



Automated constitutive modeling of isotropic hyperelasticity based on artificial neural networks

Karl A. Kalina¹ · Lennart Linden¹ · Jörg Brummund¹ · Philipp Metsch¹ · Markus Kästner¹

Received: 30 July 2021 / Accepted: 31 August 2021 / Published online: 6 October 2021
© The Author(s) 2021

Abstract

Herein, an artificial neural network (ANN)-based approach for the efficient automated modeling and simulation of isotropic hyperelastic solids is presented. Starting from a large data set comprising deformations and corresponding stresses, a simple, physically based reduction of the problem's dimensionality is performed in a data processing step. More specifically, three deformation type invariants serve as the input instead of the deformation tensor itself. In the same way, three corresponding stress coefficients replace the stress tensor in the output layer. These initially unknown values are calculated from a linear least square optimization problem for each data tuple. Using the reduced data set, an ANN-based constitutive model is trained by using standard machine learning methods. Furthermore, in order to ensure thermodynamic consistency, the previously trained network is modified by constructing a pseudo-potential within an integration step and a subsequent derivation which leads to a further ANN-based model. In the second part of this work, the proposed method is exemplarily used for the description of a highly nonlinear Ogden type material. Thereby, the necessary data set is collected from virtual experiments of discs with holes in pure plane stress modes, where influences of different loading types and specimen geometries on the resulting data sets are investigated. Afterwards, the collected data are used for the ANN training within the reduced data space, whereby an excellent approximation quality could be achieved with only one hidden layer comprising a low number of neurons. Finally, the application of the trained constitutive ANN for the simulation of two three-dimensional samples is shown. Thereby, a rather high accuracy could be achieved, although the occurring stresses are fully three-dimensional whereas the training data are taken from pure two-dimensional plane stress states.

Keywords Data-driven approach · Artificial neural networks · Hyperelasticity · Finite element simulation

1 Introduction

Continuum-based theories combined with the finite element (FE) method provide a powerful tool for the virtual design and testing of engineering components. Regarding classical solid mechanical problems, the FE formulation is based on the weak form of the balance of linear momentum which is a universal and well known physical principle corresponding to the set of material independent equations [24,49]. Furthermore, in order to simulate components which consist of a specific material, constitutive equations which describe the behavior of the considered material have to be formulated mathematically and parameterized based on experimental

or micro-mechanically generated data. Since this process is however still a challenging task for single phase materials or composites which reveal complex behavior as highly nonlinear elasticity, anisotropy or dissipative properties, it could be useful to automate or circumvent the formulation of constitutive equations. For this purpose, several approaches – generally referred to as data-based or data-driven methods – exist which have become increasingly popular in the computational mechanics community during the last years [4,47]. In the following, a brief overview of these methods is given.

Pure data-based methods in solid mechanics A first possibility to replace classical constitutive equations is the direct data-driven mechanics approach, recently proposed by Kirchdoerfer and Ortiz [31]. This method totally avoids the formulation of constitutive equations and uses sets of stress-strain tuples which characterize the material's behavior instead. Consequently, a data-driven solver seeks to minimize the distance between the searched solution (σ, ϵ)

✉ Markus Kästner
markus.kaestner@tu-dresden.de

¹ Institute of Solid Mechanics, TU Dresden, 01062 Dresden, Germany

and the material data set within a proper energy norm, while compatibility and equilibrium have to be satisfied simultaneously. In the meanwhile, several extensions and applications of the direct data-driven approach have been proposed, e. g., for noisy data sets [3,32], dynamics [33], finite strains [48], inelasticity [12], fracture mechanics [5] or bio-mechanics [51].

In contrast to classical constitutive modeling, the data-driven approach obviously requires a huge set of data characterizing the specific behavior of a material. These data could be collected from experiments or, due to the steadily increasing computing power, generated in silico from lower scale simulations. Within the data-driven formalism the latter one has been applied by Karapiperis et al. [30] for sand recently. In order to enable the collection of data from experiments, Leygue et al. [38] and Stainier et al. [55] have introduced the data-driven identification which allows to collect strain-stress tuples from displacement fields and boundary conditions in inhomogeneous samples. An application of the data-driven identification to real experiments is shown in [10], the robustness regarding incomplete input data is discussed in [11].

A further possibility to circumvent the formulation of classical constitutive equations has been proposed by Ibañez et al. [25]. Therein, the construction of constitutive manifolds from collected data is described for elasticity and inelasticity. Based on this, an application to nonlinear elasticity is given in [26]. Finally, the strict fulfillment of the 2nd law of thermodynamics, i.e., the thermodynamic consistency, during the construction of constitutive manifolds is possible by using the GENERIC paradigm (General Equation for Non-Equilibrium Reversible-Irreversible Coupling) [22].

Besides the discussed direct data-driven approach and the construction of constitutive manifolds which are comparatively new data-based methods, the application of artificial neural networks (ANNs) as a mechanical constitutive model have already been proposed in the early 90s by Ghabussi et al. [19]. Due to their easy use, straight forward integrability into FE codes [23] and the excellent ability to learn complex relationships, the usage of ANNs is the most common data-based constitutive modeling method in the computational mechanics community. Thereby, the topology of the ANN, i.e., the network architecture, plays a decisive role in its overall functioning. Thus, the so called input, hidden and output layers differ in their basic tasks. Accordingly, input variables, e. g., strains $\boldsymbol{\epsilon}$, enter into the system in the input layer and are processed further in the hidden layer. Finally, the signals that reach the output layer are the output variables, e. g., stresses $\boldsymbol{\sigma}$. In so called feedforward neural networks (FNNs) connections between the neurons are always uniquely directed. In contrast, ANNs for which feedback is explicitly desired are called recurrent neural networks (RNNs), cf. [34]. Both types of networks have been used

for the mechanical constitutive modeling intensively. For instance, to model one-dimensional viscoelasticity [1], viscoplasticity [17] or time dependent behavior of concrete [28]. Likewise, several two- and three-dimensional ANN-based models could be found, e. g., to describe the behavior of geomaterials [20] or nonlinear viscoplasticity [56]. In order to ensure thermodynamic consistency of the trained ANN-models, Masi et al. [44] recently suggested the concept of thermodynamics-based ANNs for the constitutive modeling. Thereby, a scalar energy function is approximated by the network instead of the stress $\boldsymbol{\sigma}$. In order to still enable the training with respect to $\boldsymbol{\sigma}$, customized training loops in which gradients of the energy with respect to the input variables occur in the loss, see also [9,58].

Finally, ANNs have been used to set up so called decoupled multiscale schemes [29,57] which enable computationally efficient macroscopic simulations based on the homogenized response of lower scale representative volume elements (RVEs). Thereby, in contrast to computationally expensive FE² schemes [35] which require an RVE-simulation at each quadrature point in each iteration, the small scale processes within the heterogeneous microstructure that lead to the effective macroscopic response are implicitly captured via an appropriate macro- or surrogate-model. In this context, ANN-based models are predestined for the approximation of the homogenized data basis. For instance, this has been done for the scale bridging of composites with cubic microstructures [37], the simulation of the elastic-plastic deformation behavior of open-cell foam structures [52] and to link stresses and strains or traction-separation curves obtained from molecular dynamics simulations to the continuum scale [6,14], respectively. A combination of classical FE² simulations with on-the-fly adaptive switching to ANN-based surrogate models is shown by Fritzen et al. [16].

Hybrid approaches in solid mechanics The big advantage of pure data-based methods, i.e., the potential to model the reality with high accuracy in an automatized manner, is also countered by some disadvantages: namely the loss of physical sense, the lack of data and the possibly high computational cost. According to that, a combination of classical constitutive models with data-based methods or even an enrichment with physical knowledge – the so called hybrid approach – seems to be a promising modeling strategy.

A first possibility to apply the hybrid approach is to enhance classical well established constitutive models with information from data. Thus, the stress $\boldsymbol{\sigma}$ is calculated as the sum of the model prediction $\boldsymbol{\sigma}^{\text{mod}}$ and a correction obtained from a data-based method which describes the deviation between the chosen model and the real data, i.e. $\boldsymbol{\sigma} - \boldsymbol{\sigma}^{\text{mod}}$. This technique has been used for the correction of hyperelastic models in the GENERIC formalism by González et al. [21] or for the correction of plasticity models by Ibañez et al. [27].

Likewise, it is possible to replace parts of classical models with data-based approaches as ANNs. This has been done by Settgest et al. [53] for irreversible elastic-plastic deformations and damage of metal foams, whereby the yield function and the evolution equations are described by ANNs. An application for the description of uncured natural rubber by combining the micro sphere model with FNNs and RNNs is shown by Zopf and Kaliske [59]. In Liu et al. [42], a collection of connected mechanistic building blocks with analytical homogenization solutions is used to describe complex macroscopic responses without the loss of essential physics. In the context of isotropic hyperelasticity Shen et al. [54] as well as Liang and Chandrashekhara [39] approximate the free energy function by an ANN-based model with three deformation-type invariants as the input and thus incorporate knowledge about the anisotropy class of the considered material. This methodology, i.e., the usage of invariants as the input instead of the strain, has been continued for materials with more general anisotropy [40,41].

Content Within this contribution, an ANN-based hybrid approach for the efficient automated modeling and simulation of isotropic hyperelastic solids is presented. Following Shen et al. [54], three deformation-type invariants are used as the input in order to incorporate physical knowledge. Unlike previous works, three corresponding stress coefficients which follow from simple linear least square optimizations replace the stress tensor in the output layer likewise. Thus, a fully physically-based reduction of the problem’s dimensionality is applied for both, input and output quantities, respectively. With the reduced data set, an ANN is trained by using standard machine learning methods. A combination of the predicted stress coefficients with derivatives of the invariants with respect to the deformation finally enables the calculation of the stress. Furthermore, in order to ensure thermodynamic consistency a posteriori, the previously trained network is modified by constructing a pseudo-potential within an integration step and a subsequent derivation which leads to a further ANN-based model. The proposed strategy enables the efficient application of ANNs for the constitutive modeling by using standard training algorithms. It is exemplarily applied to the description of a highly nonlinear Ogden type material [49], where only two-dimensional plane stress data are needed for the training of a fully three-dimensional model. The trained ANNs are included into an open source FE software to test their approximation quality in different three-dimensional simulation examples.

The organization of the paper is as follows: In Sect. 2, the basic equations of the finite strain continuum mechanics theory as well as general constitutive relations for isotropic hyperelastic solids are given. After this, a general framework for the ANN-based modeling of isotropic hyperelasticity is presented in Sect. 3. The developed approach is exemplarily applied to hypothetical data which were generated by using

a Ogden model in Sect. 4. Therein, the data processing and training processes as well as a validation by means of representative FE simulations are shown. After a discussion of the results, the paper is closed by concluding remarks and an outlook to necessary future work in Sect. 5.

2 Continuum solid mechanics and hyperelasticity

In this section, the main relations of hyperelastic constitutive models including the relevant basic continuum mechanical equations are summarized. For a detailed overview, the reader is referred to the textbooks of Holzapfel [24] or Ogden [49]. In order to describe the physical state of the considered solids, the space of tensors

$$\mathcal{L}_n := \underbrace{\mathbb{R}^3 \otimes \dots \otimes \mathbb{R}^3}_{n\text{-times}} \quad \forall n \in \mathbb{N}_{\geq 1}, \tag{1}$$

except for a tensor of rank zero, is defined, where \mathbb{R}^3 is the Euclidean vector space and \mathbb{N} is the set of natural numbers.

2.1 Kinematics and balance laws

Regarding the reference and current configurations of a material body given by $\mathcal{B}_0 \subset \mathbb{R}^3$ and $\mathcal{B} \subset \mathbb{R}^3$, the displacement vector $\mathbf{u} \in \mathcal{L}_1$ of a material point capturing the positions $\mathbf{X} \in \mathcal{B}_0$ at time $t_0 \in \mathbb{R}_{\geq 0}$ and $\mathbf{x} \in \mathcal{B}$ at time $t > t_0$ is given by $\mathbf{u}(\mathbf{X}, t) := \boldsymbol{\varphi}(\mathbf{X}, t) - \mathbf{X}$. Therein, $\boldsymbol{\varphi}(\mathbf{X}, t)$ represents the bijective motion function which is continuous in space and time. With that, in order to describe the state of deformation, the deformation gradient $\mathbf{F} \in \mathcal{L}_2$ and its determinant are defined by

$$\mathbf{F} := (\nabla_{\mathbf{X}} \boldsymbol{\varphi})^T \quad \text{and} \quad J := \det \mathbf{F} > 0, \tag{2}$$

where $\nabla_{\mathbf{X}} := \mathbf{E}_K \partial_{X_K}$ is the nabla-operator with respect to the reference configuration. Therein, $\mathbf{E}_K \in \mathcal{L}_1$ denotes the Cartesian basis vector. Furthermore, the positive definite right Cauchy-Green deformation tensor $\mathbf{C} := \mathbf{F}^T \cdot \mathbf{F} \in \mathcal{S}ym$ with $\mathcal{S}ym := \{\boldsymbol{\tau} \in \mathcal{L}_2 \mid \boldsymbol{\tau} = \boldsymbol{\tau}^T\}$ is introduced.

To complete the necessary general equations, relevant balance laws are given in short. Supposing mass conservation, the balance of linear momentum with neglected inertia terms is given by

$$\nabla \cdot \boldsymbol{\sigma} + \varrho \mathbf{f} = \mathbf{0} \quad \text{or} \quad \nabla_{\mathbf{X}} \cdot (\mathbf{T} \cdot \mathbf{F}^T) + \varrho_0 \mathbf{f} = \mathbf{0}, \tag{3}$$

where the symmetry of the introduced Cauchy and 2nd Piola-Kirchhoff stress tensors $\boldsymbol{\sigma} \in \mathcal{S}ym$ and $\mathbf{T} \in \mathcal{S}ym$ follows from the balance of angular momentum. These tensors are related via the equation $\mathbf{T} := J \mathbf{F}^{-1} \cdot \boldsymbol{\sigma} \cdot \mathbf{F}^{-T}$. The symbols

f , ϱ and ϱ_0 denote a mass specific force and the mass density with respect to \mathcal{B} and \mathcal{B}_0 , respectively. Finally, in order to ensure thermodynamic consistency, all physical processes have to fulfill the Clausius-Duhem-inequality. If thermal effects are neglected, it is given by

$$-\dot{\Psi} + \frac{1}{2} \mathbf{T} : \dot{\mathbf{C}} \geq 0. \tag{4}$$

In the equation above, Ψ denotes the Helmholtz free energy density with respect to $dV_0 \subset \mathcal{B}_0$ and $(\dot{\cdot})$ is the material time derivative.

2.2 Constitutive relations for isotropic hyperelastic materials

Regarding a hyperelastic solid, one finds from Eq. (4) and applying the principle of Coleman and Noll [7], that the stress tensor \mathbf{T} follows from the Helmholtz free energy density Ψ by

$$\mathbf{T} = 2 \frac{\partial \Psi}{\partial \mathbf{C}}. \tag{5}$$

If the behavior is restricted to isotropy, i.e. the relation $\Psi(\mathbf{C}) = \Psi(\mathbf{Q} \cdot \mathbf{C} \cdot \mathbf{Q}^T)$ holds for all special orthogonal tensors $\mathbf{Q} \in Orth^+ := \{\boldsymbol{\tau} \in \mathcal{L}_2 \mid \boldsymbol{\tau}^T = \boldsymbol{\tau}^{-1}, \det \boldsymbol{\tau} \equiv 1\}$, the free energy function can be given by $\Psi = \Psi(I_1, I_2, I_3)$ including

$$I_1 := \text{tr } \mathbf{C}, \quad I_2 := \frac{1}{2} (\text{tr}^2 \mathbf{C} - \text{tr } \mathbf{C}^2) \quad \text{and} \quad I_3 := \det \mathbf{C}, \tag{6}$$

where I_α with $\alpha \in \{1, 2, 3\}$ denote the invariants of \mathbf{C} . Using the definition (6), the stress follows to

$$\mathbf{T} = \sum_{\alpha=1}^3 \underbrace{2 \frac{\partial \Psi}{\partial I_\alpha}}_{=: f_\alpha} \underbrace{\frac{\partial I_\alpha}{\partial \mathbf{C}}}_{=: \mathbf{G}^\alpha} = \sum_{\alpha=1}^3 f_\alpha \mathbf{G}^\alpha. \tag{7}$$

Therein, f_α and $\mathbf{G}^\alpha \in \mathcal{L}_2$ denote stress coefficients related to the invariants I_α and corresponding tensor generators, respectively.

3 ANN-based constitutive modeling framework for isotropic elasticity

After the continuum mechanical basics, the following section introduces the main idea of this work, i.e. the way how the constitutive ANN is efficiently trained by a suitable data set $\mathcal{D} := \{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_n\}$ consisting of $n \in \mathbb{N}$ data tuples $\mathcal{D}_i := ({}^i \mathbf{C}, {}^i \mathbf{T}) \in \mathbb{R}^{6 \times 1} \times \mathbb{R}^{6 \times 1}$. The introduced vectors ${}^i \mathbf{C}$ and ${}^i \mathbf{T}$ contain the coordinates of the

symmetric deformation and stress tensors, respectively: ${}^i \mathbf{C} = ({}^i C_{11}, {}^i C_{22}, \dots, {}^i C_{12})^T$, ${}^i \mathbf{T} = ({}^i T_{11}, {}^i T_{22}, \dots, {}^i T_{12})^T$. Within the proposed framework, known physical conditions and constraints are incorporated, where – as mentioned in the beginning – a restriction to isotropic hyperelasticity is made for now. The framework is subdivided into the steps

- (a) data mining,
- (b) data processing,
- (c.1) training process, and
- (c.2) thermodynamically consistent correction.

A scheme representing the procedure of the approach is depicted in Fig. 1. The single steps are described detailed in the following.

3.1 Data mining (a)

To start with, the data basis for the training of the network has to be achieved. Thereby it is important to collect a broad set of stress strain states to ensure the correct description of the material’s behavior later on. Ideally, the data mining has to be done based on real experiments of complex shaped samples in combination with digital image correlation [10]. The evaluation of such experiments with the algorithm proposed by Leygue et al. [38] enables the extraction of a high number of multiaxial stress-strain states for the considered material. However, in want of such experiments, the data basis is generated numerically by means of virtual experiments for now. To this end, a virtual sample is tested in several FE simulations. Within these virtual experiments which are described in the examples given in Sect. 4, the required data tuples \mathcal{D}_i are collected at the quadrature points of the finite elements.

3.2 Data processing – transforming the data into the space of invariants (b)

Now it is assumed that a suitable data set \mathcal{D} of a specific material is available. In order to achieve a constitutive ANN which is able to compute the stress \mathbf{T} from a given state of deformation \mathbf{C} with high accuracy and at the same time as few neurons as possible, it is essential to reduce the dimensionality of the input and output quantities. To this end, a variety of methods exist, e.g. the locally linear embedding (LLE) technique which is used in Ibañez et al. [25].

Following the works of Shen et al. [54] or Liang and Chandrashekhara [39], such a transformation is naturally given for the input values by using the invariants according to Eq. (6) instead of the coordinates of the deformation tensor \mathbf{C} .¹

¹ In the case of non-isotropic materials, the input values can be transformed from the coordinates of \mathbf{C} into a set of invariants which contains structural tensors, see Linka et al. [41].

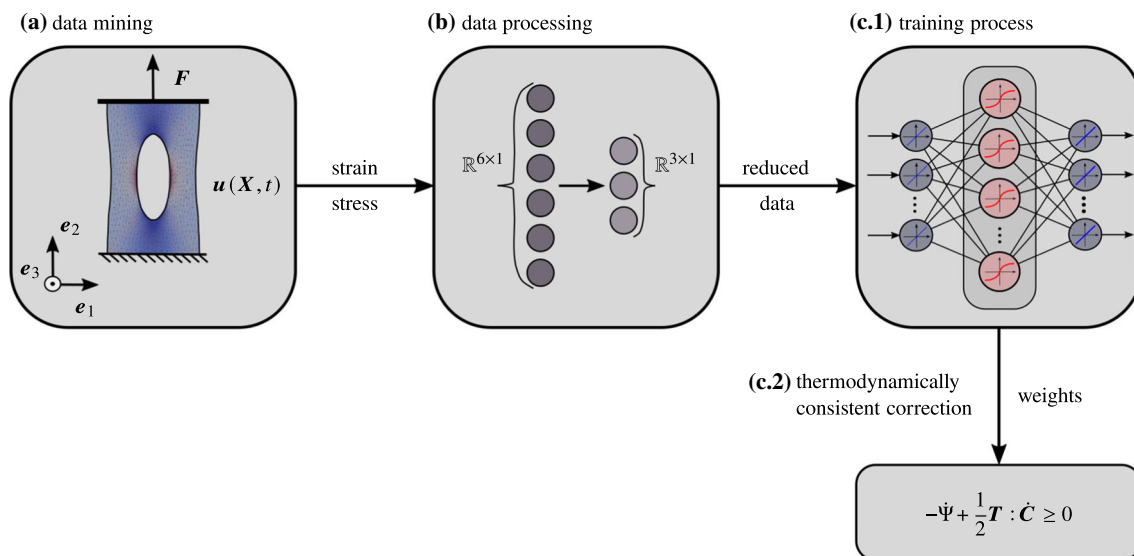


Fig. 1 Procedure of the ANN-based constitutive modeling framework for isotropic elasticity with four steps. **a** data mining: collection from stress-strain tuples from FE simulations, **b** data processing: reduction of the data set’s dimensionality, **c.1** training process: training of the constitutive ANN within the reduced data space, and **c.2** thermodynamically consistent correction: adjustment of the weights based on the Clausius-Duhem inequality

However, in contrast to these works in which the free energy density $\Psi : \mathcal{Sym} \rightarrow \mathbb{R}_{\geq 0}$ is approximated by the ANN and the stress tensor \mathbf{T} is not trained directly, quantities related to \mathbf{T} should serve as the output in the following. Consequently, the evaluation of the path integral

$$\Psi(\mathbf{C}) = \frac{1}{2} \int_{\mathcal{P}_{\mathbf{C}}} \mathbf{T}(\tilde{\mathbf{C}}) : d\tilde{\mathbf{C}} \tag{8}$$

with $\mathcal{P}_{\mathbf{C}} : [a, b] \rightarrow \mathcal{Sym}$, $\mathcal{P}_{\mathbf{C}}(a) = \mathbf{1}$, $\mathcal{P}_{\mathbf{C}}(b) = \mathbf{C}$ from the data tuples \mathcal{D}_i could be avoided.² In order to reduce the dimensionality of the output variables anyhow, the isotropy property according to Eqs. (6) and (7) is used so that the variables f_α replace the coordinates of \mathbf{T} . The ANN thus directly approximates on the level of stresses which is preferable for the solution of the balance of linear momentum (3).

In contrast to the deformation type invariants I_α , the stress coefficients f_α cannot be determined directly since the analytical expression for the free energy function Ψ is unknown. However, if it is assumed that $\Psi(I_1, I_2, I_3)$ exists, the relation

$${}^i \mathbf{T} = \sum_{\alpha=1}^3 {}^i f_\alpha {}^i \mathbf{G}^\alpha \tag{9}$$

holds for each tuple \mathcal{D}_i . Consequently, the three stress coefficients ${}^i f_\alpha$ could be determined from a simple least square optimization for each tuple. For the cases of states with different eigenvalues $\lambda_\alpha^2 \in \mathbb{R}_{\geq 0}$ of the deformation tensor \mathbf{C} , the optimization problem is given by

$${}^i \underline{\mathbf{f}} = \arg \min_{\underline{\mathbf{f}} \in \mathbb{R}^{3 \times 1}} \left\| {}^i \mathbf{T} - \sum_{\alpha=1}^3 {}^i f_\alpha {}^i \mathbf{G}^\alpha \right\|^2 \quad \forall i \in \{1, \dots, n\}, \tag{10}$$

where $\| \cdot \|$ denotes the Frobenius norm of a rank two tensor. Due to the convexity with respect to ${}^i f_\alpha$, the solution of Eq. (10) is unique and follows from a linear optimization. With that, each data tuple \mathcal{D}_i could be transformed to a tuple $\mathcal{D}_i^{\text{red}} := ({}^i \underline{\mathbf{I}}, {}^i \underline{\mathbf{f}}) \in \mathbb{R}^{3 \times 1} \times \mathbb{R}^{3 \times 1}$ of reduced dimensionality.

In Fig. 2, the transformation step is exemplarily visualized for deformation-stress states collected from an FE simulation of a uniaxially pulled disc with a hole. Thereby, the constitutive behavior is given by a hyperelastic Ogden model [49], cf. Sect. 4.1.1. In the transformation process, a transition from a rather disordered set of data to well-ordered structures in the reduced space takes place. Furthermore, as part of the transformation process, the cardinality n of the reduced data set $\mathcal{D}^{\text{red}} := \{\mathcal{D}_1^{\text{red}}, \mathcal{D}_2^{\text{red}}, \dots, \mathcal{D}_n^{\text{red}}\}$ can be decreased by several orders of magnitude compared to the original set \mathcal{D} , by just removing similar tuples within a given tolerance. Con-

² In general, the path integral given in Eq. (8) have to be evaluated numerically, cf. Shen et al. [54]. This is due to the fact that the function $\mathbf{T}(\mathbf{C})$ is unknown at this step.

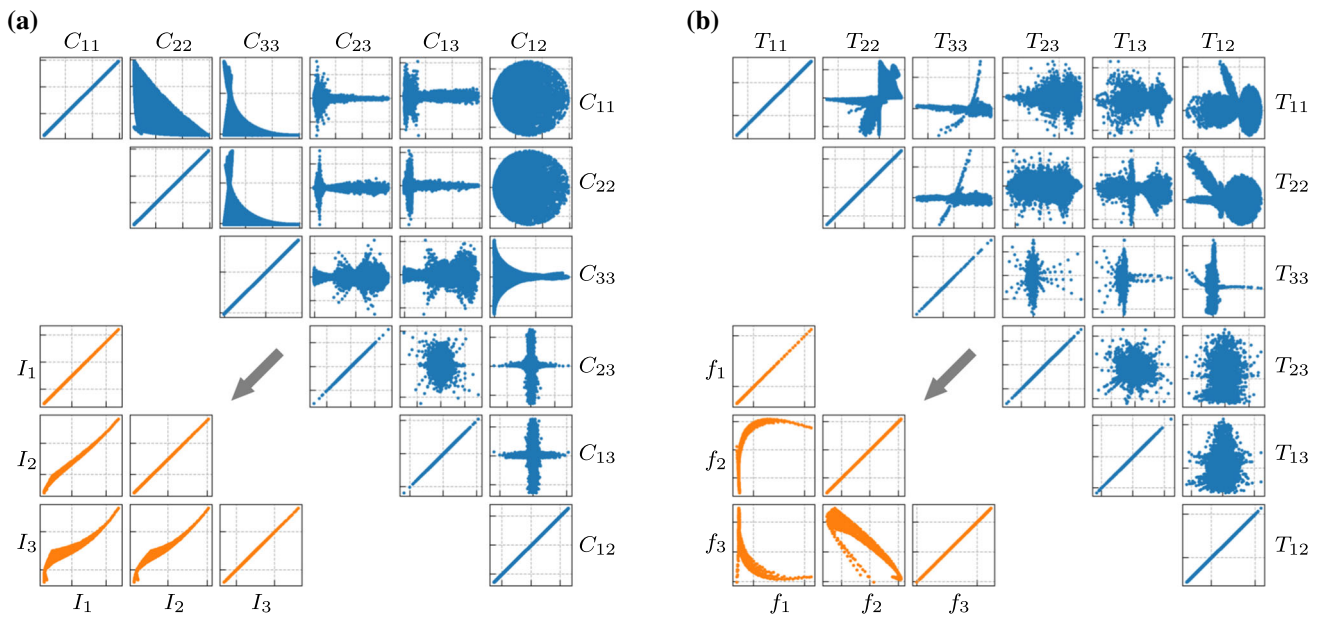


Fig. 2 Transformation of a hyperelastic stress-strain data set \mathcal{D} into the reduced space \mathcal{D}^{red} consisting of strain invariants I_α and stress coefficients f_α . Correlation plots of the transformed quantities: **a** deformation components and strain invariants, as well as **b** stress components and stress coefficients

sequently, the training of the constitutive ANN within the reduced space \mathcal{D}^{red} is much more efficient later on.

Considering deformation states with multiple eigenvalues such as the undeformed one, where $\mathbf{C} = \mathbf{1}$, the uniqueness of Eq. (10) is not given anymore. Thus, the true stress coefficients, which – according to Eq. (7) – have to follow from a free energy function, are not determinable as described above. Due to this, the sought values f_α have to be calculated by means of an interpolation. Each function $f_\alpha(I_1, I_2, I_3)$ is now approximated by a hyperplane in the invariant space. Thus, within the surrounding of the deformation state \mathbf{C}^* , the ansatz

$$f_\alpha^{\text{int}}(I_1, I_2, I_3, \underline{\kappa}) := \kappa_{\alpha 0} + \sum_{\beta=1}^3 \kappa_{\alpha\beta} I_\beta \quad \text{with} \quad (11)$$

$$\underline{\kappa} := (\kappa_{10}, \dots, \kappa_{13}, \kappa_{20}, \dots, \kappa_{32}, \kappa_{33})^T \in \mathbb{R}^{12 \times 1} \quad (12)$$

for all $\alpha \in \{1, 2, 3\}$ is chosen. The parameters $\underline{\kappa}$ can now be found from a constrained optimization problem which ensures that the corresponding state of stress \mathbf{T}^* is correct. In this case with $\beta \in \{1, 2, 3\}$ it is given by

$$\underline{\kappa} = \arg \min_{\underline{\kappa} \in \mathcal{C}} \frac{1}{2} \sum_{k=1}^n \sum_{\alpha=1}^3 \left(f_\alpha^{\text{int}}(I_\beta({}^k\mathbf{C}), \underline{\kappa}) - {}^k f_\alpha \right)^2 \quad (13)$$

with respect to the corresponding constraint subset

$$\mathcal{C} := \left\{ \underline{\kappa} \in \mathbb{R}^{12 \times 1} : \mathbf{T}^* = \sum_{\alpha=1}^3 f_\alpha^{\text{int}}(I_\beta(\mathbf{C}^*), \underline{\kappa}) \partial_{\mathbf{C}} I_\alpha(\mathbf{C}^*) \right\}. \quad (14)$$

3.3 Training of the constitutive ANN (c.1)

Based on the extracted and transformed data basis \mathcal{D}^{red} , a feed forward neural network has to be trained which could be done by using conventional training loops. In the examples given in Sect. 4, the fastest training could be achieved with the deep learning toolbox included in Matlab, where a Levenberg-Marquardt optimization procedure has been used. Since input data I_α and output data f_α are normalized to the range $[-1, 1]$, the ANN-based model for the stress coefficients is given by

$$f_\alpha^{\text{ANN}}(i_\beta) := \underbrace{\frac{f_{(\alpha)}^{\max} - f_{(\alpha)}^{\min}}{2}}_{=: X_\alpha} f_\alpha(i_\beta) + \underbrace{\frac{f_{(\alpha)}^{\max} + f_{(\alpha)}^{\min}}{2}}_{=: Y_\alpha}, \quad (15)$$

where the superscripts $(\cdot)^{\max}$ and $(\cdot)^{\min}$ denote the maximum and minimum components of a given data set, respectively. Choosing hyperbolic tangent activation functions and restricting the network architecture to only one hidden layer, the normalized stress coefficients are given by

$$f_\alpha(i_\beta) := B_\alpha + \sum_{\beta=1}^N W_{\alpha\beta} \tanh \left(\sum_{\gamma=1}^3 w_{\beta\gamma} i_\gamma + b_\beta \right), \tag{16}$$

$$i_\beta(I(\beta)) := \left[I_\beta - \frac{I_\beta^{\max} + I_\beta^{\min}}{2} \right] \frac{2}{\underbrace{I_\beta^{\max} - I_\beta^{\min}}_{=: U_\alpha^{-1}}}, \tag{17}$$

$=: V_\alpha$

where $f_\alpha, i_\beta \in [-1, 1]$. In the equations above, $W_{\alpha\beta}, w_{\alpha\beta}$ and B_α, b_β are weights and bias values, so that the total number of training parameters follows to $p = 3 + 7N$. Using the introduced Eqs. (15)–(17), the linkage of the 2nd Piola-Kirchhoff stress tensor \mathbf{T} and the right Cauchy-Green deformation tensor \mathbf{C} is modeled by

$$\mathbf{T}^{\text{ANN}}(\mathbf{C}) = \sum_{\alpha=1}^3 f_\alpha^{\text{ANN}}(I_\beta(\mathbf{C})) \mathbf{G}^\alpha(I_\gamma(\mathbf{C}), \mathbf{C}). \tag{18}$$

The model equations have been implemented within the FE toolbox *FEniCS* [2,43]. Therein, the tangent

$$\begin{aligned} \mathbb{C}^{\text{ANN}} &:= 2 \frac{\partial \mathbf{T}}{\partial \mathbf{C}} \in \mathcal{L}_4 \\ &= 2 \sum_{\alpha=1}^3 \sum_{\beta=1}^3 \frac{\partial f_\alpha^{\text{ANN}}}{\partial I_\beta} \frac{\partial I_\alpha}{\partial \mathbf{C}} \otimes \frac{\partial I_\beta}{\partial \mathbf{C}} + 2 \sum_{\alpha=1}^3 f_\alpha^{\text{ANN}} \frac{\partial^2 I_\alpha}{\partial \mathbf{C} \partial \mathbf{C}}, \end{aligned} \tag{19}$$

which is required within the solution via a Newton–Raphson scheme, is calculated automatically by means of automatic differentiation. The reader is referred to Hashash et al. [23] for details concerning the implementation into classical FE codes.

3.4 Thermodynamically consistent correction (c.2)

With the conventional training of the constitutive ANN, an accurate approximation of the deformation-stress relation is possible. However, the thermodynamic consistency of the ANN-based model is not ensured in any case, since this would require the existence of a free energy density Ψ that is related to the stresses via Eq. (5). Thus, in order to satisfy thermodynamic consistency a priori, customized training loops which optimize the network not only with respect to the output which have to be chosen to Ψ , but also with respect to gradients $\partial_{\mathbf{C}} \Psi$ of the output quantities have to be used, cf. [9,41,44,58].

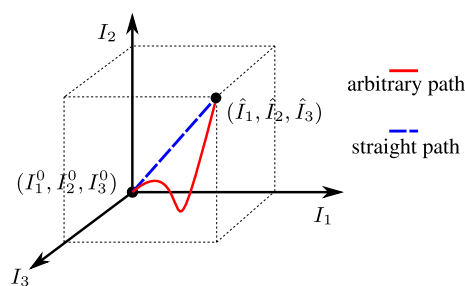


Fig. 3 Possible integration paths from the undeformed state \mathbf{C}^0 to a deformed state $\hat{\mathbf{C}}$ within the invariant space. The chosen straight path is given by $\tilde{I}_\alpha = (\hat{I}_\alpha - I_\alpha^0)p + I_\alpha^0$

Within this work, an alternative approach which allows to easily use standard machine learning algorithms is presented. It obligatorily enforces the fulfillment of the 2nd law of thermodynamics a posteriori and is based on the calculation of a pseudo-potential from a network which has been previously trained according to (c.1).

In a first step, the mechanical work W along a piecewise smooth path $\mathcal{P}_{\mathbf{C}} : [a, b] \rightarrow \mathcal{S}ym$ has to be calculated by the integration

$$W = \frac{1}{2} \int_{\mathcal{P}_{\mathbf{C}}} \mathbf{T}(\tilde{\mathbf{C}}) : d\tilde{\mathbf{C}} \tag{20}$$

using the approximated relation (18). By exploiting the differential relationship $dI_\alpha = \partial_{\mathbf{C}} I_\alpha : d\mathbf{C} = \mathbf{G}^\alpha : d\mathbf{C}$, the introduced integral (20) can be transformed into the space of invariants. Consequently, the path $\mathcal{P}_{\mathbf{C}}$ is now replaced by $\mathcal{P}_I : [a, b] \rightarrow \mathbb{R}^{3 \times 1}$ and it follows

$$\begin{aligned} W &= \sum_{\alpha=1}^3 \int_{\mathcal{P}_I} f_\alpha(\tilde{I}_1, \tilde{I}_2, \tilde{I}_3) d\tilde{I}_\alpha \\ &= \sum_{\alpha=1}^3 \int_a^b f_\alpha(\tilde{I}_1(p), \tilde{I}_2(p), \tilde{I}_3(p)) \tilde{I}'_\alpha(p) dp. \end{aligned} \tag{21}$$

To facilitate the calculation of Eq. (21), a path given by the parametrization $\tilde{I}_\alpha = (\hat{I}_\alpha - I_\alpha^0)p + I_\alpha^0$ with $p \in [0, 1]$ is chosen in the following, where I_α^0 denotes the values of the invariants according to Eq. (6) at $\mathbf{C} = \mathbf{1}$. A graphical visualization of an arbitrary and the chosen path is given in Fig. 3

Considering a trained constitutive ANN with only one hidden layer, the integral could be solved analytically. By using Eqs. (15)–(17), the mechanical work W follows to

$$W = \frac{1}{2} \sum_{\alpha=1}^3 \left[\sum_{\beta=1}^N \frac{(\hat{I}_\alpha - I_\alpha^0) X_\alpha W_{\alpha\beta}}{\sum_{\mu=1}^3 \frac{w_{\beta\mu}}{U_\mu} (\hat{I}_\mu - I_\mu^0)} \Lambda_\beta + B_\alpha X_\alpha + Y_\alpha \right],$$

$$\Lambda_\beta := \ln \left(\frac{\cosh \left(\sum_{\gamma=1}^3 \frac{w_{\beta\gamma}}{U_\gamma} (\hat{I}_\gamma - V_\gamma) + b_\beta \right)}{\cosh \left(\sum_{\kappa=1}^3 \frac{w_{\beta\kappa}}{U_\kappa} (I_\kappa^0 - V_\kappa) + b_\beta \right)} \right).$$

(22)

Now, the determined work W is assumed to define a pseudo-potential $\Psi^* := W^3$, so that an alternative model for the stress response can be obtained by $\mathbf{T}^* = 2\partial_C \Psi^*$. Since the corrected ANN-based model fulfills Eq. (5), the thermodynamic consistency is guaranteed in any case. The updated stress coefficients are given by

$$f_\varepsilon^{\text{ANN}^*} = Y_\varepsilon + X_{(\varepsilon)} B_\varepsilon + \sum_{\beta=1}^N \frac{X_{(\varepsilon)} W_{\varepsilon\beta}}{\sum_{\mu=1}^3 \frac{w_{\beta\mu}}{U_\mu} (I_\mu - I_\mu^0)} \Lambda_\beta$$

$$+ \sum_{\alpha=1}^3 \sum_{\beta=1}^N \frac{(I_\alpha - I_\alpha^0) \frac{w_{\beta\varepsilon}}{U_{(\varepsilon)}} X_\alpha W_{\alpha\beta}}{\sum_{\kappa=1}^3 \frac{w_{\beta\kappa}}{U_\kappa} (I_\kappa - I_\kappa^0)} \Phi_\beta$$

$$- \sum_{\alpha=1}^3 \sum_{\beta=1}^N \frac{(I_\alpha - I_\alpha^0) \frac{w_{\beta\varepsilon}}{U_{(\varepsilon)}} X_\alpha W_{\alpha\beta}}{\left(\sum_{\kappa=1}^3 \frac{w_{\beta\kappa}}{U_\kappa} (I_\kappa - I_\kappa^0) \right)^2} \Lambda_\beta,$$

$$\Phi_\beta := \tanh \left(\sum_{\gamma=1}^3 \frac{w_{\beta\gamma}}{U_\gamma} (I_\gamma - V_\gamma) + b_\beta \right)$$

(23)

including normalization weights X_α , Y_α , U_α and V_α . This results in the corrected ANN-based model

$$\mathbf{T}^{\text{ANN}^*}(\mathbf{C}) = \sum_{\alpha=1}^3 f_\alpha^{\text{ANN}^*}(I_\beta(\mathbf{C})) \mathbf{G}^\alpha(I_\gamma(\mathbf{C}), \mathbf{C}),$$

(24)

where the weights b_α , B_α , $w_{\alpha\beta}$ and $W_{\alpha\beta}$ – determined in the training step (c.1) – stay unchanged. Consequently, no customized training loops have to be designed for the application of the proposed approach. Similar to the original model (18), the thermodynamically corrected equations are implemented into *FEniCS*.

³ The scalar Ψ^* is termed a pseudo-potential since the underlying stress coefficients f_α^{ANN} do not have a conservative character in the general case. As a consequence of this, the Ψ^* naturally depends on the chosen integration path. However, in the case of a highly accurate stress prediction by the ANN, the path dependence should be nominal.

4 Examples

4.1 Data generation

In a first step, the data set for the training of the ANN has to be achieved. As mentioned above, the data basis is generated numerically in want of real experiments for now. To this end, a virtual sample is tested in several FE simulations. Within these virtual experiments, the required data set \mathcal{D} consisting of the tuples \mathcal{D}_i is collected at the quadrature points of the finite elements.

4.1.1 Constitutive behavior

In order to demonstrate the ability of the proposed ANN-based method, a highly nonlinear stress-strain relation is chosen for the constitutive behavior of the sample’s material. To this end, an Ogden model [49] which is given by the free energy density function

$$\Psi := \sum_{p=1}^{N_O} \frac{\mu_p}{\alpha_p} \left(\sum_{\beta=1}^{N_\lambda} \nu_\beta \bar{\lambda}_\beta^{\alpha_p} - 3 \right) + \frac{\kappa}{4} (J^2 - 2 \ln J - 1)$$

(25)

is used. In the equation above, μ_p , α_p and κ denote parameters of the Ogden model and the compression modulus. The symbols $\nu_\beta \in \{1, 2, 3\}$ and $N_\lambda \in \{1, 2, 3\}$ are the algebraic multiplicity of λ_β and the number of independent eigenvalues. Furthermore, $\bar{\lambda}_\beta := J^{-1/3} \lambda_\beta$ are the isochoric principal stretches which follow from the Flory split [15]

$$\mathbf{F} = J^{1/3} \mathbf{1} \cdot \bar{\mathbf{F}} \quad \text{with} \quad \bar{\mathbf{F}} = J^{-1/3} \mathbf{F} \quad \text{and} \quad \det \bar{\mathbf{F}} \equiv 1,$$

(26)

i.e. the decomposition of the deformation gradient into volumetric $J^{1/3} \mathbf{1}$ and isochoric $\bar{\mathbf{F}}$ parts. The stress of the Ogden model is given by

$$\mathbf{T} = \sum_{\beta=1}^{N_\lambda} \left[\frac{1}{\lambda_\beta^2} \sum_{p=1}^{N_O} \mu_p \left(\bar{\lambda}_\beta^{\alpha_p} - \frac{1}{3} \sum_{\gamma=1}^{N_\lambda} \nu_\gamma \bar{\lambda}_\gamma^{\alpha_p} \right) + \frac{\kappa}{2} \lambda_\beta^{-2} (J^2 - 1) \right] \mathbf{M}^\beta,$$

(27)

where $\mathbf{M}^\beta \in \mathcal{L}_2$ denotes the projection tensors of the right Cauchy-Green deformation tensor:

$$\mathbf{C} = \sum_{\beta=1}^{N_\lambda} \lambda_\beta^2 \mathbf{M}^\beta \quad \text{with} \quad \mathbf{M}^\beta := \delta_{1N_\lambda} \mathbf{1} + \prod_{\beta \neq \alpha} \frac{\mathbf{C} - \lambda_\beta^2 \mathbf{1}}{\lambda_\alpha^2 - \lambda_\beta^2}.$$

(28)

Table 1 Constitutive parameters of the Ogden model (27) within the virtual experiments. Initial shear modulus G^{init} and Poisson’s ratio ν^{init} as well as parameter sets μ_p, α_p and κ

G^{init} /kPa	ν^{init} /–	μ_1 /kPa	μ_2 /kPa	μ_3 /kPa	α_1 /–	α_2 /–	α_3 /–	κ /kPa
100.0	0.44	–26.62	29.04	0.0098	–5.0	2.3	12.0	800.0

For further details on the calculation of isotropic tensor functions, the interested reader is referred to Miehe [45,46].

The material parameters chosen for the virtual experiments are given in Table 1, where they are related to initial shear modulus G^{init} and Poisson’s ratio ν^{init} via

$$G^{\text{init}} = \frac{1}{2} \sum_{p=1}^{N_0} \alpha_p \mu_p \quad \text{and} \quad \kappa = \frac{2}{3} G^{\text{init}} \frac{1 + \nu^{\text{init}}}{1 - 2\nu^{\text{init}}}. \quad (29)$$

Regarding the stretch-stress curves for uniaxial loading, a highly nonlinear response is achieved with this parameter set, cf. Fig. 4. Here, the quantity $\mathbf{P} := J\mathbf{F}^{-1} \cdot \boldsymbol{\sigma} \in \mathcal{L}_2$ is the non-symmetric 1st Piola–Kirchhoff stress tensor.

4.1.2 Virtual experiments

In order to enable the training process of the constitutive ANN within a broad subset of the invariant space $\mathcal{I} \subset \mathbb{R}^{3 \times 1}$, the deformation states of the virtual samples have to be as heterogeneous as possible. With regard to real experiments [10,50], non-uniform thin discs, e. g., with holes, are chosen for this purpose. Thereby it has to be taken into account that only displacements on the surface of the sample are measurable in an experimental setting. A thin sample which could be loaded into a close to plane stress state is thus preferable.

Consequently, in order to obtain perfect plane stress data, the virtual experiments are realized within a two-dimensional

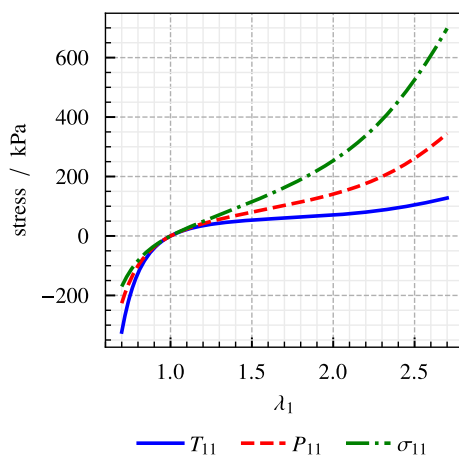


Fig. 4 Uniaxial tension test of the Ogden model (27) for the parameters given in Table 1: Stretch-stress curves for the Cauchy stress σ_{11} as well as the 1st and 2nd Piola–Kirchhoff stresses P_{11} and T_{11}

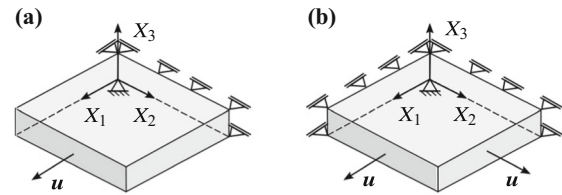


Fig. 5 Specimen loading with bearings and prescribed displacement u within the virtual experiments: **a** uniaxial tension and **b** equibiaxial tension

setting.⁴ A study analyzing two main influences – loading and hole geometry – with respect to the resulting sets of invariants is presented in the following. Thereby, the FE meshes were generated by using *Gmsh* [18].

Influence of the specimen loading on the data set In a first study, the two load cases (i) uniaxial tension and (ii) equibiaxial tension are considered for a specimen with circular hole (sample I). As shown in the schematic representations given in Fig. 5, displacement boundary conditions with a maximum value of $\hat{u} \leq 60$ mm for the uniaxial tension and a maximum value of $\hat{u} \leq 40$ mm for the equi-biaxial tension are prescribed within 40 increments. Each data set is depicted in the three invariant-planes I_1 – I_2 , I_1 – I_3 and I_2 – I_3 within Fig. 6a, whereby the homogeneous uniaxial as well as equi-biaxial tension and compression load cases are included for reasons of orientation. For the case of incompressible solids, where $I_3 \equiv 1$ holds, these curves are independent of the material’s behavior and delimit the area of admissible invariant sets within the I_1 – I_2 -plane [8,41]. However, these statements do not hold without restriction for the considered case of a compressible material.

Regarding the collected data sets in the invariant space, it can be seen that \mathcal{I} is widely sampled between the plotted curves for homogeneous uniaxial as well as equi-biaxial tension and compression load cases, see Fig. 6a. Furthermore, although only tensile tests were performed, data points in the compression range with $I_3 < 1$ can also be found due to the inhomogeneous stress and strain states occurring in the specimen. Finally, it is observed that \mathcal{I} is sampled over a much wider range by the equi-biaxial tensile test compared to the uniaxial one. As a resulting consequence for the investigated disc containing a circular hole geometry, the calibration of

⁴ The out of plane stress is enforced to zero by locally searching the out of plane stretch λ_3 from the implicit constraint $T_{33}(\lambda_1, \lambda_2, \lambda_3) = 0$, cf. Eq. (27).

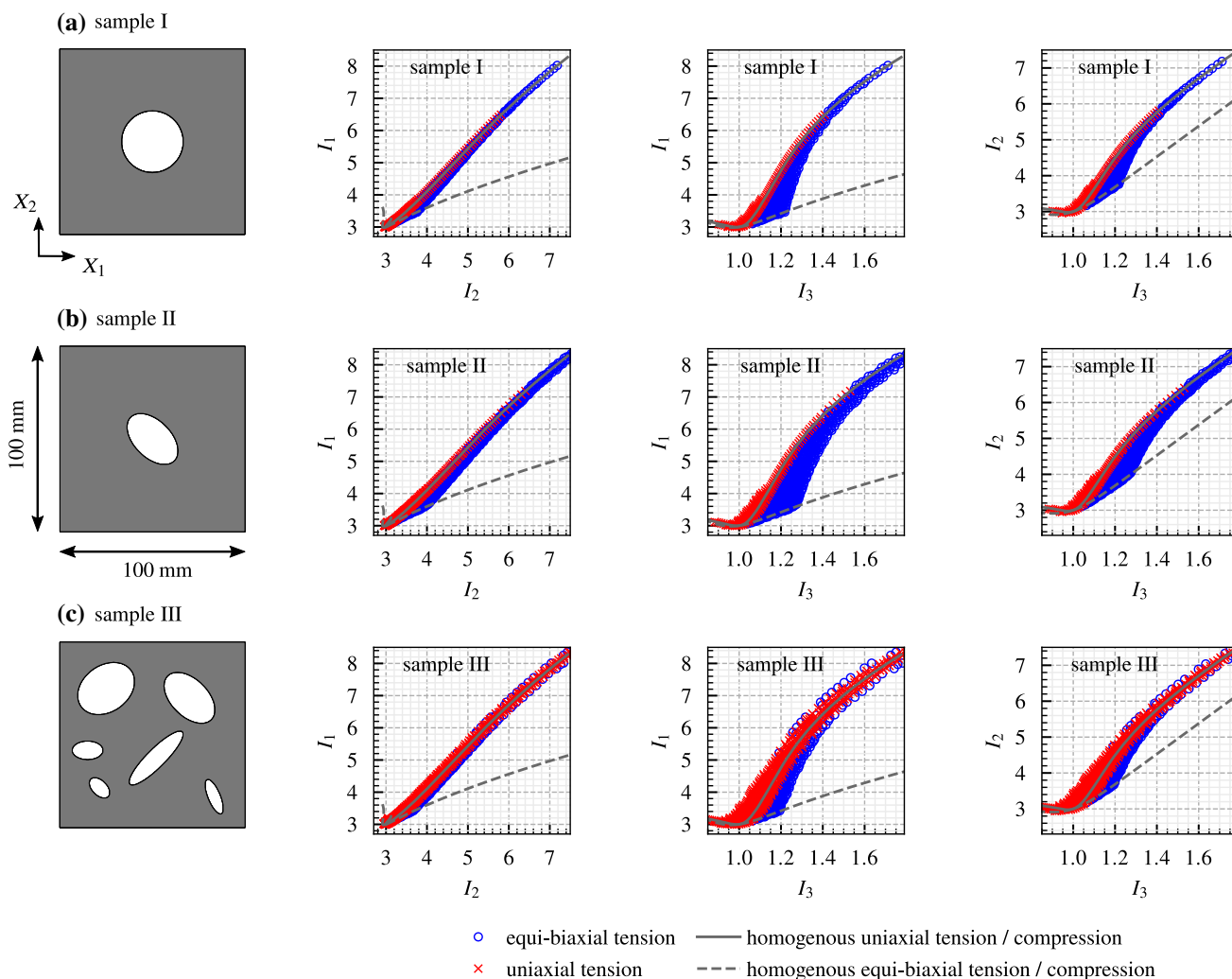


Fig. 6 Representation of the extracted data sets by means of the strain invariants I_α for the analyzed plane stress samples: **a** sample I with a circular hole, **b** sample II with an elliptical hole, and **c** sample III with arbitrarily arranged elliptical holes. The plotted gray lines represent the homogeneous uniaxial as well as equi-biaxial tension and compression states for the considered material

ANN-based models can be significantly improved if equi-biaxial tension tests are available.

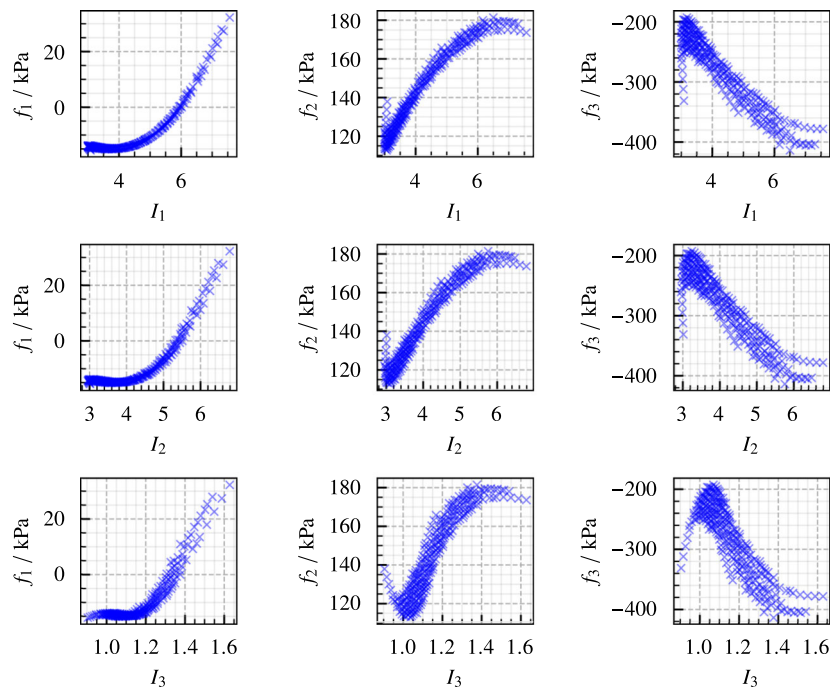
Influence of the hole geometry on the data set Besides the loading type, the two-dimensional specimen geometry is varied in the following second study. For better comparability, the plots for sample II and III given in the Fig. 6b, c are truncated with respect to those of specimen I.

A comparison of the simulation results for sample I and II reveals that a replacement of the circular by a more complex elliptical hole geometry leads to an improved sampling of \mathcal{I} . The ratios of the collected uniaxial and biaxial data sets however stay approximately unchanged. Considering the data for sample III which includes several randomly distributed ellipses, so once again more complex than sample II, the information content of the simple uniaxial tension test is clearly increased compared to the samples I and II. Here,

much more data points could be collected, especially in the compression range.

Accordingly, in order to test the proposed method schematized in Fig. 1 for the most easily realizable experimental setup, only the uniaxial tension data of sample III will be used for the ANN training process later on. The highly nonlinear character of the investigated Ogden-type material manifests in the corresponding stress coefficients f_α shown in Fig. 7. To close the virtual experiments study, it is mentioned again that the collected data set consists of true plane stress tuples with ${}^i\mathbf{T} = ({}^i T_{11}, {}^i T_{22}, 0, 0, 0, {}^i T_{12})^T$, since a two-dimensional simulation setting were used during the virtual experiments. In three-dimensional specimens, of course, gradients in the thickness direction will always appear under real loading conditions. However, the resulting deformation and stress components on the sample’s surface are in good agreement

Fig. 7 Visualization of the reduced Ogden-type elastic data set used for the training of the constitutive ANN. The data set is shown by means of the three stress coefficients f_α in dependence of the three strain invariants I_α



compared to a realistic three-dimensional setting. This is shown in the Appendix A for the considered example, Fig. 14.

4.2 Training of the constitutive ANN

After the selection of a suitable data set \mathcal{D} and the transformation to \mathcal{D}^{red} , the ANN training process according to Sect. 3 (c.1) – i.e. the calibration of the normalized stress coefficients f_α defined in Eq. (15) – is performed. In order to find a network which contains a minimum number of parameters while approximating $f_\alpha(i_\beta)$ with sufficient quality, the number of neurons $N \in \mathbb{N}$ in the only hidden layer is systematically varied within the range $[1, 20] \cap \mathbb{N}$. For each particular neuron number, the respective ANN is trained 100 times, where the weights of the best achieved training state are stored at the end.⁵ For this purpose, the reduced data set \mathcal{D}^{red} is randomly divided into training (70%), validation (15%) and test (15%) data. These subsets serve for the calibration in the training, generalization after each training progress and for the evaluation of the network’s quality afterwards, respectively. In order to evaluate the trained networks with respect to the prediction of the stress tensor \mathbf{T} , the fol-

lowing mean and maximum error measures are used:

$$i\epsilon_{\text{mean}} := \frac{({}^i T_{11}^{\text{ANN}} - {}^i T_{11}) + \dots + ({}^i T_{12}^{\text{ANN}} - {}^i T_{12})}{6\|{}^i \mathbf{T}\|_\infty} \text{ and} \tag{30}$$

$$i\epsilon_{\text{max}} := \frac{\|{}^i \mathbf{T}^{\text{ANN}} - {}^i \mathbf{T}\|_\infty}{6\|{}^i \mathbf{T}\|_\infty}. \tag{31}$$

In the equations above, $\|\cdot\|_\infty$ denotes the maximum norm of a vector valued quantity. To avoid singularities, the introduced errors are set to zero if $\|{}^i \mathbf{T}\|_\infty \leq 10^{-4}$ kPa holds which is a negligible stress value according to the material’s shear modulus $G^{\text{init}} = 100$ kPa.

The performance of the trained ANNs measured by the maximum values $\max |i\epsilon_{\text{mean}}|$ and $\max i\epsilon_{\text{max}}$ is given in Fig. 8. According to that, a clear tendency to decrease the errors while increasing the number of neurons N in the hidden layer is visible. Defining the criteria $\max_i |i\epsilon_{\text{mean}}| \leq 0.015\%$ and $\max_i i\epsilon_{\text{max}} \leq 0.015\%$ lead to the selection of an ANN with $N = 11$. Regarding other ANN-based approaches, this is a rather small network, see [6,13]. Moreover, the transformation step $\mathcal{D} \rightarrow \mathcal{D}^{\text{red}}$ enables to use a set containing only $n \approx 300$ tuples $\mathcal{D}_i^{\text{red}}$ for the training which is, again, comparatively small, cf. [37,41].

Finally, the predictions of the chosen network are compared with the true plane stress values generated by using the Ogden model in Fig. 9. Analogously, this comparison is depicted for the thermodynamically corrected model, cf. the formalism given in Sect. 3 (c.2), within Fig. 10. Summarizing the approximation quality, both ANN-based models provide

⁵ Due to local minima in the loss function, the optimization procedure which is applied within the training process depends on the starting values of the weights. To overcome this, the network is trained several times.

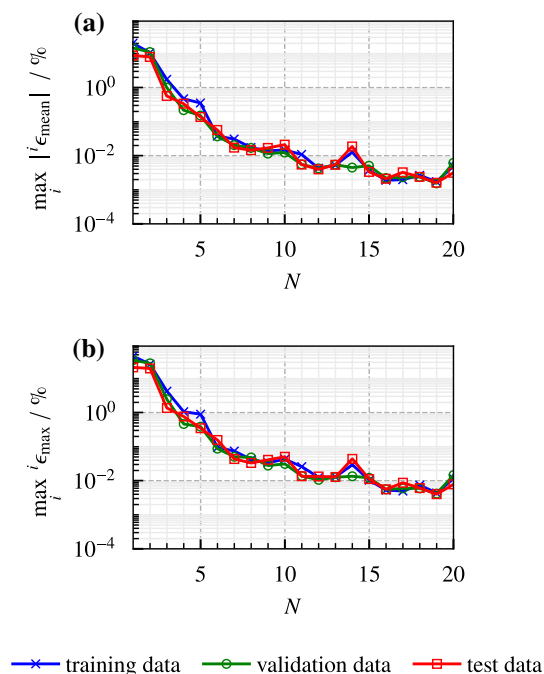


Fig. 8 Magnitude of the **a** maximum error measures $\max_i |\epsilon_{\text{mean}}|$ and **b** $\max_i |\epsilon_{\text{max}}|$ for training, validation data and test data in relation to the number of neurons in hidden layer

nearly perfect predictions for the in-plane stress components. However, for the out-of-plane components, the corrected model performs slightly worse compared to the simple original ANN model, although the predictions are still small in magnitude compared to the in-plane components.

4.3 Validation

In order to prove the suitability of both ANN-based models for the simulation of complex shaped samples and components, a comparison to reference results generated with the Ogden model (27) which has been used for the calibration is shown in the following. Thereby, two three-dimensional problems are considered: the uniaxial tension of a cuboid sample, and the torsion of a prismatic torsional sample, cf. the sketches given in Fig. 11a, b.

Cuboid under uniaxial tension To start with, the three-dimensional cuboid with a thickness-length ratio of 1/4 is considered. It is loaded under uniaxial tension, where a displacement of $\hat{u} = 60$ mm – this is equal to 60% effective strain of the sample – is prescribed on the top surface. Although the overall strain is identical to the utilized training load case, complex three-dimensional stress states are expected within the cuboid. Thus, regarding the highly non-linear material response of the reference constitutive law, this

example is a good possibility to validate and explore potential limits of the ANN-based approaches.

For all of the three models, the global reaction force depending on the prescribed displacement has been calculated within the simulations. Considering these curves which are given in Fig. 11a, a nearly perfect alignment of both ANN-solutions with the reference curve is observed.

Furthermore, in order to evaluate the ANN-predictions of local deformation \mathbf{C} and stress fields \mathbf{P} , the simulated 11-components within the cuboid domain are compared to each other in Fig. 12a, b, respectively. According to that, relative errors below 0.06% and 0.16% are achievable with the original and the corrected ANN-based model, respectively. Slightly increased deviations below 0.47% and 0.42% are observed for the stress component P_{11} , whereby – in order to avoid singularities – the absolute errors $|P_{11} - P_{11}^{\text{ANN}}|$ and $|P_{11} - P_{11}^{\text{ANN}^*}|$ are related to the maximum value of $|P_{11}|$. As expected, the predictive performance of the corrected model (24) is of similar quality compared to the initial model (18). However, a significantly aggravated convergence behavior within the Newton iteration is observed within the performed FE simulation for the corrected model. This behavior is supposedly attributable to the singular points occurring in the corrected stress coefficients, cf. Eq. (23). In total, slightly twice as many iterations are required compared to the uncorrected ANN model.

Altogether, a rather high accuracy could be achieved for \mathbf{C} and \mathbf{P} , although the occurring maximum stress of $P_{11} = 237$ kPa is increased compared to the training data set which is moreover purely two-dimensional with respect to the stresses, cf. Fig. 9a. In view of the inability of ANNs to extrapolate, this is a quite astonishing result. However, since the trained network lives in the reduced $I_\alpha - f_\beta$ -space and the final stress prediction is obtained in combination with the tensor-valued generators \mathbf{G}^α , the network actually does not, or only slightly, have to extrapolate at all. More precisely, the deformation-type invariant set of the considered cuboid under uniaxial loading is almost completely included within the data set used for the training process. This is shown in Fig. 15a within the Appendix B. However, since the cuboid's data set within the space \mathcal{I} is not completely covered by the training data set, extrapolation is at least partially necessary.

Torsional sample In the second validation setup, a three-dimensional torsion sample with three circular holes is loaded by specifying a distortion of $\hat{\phi} = 45^\circ$.

Similar to the previous validation case, the global reaction torque depending on the prescribed distortion has been calculated. Again, a nearly perfect alignment of both ANN-solutions with the reference curve is visible in Fig. 11b. Likewise, a comparison of the predictions for local strain C_{12} and stress P_{12} reveal relative errors in a similar range as for the first example, see Fig. 13. The stress prediction of the corrected model is now apparently worth

Fig. 9 Verification of the ANN-based stress prediction (18) for a model with $N = 11$ neurons trained with pure plane stress data: **a** in-plane components and **b** out-of-plane components

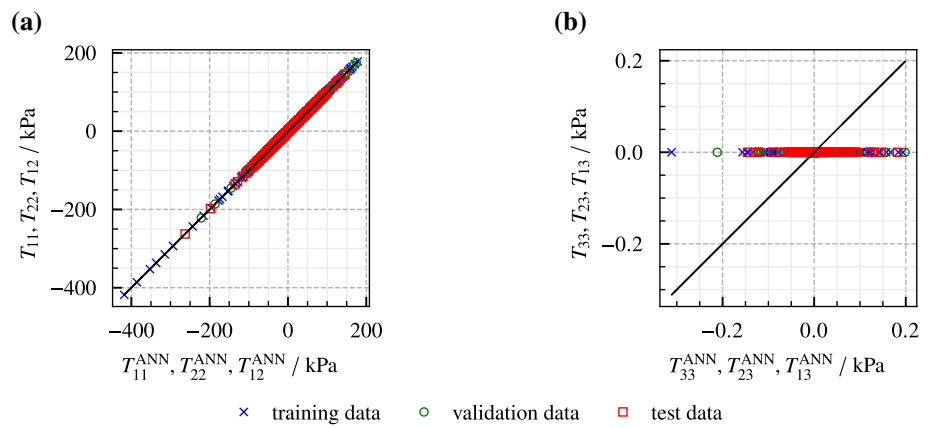


Fig. 10 Verification of the thermodynamically corrected stress prediction for an ANN-based model with $N = 11$ neurons trained with pure plane stress data: **a** in-plane components and **b** out-of-plane components

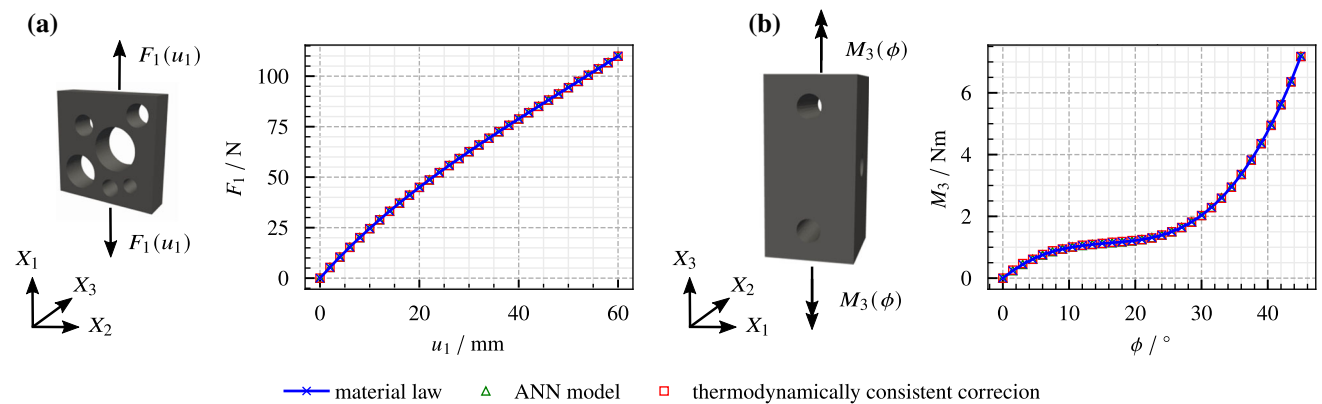
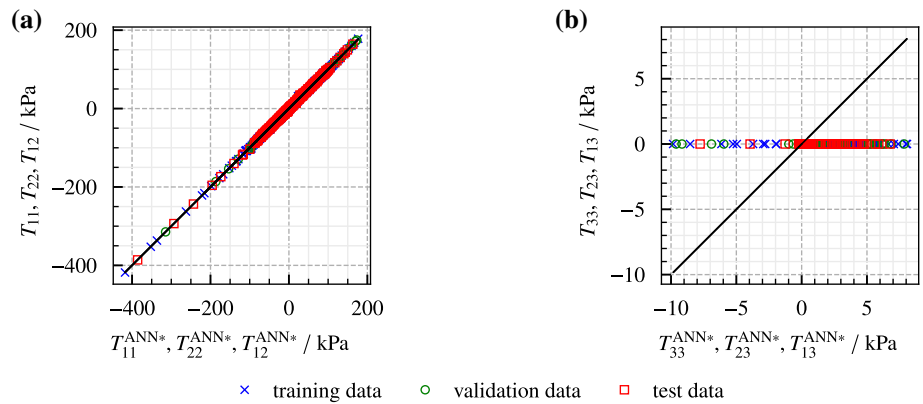


Fig. 11 Reaction loads for the considered validation cases: **a** force-displacement curves of the cuboid sample with edge length $100 \text{ mm} \times 100 \text{ mm} \times 25 \text{ mm}$ under uniaxial tension, and **b** torque-angle curves of the torsional sample with edge length $200 \text{ mm} \times 100 \text{ mm} \times 100 \text{ mm}$

compared to the original one but however still below 0.46%.

Thus, also in this example, the ANN-based approach is very well able to describe the trained nonlinear constitutive behavior within the FE simulation of a comparatively complex load case. This is once again due to the fact that the deformation-type invariant space of the torsion sample is included into the used training data set within wide areas which is shown in the It has to be Appendix B, Fig. 15b. Thereby, the used ANN with not more than $N = 11$ neurons

in only one hidden layer is very small compared to other ANN-based constitutive models [41].

5 Conclusions

In this work, a novel ANN-based approach for the automated constitutive modeling of isotropic hyperelastic solids is presented. It is based on a physically motivated reduction of the problem’s dimensionality for the deformation and stress type

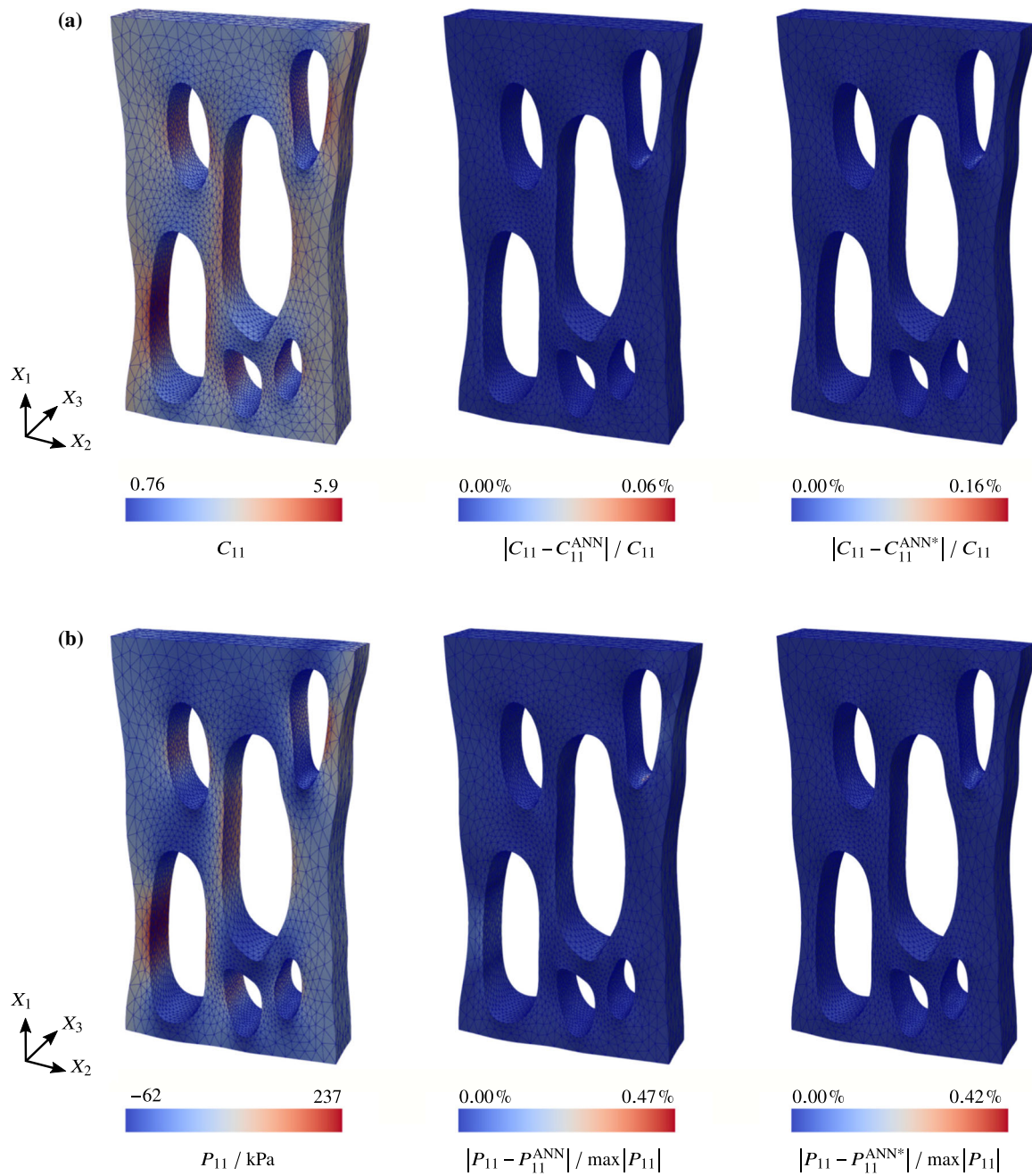


Fig. 12 Simulated local distribution of secondary field quantities for the cuboid under uniaxial tension as well as relative errors of the ANN-based models (18) and (24) with respect to the reference Ogden model: 11-components of **a** right Cauchy-Green deformation tensor C , and **b** 1st Piola-Kirchhoff stress tensor P . The surface plots are given on the deformed configuration \mathcal{B}

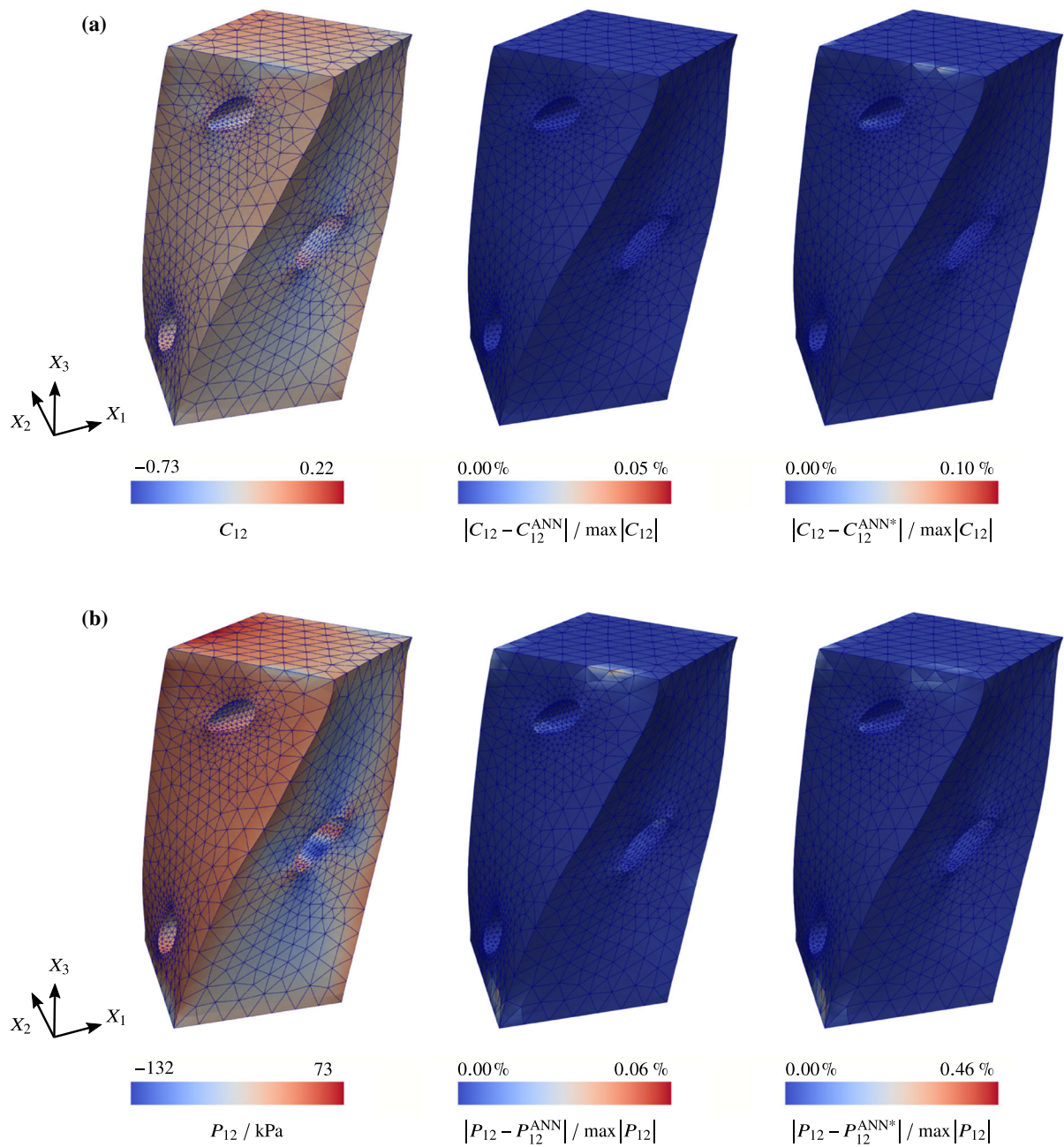


Fig. 13 Simulated local distribution of secondary field quantities for the torsional sample as well as relative errors of the ANN-based models (18) and (24) with respect to the reference Ogden model: **a** right Cauchy-Green deformation tensor C , and **b** 1st Piola-Kirchhoff stress tensor P . The surface plots are given on the deformed configuration \mathcal{B}

quantities. In contrast to the most ANN-based constitutive models, the proposed procedure enables a highly accurate stress prediction with only small network architectures. Furthermore, it is possible to resolve fully three-dimensional stress fields while only two-dimensional plane stress data are used for the training process of the ANN. Starting from basic continuum mechanical equations, a short revision of hyperelastic constitutive models is given. Based on this, the developed ANN-based modeling approach is

presented, where four steps – (a) data mining, (b) data processing, (c.1) training process, and (c.2) thermodynamically consistent correction – are discussed in detail. This approach is exemplarily applied to data collected from virtual experiments with samples comprising a highly nonlinear Ogden type material behavior. Thereby, different influences on the resulting data sets are investigated. In view of realistic experiments, where only information on the sample’s surface are available and thin specimens are thus preferable, the virtual

samples are loaded in pure plane stress states. Finally, the trained ANN-based models are used for the simulation of two three-dimensional samples, a cuboid under uniaxial tension and a prismatic torsional sample. A comparison of these simulations with the results achieved with the Ogden type model shows a rather high accuracy for both ANN-based models with a small architecture consisting of only 11 neurons in one hidden layer.

Altogether, the presented ANN-based approach has shown to be an efficient tool for the description of isotropic hyperelastic solids which reveal a strongly nonlinear stress-strain-response. Due to the use of standard machine learning toolboxes, it is easy to use and could be integrated as a user subroutine into commercial or non-commercial FE codes.

In order to enable the robust application of the proposed ANN-based procedure for the analysis and design of complex engineering components, several extensions have to be made in the future. For instance, an extension for the processing of noisy data sets [3,36] which enables to use data collected from real experiments is essential. Furthermore, to overcome the poor convergence behavior of the a posteriori corrected model, it is favorable to approximate the free energy function Ψ instead of variables on the stress level with the ANN. With that in combination with customized training processes which incorporate gradients $\partial_C \Psi$ in the loss, an a priori thermodynamically consistent training process is possible [41,44,58]. Finally, an extension to more general anisotropy classes [13,41] or dissipative constitutive behavior [44,56,59] is needed.

Acknowledgements All presented computations were performed on a PC-Cluster at the Center for Information Services and High Performance Computing (ZIH) at TU Dresden. The authors thus thank the ZIH for generous allocations of computer time.

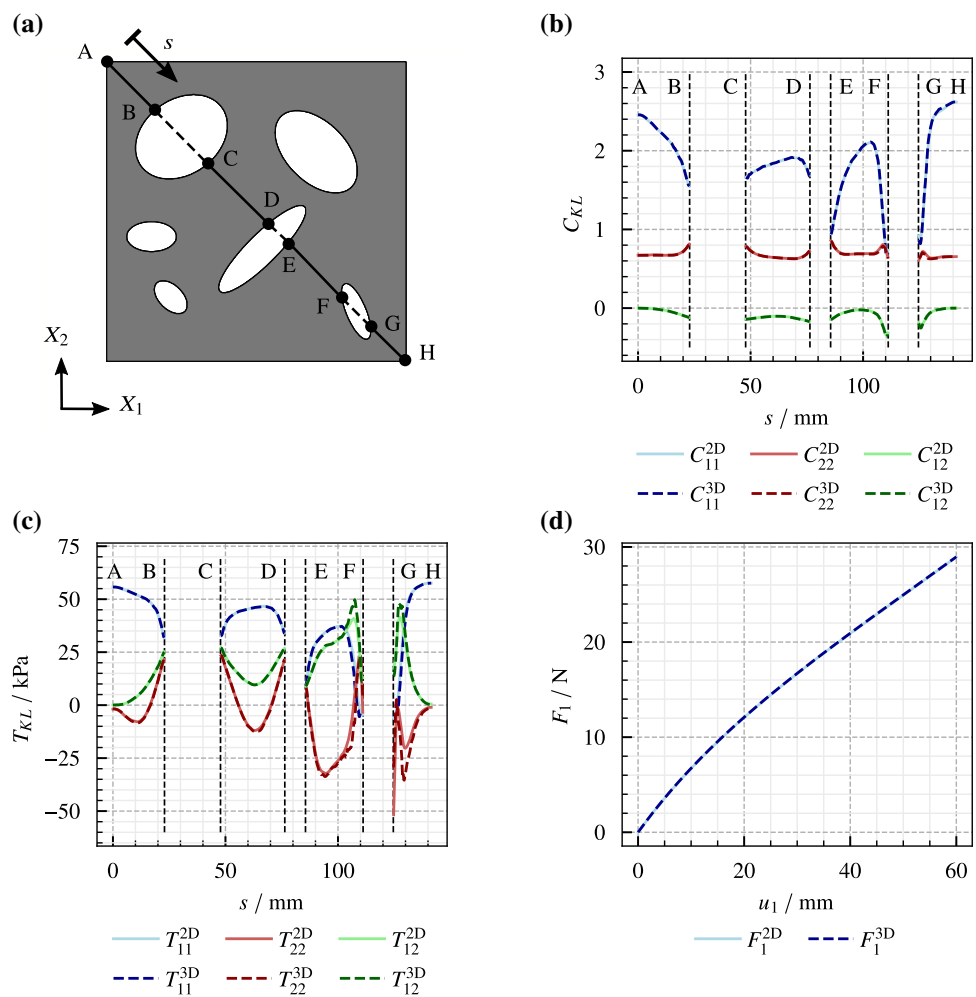
Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

A Comparison of two- and three-dimensional simulation settings

Into this appended section, deformation and stress fields within a thin sample which were computed by using the plane stress assumption are considered. These plane stress results are compared to a fully three-dimensional simulation of the same sample with 5 mm thickness and 100 mm edge length shown in Fig. 14a. Since gradients in the thickness direction will always appear under real loading conditions, only the surface of the three-dimensional setup is of interest. As shown in Fig. 14b, the in-plane components of C are perfectly aligned along a diagonal path across the sample for both simulations. The in-plane stress components depicted in Fig. 14c are likewise in good agreement in the main part of the path. Only close to the points F and G larger deviations occur. Moreover, almost no differences can be identified for the reaction force in dependence of the prescribed displacement, cf. Fig. 14d.

Fig. 14 Comparison of plane stress and full three dimensional FE simulations for a thin disc with 5 mm thickness and 100 mm edge length: **a** path along the specimen surface, **b, c** calculated deformation and stress components, as well as **d** resulting force displacement curves



B Comparison of training and validation invariant sets

Within this appended section, the data sets \mathcal{D}_{cub} and \mathcal{D}_{tor} related to the cuboid under uniaxial tension and the torsional sample are compared to the training data set $\mathcal{D}_{\text{train}}$ which has

been used for the ANN calibration process. To this end, the deformation-type invariants I_1 – I_3 of the three reduced sets are visualized in Fig. 15. As shown therein, the set $\mathcal{D}_{\text{train}}$ which contains only pure plane stress states, covers almost the entire data sets of both samples within the invariant space.

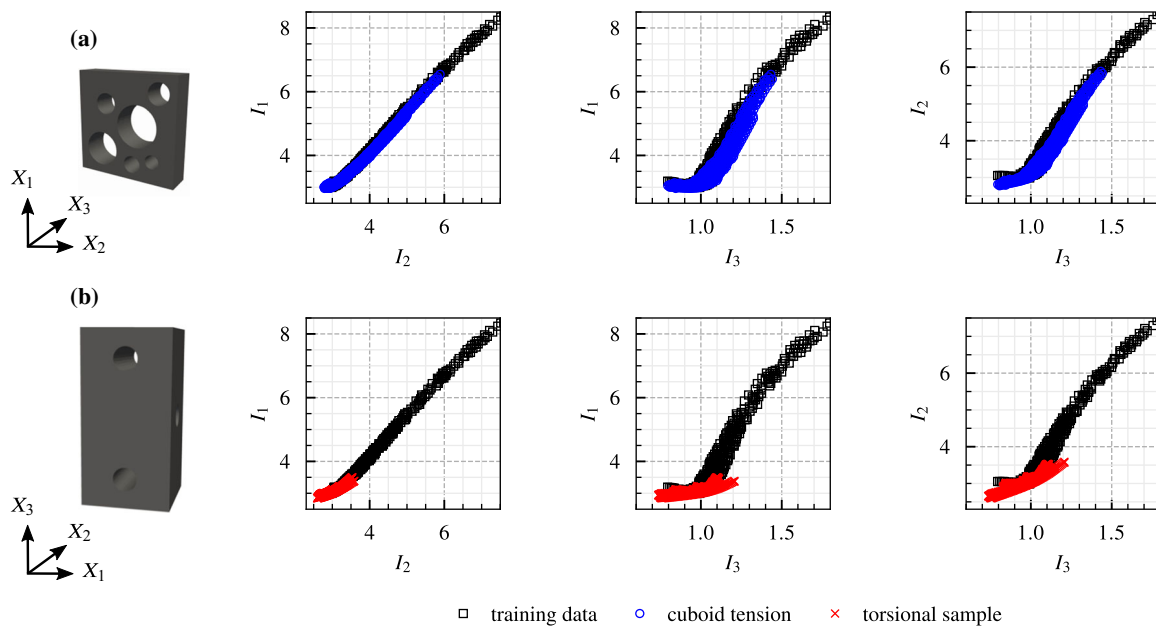


Fig. 15 Representation of the extracted data sets by means of the strain invariants I_α for the validation cases: **a** cuboid with arbitrary arranged circular holes and edge length 100 mm \times 100 mm \times 25 mm under uniaxial tension and **b** torsional sample with three circular holes and edge length 100 mm \times 100 mm \times 200 mm

References

1. Al-Haik M, Hussaini M, Garmestani H (2006) Prediction of nonlinear viscoelastic behavior of polymeric composites using an artificial neural network. *Int J Plast* 22(7):1367–1392. <https://doi.org/10.1016/j.ijplas.2005.09.002>
2. Alnæs M, Blechta J, Hake J, Johansson A, Kehlet B, Logg A, Richardson C, Ring J, Rognes ME, Wells GN (2015) The FEniCS project version 1.5. *Arch Numer Softw* 3(100):9–23
3. Ayensa-Jiménez J, Doweidar MH, Sanz-Herrera JA, Doblaré M (2018) A new reliability-based data-driven approach for noisy experimental data with physical constraints. *Comput Methods Appl Mech Eng* 328:752–774. <https://doi.org/10.1016/j.cma.2017.08.027>
4. Bock FE, Aydin RC, Cyron CJ, Huber N, Kalidindi SR, Klusemann B (2019) A review of the application of machine learning and data mining approaches in continuum materials mechanics. *Front Mater* 6:110. <https://doi.org/10.3389/fmats.2019.00110>
5. Carrara P, De Lorenzis L, Stainier L, Ortiz M (2020) Data-driven fracture mechanics. *Comput Methods Appl Mech Eng* 372:113390. <https://doi.org/10.1016/j.cma.2020.113390>
6. Chung I, Im S, Cho M (2021) A neural network constitutive model for hyperelasticity based on molecular dynamics simulations. *Int J Numer Methods Eng* 122(1):5–24. <https://doi.org/10.1002/nme.6459>
7. Coleman BD, Noll W (1963) The thermodynamics of elastic materials with heat conduction and viscosity. *Arch Ration Mech Anal* 13(1):167–178
8. Criscione JC, Humphrey JD, Douglas AS, Hunter WC (2000) An invariant basis for natural strain which yields orthogonal stress response terms in isotropic hyperelasticity. *J Mech Phys Solids* 48(12):2445–2465. [https://doi.org/10.1016/S0022-5096\(00\)00023-5](https://doi.org/10.1016/S0022-5096(00)00023-5)
9. Czarnecki WM, Osindero S, Jaderberg M, Swirszcz G, Pascanu R (2017) Sobolev training for neural networks, p 10. <https://arxiv.org/abs/1706.04859>
10. Dalémat M, Coret M, Leygue A, Verron E (2019) Measuring stress field without constitutive equation. *Mech Mater* 136:103087
11. Dalémat M, Coret M, Leygue A, Verron E (2021) Robustness of the data-driven identification algorithm with incomplete input data, p 22. <https://hal.inria.fr/hal-03028848/>
12. Eggersmann R, Kirchdoerfer T, Reese S, Stainier L, Ortiz M (2019) Model-free data-driven inelasticity. *Comput Methods Appl Mech Eng* 350:81–99. <https://doi.org/10.1016/j.cma.2019.02.016>
13. Fernández M, Jamshidian M, Böhlke T, Kersting K, Weeger O (2020) Anisotropic hyperelastic constitutive models for finite deformations combining material theory and data-driven approaches with application to cubic lattice metamaterials. *Comput Mech*. <https://doi.org/10.1007/s00466-020-01954-7>
14. Fernández M, Rezaei S, Rezaei Mianroodi J, Fritzen F, Reese S (2020) Application of artificial neural networks for the prediction of interface mechanics: a study on grain boundary constitutive behavior. *Adv Model Simul Eng Sci* 7(1):1. <https://doi.org/10.1186/s40323-019-0138-7>
15. Flory P (1961) Thermodynamic relations for high elastic materials. *Trans Faraday Soc* 57:829–838
16. Fritzen F, Fernández M, Larsson F (2019) On-the-fly adaptivity for nonlinear twoscale simulations using artificial neural networks and reduced order modeling. *Front Mater* 6:75. <https://doi.org/10.3389/fmats.2019.00075>
17. Furukawa T, Yagawa G (1998) Implicit constitutive modelling for viscoplasticity using neural networks. *Int J Numer Methods Eng* 43(2):195–219. [https://doi.org/10.1002/\(SICI\)1097-0207\(19980930\)43:2<195::AID-NME418>3.0.CO;2-6](https://doi.org/10.1002/(SICI)1097-0207(19980930)43:2<195::AID-NME418>3.0.CO;2-6)
18. Geuzaine C, Remacle JF (2009) Gmsh: a 3-D finite element mesh generator with built-in pre- and post-processing facilities. *Int J Numer Methods Eng* 79(11):1309–1331. <https://doi.org/10.1002/nme.2579>
19. Ghaboussi J, Garrett JH, Wu X (1991) Knowledge-based modeling of material behavior with neural networks. *J Eng Mech* 117(1):132–153. [https://doi.org/10.1061/\(ASCE\)0733-9399\(1991\)117:1\(132\)](https://doi.org/10.1061/(ASCE)0733-9399(1991)117:1(132))

20. Ghaboussi J, Sidarta D (1998) New nested adaptive neural networks (NANN) for constitutive modeling. *Comput Geotech* 22(1):29–52. [https://doi.org/10.1016/S0266-352X\(97\)00034-7](https://doi.org/10.1016/S0266-352X(97)00034-7)
21. González D, Chinesta F, Cueto E (2019) Learning corrections for hyperelastic models from data. *Front Mater* 6:14. <https://doi.org/10.3389/fmats.2019.00014>
22. González D, Chinesta F, Cueto E (2019) Thermodynamically consistent data-driven computational mechanics. *Continuum Mech Thermodyn* 31(1):239–253. <https://doi.org/10.1007/s00161-018-0677-z>
23. Hashash YMA, Jung S, Ghaboussi J (2004) Numerical implementation of a neural network based material model in finite element analysis: neural network based material model. *Int J Numer Methods Eng* 59(7):989–1005. <https://doi.org/10.1002/nme.905>
24. Holzapfel GA (2000) *Nonlinear solid mechanics—a continuum approach for engineering*. Wiley, Chichester
25. Ibañez R, Abisset-Chavanne E, Aguado JV, Gonzalez D, Cueto E, Chinesta F (2018) A manifold learning approach to data-driven computational elasticity and inelasticity. *Arch Comput Methods Eng* 25(1):47–57. <https://doi.org/10.1007/s11831-016-9197-9>
26. Ibañez R, Borzacchiello D, Aguado JV, Abisset-Chavanne E, Cueto E, Ladeveze P, Chinesta F (2017) Data-driven non-linear elasticity: constitutive manifold construction and problem discretization. *Comput Mech* 60(5):813–826. <https://doi.org/10.1007/s00466-017-1440-1>
27. Ibañez R, Abisset-Chavanne E, González D, Duval JL, Cueto E, Chinesta F (2019) Hybrid constitutive modeling: data-driven learning of corrections to plasticity models. *Int J Mater Form* 12(4):717–725. <https://doi.org/10.1007/s12289-018-1448-x>
28. Jung S, Ghaboussi J (2006) Neural network constitutive model for rate-dependent materials. *Comput Struct* 84(15–16):955–963. <https://doi.org/10.1016/j.compstruc.2006.02.015>
29. Kalina KA, Rassloff A, Wollner M, Metsch P, Brummund J, Kästner M (2020) Multiscale modeling and simulation of magnetoactive elastomers based on experimental data. *Phys Sci Rev*. <https://doi.org/10.1515/psr-2020-0012>
30. Karapiperis K, Stainier L, Ortiz M, Andrade J (2021) Data-driven multiscale modeling in mechanics. *J Mech Phys Solids* 147:104239
31. Kirchdoerfer T, Ortiz M (2016) Data-driven computational mechanics. *Comput Methods Appl Mech Eng* 304:81–101. <https://doi.org/10.1016/j.cma.2016.02.001>
32. Kirchdoerfer T, Ortiz M (2017) Data driven computing with noisy material data sets. *Comput Methods Appl Mech Eng* 326:622–641. <https://doi.org/10.1016/j.cma.2017.07.039>
33. Kirchdoerfer T, Ortiz M (2018) Data-driven computing in dynamics: data-driven computing in dynamics. *Int J Numer Methods Eng* 113(11):1697–1710. <https://doi.org/10.1002/nme.5716>
34. Kruse R, Borgelt C, Braune C, Mostaghim S, Steinbrecher M (2016) *Computational intelligence. Texts in computer science*. Springer, London. <https://doi.org/10.1007/978-1-4471-7296-3>
35. Lange N, Hütter G, Kiefer B (2021) An efficient monolithic solution scheme for FE2 problems. *Comput Methods Appl Mech Eng* 382:113886
36. Latorre M, Montáns FJ (2020) Experimental data reduction for hyperelasticity. *Comput Struct* 232:105919
37. Le BA, Yvonnet J, He QC (2015) Computational homogenization of nonlinear elastic materials using neural networks: neural networks-based computational homogenization. *Int J Numer Methods Eng* 104(12):1061–1084. <https://doi.org/10.1002/nme.4953>
38. Leygue A, Coret M, Réthoré J, Stainier L, Verron E (2018) Data-based derivation of material response. *Comput Methods Appl Mech Eng* 331:184–196. <https://doi.org/10.1016/j.cma.2017.11.013>
39. Liang G, Chandrashekhara K (2008) Neural network based constitutive model for elastomeric foams. *Eng Struct* 30(7):2002–2011. <https://doi.org/10.1016/j.engstruct.2007.12.021>
40. Ling J, Jones R, Templeton J (2016) Machine learning strategies for systems with invariance properties. *J Comput Phys* 318:22–35. <https://doi.org/10.1016/j.jcp.2016.05.003>
41. Linka K, Hillgärtner M, Abdolazizi KP, Aydin RC, Itskov M, Cyron CJ (2021) Constitutive artificial neural networks: a fast and general approach to predictive data-driven constitutive modeling by deep learning. *J Comput Phys* 429:110010
42. Liu Z, Wu C, Koishi M (2019) A deep material network for multiscale topology learning and accelerated nonlinear modeling of heterogeneous materials. *Comput Methods Appl Mech Eng* 345:1138–1168. <https://doi.org/10.1016/j.cma.2018.09.020>
43. Logg A, Mardal KA, Wells G (2012) *Automated solution of differential equations by the finite element method: the FEniCS book, vol 84*. Springer, Berlin
44. Masi F, Stefanou I, Vannucci P, Maffi-Berthier V (2021) Thermodynamics-based artificial neural networks for constitutive modeling. *J Mech Phys Solids* 147:104277
45. Miehe C (1993) Computation of isotropic tensor functions. *Commun Numer Methods Eng* 9(11):889–896. <https://doi.org/10.1002/cnm.1640091105>
46. Miehe C (1998) Comparison of two algorithms for the computation of fourth-order isotropic tensor functions. *Comput Struct* 66(1):37–43
47. Montáns FJ, Chinesta F, Gómez-Bombarelli R, Kutz JN (2019) Data-driven modeling and learning in science and engineering. *Comptes Rendus Mécanique* 347(11):845–855. <https://doi.org/10.1016/j.crme.2019.11.009>
48. Nguyen LTK, Keip MA (2018) A data-driven approach to nonlinear elasticity. *Comput Struct* 194:97–115. <https://doi.org/10.1016/j.compstruc.2017.07.031>
49. Ogden RW (1997) *Non-linear elastic deformations*. Dover Publications, Mineola
50. Pierron F, Grédiac M (2020) Towards material testing 2.0. A review of test design for identification of constitutive parameters from full-field measurements. *Strain*. <https://doi.org/10.1111/str.12370>
51. Sanz-Herrera JA, Mora-Macías J, Ayensa-Jiménez J, Reina-Romo E, Doweidar MH, Domínguez J, Doblaré M (2020) Data-driven computational simulation in bone mechanics. *Ann Biomed Eng*. <https://doi.org/10.1007/s10439-020-02550-9>
52. Settigast C, Abendroth M, Kuna M (2019) Constitutive modeling of plastic deformation behavior of open-cell foam structures using neural networks. *Mech Mater* 131:1–10. <https://doi.org/10.1016/j.mechmat.2019.01.015>
53. Settigast C, Hütter G, Kuna M, Abendroth M (2020) A hybrid approach to simulate the homogenized irreversible elastic–plastic deformations and damage of foams by neural networks. *Int J Plast* 126:102624
54. Shen Y, Chandrashekhara K, Breig WF, Oliver LR (2004) Neural network based constitutive model for rubber material. *Rubber Chem Technol* 77(2):257–277. <https://doi.org/10.5254/1.3547822>
55. Stainier L, Leygue A, Ortiz M (2019) Model-free data-driven methods in mechanics: material data identification and solvers. *Comput Mech* 64(2):381–393. <https://doi.org/10.1007/s00466-019-01731-1>
56. Stoffel M, Bamer F, Markert B (2019) Neural network based constitutive modeling of nonlinear viscoplastic structural response. *Mech Res Commun* 95:85–88. <https://doi.org/10.1016/j.mechrescom.2019.01.004>
57. Terada K, Kato J, Hirayama N, Inugai T, Yamamoto K (2013) A method of two-scale analysis with micro–macro decoupling scheme: application to hyperelastic composite materials. *Comput Mech* 52(5):1199–1219. <https://doi.org/10.1007/s00466-013-0872-5>
58. Vlassis NN, Ma R, Sun W (2020) Geometric deep learning for computational mechanics Part I: anisotropic hyperelasticity. *Comput Methods Appl Mech Eng* 371:113299

59. Zopf C, Kaliske M (2017) Numerical characterisation of uncured elastomers by a neural network based approach. *Comput Struct* 182:504–525. <https://doi.org/10.1016/j.compstruc.2016.12.012>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.