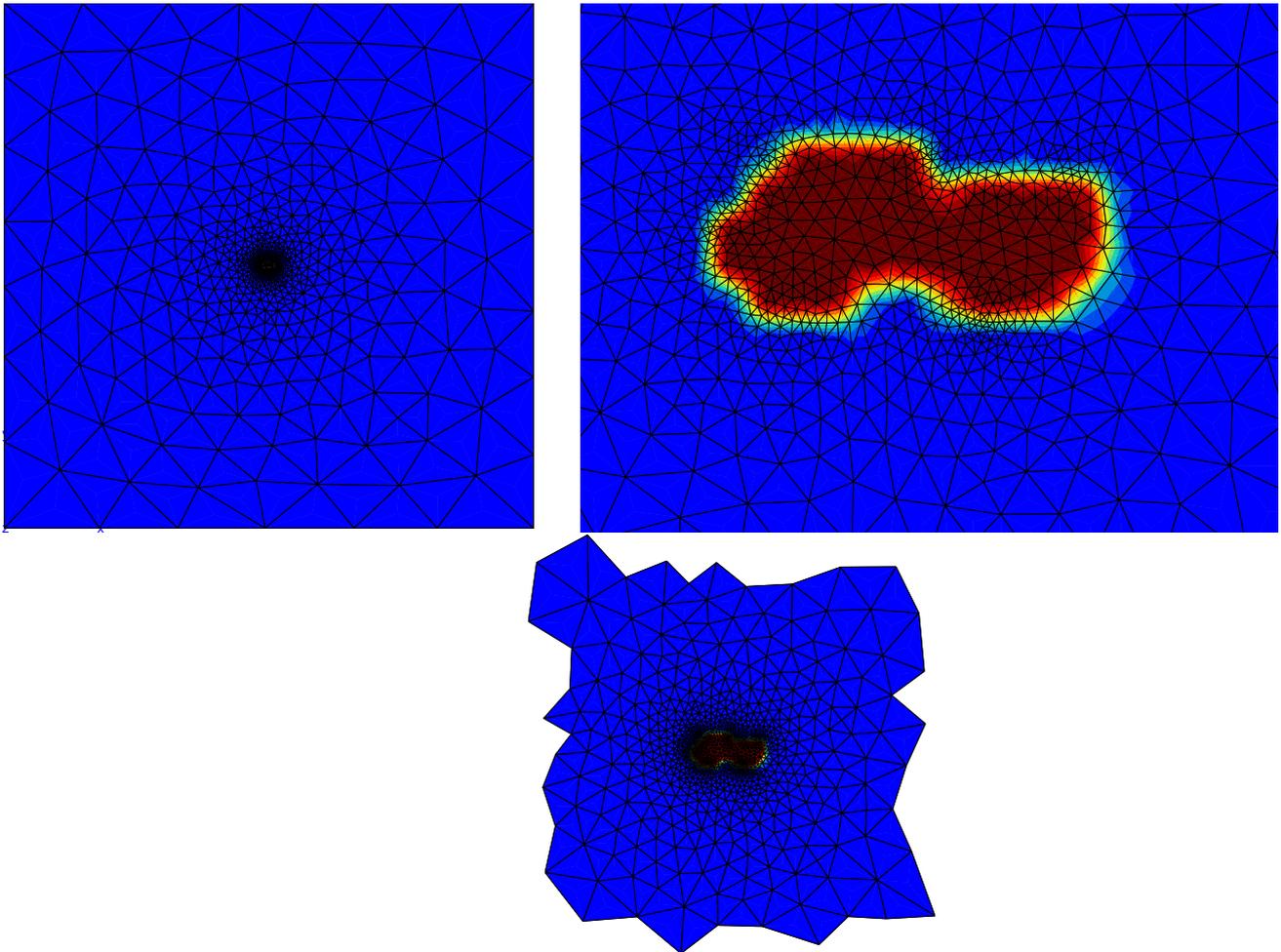


Figure 1. The finite element mesh of one digital twin (on the left), a focus on the phase field of the void at the center of this mesh (center), the mesh of the reduced domain Ω_R (right).

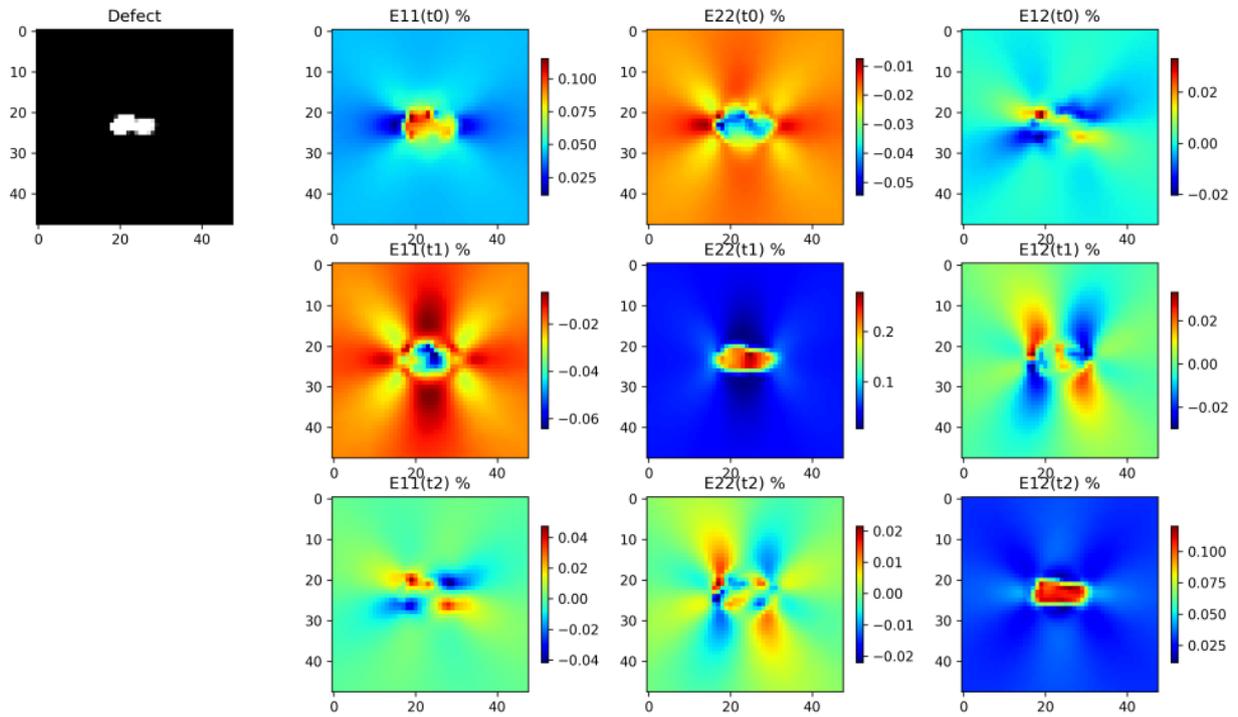


Figure 2. Multimodal data for digital twin $i = 1$, on top left $\mu_{\square}^{(1)}$, on the right the 3 components of the strain tensor $\varepsilon(\mathbf{u}^{(1)})$ at 3 time instants t_0, t_1, t_2 .

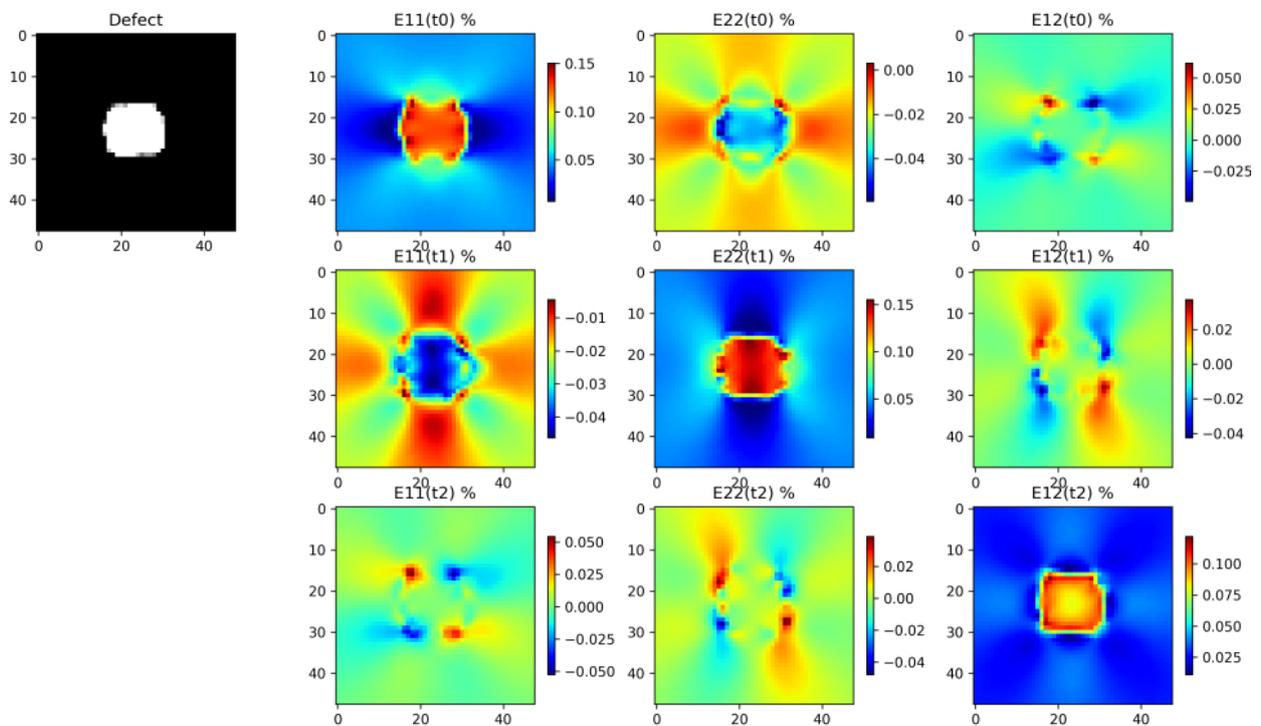


Figure 3. Multimodal data for digital twin $i = 2$, on top left $\mu_{\square}^{(2)}$, on the right the 3 components of the strain tensor $\varepsilon(\mathbf{u}^{(2)})$ at 3 time instants t_0, t_1, t_2 .

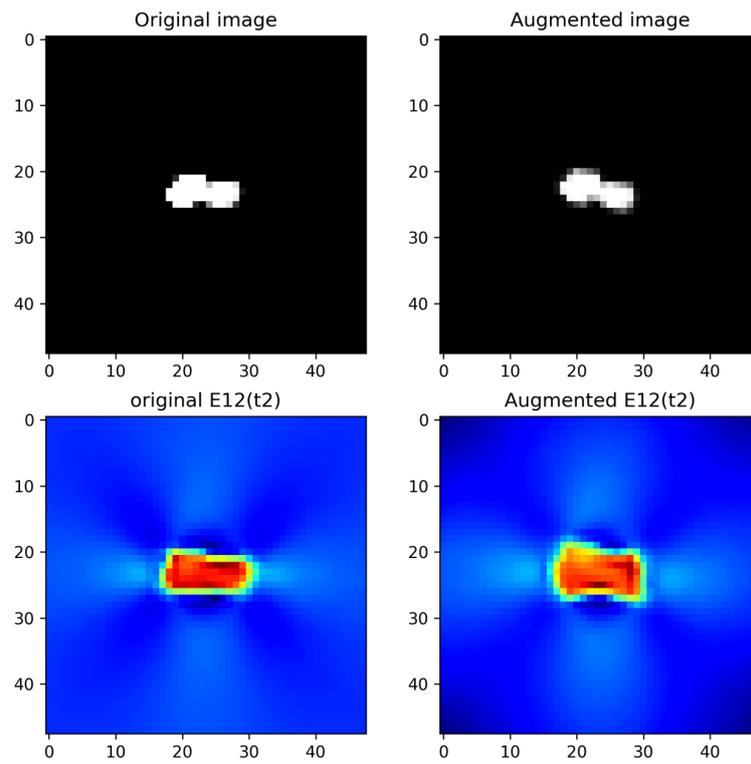


Figure 4. Example of augmented multimodal data for digital twin $i = 1$, on the left original data, on the right augmented data, on first row μ_{\square} , second row the shear strain at time $t2$.

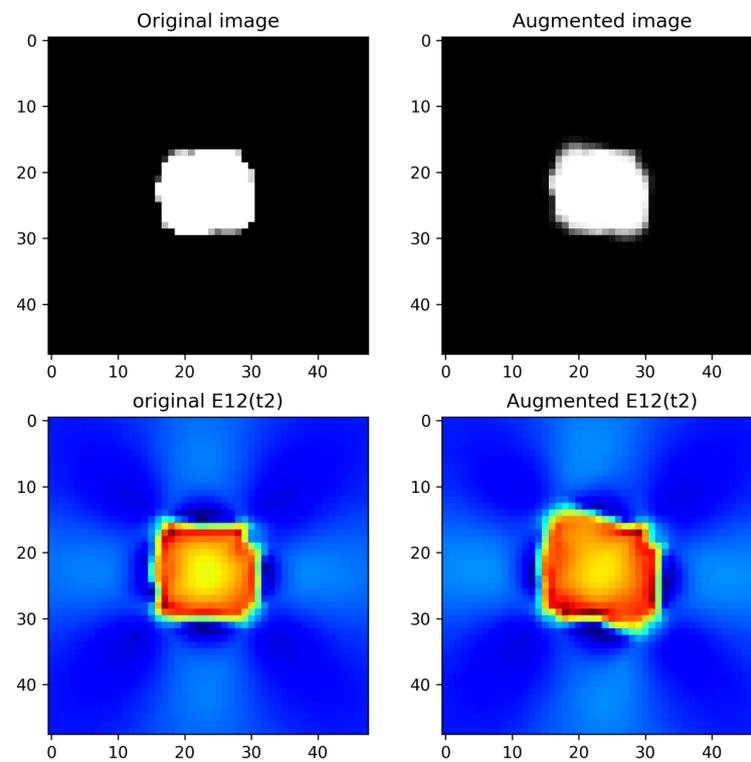


Figure 5. Example of augmented multimodal data for digital twin $i = 2$, on the left original data, on the right augmented data, on first row μ_{\square} , second row the shear strain at time $t2$.

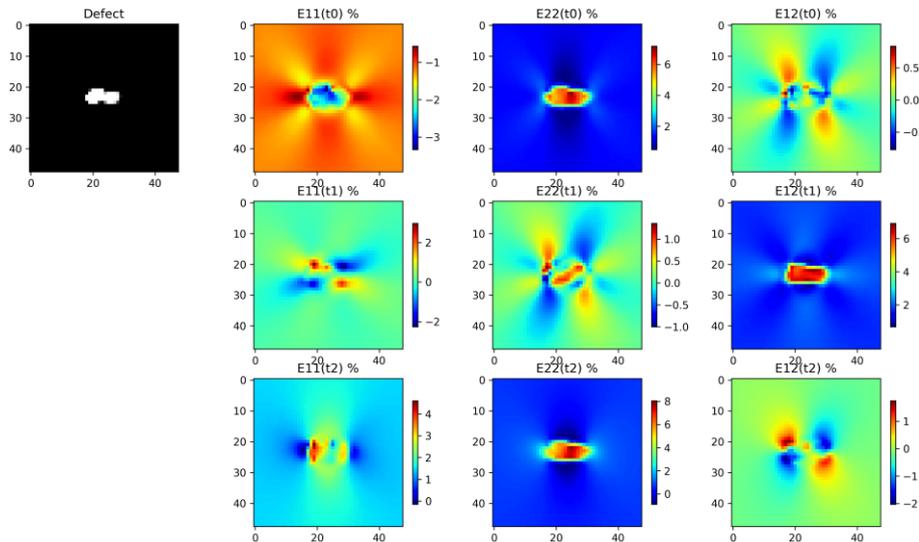


Figure 6. Strain components of the LROM for digital twin $i = 1$, on top left μ_{\square} , on the right $\varepsilon(\mathbf{V}^{(1)})$. Here, t_0, t_1, t_2 stand for the indices of the vectors that span the LROM.

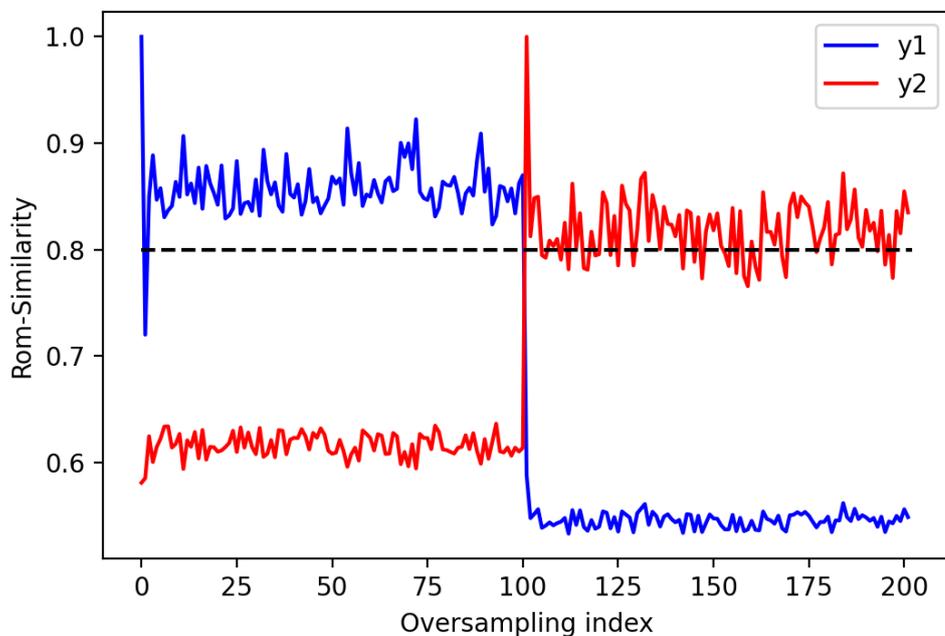


Figure 7. ROM-Similarity for all instances of augmented data in $\tilde{\mathcal{D}}^{(i)}$. Augmented data related to digital twin 1 have indexes ranging from 1 to 100, Augmented data related to digital twin 2 have indexes ranging from 101 to 200. The dashed line is the constrain $\eta^{(i)}(\mathbf{u}') < \alpha \epsilon_{tol}$.

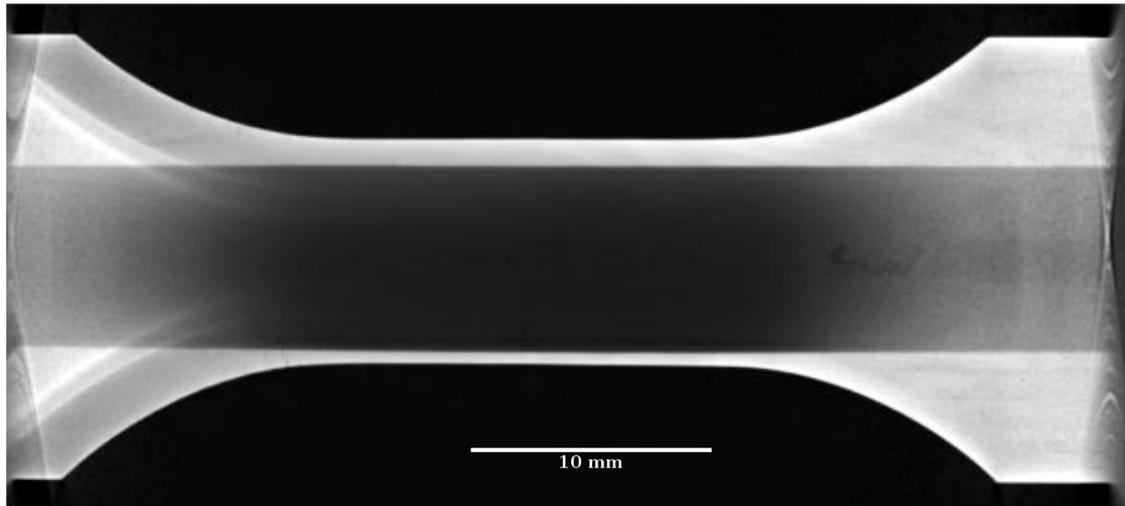


Figure 8. X-Ray computer tomography image of the mechanical specimen.

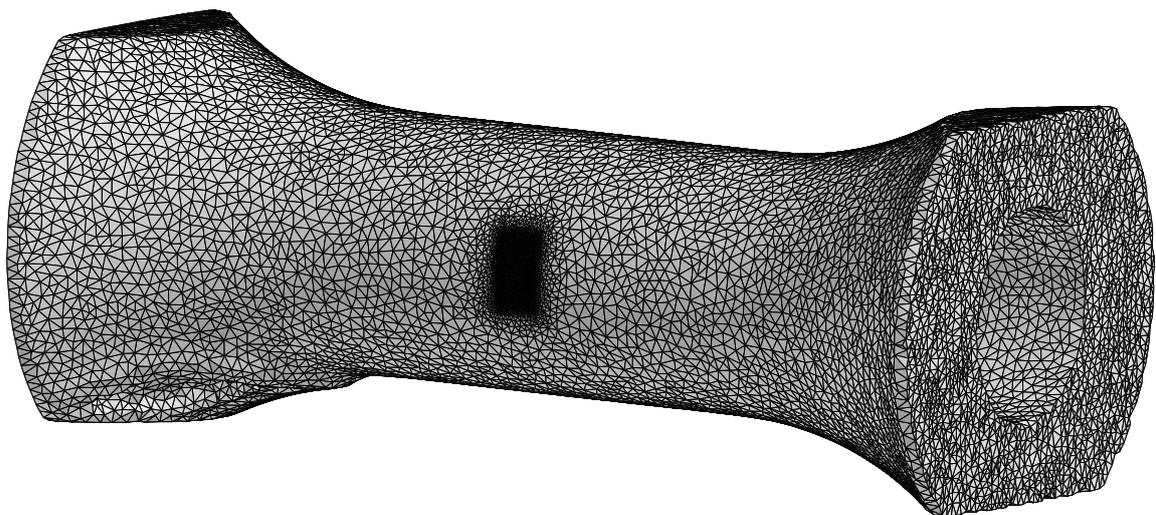


Figure 9. 3D finite element mesh based on the X-Ray computer tomography image.

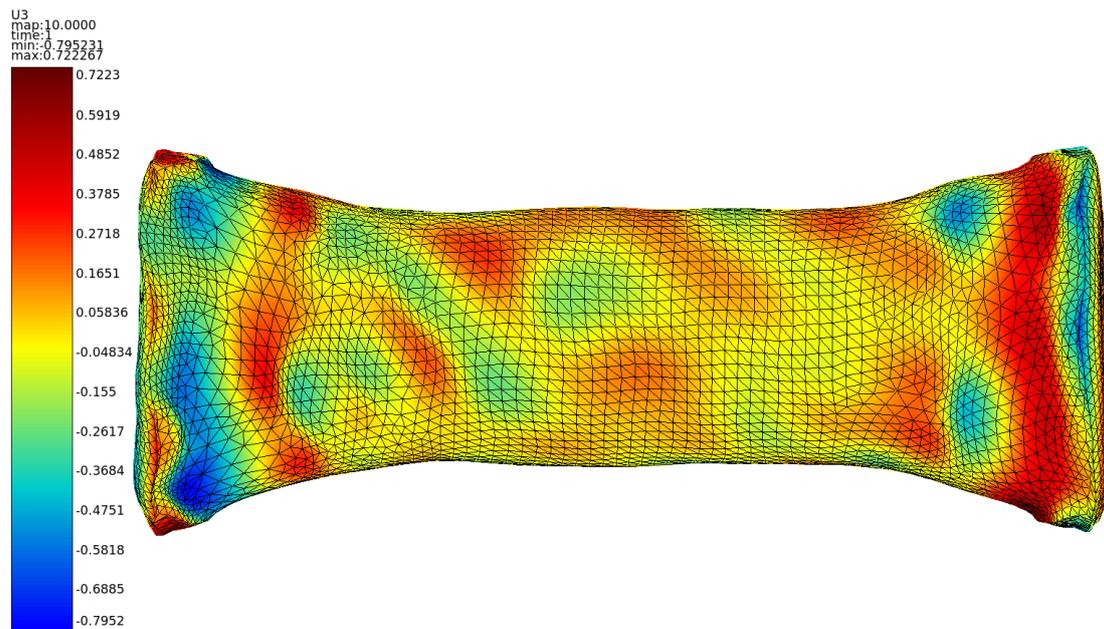


Figure 10. Example of node displacements along axial direction for mesh morphing according to augmented data for geometrical variations $\Delta\mu_{\square}$.

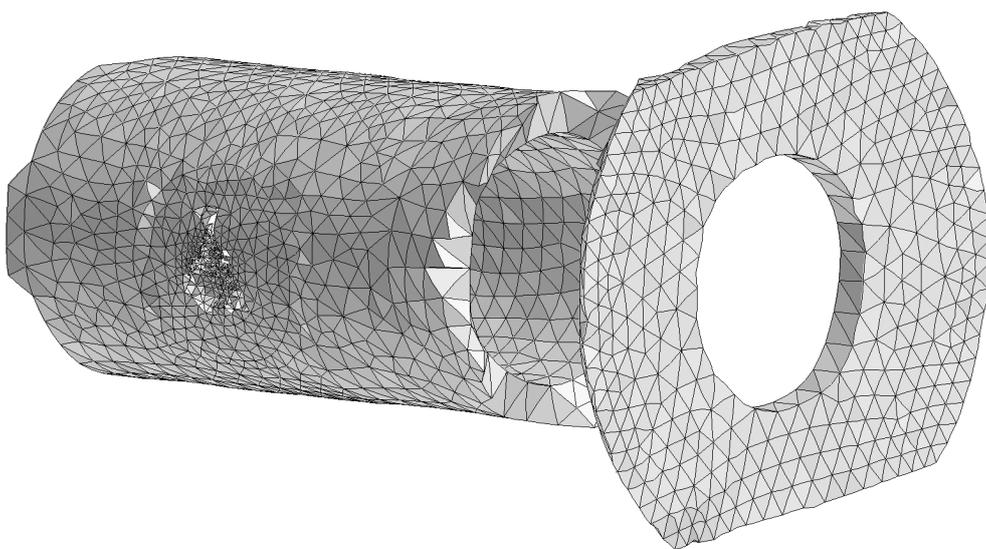


Figure 11. Finite element mesh restrained to Ω_R ($\mathcal{N}/\text{card}(\overline{\mathcal{F}}) = 2$). Most removed elements are located around the hole.

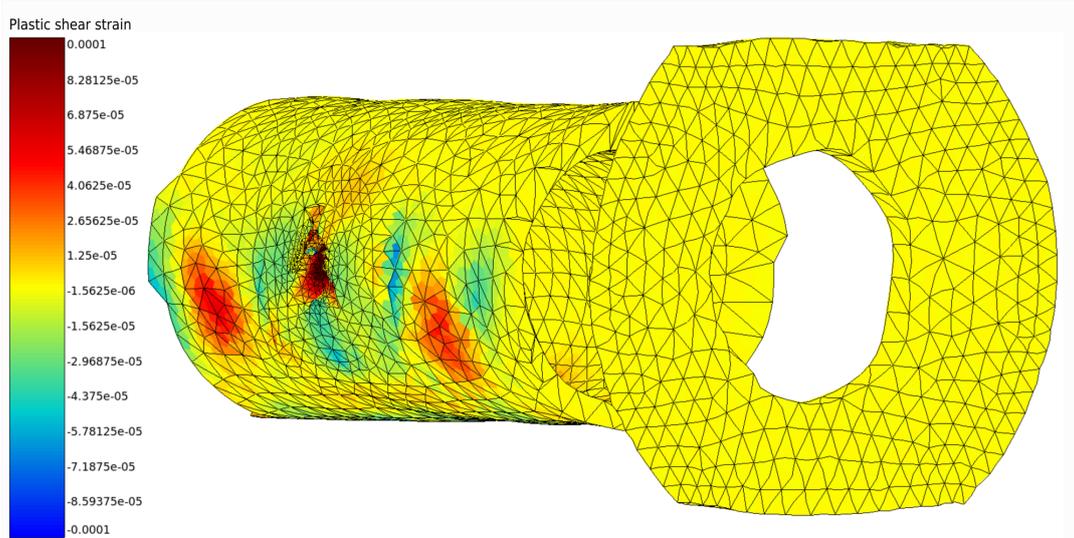


Figure 12. Shear strain forecast by the finite element model, in Ω_R , at the maximum mechanical loading of the specimen.

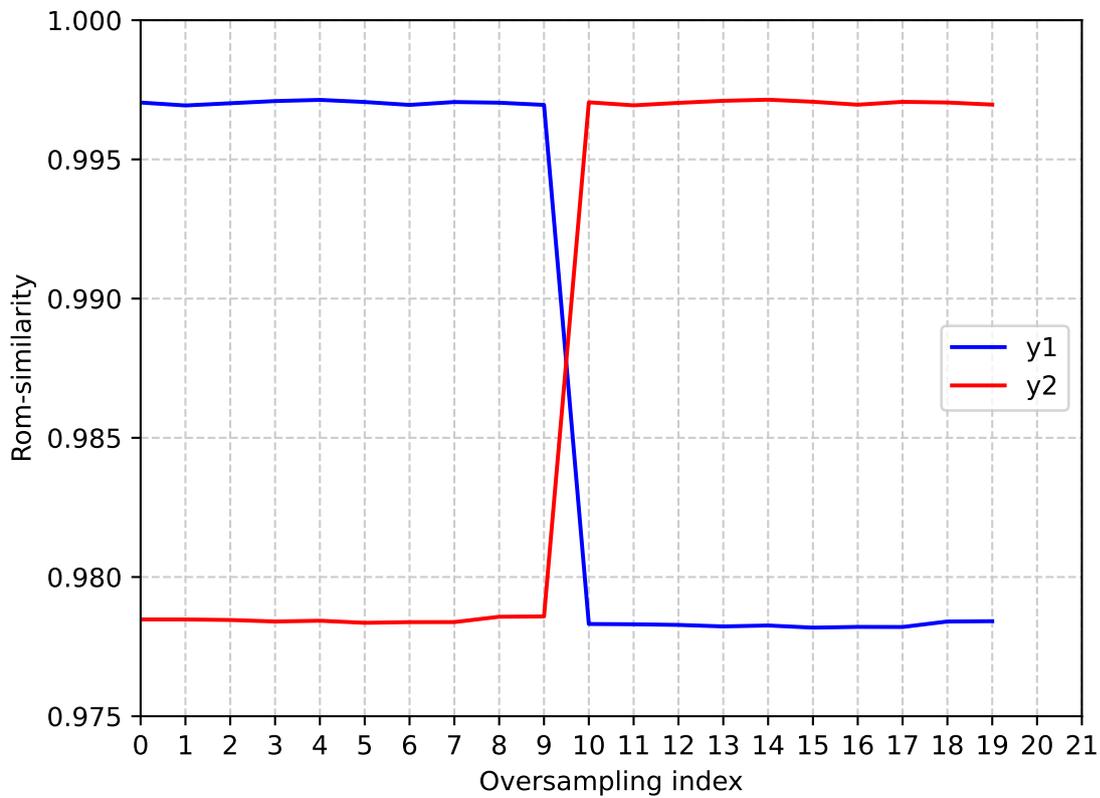


Figure 13. ROM-Similarity for all instances of augmented data in $\tilde{\mathcal{D}}^{(i)}$. Augmented data related to digital twin 1 have indexes ranging from 0 to 9, Augmented data related to digital twin 2 have indexes ranging from 10 to 19.