A novel inverse methodology for the extraction of bulk elasto-plastic tensile properties of metals using spherical instrumented indentation

Gabrielle Turcot [a,b,*], Daniel Paquet [b], Martin Lévesque [a], Sylvain Turenne [a]

[a] *École Polytehnique de Montréal, Department of Mechanical Engineering, 2500 Chemin de Polytechnique, Montréal, H3T 1J4, Québec, Canada*
[b] *Institut de Recherche d'Hydro-Québec, 1800 Boulevard Lionel-Boulet, Varennes, J3X 1S1, Québec, Canada*

A B S T R A C T

This paper presents an inverse methodology relying on finite element simulation based optimization to extract material constitutive parameters from a single spherical indentation, which are used to estimate its bulk true stress–true strain curve. Methods of this type usually assume that the tensile curve of the studied material can be modelled by a power-law involving, for example, Young's modulus, yield stress and a hardening coefficient. Since the tensile behaviour of some materials cannot be adequately fitted by such models, our approach rather relies on an optimization procedure which extracts a group of six points on the true stress–true strain curve, distributed as per a geometric progression in the strain space, as well as the elastic modulus. A direct search black-box optimization algorithm is used and is shown to be capable of eluding local minima. A surrogate step is introduced in the methodology, which is a simplified version of the inverse problem, to find a suitable starting point. The performance of the method is first investigated in a numerical study in which the estimated true stress–true strain curves lie within a maximum error of 5.7% from the corresponding target tensile curves, for four materials with different hardening behaviours. The maximum errors on the extracted elastic modulus and yield stress are 0.5% and 11.1%, respectively. A study of the convergence behaviour of the proposed method demonstrates an expected significant increase in computational times as compared to when using a hardening model. Finally, an experimental application of the method to two materials with a plastic plateau leads to true stress–true strain curves estimations with average differences of 3.9% and 0.9% with the macroscopic experimental tensile curve, respectively. The method is thus found to be accurate even in the presence of modelling and experimental errors.

1. Introduction

The knowledge of local true stress–true strain curves in components featuring microstructure gradients is essential to the prediction of their macroscopic plasticity response in service. For instance, the heat affected zone of a welded joint is characterized by spatial variations of the yield stress and hardening behaviours. These variations are due to an evolving microstructure which appears with the differing cooling rates involved, as well position dependent strain hardening taking place during the welding process. When studying fatigue behaviour, the knowledge of local true stress–true strain curves is required to understand, model, and predict the plastic response of the material through the applied load cycles.

Hydro-electric turbines feature welded joints that are subjected to fatigue loading. Fatigue failure in these turbines greatly impacts electricity production and unexpected service interruptions generate significant costs and profit loss during the necessary repairs. A reliable method capable of estimating local tensile behaviour would offer an evaluation of the integrity of the welded joints and contribute to avoid such unexpected service interruptions. It would also be a tool to evaluate the effectiveness of performed repairs and guide the development of new repairs, leading to increased weld lifetimes.

Methods such as milli-tension tests and instrumented indentation can be used to obtain bulk local true stress–true strain curves. The former method directly provides measurements of the true stress–true strain response of the material since it measures the applied load and displacement to a uni-axial specimen. It however requires very precise machining of specimens at the millimetre scale and it is destructive. On the other hand, instrumented indentation tests are comparatively simpler to implement since the specimen preparation is standardized and it is a non-destructive testing method. The challenge with

instrumented indentation rests upon the analysis of the results since the stress and strain fields induced during an indentation test are non-uniform and evolve during the indentation process. Hence, there is no direct correlation between indentation data and the material true stress–true strain curve. The development of a robust methodology for estimating tensile curves from instrumented indentation data has received a considerable amount of research interest in the past two decades (*e.g.* Bocciarelli et al., 2005; Dean and Clyne, 2017; Wang et al., 2017).

A popular approach is the use of numerical optimization to find the tensile properties which minimize the difference between indentation data obtained from experiments and that obtained by finite element simulations of the same experiments (Wang et al., 2016; Bolzon et al., 2004; Bocciarelli et al., 2005; Campbell et al., 2018). A prominent issue of these methods is the difficulty of ensuring the unicity of the solution when a single indentation curve obtained with a sharp indenter is used (Chen et al., 2007). In this work, the risk of non-uniqueness is mitigated by using a spherical indenter as well as the residual imprint of indentation as supplemental experimental information.

Optimization inverse methods previously proposed in the literature for the estimation of true stress–true strain curves by indentation approach the problem by assuming a pre-defined shape for the tensile curve (*e.g.* power law (Zhao et al., 2006; Liu et al., 2009), bi-linear (Pelletier, 2006), presence of a Lüders plateau (Pham et al., 2015; Pham and Kim, 2015)). The parameters of these laws are then optimized to estimate the tensile curve. This approach is efficient and rational, but the maximum precision of the true stress–true strain curve obtained with such methods is dependent upon the hardening model's capacity to accurately fit the experimentally observed behaviour.

An optimization inverse method which avoided the assumption of a hardening model was developed by (Bouzakis et al., 2001; Bouzakis and Michailidis, 2004, 2006) for a Berkovich indenter. The indentation curve was divided into small increments, for each of which a point was found on the tensile curve by optimizing the slope $d\sigma/d\epsilon$ after the previous point computed. No application of the method to materials with known properties to validate its performance was presented by the authors, making it difficult to evaluate the reliability and robustness of this method. Also, to use the residual imprint in such an approach, one would need to indent up to a certain depth, unload, take a topography measurement, and repeat. This process would represent a great experimental challenge because of how difficult it would be to realign the indenter in the exact same position at each cycle. The lack of this supplementary experimental information, particularly since a sharp indenter is used, means that the method could suffer from non-uniqueness issues (Chen et al., 2007). Except for that of Bouzakis et al. to the knowledge of the authors, no optimization method which does not suppose a hardening model was developed in the past.

This work presents a new inverse method, relying on finite element simulation based numerical optimization, to estimate local true stress–true strain curves of metals from an instrumented indentation test that does not rely on a predefined hardening model. The main objective is to propose a methodology which is more versatile than popular methods that assume a predefined hardening model. To do so, the variables optimized are material parameters which estimate the position of six points on the stress–strain curve. Such material parameters can therefore lead to tensile curves of any physically acceptable shape. A comparison of the performance of the proposed method to the use of the Hollomon and Ramberg–Osgood hardening models is presented to demonstrate its increased versatility. A third model is developed, called the hybrid model, which includes a plastic plateau followed by a Ramberg–Osgood power law, and its use is also compared to the proposed methodology.

This paper is organized as follows: Section 2 presents relevant background information, Section 3 describes the experimental methodology used in this study, Section 4 details the finite element models used, and Section 5 provides a complete description of the inverse methodology. The performance analysis through a numerical study and the experimental validation of the method are then shown in Section 6 followed by a brief conclusion in Section 7.

2. Background

This section presents background material relevant to this work. First, Section 2.1 describes non-uniqueness problems often observed in indentation inverse methods as well as techniques to reduce them, followed by some considerations regarding the scale of indentation in Section 2.2. Section 2.3 then lists optimization algorithms used in indentation inverse methods, states the algorithm chosen in this work and gives some context on strategies for its implementation. Section 2.4 then presents Tabor's indentation strain, a useful parameter in this work. Finally, Section 2.5 describes popular hardening models used in indentation inverse methods which will be used in this work as a comparison to the proposed methodology.

2.1. Non-uniqueness issues in the identification of constitutive model parameters by an inverse method

Researchers have shown that two materials exhibiting different tensile properties can yield almost identical indentation curves (Chen et al., 2007). This observation suggests that two different sets of parameters for a given constitutive model could be identified from the same indentation curve. Proposed improvements to reduce the risk of solution non-uniqueness in indentation inverse methods are the use of the residual imprint as supplementary experimental information (Bolzon et al., 2004; Bocciarelli and Maier, 2007), and the use of a spherical indenter (Campbell et al., 2018). Both these strategies are used in this work, to develop a robust framework for which the risks of non-unicity are reduced.

The use of the residual imprint topography, in addition to the indentation curve, enriches the experimental information extracted from the test (Bolzon et al., 2004; Bocciarelli and Maier, 2007; Meng et al., 2019). An example of the effectiveness of this strategy was presented by Renner et al. (Renner et al., 2020). They developed an identifiability index with which they demonstrated that the identifiability of material parameters of the Méric Cailletaud crystal plasticity model increased by over 40% when using the residual imprint as opposed to the indentation curve.

Recently, a new trend appeared in optimization-based methods, which is the use of exclusively the residual imprint topography as indentation data in the objective function (Wang et al., 2017; Wang and Wu, 2018; Campbell et al., 2019). This has only been done with spherical indenters. In these methods, the indentation curve is completely ignored, which greatly simplifies the data acquisition during the test. Indeed, it is difficult to obtain a high precision on the depth measurement during an indentation test (Campbell et al., 2019). This is mainly due to the deformation of the testing apparatus during the test, which is measured by the displacement sensor additionally to the actual penetration depth. The extent of this effect depends on the machine compliance, C_f, a parameter which is challenging to estimate.

In this work, since the elasto-plastic behaviour of metals is sought, the indentation curve cannot be neglected as the unloading curve remains the most sensitive part of the indentation data to the tensile modulus.

The use of a spherical indenter has a similar effect to including the residual imprint in the target indentation data, since the strain field generated during the indentation with such geometries of indenters varies with indentation depth. Hence, the maximum strain reached during the indentation experiment depends on the maximum indentation depth, which is known as geometrical similarity (Campbell et al., 2018). This is not the case for sharp indenters, for which the maximum strain remains constant with indentation depth (Tabor, 1951). The use of a spherical indenter is thus equivalent to using multiple sharp indenters of different tip angles, leading to enriched experimental information. The use of a spherical indenter also leads to other advantages (Campbell et al., 2018; Clyne et al., 2021):

- Indenter wear is slower than for pyramidal indenters as the spherical geometry is smoother;
- Indenter manufacturing is easier since no sharp edges are present;
- Finite element simulations converge better and contact detection is less problematic as compared to sharp indenters;
- Finite element simulations can take advantage of the axisymmetrical geometry, leading to considerably faster computational times.

However, it was shown that, for methods assuming a hardening model, the increased sensitivity to material parameters when using a spherical indenters is linked to the value of the ratio of the maximum indentation depth to indenter radius, h_{max}/R. It was shown that a value of 0.2 for this ratio is expected to be sufficient to obtain the necessary sensitivity to the hardening behaviour of various materials (Clyne et al., 2021; Dean and Clyne, 2017; Campbell et al., 2018).

2.2. Scale of indentation

In this work, we are seeking to estimate the bulk properties of metals. To achieve this, the scale of indentation must be sufficiently high as to sample a representative volume of material. Particularly, a minimum number of grains must be within the deformed zone as to obtain an averaged behaviour, representative of the macroscopic tensile curve. Previous studies have estimated that a deformed zone including a minimum of 12 grains is required to represent the bulk behaviour of most materials, but a higher number of grains is preferable (Campbell et al., 2018; Clyne et al., 2021). Indentation data obtained on populations of grains which are too small will be influenced by crystallographic orientation and is not expected to result in accurate estimations of the bulk elasto-plastic behaviour. The interaction between grains would also not be well sampled in this case.

2.3. Numerical optimization

The proposed methodology is based on an optimization problem. In general, such a problem can be cast as:

$$\underset{\mathbf{X}}{\arg\min} \quad f(\mathbf{X}) \tag{1a}$$

$$\text{subjected to} \quad c_k(\mathbf{X}) \leq 0, \ \forall \ k \tag{1b}$$

where $\mathbf{X}$ is the variable vector, $f(\mathbf{X})$ is the objective function, and $c_k(\mathbf{X})$ are the different constraints of the problem. The constraints are defined by inequalities: if $c_k(\mathbf{X}) \leq 0$, then constraint k is respected and variables $\mathbf{X}$ form an acceptable solution candidate.

The problem studied in this work is a complex black-box optimization problem, *i.e.* a problem in which the objective function cannot be analytically derived. The objective function depends on a finite element simulation. Though numerical derivatives could be computed, this would be prohibitive due to the time required to perform this calculation. Also, since the optimization problem is constrained and local minima might exist, the numerical derivatives may lead the optimization to such a local minimum. A derivative free approach is thus preferred for this type of problem.

2.3.1. Optimization algorithm

Different types of optimization algorithms have been used to solve black box optimization problems aiming to estimate the material true stress–true strain curve from indentation tests, such as the interior-point (Wang et al., 2017; Wang and Wu, 2018) and the Nelder–Mead simplex (Chakraborty and Eisenlohr, 2017; Campbell et al., 2018, 2019) algorithms. Trust region algorithms (Bocciarelli et al., 2005; Bolzon et al., 2004; Kang et al., 2012; Beghini et al., 2006), well known in the field of non-linear optimization, are the most commonly used. Even though these algorithms use strategies to escape local minima, it was shown that there exist situations where these strategies can fail

and the algorithm still converges to a local minimum (Audet and Hare, 2017). These algorithms also do not perform well in the presence of optimization constraints.

The algorithm used in this work is the Mesh-Adaptive-Direct-Search algorithm using orthogonal directions (OrthoMADS) implemented using the software NOMAD interfaced on Matlab (Abramson et al., 2009; Audet and Hare, 2017; Abramson et al., 2018; Audet et al., 2009). The choice of this algorithm is justified by its aptitude to escape of local minima and its high performance even in the presence of highly constrained problems. To the knowledge of the authors, direct-search algorithms have not been applied to indentation inverse problems. The algorithm is not described here for sake of brevity, the reader is referred to cited papers for details.

2.3.2. Handling of constraints

The OrthoMADS algorithm is well equipped to deal with constraints. Two types of constraints exist: non-relaxable and relaxable (Audet and Hare, 2017). When a non-relaxable constraint is violated, the objective function, $f(\mathbf{X})$, cannot be evaluated from the variables. For example, an error could result from the finite element model, and therefore, no $f(\mathbf{X})$ value can be obtained. These constraints are dealt with by the use of the extreme barrier algorithm, which simply gives a value of infinity to $f(\mathbf{X})$ when the constraint is violated. This creates jumps in $f(\mathbf{X})$ and leads to additional difficulties for the algorithm to navigate the variable space.

On the other hand, when a relaxable constraint is violated, $f(\mathbf{X})$ can still be evaluated, but the set of variables cannot be the optimal solution. When such constraints are violated, the computed values of the objective function for this set of variables can help the algorithm reach a physically acceptable solution in a faster manner. The value of the relaxable constraint, $c_k(\mathbf{X})$, quantifies the degree to which the variables are far from being acceptable. The progressive barrier algorithm is then used to minimize $c_k(\mathbf{X})$ and $f(\mathbf{X})$ simultaneously until the constraint is respected (Audet and Dennis, 2009).

2.3.3. Optimization using quadratic models

By default, the OrthoMADS algorithm builds quadratic models based on the numerical derivatives of the function to obtain the trial points which will form the initial simplex in the search step. The use of such models should improve the performance of the optimization algorithm. However, in the presence of local minima, these models sometimes lead the algorithm in wrong directions. Also, when non-relaxable constraints are not respected, jumps appear in the objective function. The derivatives can therefore slow down the optimization procedure instead of making it more efficient. This feature can thus be kept activated or disabled depending on the problem studied. Both options are used in different steps of the methodology proposed in this work.

2.3.4. Surrogate function strategy

A common strategy in black box optimization is the use of a surrogate function. Such a function behaves in a similar fashion to the objective function, while requiring a shorter computation time (typically at least 10 times shorter) (Audet and Hare, 2017). When finite element simulations are used in the objective function, the surrogate function can be created, for instance, by using a simplified version of the finite element model. While the value of this function remains an approximation of the actual objective function, it is assumed that the variables generating the local minima and the global minimum should be similar for both functions. In this work, a surrogate function is used to obtain a suitable starting point for the optimization procedure.

2.4. Tabor's indentation strain

Tabor's indentation strain is obtained from an empirical relationship developed to describe the strain field under the indentation with a single value, representative of the general strain state under the indenter (Tabor, 1951). It is defined as:

$$\varepsilon_{ind} = \frac{0.2a}{R}, \tag{2}$$

where a is the contact radius of the indentation and R is the radius of the indenter.

Tabor's indentation strain is used in this work to help limit the strain range of the estimated true stress–true strain curve (see Section 5.2.1).

2.5. Pre-defined hardening models

To compare the performance of the proposed methodology with methods assuming a hardening model, we modified our method to use such models and applied it to the materials studied in this work. Since the most commonly used models for the estimation of true stress–true strain curves by indentation are power laws, the models investigated in this work are those developed by Hollomon (used for example in (Dean and Clyne, 2017)) and Ramberg–Osgood (used for example in (Zhao et al., 2006; Liu et al., 2009; Wang et al., 2016)).

Both models rely on a separation of elastic and elasto-plastic strains as per:

$$\varepsilon = \varepsilon_{el} + \varepsilon_p, \tag{3}$$

where ε_{el} and ε_p are the elastic and plastic parts of the total strain, ε, respectively. For both models, ε_{el} is defined with Hooke's law:

$$\varepsilon_{el} = \frac{\sigma}{E}, \tag{4}$$

where E is the elastic modulus and σ is the applied stress.

For the Hollomon model, the stress in the elastic–plastic region of the tensile curve is defined by:

$$\sigma = \sigma_y \left(1 + \frac{E}{\sigma_y}\varepsilon_p\right)^n, \tag{5}$$

where σ_y is the yield stress and n is the hardening coefficient.

The plastic behaviour following the Ramberg–Osgood model is described by:

$$\sigma = \sigma_y + K\varepsilon_p^n, \tag{6}$$

where K is a plasticity constant.

3. Experimental methodology

3.1. Materials

Four steels with different tensile behaviours were investigated in this study: the eutectoïd steel SAE 1080, the carbon steel ASTM A516, the martensitic stainless steel AISI 415 and the austenitic stainless steel AISI 304L. The general characteristics of these steels are given in Table 1. The true stress–true strain curves were obtained by conventional tensile tests with an Instron 1332 hydraulic universal testing machine in accordance with the ASTM E8/E8M-11 standard (Anon, 2011). The specimens were cylindrical with a diameter of 9.5 mm and a length of 57.2 mm. An extensometer was used with a gauge length of 25 mm to accurately measure the strain. The displacement rate applied by the tensile machine was 1 mm/min.

3.2. Instrumented indentation tests

3.2.1. Indentation curves

The micro-indentation experiments were conducted using the MHT micro-indentation tester, manufactured by Anton-Paar. A calibration procedure developed by the authors in unpublished work led to a machine compliance of 2.99×10^{-2} μm/N.

Spherical indenters compatible with the indentation machine are of sphero-conical shape. Therefore, a diamond sphero-conical indenter with a cone half-angle of 45° and a theoretical tip radius, R, of 50 μm was used. The actual geometry of the indenter tip was measured, for accurate modelling in the finite element simulations, using an Olympus LEXT OLS4100 laser scanning 3D microscope. It was estimated that the transition from a spherical to a conical geometry begins around an indentation depth of 12 μm.

The indentation tests were load-controlled and 20 indents were performed in each of the four materials applying a maximum load of 5 N. This load was chosen as it led to a good repeatability between indentation curves and averaged residual imprints (see Section 3.2.2) for all materials, pointing to the sampling of the bulk properties, while creating indents which were small enough to enable a spacing between indents of 1 mm. At this load, the contact region remained in the purely spherical part of the indenter. The loading and unloading durations were set to 30 s and a dwell period of 30 s was included at the maximum load to prevent any creep effect on the unloading portion of the indentation curve.

The specimens were prepared by electric discharge machining (EDM) to avoid strain hardening the material in the region of interest. The surfaces were first ground sequentially with silicon carbide paper with grits of 600 and 800 particles/in^2. The surfaces were then polished sequentially with 6 μm, 3 μm and 1 μm diamond suspensions deposited on Struers MD-Pan, MD-Mol and MD-Nap cloths, respectively. The final step of the polishing procedure was performed with Mastermet 0.05 μm colloidal silica deposited onto a Struers MD-Chem cloth.

3.2.2. Measurement of residual imprints

The residual imprints were measured with an Olympus LEXT OLS4100 laser scanning 3D microscope. The resolution of the height measurement for this equipment is 10 nm. The step size between height samples was set to 0.25 μm.

Fig. 1(a) shows a typical residual imprint measurement. The topography is not perfectly axisymmetrical. This could be due to material anisotropy. However, isotropy was observed for all steels, by conducting Brinell hardness tests in the longitudinal, long-transverse and short-transverse directions of the four steel plates used in this work. The asymmetry of the imprint could also be caused by heterogeneity at the micrometre scale. However, the indent presented in Fig. 1 was conducted in ASTM A516, and for this material, a circle approximating the size of the plastically deformed zone with a radius of 80 μm on the surface contained approximately 80 grains. This number should be enough to limit the effect of material heterogeneity on the pile-up height. What can be seen in Fig. 1(a) is instead caused mainly by unavoidable setup misalignment between the indenter and the surface. Indeed, when observing the imprints, a preferential direction was noticed (in the north-east quadrant of the indents) in which the pile-up was always higher. This did not change upon rotating the specimen, and rotating the indenter, which points to an inherent misalignment in the setup. Using shims to align the specimens, the angle characterizing this misalignment was estimated to be under 0.5°. These shims could unfortunately not be used to conduct the indentation tests as this led to an increased compliance since the specimens were no longer well supported on the specimen holder of the testing machine.

We therefore obtained the average experimental profile to enable a comparison with the axisymmetrical profile generated by the finite element simulations. This was found to be a satisfactory approximation by comparison with simulated imprints for ASTM A516 steel. Further

Table 1
General characteristics of the four steels studied in this work. The hardening coefficient is defined by the slope of a linear fit of the logarithm of the true stress as a function of the logarithm of the true strain as measured by the tensile tests in the elasto-plastic region.

Material	Eutectoïd steel SAE 1080	Carbon steel ASTM A516	Stainless steel AISI 415	Stainless steel AISI 304L
Microstructure	Pearlitic	Ferrito-pearlitic	Martensitic	Austenitic
Metallurgical state	Normalized	Normalized	Double tempered	Normalized
Hardening coefficient	0.284	0.226	0.151	0.377
Yield stress (MPa)	498	345	500	200
Lüders plateau	Present	Present	Absent	Absent

(a)

(b)

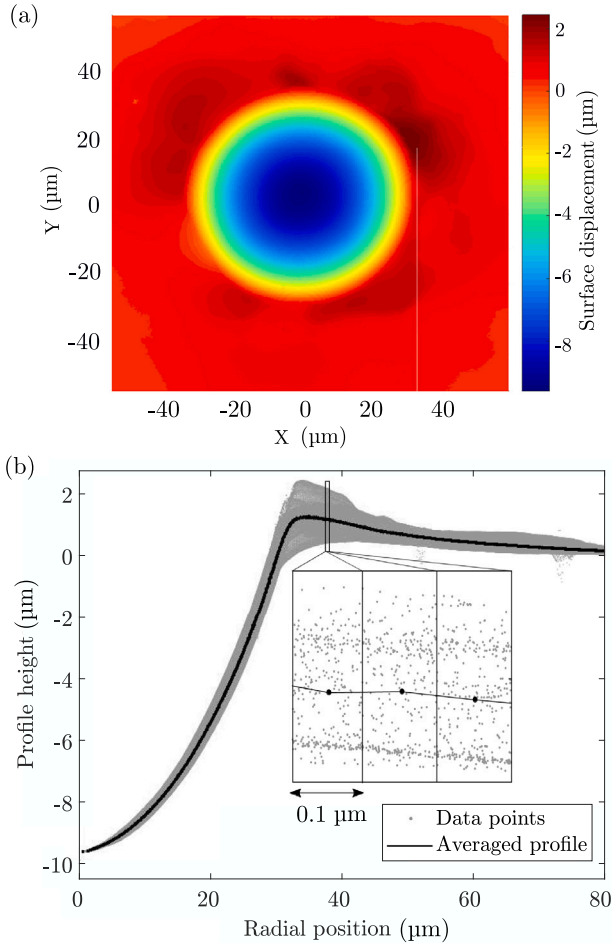**Fig. 1.** Example of the residual imprint of an indent performed with a sphero-conical indenter of radius $R = 50$ μm and a maximum load of 5 N in an ASTM A516 steel specimen: (a) full imprint topography; and (b) corresponding projection of the data points on the r, z plane of a 3D cylindrical coordinates system centred at the tip of the indent. Also shown is the corresponding average profile, obtained by averaging the height of all the projected data points in intervals of 0.1 μm of the radial position.

details are given in Section 6.3. To obtain the average profile, 3D cylindrical coordinates centred at the tip of the indent were projected on the r, z plane. This allowed plotting all the measured data on a single 2D graph, amounting to over 400 000 points per indent, as shown in Fig. 1(b). The radial axis was separated into intervals of 0.1 μm and the average height of all points within each interval was computed to obtain an average profile which was accurate for all radial positions. Such an averaged profile is shown in Fig. 1(b).

4. Finite element modelling

The inverse methodology developed in this paper requires two finite element models:

1. The detailed model, which is accurate and represents an experimental indentation test as closely as possible;
2. The surrogate model, which is a simplified version of the detailed model. It is therefore less accurate, but requires shorter computational times than the detailed model.

4.1. Detailed model

The spherical indentation test was modelled through axisymmetric finite element modelling using the commercial software ANSYS Mechanical APDL 19.2. A large strain kinematics with an updated Lagrangian formulation was used. The load was applied through displacement control of the indenter, which leads to better convergences, particularly when Lüders plateaus are involved.

Geometry
The specimen was modelled as cylindrical with a height $h_d = 323.7$ μm and radius $r_d = 522.6$ μm. These dimensions, determined through a convergence study, were large enough for the specimen to respond to indentation tests as a semi-infinite solid would. The indenter's geometry was modelled to match the measured indenter profile.

Mesh refinement
The mesh comprised of 3801 4-node general axisymmetric elements (SOLID272) and is shown in Fig. 2(a). The smallest element size in the specimen, s_d, was 0.65 μm. The element size in the most refined region of the indenter, $s_{i,d}$, was set to be twice as large as that in the specimen, *i.e.* 1.3 μm. This ratio is believed to lead to better convergences of finite element simulations when contact pairs are involved (Anon, 2018). The most refined region of the mesh contained $60 \times 40 = 2400$ elements and the size of the elements increased gradually as moving away from this refined region to reduce the required computational time.

Boundary conditions
The vertical displacement of the bottom nodes of the specimen was restricted but not their radial displacement.

Material constitutive model
The constitutive behaviour of the specimen was modelled as a rate-independent elasto-plastic material with isotropic hardening. The von Mises yield criterion was used and an associated flow rule plasticity was adopted. The true stress–true strain curve was entered as an input through ANSYS's *Multilinear Isotropic Hardening Model* to eliminate the assumption of a pre-defined hardening model. This was done by prescribing several points of the true stress–true strain curve *i.e.* the optimization variables (defined in Section 5.2.1). The Poisson coefficient was set to $\nu = 0.33$.

To compare the performance of the method with the same method but assuming a predefined hardening model, additional simulations are conducted using the parameters of the Hollomon or Ramberg–Osgood hardening models, previously described in Section 2.5, to input the true stress–true strain curve into ANSYS.

Some materials exhibit tensile curves that cannot be approximated by a power law, like structural steels showing a Lüders plateau. A hybrid model was therefore developed to incorporate a plastic plateau
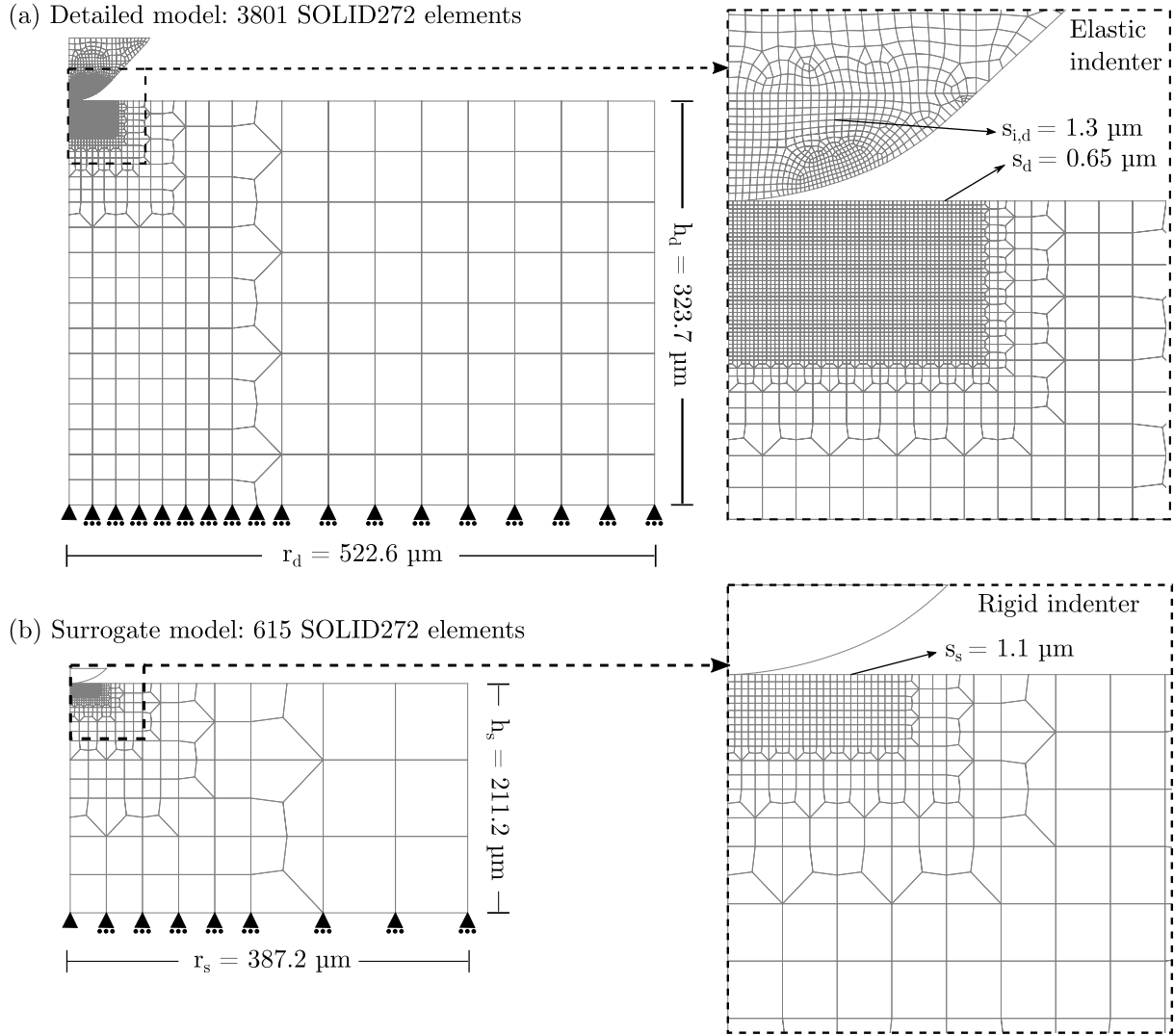
(a) Detailed model: 3801 SOLID272 elements

Elastic indenter

$s_{i,d} = 1.3$ µm
$s_d = 0.65$ µm

$h_d = 323.7$ µm

$r_d = 522.6$ µm

(b) Surrogate model: 615 SOLID272 elements

$h_s = 211.2$ µm

$r_s = 387.2$ µm

Rigid indenter

$s_s = 1.1$ µm

Fig. 2. Finite element meshes for spherical indentation tests with $R = 50$ µm, used in the (a) detailed model; and (b) surrogate model. For each model, the height, h_d or h_s, and radius, r_d or r_s, of the specimen are shown as well as the smallest element size in the specimen, s_d or s_s. The element size set in the most refined region of the indenter, $s_{i,d}$, is also shown for the detailed model.

in the plastic behaviour of the Ramberg–Osgood model. The Ramberg–Osgood model was chosen because it led to the best fit of the hardening region of the tensile curve, defined by the region where strains are larger than the yield strain, for the studied steels ASTM A516 and SAE 1080. In this hybrid model, the strain is divided into elastic and plastic parts, as in Eq. (3). The elastic strain is again defined by Hooke's law, Eq. (4). The plastic behaviour of the hybrid model is described by:

$$\sigma = \begin{cases} \sigma_y, & \text{if } \varepsilon_y \leq \varepsilon_p \leq \varepsilon_{pl} \\ \sigma_y + K(\varepsilon_p - \varepsilon_{pl})^n, & \text{otherwise} \end{cases} \quad (7)$$

where ε_y is the yield strain and ε_{pl} is the strain at the end of the Lüders plateau. The length of the plateau, called the yield point elongation, is the defined by $\varepsilon_{pl} - \varepsilon_y$.

The material behaviour of the indenter was modelled as linearly elastic with diamond's properties: Young's modulus $E = 1\,141$ GPa and Poisson's ratio $v = 0.07$.

Contact and friction

The interaction between the indenter and the specimen was modelled through the augmented Lagrangian method available in ANSYS for node-to-surface contact problems. The indenter surface was defined as the target surface and meshed using contact elements TARGE169,

which overlay the solid elements at the boundary of the indenter. The contact surface was defined on the top of the specimen to extend up to a radial position equal to the radius of the indenter. This contact surface was meshed using contact elements CONTA175, each defined by one node associated to an existing node on the specimen surface.

The effect of friction was incorporated in the model by adopting the Coulomb friction law. The exact value of the friction coefficient, μ, between the indenter and the surface is not known, and it is challenging to measure this quantity experimentally. However, studies have shown that the effect of friction generally saturates at a value larger than $\mu = 0.2$ and many authors use this limiting value as an estimate for the friction coefficient (Heinrich et al., 2009; Campbell et al., 2018). Values of μ between 0.1 and 0.3, at intervals of 0.025, were tested for all materials. It was found that the effect of varying μ on the indentation curves and on the contact region of the residual imprint was negligible. The difference between the maximum loads and residual depths obtained with $\mu = 0.1$ and $\mu = 0.3$ were under 2% and 1%, respectively, for all materials. The effect was more pronounced in the pile-up region of the residual imprint. Increasing μ from 0.1 to 0.3 resulted in a decrease of 17% to 27% in the maximum pile-up height, depending on the material. The effect of varying μ decreased with increasing μ, saturating around a value of 0.225 ± 0.025, depending on materials. A value of $\mu = 0.2$ led to the best agreement between simulation and experimental indentation
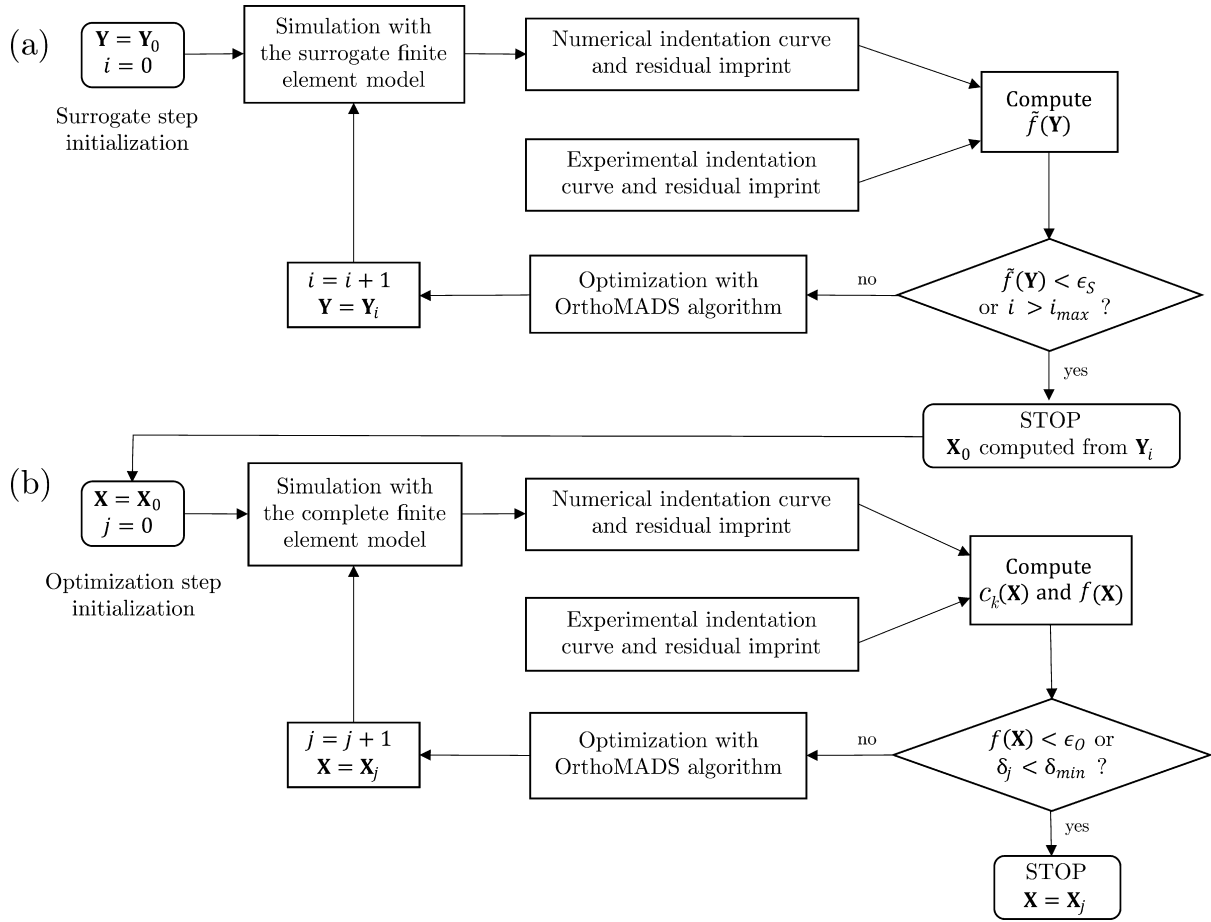
Fig. 3. Flowchart describing the inverse methodology proposed in this paper: (a) The surrogate step, a loop in which the surrogate finite element model is used to obtain the numerical indentation curve and imprint for a set of variables **Y** which is then used in combination with the experimental indentation curve and residual imprint to compute the surrogate function, $\tilde{f}(\mathbf{Y})$. If one of the stopping criteria is met, *i.e.* either $\tilde{f}(\mathbf{Y})$ is less than ϵ_S, the tolerance for the surrogate function or the number of function evaluations is higher than a predetermined maximum, i_{max}, the variable vector **Y** is the solution of this step. Otherwise the OrthoMADS optimization algorithm is used to generate a new variable vector and the process is repeated until convergence. $\mathbf{Y_0}$ is the starting point of this step. The solution of the surrogate step is used to obtain the starting point, $\mathbf{X_0}$, for the optimization step; (b) The optimization step, a loop which follows the same structure as the surrogate step. The differences are that the variables optimized are in the vector **X**, that the objective function $f(\mathbf{X})$ is computed instead of $\tilde{f}(\mathbf{Y})$ and the constraints, $c_k(\mathbf{X})$, are also computed. ϵ_O is the tolerance for the objective function and δ_{max} is the minimum mesh size which is assumed for algorithmic convergence.

residual imprints, and this value was thus used in the finite element models.

4.2. Surrogate model

The surrogate model is a simplified version of the detailed model, which uses a pre-defined hardening model to describe material behaviour.

4.2.1. Simplifications to the detailed model

We looked for simplifications that reduced the computational time while inducing variations lower than a threshold of 20% in important indentation parameters for simulated indentations with a maximum load of 5 N on all materials. These parameters are the maximum load, P_{max}, the contact stiffness (the slope at the onset of unloading in the indentation curve), S, the residual depth, h_r, and the maximum pile-up height, u_{max}. They are illustrated in Fig. 4.

The first simplification that was investigated was the assumption of a frictionless contact, which resulted in an unacceptable error in the residual imprints profiles.

Two investigated simplifications had the desired effect of reducing the computation time while not deteriorating excessively the solution. These simplifications are:

1. Modelling the indenter as a rigid surface instead of as a deformable body;
2. Reducing the number of elements in the mesh from 3 801 SOLID272 elements to 615 by:
   (a) increasing the smallest element size from $s_d = 0.65$ to $s_s = 1.1$ µm in the specimen;
   (b) reducing the size of the most refined region from 60×40 to 26×10 elements;
   (c) reducing the dimensions of the specimen from $h_d = 323.7$ µm and $r_d = 522.6$ µm to $h_s = 211.2$ µm and $r_s = 387.2$ µm.

The mesh of the surrogate model is shown in Fig. 2(b). The computational time for this model was between 8 to 10 times lower than the detailed model, depending on the material modelled.

4.2.2. Material elasto-plastic behaviour

In this surrogate model, the elasto-plastic behaviour of the specimen was modelled using a Ramberg–Osgood power-law (Eq. (6)) instead of using the *Multilinear Isotropic Hardening Model* to reduce the number of variables and eliminate the constraints. This particular power law model was chosen since it leads to the best preliminary approximations for all four studied materials.
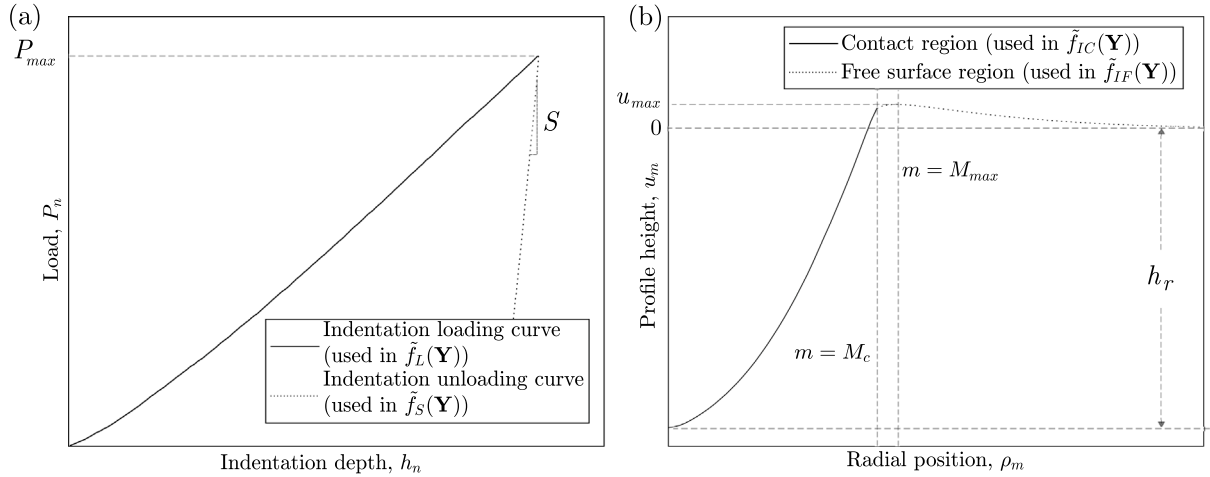
Fig. 4. Typical indentation data: (a) indentation curve comprised of loading and unloading parts and (b) indentation residual imprint profile separated in contact and free surface regions. The maximum pile up height, u_{max}, is at the M_{max}^{th} nodal position, which is used to find the M_c^{th} node by using Eq. (12) to approximate the separation of the contact and free surface zones of the residual imprint.

5. Inverse methodology for the estimation of the true stress–true strain curves

The inverse methodology for estimating the true stress–true strain curves proposed in this paper is summarized in the flow chart shown in Fig. 3. It is separated into two steps: the surrogate step and the optimization step. The details of both steps of the optimization procedure are described in this section.

The surrogate step is a loop in which, for each iteration i, the surrogate finite element model is used to obtain the numerical indentation curve and imprint for a set of variables $\mathbf{Y} = \mathbf{Y_i}$, containing the material parameters of a Ramberg–Osgood model. The obtained numerical indentation data is then used in combination with the experimental indentation curve and residual imprint to compute the surrogate function, $\tilde{f}(\mathbf{Y})$, to be minimized by optimization. The problem defined in the surrogate step is unconstrained. If one of the stopping criteria is met, i.e. either $\tilde{f}(\mathbf{Y})$ is less than ϵ_S, the tolerance for the surrogate function, or the number of function evaluations is higher than a predetermined maximum, the variable vector $\mathbf{Y}$ is the solution of this first step. Otherwise, the OrthoMADS optimization algorithm moves to iteration $i+1$, generates a new variable vector $\mathbf{Y} = \mathbf{Y_{i+1}}$ and the process is repeated until one of the stopping criteria is met.

The solution of the surrogate step is used to generate the starting point, $\mathbf{X_0}$, for the optimization step, which is another loop following the same structure as the surrogate step. For iteration j, the variables in the optimization step, $\mathbf{X} = \mathbf{X_j}$, define a group of 6 points on the true stress–true strain curve and the elastic modulus. The set of constraints, $C(\mathbf{X})$, is computed to verify that the variables are physically acceptable and will not cause convergence problems for the finite element simulations. The numerical indentation curve and imprints are obtained with the detailed finite element model using $\mathbf{X}$ to define material behaviour. From the simulated and experimental indentation data, the objective function $f(\mathbf{X})$ is computed and the algorithm verifies if a stopping criterion is met. In the affirmative, $\mathbf{X}$ defines the final estimate of the material true stress–true strain curve and in the negative, a new variable vector, $\mathbf{X} = \mathbf{X_{j+1}}$, is generated to begin the next iteration $j+1$. The stopping criteria for the optimization step are a tolerance on $f(\mathbf{X})$, ϵ_O, and a maximum size of the algorithm mesh, δ_{max}.

5.1. Surrogate step: determination of suitable starting point for subsequent optimization step

5.1.1. Optimization variables of the surrogate step
Using a Ramberg–Osgood power law requires only four variables to be optimized, namely, the elastic modulus, E, the yield stress, σ_y, the

hardening coefficient, n, and the hardening constant, K, forming the vector:

$$\mathbf{Y} = \left[E, \sigma_y, n, K \right]. \tag{8}$$

The lower and upper bounds, $\mathbf{Y}_l$ and $\mathbf{Y}_u$, applied to vector $\mathbf{Y}$, as well as the starting point of this surrogate optimization step, $\mathbf{Y_0}$, are set to the following values:

$$\mathbf{Y}_l = \left[170 \text{ GPa}, 100 \text{ MPa}, 0.1, 500 \right], \tag{9a}$$

$$\mathbf{Y}_u = \left[230 \text{ GPa}, 1\,000 \text{ MPa}, 0.9, 2\,500 \right], \tag{9b}$$

$$\mathbf{Y_0} = \left[200 \text{ GPa}, 400 \text{ MPa}, 0.5, 1\,500 \right]. \tag{9c}$$

Since the elastic modulus of steels is known to be around 200 GPa, tight bounds were set, corresponding to a +/- 15% variation around this expected value. The bounds on σ_y were chosen to include the tensile behaviour of different types of steels. The hardening coefficient, n, can theoretically take values ranging from 0 to 1. However, values of n too close to these theoretical limits often result in convergence problems in the finite element simulation. Bounds of 0.1 and 0.9 where found to be the limits which avoided such convergence issues. Finally, the bounds on K were chosen in a similar fashion to those on n, to include as many materials as possible while avoiding convergence issues. The starting point, $\mathbf{Y_0}$, was chosen with values in the middle of the intervals.

5.1.2. Surrogate function
The surrogate function $\tilde{f}(\mathbf{Y})$ quantifies the difference between indentation data obtained from simulation with the surrogate model and the target experimental indentation data. It is defined as the weighted sum of four sub-functions:

$$\tilde{f}(\mathbf{Y}) = w_L \tilde{f}_L(\mathbf{Y}) + w_{IC} \tilde{f}_{IC}(\mathbf{Y}) + w_{IF} \tilde{f}_{IF}(\mathbf{Y}) + w_S \tilde{f}_S(\mathbf{Y}), \tag{10}$$

where w_L, w_{IC}, w_{IS} and w_S are the weights attributed to the sub-functions accounting for the indentation loading curve (L), the residual imprint in the contact region (IC), the residual imprint in the free surface region (IF), and the contact stiffness (S). Each of these four sub-functions compare a different portion of the indentation experimental and numerical data, which are illustrated in Fig. 4.

Indentation loading curve, $\tilde{f}_L(\mathbf{Y})$
$\tilde{f}_L(\mathbf{Y})$ quantifies the difference between the numerical and experimental indentation loading curves and is defined by a normalized, dimensionless, sum of squares, as follows:

$$\tilde{f}_L(\mathbf{Y}) = \sum_{n=1}^{N} \frac{1}{N} \left(\frac{P_{exp,n} - P_{num,n}(\mathbf{Y})}{P_{max}} \right)^2, \tag{11}$$

where $P_{num,n}(\mathbf{Y})$ is the load computed at the nth depth of the numerical indentation curve and $P_{exp,n}$ is the load interpolated at the same depth on the target experimental curve. P_{max} is the maximum load measured experimentally and N is the total number of points being investigated on the numerical curve. N depends on the number of sub-steps used by the finite element simulation to achieve convergence.

Residual imprint, $\tilde{f}_{IC}(\mathbf{Y})$ and $\tilde{f}_{IF}(\mathbf{Y})$

The sub-functions $\tilde{f}_{IC}(\mathbf{Y})$ and $\tilde{f}_{IF}(\mathbf{Y})$ quantify the difference between the experimental and numerical residual imprints in the contact region and in the free surface region, respectively. To separate these regions, the M_{max}^{th} nodal position is first derived as the nodal position at which the maximal pile-up height of the numerical imprint is observed. None of the studied materials presented a sink-in behaviour in their residual imprints so this situation was not included in the method at this time. The radial nodal position approximating the separation of the contact and free surface zones of the residual imprint, the M_C^{th} node, is defined by:

$$M_C = M_{max} - 4. \tag{12}$$

This definition was found by finite element simulations to lead to the best compromise to approximate the position of the separation between contact and free-surface regions for all the materials studied in this paper. Fig. 4 shows an example of the M_{max}^{th} and M_C^{th} node radial positions.

$\tilde{f}_{IC}(\mathbf{Y})$ and $\tilde{f}_{IF}(\mathbf{Y})$ are defined by normalized, dimensionless, sums of squares, as follows:

$$\tilde{f}_{IC}(\mathbf{Y}) = \sum_{m=1}^{M_C} \frac{1}{M_C} \left(\frac{u_{exp,m} - u_{num,m}(\mathbf{Y})}{h_r} \right)^2, \tag{13}$$

$$\tilde{f}_{IF}(\mathbf{Y}) = \sum_{m=M_C+1}^{M} \frac{1}{M - M_C} \left(\frac{u_{exp,m} - u_{num,m}(\mathbf{Y})}{u_{max}} \right)^2, \tag{14}$$

where $u_{exp,m}$ and $u_{num,m}(\mathbf{Y})$ are the heights of the experimental and numerical residual imprint profiles, at the mth radial distance from their centre, respectively. M is the total number of points contained in the numerical profile. These M locations correspond to the coordinates of the nodes on the surface of the specimen in the finite element model. The experimental profile is interpolated at the same M radial locations. Parameters h_r and u_{max} are respectively the residual depth, and maximum pile-up height, of the experimental residual imprint profile.

The sub-functions $\tilde{f}_{IC}(\mathbf{Y})$ and $\tilde{f}_{IF}(\mathbf{Y})$ were not combined into a single sub-function because such a combination results in the contact region driving most of the value of the overall sub-function. Indeed, the difference in height between experimental and numerical imprints in both regions would then be normalized by the same value (h_r) and the absolute difference is larger in the contact region, giving this region a higher weight in the sub-function. This would have a detrimental effect on the estimated strain hardening behaviour of the analysed materials because the pile-up region of the residual imprint is very sensitive to the material hardening behaviour. By normalizing the separated sub-functions by their respective peak values, the sub-functions are more representative of the relative differences in height for both regions.

Another approach which was attempted was to obtain the relative error in each point, by normalizing the squared difference ($u_{exp,m} - u_{num,m}$)^2 with the squared experimental height at the point of interest, $u_{exp,m}^2$. However, this led to a substantial increase in relative difference for points located at a height close to zero, taking up most of the value of the sub-function, and a significant decrease in the performance of the method.

Contact stiffness, $\tilde{f}_S(\mathbf{Y})$

$\tilde{f}_S(\mathbf{Y})$ is defined as:

$$\tilde{f}_S(\mathbf{Y}) = \left(\frac{S_{exp} - S_{num}(\mathbf{Y})}{S_{exp}} \right)^2, \tag{15}$$

where S_{exp} and $S_{num}(\mathbf{Y})$ are the contact stiffness of the experimental and numerical indentation curves, respectively.

Weight values

The weights were attributed to the sub-functions after trial and error runs with different weight values combinations with the objective to obtain starting points which were the closest to the target curves for a limited amount of function evaluations. The target indentation data used in these runs were obtained by finite element simulation with the detailed model and the tensile curves of the four studied materials. These weights were $w_L = 50$, $w_{IC} = w_{IF} = 10$ and $w_S = 1$. Since the tensile behaviour of the four studied materials are quite different (see Table 1) it is hypothesized that these weights could be applied to other types of materials.

5.1.3. Optimization algorithm

The unconstrained black-box optimization problem for this step can be written as:

$$\underset{\mathbf{Y}=[E,\sigma_y,n,K]}{\arg \min} \tilde{f}(\mathbf{Y}). \tag{16}$$

The OrthoMADS algorithm (previously described in Section 2.3) is used to solve this problem. The quadratic model feature was kept activated because doing so resulted in starting points which were the closer to the target curves. The first stopping criterion is a convergence value for $\tilde{f}(\mathbf{Y})$ of $\epsilon_S = 10^{-3}$. If the algorithm cannot find the variables $\mathbf{Y}$ which result in $\tilde{f}(\mathbf{Y}) < \epsilon_S$, a maximum number of function evaluations, i_{max}, is set to 150. In this step, the maximum number of evaluations is expected to be the limiting factor. The tolerance ϵ_S is set to a low value to let the algorithm reach the smallest value it can in the allowed number of surrogate function evaluations. For the four materials studied, increasing i_{max} did not lead to significant improvements in the obtained starting points for the optimization step.

5.2. Optimization step: final estimation of true stress–true strain curves

5.2.1. Optimization variables used in the optimization step

A set of pre-defined true strain values is computed prior to the optimization process in such a way that only the stress values at these pre-defined strains are optimized along with the elastic modulus to obtain a group of six points on the estimated true stress–true strain curve. The pre-defined strains are determined in two steps, the first is the determination of the complete range of strains, followed by the positioning of the intermediate strains within this interval.

The range of strains of the estimated tensile curve is the interval between the yield strain, ϵ_y, and a maximum strain value, ϵ_{max}, defining the strain position of the last point extracted. The maximum plastic strain values, ϵ_{max}, were obtained for the four steels studied in this paper with finite element simulations of an indentation with a maximum load of 5 N, using their experimental true stress–true strain curve obtained from macroscopic tensile tests. Since the actual tensile curves of the materials are unknown when the method is applied, it must be estimated from indentation parameters. To do so, Tabor's indentation strain, ϵ_{ind}, defined in Eq. (2), was computed for all materials using the contact radius obtained from the simulated residual imprint (Tabor, 1951). An approximate correlation between the values of ϵ_{max} obtained by finite element simulations and the values of ϵ_{ind} calculated from the simulated indentation data was found to be:

$$\epsilon_{max} \approx \alpha \, \epsilon_{ind}, \tag{17}$$

where, α is a constant. The average value of α obtained by finite element simulations for the materials studied in this work was 3.5, with a relative standard deviation of 16%. Since the tensile behaviour of the four studied steels are quite different, a value of $\alpha = 3.5$ is assumed to be a good approximation for most metals and is used in the proposed methodology. The relationship $\epsilon_{max} \approx 3.5 \, \epsilon_{ind}$ can then be used to estimate the position of the last point to extract on the tensile curve, as ϵ_{ind} can be obtained from experimental indentation data.

The yield strain, defined as $\epsilon_y = \sigma_y/E$, and the value of ϵ_{max} are then limits of the total interval of the strain values of the points to
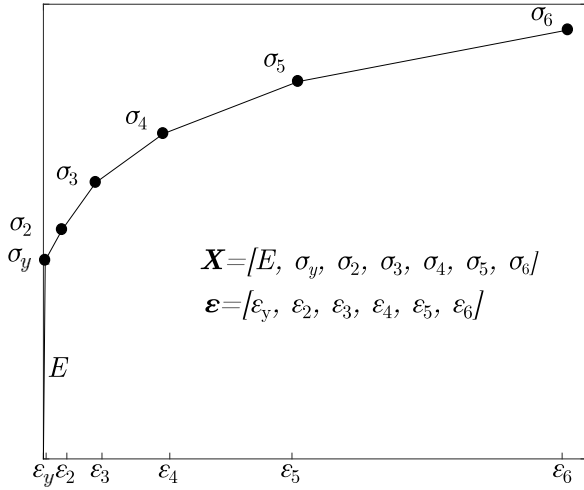
Fig. 5. Illustration of the variables for the optimization step. The variable vector, $\mathbf{X}$, contains the elastic modulus, the yield stress and 5 values of true stress associated with pre-defined strains, obtained with Eqs (17) to (19), that define a tensile curve formed of 6 data points. Linear interpolation is used between estimated points.
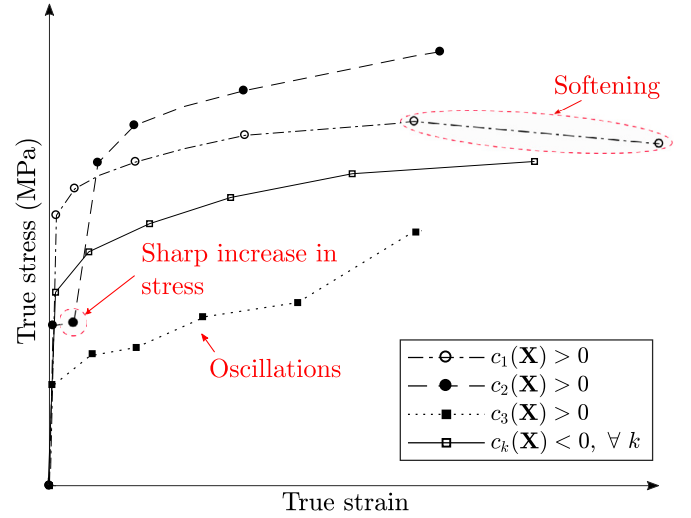

Fig. 6. Examples of true stress–true strain curves which do not respect each of the constraints, in which case the value of the constraint is greater than 0: $c_1(\mathbf{X})$, which serves to avoid softening behaviour in the tensile curves; $c_2(\mathbf{X})$, which serves to avoid sharp increases in stress resulting in convergence failure of finite element simulation; and $c_3(\mathbf{X})$, which serves to avoid oscillations in the tensile curve; as compared to a true stress–true strain curve which respects all constraints ($c_k(\mathbf{X}) < 0 \ \forall \ k$).

be estimated on the tensile curve. From this interval, the intermediate strain values, representing the positions of the 2nd to 5th points to estimate, are calculated by adopting a geometric progression. This choice of progression serves to obtain a higher density of points on the true stress–true strain curve at lower strains and a lower density of points at higher strains. This approach was used to capture more features near the yield stress, like abrupt changes in hardening or a Lüders plateau. The geometric progression used is defined by:

$$\varepsilon_i = \varepsilon_y + \sum_{j=2}^{i} \gamma 2^{j-2}, \quad i = 2, 3, 4, 5, \tag{18}$$

where

$$\gamma = \frac{\varepsilon_{max} - \varepsilon_y}{\sum_{i=2}^{6} 2^{i-2}}. \tag{19}$$

Hence, the material parameters to be identified are the elastic modulus, E, and the yield stress, σ_y, which together give the yield strain, ε_y, as well as the true stress values, σ_2 to σ_6, at pre-defined strains, ε_2 to ε_{max}. The variable vector $\mathbf{X}$ and the predefined strain vector, $\boldsymbol{\varepsilon}$, are:

$$\mathbf{X} = [E, \sigma_y, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6], \tag{20a}$$

$$\boldsymbol{\varepsilon} = \left[\frac{\sigma_y}{E}, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5, \varepsilon_{max} \right], \tag{20b}$$

corresponding to an input true stress–true strain curve with six data points and an optimization problem comprising seven variables. This number of points was found to be the best compromise between stability and the capability to represent the shape of tensile curves accurately. Using these variables, the tensile curve is estimated between the points by linear interpolation. A graphical representation of these vectors and the true stress–true strain curve obtained is shown in Fig. 5.

From the power law parameters obtained by the surrogate step, the stresses corresponding to the predefined strains, $\boldsymbol{\varepsilon}$, are computed to obtain the vector $\mathbf{X}_0$, which is the starting point for the optimization step. The vectors $\mathbf{X}_L$ and $\mathbf{X}_U$, representing the lower and upper bounds of the variables are defined as:

$$\mathbf{X}_L = [170\,000, 100, 100, 100, 100, 100, 100, 100, 100] \text{ MPa}, \tag{21a}$$

$$\mathbf{X}_U = [230\,000, 1\,000, 1\,000, 1\,000, 1\,500, 1\,500, 1\,500, 1\,500, 1\,500] \text{ MPa}. \tag{21b}$$

The same bounds as in the surrogate step were applied to the values of E and σ_y. The bounds on σ_2 to σ_6 were chosen to be sufficiently wide

for the variables to yield materials with various resistances and strain hardening behaviours.

5.2.2. Optimization constraints

Two non-relaxable constraints, $c_1(\mathbf{X})$ and $c_2(\mathbf{X})$, and one relaxable constraint, $c_3(\mathbf{X})$, are applicable to this optimization step and are defined below. Examples of true stress–true strain curves which do not respect each of these constraints are illustrated in Fig. 6, as well as a tensile curve which respects all constraints.

Non-relaxable constraint $c_1(\mathbf{X})$: avoid softening

The first non-relaxable constraint, $c_1(\mathbf{X})$, is respected if all the tangent moduli, $(E_T)_i$, in the plastic portion of the true stress–true strain curve are positive, i.e., when there is no softening. The material is then said to be stable in a Drucker's sense (Drucker, 1957). The tangent moduli, $(E_T)_i$ are defined by:

$$(E_T)_i = \frac{X_{i+2} - X_{i+1}}{\varepsilon_{i+1} - \varepsilon_i}, \quad i = 1, 2, 3, 4, 5, \tag{22}$$

where $\mathbf{X}$ and $\boldsymbol{\varepsilon}$ are the variables and pre-defined strains vectors, respectively. For example, $(E_T)_1$ is the slope of the first segment after the yield strain.

This constraint is non-relaxable because an error results from the finite element software if this constraint is not respected. $c_1(\mathbf{X})$ is expressed as follows:

$$c_1(\mathbf{X}) = \begin{cases} 0, & \text{if } (E_T)_i > 0, \ \forall \ i \\ 1, & \text{otherwise.} \end{cases} \tag{23}$$

Non-relaxable constraint $c_2(\mathbf{X})$: avoid sharp increases in stress

Sharp increases in stress on the tensile curve lead the simulation towards a convergence failure. This happens when E_T after a plastic plateau is very high, making the change in the slope of the tensile curve very abrupt at the end of the plateau. Non-relaxable constraint $c_2(\mathbf{X})$ thus sets a limit of 30 GPa on the change in E_T. This value was found, by trial and error, to allow the finite element simulations to converge. Furthermore, abrupt increases of over 30 GPa in the tangent modulus at the end of a plastic plateau is not physically observed in metals.

Defining the vector $\mathbf{\Delta E_T}$ as the changes in E_T at each applicable discrete point on the curve:

$$\Delta(E_T)_j = (E_T)_{j+1} - (E_T)_j, \quad j = 1, 2, 3, 4, \tag{24}$$

$c_2(\mathbf{X})$ is then expressed as:

$$c_2(\mathbf{X}) = \begin{cases} 0, & \text{if } \Delta(E_T)_j < 30 \text{ GPa}, \ \forall \ j \\ 1, & \text{otherwise.} \end{cases} \tag{25}$$

Relaxable constraint $c_3(\mathbf{X})$: avoid oscillations

The relaxable constraint was implemented to remove oscillations in the optimized tensile curve. To do so, E_T must always decrease as strain increases, and the vector $\mathbf{\Delta E_T}$ should as a consequence only contain negative values, unless a plastic plateau exists. Indeed, the existence of a plateau would imply an increase of E_T at its end.

The first step in creating a mathematical description for this constraint is the definition of a new vector, $\mathbf{\Delta E_T^+}$, which is obtained from the vector $\mathbf{\Delta E_T}$ by changing the negative values to zero. The sum of the elements in $\mathbf{\Delta E_T^+}$ can then serve as the value of $c_3(\mathbf{X})$ if no plastic plateau is present. For example, if: $\mathbf{\Delta E_T} = \begin{bmatrix} -2 & 3 & -4 & 2 \end{bmatrix}$ GPa, then $\mathbf{\Delta E_T^+} = \begin{bmatrix} 0 & 3 & 0 & 2 \end{bmatrix}$ GPa and $c_3(\mathbf{X}) = 5$ GPa. The larger the value of $c_3(\mathbf{X})$, the more pronounced are the oscillations observed in the tensile curve formed by the variables. If a plastic plateau exists, only the values in $\mathbf{\Delta E_T^+}$ associated to points after the strain marking its end, ε_{pl}, would be summed.

It must then be determined if a plastic plateau exists in the tensile curve defined by $\mathbf{X}$. The limit of the slope in the plastic plateau for the steels studied in this work was 1.25 GPa which is also applicable to most steels presenting a plastic plateau. The initial values of E_T are then compared to this threshold. However, a difficulty arises in the determination of the strain value that marks the end of the plateau, ε_{pl}, since it could, in reality, lie in an interval between two predefined strains. We therefore find the limits of this interval to complete the definition of $c_3(\mathbf{X})$. The lower limit of this interval is the strain value after which E_T increases over the limit of 1.25 GPa. The higher limit of the interval is simply the next point on the curve. With these limits known, the index, j_0, of the first $(\Delta E_T^+)_j$ appearing after the plastic plateau can be obtained, which corresponds to the point on the tensile curve after the higher limit of the interval containing ε_{pl}.

$c_3(\mathbf{X})$ is then defined by:

$$c_3(\mathbf{X}) = \sum_{j=j_0}^{4} (\Delta E_T^+)_j. \tag{26}$$

Examples of tensile curves with ε_{pl} in three different intervals are presented in Fig. 7. In each case, the corresponding value of j_0 is shown. This figure also shows how these curves could be adjusted, if needed, to estimate the true position of ε_{pl} once the optimization is completed.

5.2.3. Objective function, $f(\mathbf{X})$

The objective function, $f(\mathbf{X})$, is constructed in the same way as the surrogate function, $\tilde{f}(\mathbf{Y})$, except that the numerical information comes from the detailed finite element model instead of the surrogate model. The objective function is therefore the sum of four sub-functions:

$$f(\mathbf{X}) = w_L f_L(\mathbf{X}) + w_{IC} f_{IC}(\mathbf{X}) + w_{IF} f_{IF}(\mathbf{X}) + w_S f_S(\mathbf{X}), \tag{27}$$

where $f_L(\mathbf{X})$, $f_{IC}(\mathbf{X})$, $f_{IF}(\mathbf{X})$ and $f_S(\mathbf{X})$ are the sub-functions accounting for the indentation loading curve (L), the residual imprint in the contact (IC) and free surface (IF) regions, and the contact stiffness (S), respectively. The constants w_L, w_{IC}, w_{IF} and w_S correspond to the weights attributed to the four parts of the function. Contrary to the surrogate step, a value of 1 for all four weights led to an efficient identification of the true stress–true stress curves. Other weight combinations were tested through the development phase of the method, but the best performance was observed with unity weights. In future works, more rigorous sensitivity and performance studies should be conducted
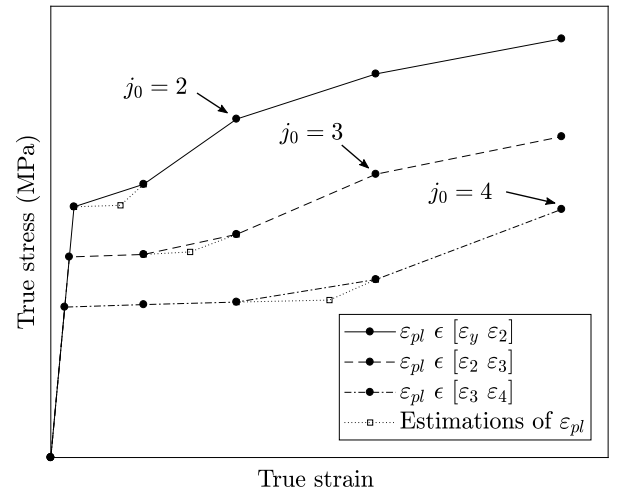

Fig. 7. Examples of tensile curves obtained with different $\mathbf{X}$ in which ε_{pl} lies within the intervals $[\varepsilon_y, \varepsilon_1]$, $]\varepsilon_1, \varepsilon_2]$ and $]\varepsilon_2, \varepsilon_3]$ and extrapolations to estimate the actual position of ε_{pl}.

to find optimal weights which could lead to an increased efficiency of the proposed method.

The difference in weights between the surrogate and objective functions could be explained by the varying influence of the simplifications involved in the surrogate model to different indentation data. The weights used in the surrogate function thus changed to compensate for that effect and the indentation data being less influenced by the simplifications, and thus being more precise, received higher weights. This is not necessary for the objective function, which uses the detailed model.

5.2.4. Optimization algorithm

For the optimization step, the constrained black-box optimization problem can be written as:

$$\underset{\mathbf{X} = [E, \sigma_y, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6]}{\arg \min} f(\mathbf{X}) \tag{28a}$$

$$\text{subjected to} \quad c_1(\mathbf{X}) \leq 0, \ c_2(\mathbf{X}) \leq 0, \ c_3(\mathbf{X}) \leq 0. \tag{28b}$$

This optimization problem is solved using the OrthoMADS algorithm, but in this case the quadratic models feature was disabled. The constraints involved in this step made the quadratic models lead the algorithm in the wrong direction. Two stopping criteria were used for this step. The convergence value of $f(\mathbf{X})$ is set to $\epsilon_O = 10^{-5}$. If this limit is not reached, a minimum mesh size, δ_{min}, is set to 100 MPa for the elastic modulus, E, and 10^{-2} MPa for the stress values σ_y to σ_6.

Table 2 shows a summary of the parameters used in the surrogate and optimization steps which together constitute the proposed methodology.

6. Results and discussion

6.1. Power law curve fitting

This section serves to demonstrate the ability of hardening models to fit experimental tensile data. No indentation data is used in this section. Fig. 8 shows the experimental true stress–true strain curves for the four studied materials as well as the best fitting Ramberg–Osgood, Hollomon and hybrid hardening models for these materials.

It can be observed that different models fit the tensile curves better for the different materials. For instance, AISI 304L can be very accurately represented by a Ramberg–Osgood model, with an average error

Table 2

Summary of the parameters defining the optimization problems to be solved in the surrogate and optimization steps of the proposed methodology.

|	Surrogate step	Optimization step
Variables	Ramberg–Osgood power law : $\mathbf{Y} = [E, \sigma_y, n, K]$	No pre-defined hardening model: $\mathbf{X} = [E, \sigma_y, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6]$
Constraints	None	$c_1(\mathbf{X})$, $c_2(\mathbf{X})$, $c_3(\mathbf{X})$
Weights	$w_L = 50$, $w_{IC} = 10$, $w_{IS} = 10$, $w_S = 1$	$w_L = 1$, $w_{IC} = 1$, $w_{IS} = 1$, $w_S = 1$
Starting point	$\mathbf{Y_0} = [200 \text{ GPa}, 400 \text{ MPa}, 0.5, 1500]$	$\mathbf{X_0}$ derived from the solution of surrogate step
Finite element model	Surrogate model	Detailed model
Stopping criteria	$\epsilon_S = 10^{-3}$, $i_{max} = 150$	$\epsilon_O = 10^{-5}$, $\delta_{min}(E) = 100$ MPa, $\delta_{min}(\sigma_y \text{ to } \sigma_6) = 10^{-2}$ MPa
Use of quadratic models	Enabled	Disabled
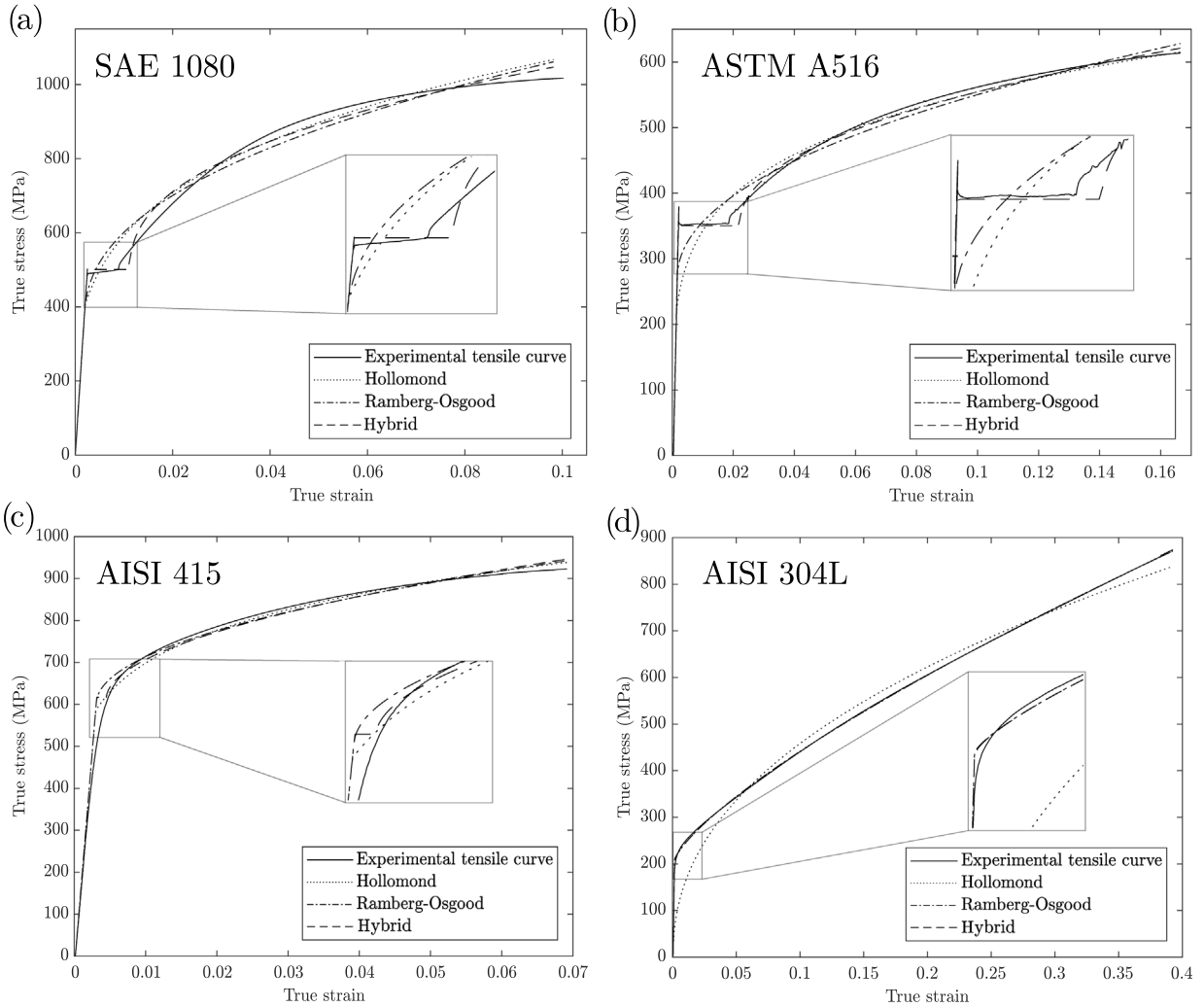
Fig. 8. Experimental true stress–true strain behaviour of four steels compared to the tensile curves obtained by fitting the models developed by Hollomon, Ramberg–Osgood and the hybrid model, to the experimental tensile curves: (a) SAE 1080, (b) ASTM A516, (c) AISI 415 and (d) AISI 304L. This figure shows that some hardening models can better approximate the experimental true stress–true strain curves than others depending on the material and it is concluded that no hardening model can represent the whole tensile curve of all materials.

of 0.35% across the hardening region, but not by a Hollomon model, in which case the yield stress is underestimated by 87% to compensate the inability of this model to fit a linear hardening behaviour. The fit of the hybrid model to the AISI 304L experimental data found a plateau characterized by a yield point elongation of $0.15\varepsilon_y = 1.56 \times 10^{-4}$, which is not significant, and therefore the fit was very similar to that of the Ramberg–Osgood model.

The strain hardening behaviour of AISI 415 can be approximated within an error of 2% with all three models after a strain value of 5×10^{-3}, but there is always an overestimation of the yield stress, ranging from 17%, for the Hollomon model, to 24%, for the hybrid model. In addition, when the hybrid model is used, a non-negligible plateau characterized by a yield point elongation of $0.45\varepsilon_y = 1.42 \times 10^{-3}$ appears. The plateau improves the fit to the experimental data in the hardening region by shifting the onset of strain hardening. The

Table 3

Ramberg–Osgood parameters obtained by the surrogate step of the optimization procedure and target values of elastic modulus, E, and yield stress, σ_y for comparison. These parameters were obtained using numerical indentation data.

	SAE 1080				ASTM A516			
	E (GPa)	σ_y (MPa)	K (MPa)	n	E (GPa)	σ_y (MPa)	K (MPa)	n
Surrogate step	206.5	520	1693	0.59	209	350.0	618	0.660
Target values	205	498	N.A.	N.A.	208	345	N.A.	N.A.
	AISI 415				AISI 304L			
	E (GPa)	σ_y (MPa)	K (MPa)	n	E (GPa)	σ_y (MPa)	K (MPa)	n
Surrogate step	192	620	700	0.36	200	230.0	1250	0.821
Target values	195	500	N.A.	N.A.	199	200	N.A.	N.A.

apparition of a plateau that does not actually exist demonstrates the inability of this hardening model to be used with materials exempt of a plastic plateau.

As expected, the tensile curves of the two steels which actually have a Lüders plateau in their true stress–true strain curves, *i.e.* SAE 1080 and ASTM A516, cannot be well represented by the Ramberg–Osgood and Hollomon models, particularly for strains below 3×10^{-2}. The use of the hybrid model improves the fit, but, for both materials, the yield point elongation is overestimated by at least 20%. For ASTM A516, the strain hardening behaviour is better captured with the hybrid model as compared to the other two models. This is not the case for SAE 1080 steel, whose strain hardening cannot be accurately captured by the Ramberg–Osgood part of the hybrid model.

With these results, it can be concluded that none of these models can accurately fit the entire tensile curve of the four materials studied in this paper. Assuming one of these models beforehand could therefore lead to errors of varying magnitudes in the identified tensile cures, depending on the material/model combination.

It must be noted that it is possible that new models developed in the future may be able to fit more materials with more accuracy, especially regarding materials including a plastic plateau, for which hardening models are not commonly available at this time.

6.2. Numerical study

The first step in assessing the performance of the proposed methodology is a numerical study, which eliminates the influence of modelling errors, namely the errors induced by the finite element modelling hypotheses, and experimental errors. The numerical study serves to test the efficiency of the algorithm chosen to solve the optimization problem, and to quantify the precision with which tensile curves of different materials can be estimated with the method under ideal conditions.

This is undertaken by using indentation test data generated by finite element simulations, rather than experiments, as the input to the inverse method. The true stress–true strain curves used to numerically generate the target indentation data were obtained from extrapolations of the experimental tensile curves and are called the target tensile curves. In the case of ASTM A516 and SAE 1080 steels, the plastic plateau region was modelled by a straight line, eliminating experimental fluctuations and the instability at the upper yield point.

6.2.1. Estimated tensile curves

Table 3 shows the Ramberg–Osgood parameters obtained by the surrogate step of optimization. The target values for the elastic modulus, E, and yield stress, σ_y, are also shown for comparison. The obtained starting points are within +/- 1.5% and 24% of the target tensile properties E and σ_y, respectively. This demonstrates that a suitable starting point for the optimization step is found through the surrogate step and that the surrogate finite element model is effective.

The true stress–true strain curves generated by the surrogate step and upon completion of both steps of the proposed methodology are shown in Fig. 9 for the four steels studied. For comparison purposes,

Fig. 9 also shows the estimated true stress–true strain curves if instead of optimizing the variables in **X**, the parameters of a Hollomon (Eq. (5)), Ramberg–Osgood (Eq. (6)), or hybrid (Eq. (7)) hardening model were used as variables in the optimization step.

With the proposed methodology, the estimated curves lie very close to the target curves for all four studied materials. This demonstrates the capability of the proposed method to obtain different shapes of tensile curves with a good precision. It is also noticeable that assuming a hardening model either deteriorates the estimation of the overall shape of the tensile curve (*e.g.* SAE 1080 steel) or gives equivalent results (*e.g.* hybrid model for ASTM 516 and Ramberg–Osgood model for AISI 304L). The proposed methodology is therefore more versatile because it can lead to accurate results for materials with varying tensile behaviours.

Table 4 lists the errors on elastic modulus, ΔE, and yield stress, $\Delta\sigma_y$, associated with the estimated tensile curves shown in Fig. 9. The proposed methodology evaluates E and σ_y with maximum errors of 0.5% (AISI 304L) and 11.1% (AISI 415), respectively. Assuming a hardening model in the optimization step either increases the error on σ_y, or leads to similar errors, and has no significant impact on the estimation of E. This can be explained by the increased error in the strain hardening region of the estimated curve when using a hardening model that cannot fit this region properly. The yield stress is then altered to compensate that effect.

The precision of the estimated true stress–true strain curves with regard to the target curves in the hardening region is quantified through $|\overline{\Delta\sigma_p}|$, which is the average of the absolute value of the error over the hardening region. This average was obtained with interpolations of the estimated and target true stress–true strain curves at true strain values incremented by 5×10^{-4}, leading to a number of data points higher than 600 for all materials. The values of $|\overline{\Delta\sigma_p}|$ are shown in Table 4. For all materials, this average error is lowest when using the proposed methodology, reaching a maximum value of 2.42% in the case of ASTM A516.

The presence of a plastic plateau was successfully identified in the estimated tensile curves of ASTM A516 and SAE 1080 steels. For ASTM A516, a plateau with a tangent modulus below the limit of 1.25 GPa set in constraint $c_3(\boldsymbol{X})$ is found, so not further manipulation is necessary. However, for SAE 1080, the strain value marking the end of the plateau, ε_{pl}, was not precisely identified. Fig. 10 displays the results obtained by extrapolation of the line between σ_2 and σ_3 to create a horizontal plateau. The yield point elongation then matches very well with that of the target curve. However, since the position of the strains are not optimized, an important error can arise in the estimation of ε_{pl}: for ASTM A516, this value is underestimated by 19.1%.

It was attempted to add the strain values as optimization variables, but the corresponding increase in the number of degrees of freedom made the determination of the tensile curve parameters unstable. For example, the extracted material parameters obtained when using the same method, but adding a single variable to modify the strain values by scaling the geometric progression, led to a tensile curve resulting in an averaged difference with the strain hardening region of the

Table 4

Errors on the estimated elastic modulus, ΔE, and yield stress, $\Delta \sigma_y$, and the average of the absolute value of the error over the hardening region of the true stress–strain curve derived from the extracted material parameters, $\overline{|\Delta \sigma_p|}$, obtained by the optimization step when using the proposed methodology and when assuming either the Ramberg–Osgood, Hollomon or hybrid hardening model. The optimization was conducted using numerically generated indentation data.

|	SAE 1080			ASTM A516		
	ΔE (%)	$\Delta \sigma_y$ (%)	$\overline{|\Delta \sigma_p|}$ (%)	ΔE (%)	$\Delta \sigma_y$ (%)	$\overline{|\Delta \sigma_p|}$ (%)
Proposed methodology	0.2	−2.6	1.24	0.3	1.8	2.42
Assuming Ramberg–Osgood model	0.2	1.7	13.4	0.1	−1.1	9.15
Assuming Hollomon model	−0.2	−9.2	6.12	−0.4	−17.0	4.31
Assuming hybrid model	0.3	7.6	9.19	−0.1	1.5	3.19
	AISI 415			AISI 304L		
	ΔE (%)	$\Delta \sigma_y$ (%)	$\overline{|\Delta \sigma_p|}$ (%)	ΔE (%)	$\Delta \sigma_y$ (%)	$\overline{|\Delta \sigma_p|}$ (%)
Proposed methodology	0.0	11.1	0.96	−0.5	0.2	1.53
Assuming Ramberg–Osgood model	−0.3	28.0	6.67	−0.5	0.5	1.78
Assuming Hollomon model	−0.5	20.8	3.21	−1.5	−24.0	15.08
Assuming hybrid model	0.1	39.6	10.17	−0.6	1.9	3.28
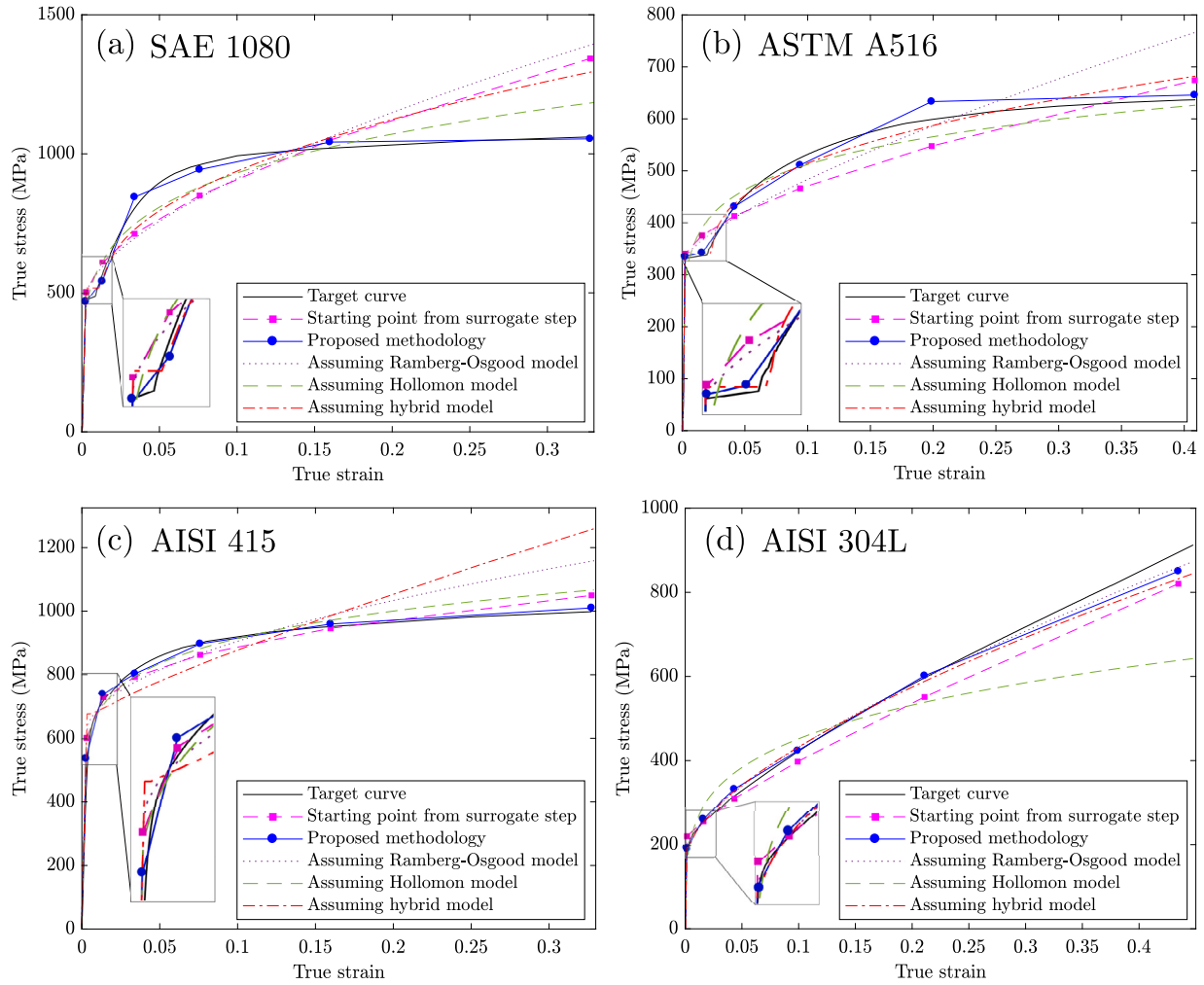
Fig. 9. Comparison of the target true stress–true strain curves to the starting points obtained from the surrogate step, and those estimated with the optimization step when using the proposed methodology and when assuming either the Ramberg–Osgood, Hollomon or hybrid hardening model. The target indentation data used was numerically generated. This figure shows the capacity of the proposed method to approach the general shape of the sought true stress–true strain curves for all materials, which is not the case when assuming hardening models.

experimental tensile curve which was 10 times higher as compared to when strains were pre-defined for a test done on ASTM A516 steel.

Another issue to point out about the strain values is the limited strain range of the estimated tensile curve. Applications may require the knowledge of the plastic behaviour at strains higher than ε_{max}. The estimated true stress–true strain curve could then be extrapolated using a linear or power-law equation, depending on the results. In this work, the estimated tensile curves of ASTM A516, SAE 1080 and AISI 415 tended to perfect plasticity at high strains, and that of AISI 304L displayed a hardening behaviour very close to linear. For these materials, a linear extrapolation of the last slope of the estimated curve could be used to approximate the true stress–true strain behaviour
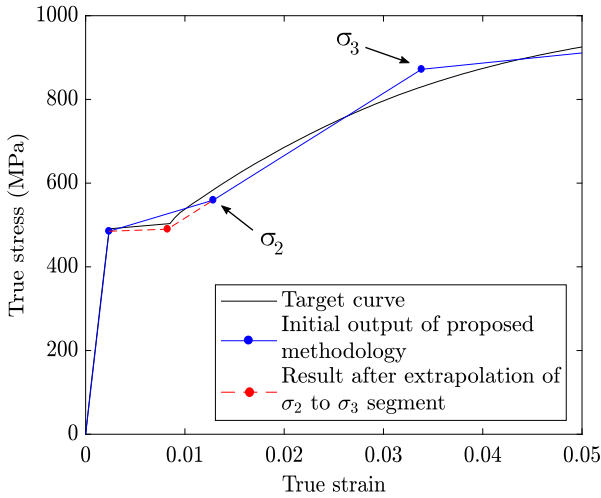
Fig. 10. Estimation of the actual yield point elongation by extrapolation of the σ_2 to σ_3 segment of the true stress–true strain curve obtained for SAE 1080 steel. This figure shows how the value of the strain at the end of the plateau can be well estimated even when it falls in between two values of pre-defined strains.

at higher strains. Other relationships might be better suited for other materials.

Fig. 11 shows the simulated indentation data obtained with the target curve compared to that obtained with the true stress–true strain curves estimated by the proposed methodology or when assuming either the Ramberg–Osgood, Hollomon or hybrid hardening model for ASTM A516 steel. For this material, the difference between the estimated true stress–true strain curves and the target curves varied with the assumed model or lack thereof. Particularly, the error on the yield stress ranged from −1.1% to −17% and the value of $|\overline{\Delta\sigma_p}|$ ranged from 2.42% to 9.15%. However, these variations in the tensile curves did not lead to very large differences in the simulated indentation curves and imprints. Indeed, the errors on the maximum load P_{max} ranged from −0.31% to 1.0% and the errors on u_{max} were all below 1%, except when the true stress–true strain curve estimated assuming a Hollomon hardening model is used, in which case the error reached 2.7%. The similarity in the simulated indentation curves and imprints obtained with the different estimated tensile curves points to the possibility of non-uniqueness issues. A higher depth could perhaps be used in further studies to increase the value of the ratio h_{max}/R and improve the sensitivity of the indentation data to the material parameters used to estimate the tensile curve.

6.2.2. Performance profiles

The convergence behaviour of the method must also be analysed to infer its overall performance. Fig. 12 presents the performance profiles of the surrogate and optimization steps of the proposed method for all four materials. These stairstep curves display the value of $\tilde{f}(\mathbf{Y})$ or $f(\mathbf{X})$ corresponding to the incumbent solution, as a function of number of function evaluations.

For the surrogate step, the maximum number of $\tilde{f}(\mathbf{Y})$ evaluations is reached for all materials before the tolerance based stopping criterion $\tilde{f}(\mathbf{Y}) < \epsilon_S = 10^{-3}$.

For the optimization step, the tolerance based stopping criterion ($f(\mathbf{X}) < \epsilon_O = 10^{-5}$) was reached for all materials, except for AISI 304L, for which the minimum mesh size was reached beforehand. The proposed methodology performed the best for AISI 415 steel, reaching convergence in 594 $f(\mathbf{X})$ evaluations. The performance profile of the algorithm for SAE 1080 steel particularly illustrates the capability of the Ortho-MADS optimization algorithm to get out of local minima. Indeed, the value of $f(\mathbf{X})$ does not improve significantly between 350 and 600 evaluations for that material, since the solution is then trapped into

the neighbourhood of a local minimum. After 600 evaluations however, the algorithm successfully escapes this local minimum, and the value of $f(\mathbf{X})$ starts to decrease effectively and reaches the convergence criterion after an additional 200 evaluations.

To evaluate the use of adding the surrogate step, the optimization step was launched directly with the starting point of the surrogate step (given in Eq. (9c)). The performance profiles show that omitting the surrogate step has an important effect for ASTM A516 and AISI 304L, for which the number of necessary function evaluations to reach one of the stopping criteria is increased by a factor of 2.6 and 2.5, respectively. Also, in the case of ASTM A516, the tolerance based stopping criterion is not reached, meaning that the obtained solution is farther from the target curve obtained using the surrogate step. For AISI 415 and SAE 1080, the number of function evaluations necessary was 2.9% higher and 0.3% smaller, respectively. These differences are considered negligible. Even though only two of the four studied materials were affected by the use of the surrogate step, it is still imperative to include it in the method to ensure the performance is optimized for all materials.

Fig. 13 displays the performance profiles corresponding of the optimization step when using the proposed methodology and when assuming either the Ramberg–Osgood, Hollomon or hybrid hardening models. A general, expected, increase in computational time is observed for the proposed methodology since it optimizes more variables. The worst decrease in computational efficiency is observed for SAE 1080, for which the use of all three hardening models led to a value of $f(\mathbf{X})$ very close to their respective converged result within 80 function evaluations. The proposed method required 819, an increase by a factor of ten. However, the final value of $f(\mathbf{X})$ is below the tolerance ϵ_O for the proposed method while it is not the case for when using the hardening models, which converged by reaching the minimum mesh size values. A lower $f(\mathbf{X})$ was observed with the proposed method for all materials, except for the Ramberg–Osgood model used with AISI 304L in which case the results are comparable. This is in line with results presented in Fig. 9, since a lower $f(\mathbf{X})$ is associated with a better estimation of the true stress–true strain curve.

From these results, it is apparent that the proposed method is more time consuming than when assuming a hardening model. It does however render more accurate estimations of tensile curves in cases when the experimental true stress–true strain curve is not adequately fitted by popular hardening models. The added value of using this novel method thus depends upon the requirements of the user in terms of precision and restrictions in terms of computation time. This method could therefore be better suited for a research environment rather than an industrial environment.

6.3. Experimental validation of the detailed finite element model

Fig. 14 shows a comparison between experimental and numerical indentation data used to validate the detailed finite element model. The simulated indentation data were obtained using the experimental true stress–true strain curves of each material as the input material behaviour to observe how well the model can reproduce the indentation experiments when the correct tensile curves are used.

For further comparison, Table 5 quantifies the modelling errors for the indentation parameters P_{max}, S, h_r and u_{max}. It can be observed that the finite element model captures the material behaviour during the indentation test with a better precision in the case of ASTM A516 and SAE 1080 steels as compared to the two others. The error on u_{max} for AISI 415 reaches a value of −47% and the error on S and P_{max} for AISI 304L reach values of 20% and 10.6%, respectively. This can perhaps be explained by strain-induced phase transformations from austenite to martensite which could take place at higher strains which are locally present under the indenter during the indentation process (Perdahcıoğlu and Geijselaers, 2012). Since martensite is a harder phase than austenite and has a higher volume, its emergence
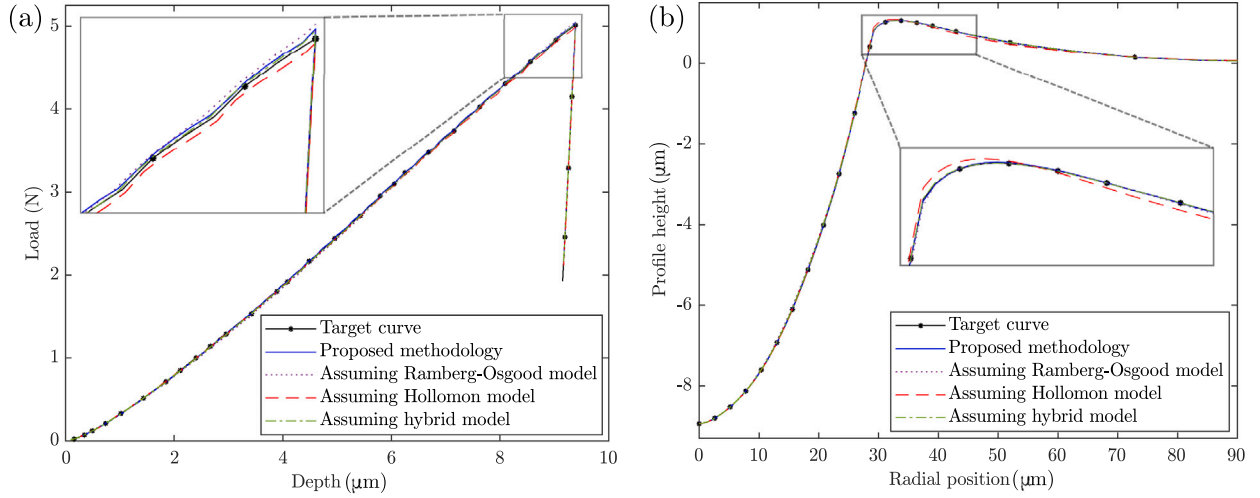
Fig. 11. Indentation data obtained by finite element simulation using the target curve, and the estimated true stress–true strain curves using the proposed methodology, and when assuming either the Ramberg–Osgood, Hollomon or hybrid hardening model for ASTM A516 steel. It can be observed that even though the estimated true stress–true strain curves are different (see Fig. 9(b)), the simulated indentation data are similar, pointing towards possible non-uniqueness issues.
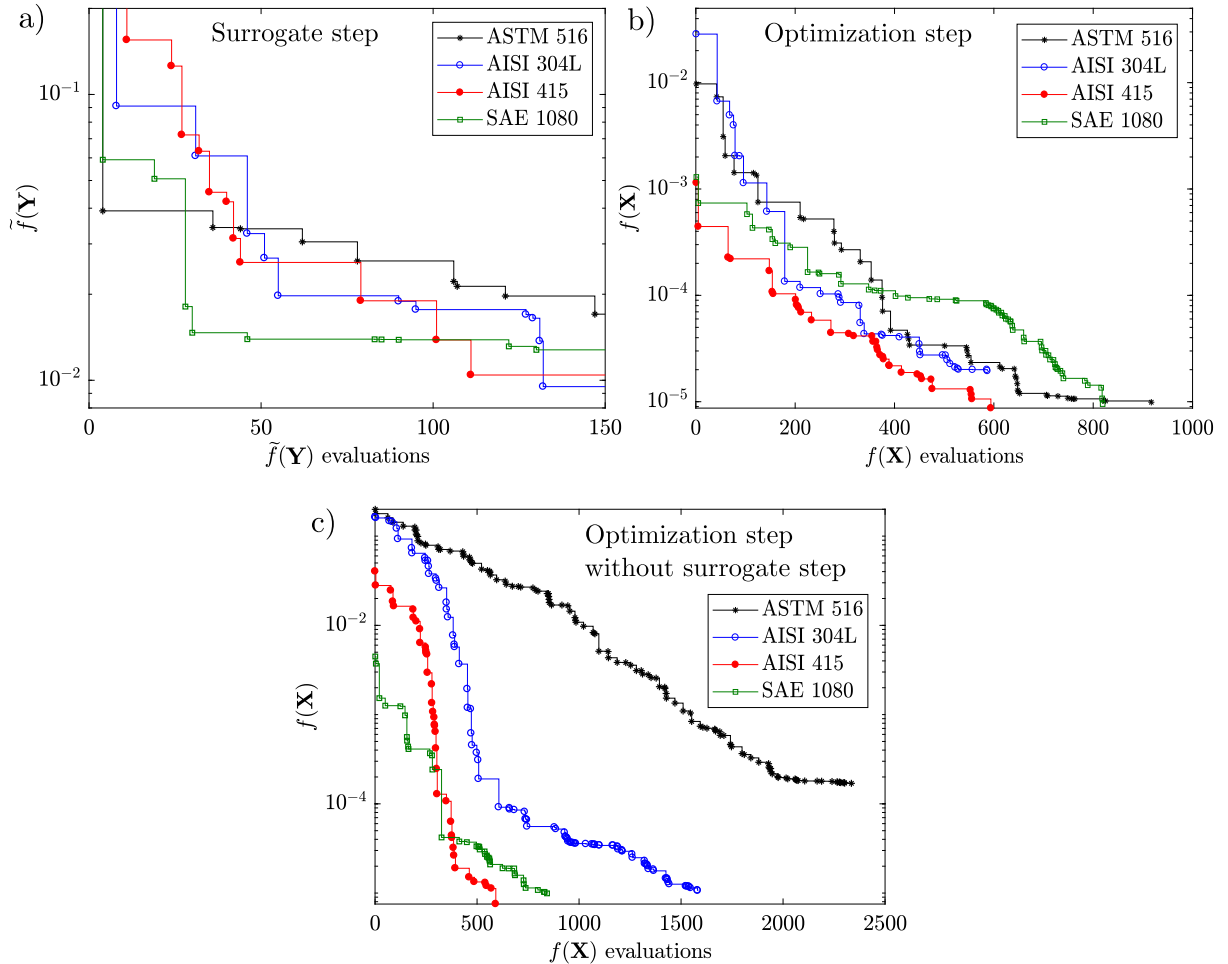

Fig. 12. Performance profiles of the proposed method for four materials using numerically obtained target indentation data for (a) the surrogate step; (b) the optimization step and (c) the optimization step without the surrogate step. This figure shows the efficient convergence behaviour of the chosen optimization algorithm for all materials and the importance of including the surrogate step in the methodology.

would increase P_{max} as compared to the simulated curve, in which this phase transformation is not modelled. The effect on u_{max} and S of this phase transformation are more difficult to analyse and will be studied in future works.

AISI 304L has an austenitic microstructure and a fraction of about 15%–20% of residual or reformed austenite is present in the AISI 415 used in this study, as measured by XRD in a previous study (Fréchette, 2018). ASTM A516 and SAE 1080 do not contain any austenite,
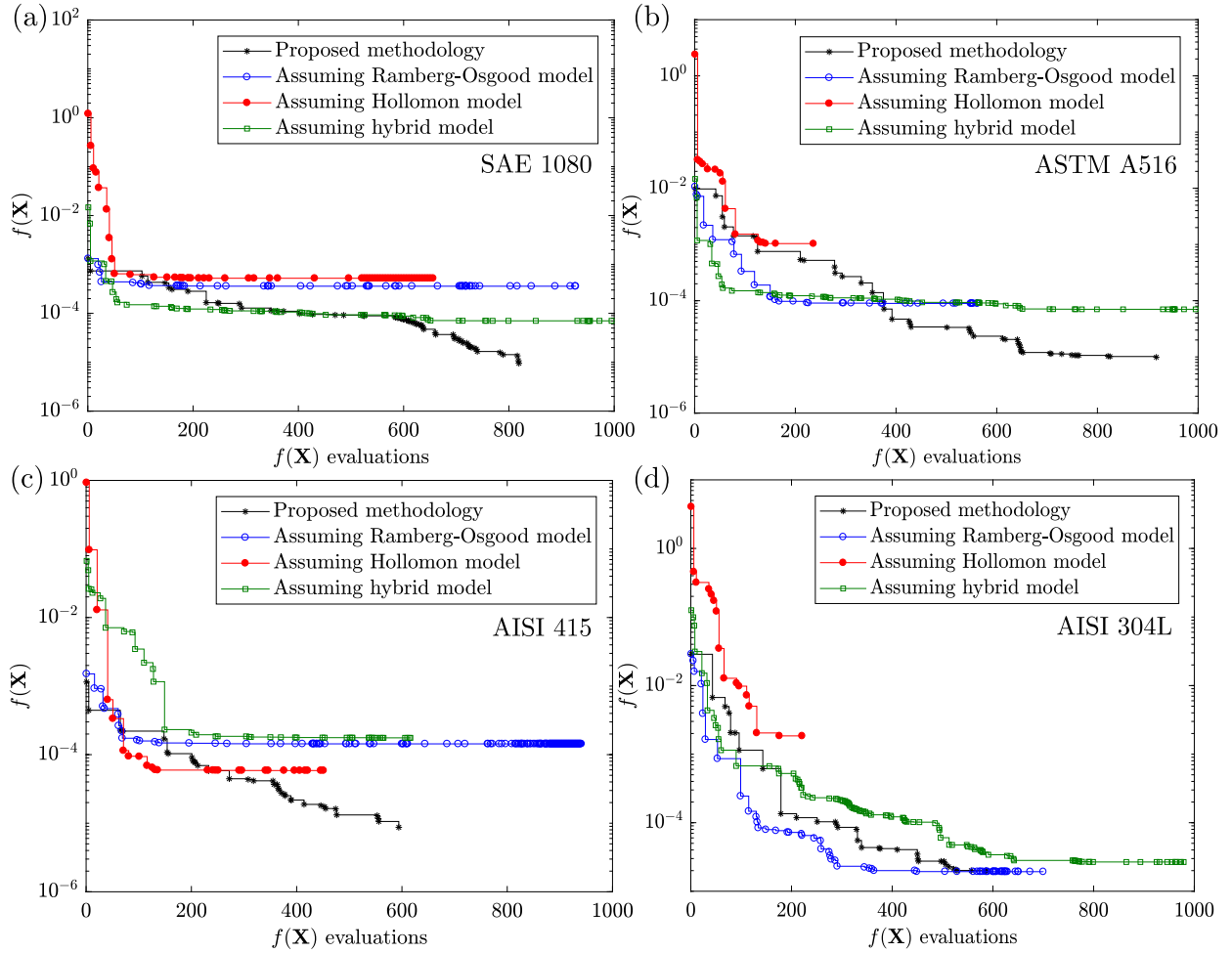
Fig. 13. Comparison of the performance profiles of the optimization step when using the proposed methodology and when assuming either the Ramberg–Osgood, Hollomon or hybrid hardening model. The target indentation data used was numerically generated. This figure shows a general increase in computation time using the proposed method, due to the increased number of variables. However, it is also apparent that the value of $f(\mathbf{X})$ reached by the optimization is lower with the proposed method, meaning a more precise estimation of the true stress–true strain curve.

and therefore cannot be affected by this phenomenon. Austenite to martensite strain-induced transformations are highly dependent upon the stress state (tension vs. compression, uni-axial vs. triaxial) (Per-dahcıoğlu and Geijselaers, 2012). This effect cannot therefore be captured in a single tensile test. For such materials, improvements must be done to the finite element models to include such constitutive behaviour to make this method more accurate. The experimental validation of the proposed methodology is thus performed only for ASTM A516 and SAE 1080 steels.

Another important factor to consider is the scale of indentation. Through metallographic observations of the undeformed specimen surfaces of each steel, the number of grains or martensite blocks included in the plastic zone was estimated, as shown in Table 6. The size of the plastic zones were approximated from the residual imprint profiles as a circle extending to the radial position at which the pile-up height had reduced by half of its maximum value.

Keeping in mind that layers of grains are present in the plastic zone below the surface, ASTM A516 contains an appreciable number of grains. As a minimum number of 12 grains is suggested in the literature (Campbell et al., 2018; Clyne et al., 2021), the indentation results are considered representative of the bulk behaviour for this material. This is in line with the correspondence of the finite element model and the experimental indentation data. The plastic zone of SAE 1080 is also considered to contain a fair amount of grains and with the finite element results, we conclude that the bulk properties are sampled.

Table 5

Differences between indentation maximum load, P_{max}, contact stiffness, S, residual depth, h_r and maximum pile-up height, u_{max}, obtained from experiments and those obtained by the detailed finite element model. Errors are particularly high in regards to u_{max} for AISI 415 steel as well as S and P_{max} for AISI 304L.

Material	ΔP_{max} (%)	ΔS (%)	Δh_r (%)	Δu_{max} (%)
SAE 1080	−3.8	1.7	9.2	−14.2
ASTM A516	4.6	5.0	3.3	9.4
AISI 415	−6.8	−2.5	4.0	−47.0
AISI 304L	−10.6	20.6	3.7	14.6

The same cannot be said for the indentations in AISI 304L and AISI 415 steels, which both contain less than 10 grains or martensite blocks. This could contribute to the errors observed in Fig. 14 and Table 5. However, the indentation results for these two materials were very repeatable, for both the indentation curves and residual imprints after averaging, which could indicate a bulk response. A larger indenter along with higher loads could be used in further studies with these two materials to verify if a better correspondence with the detailed finite element model can be achieved.

6.4. Application of the inverse methodology to experimental indentation data

The true stress–true strain curves obtained by the surrogate and optimization steps of the proposed methodology using experimental
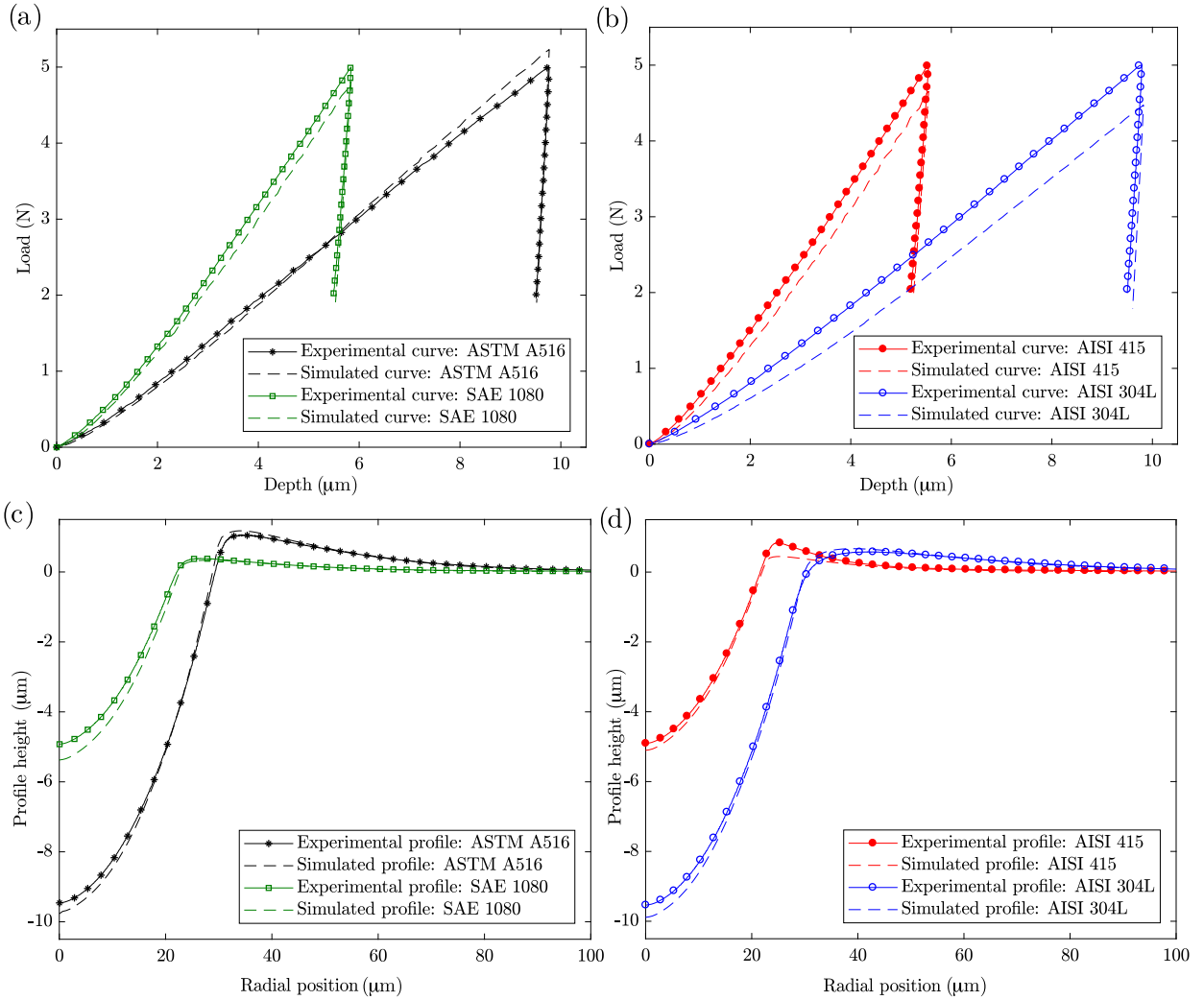
Fig. 14. Comparison of numerical and experimental indentation data obtained with a sphero-conical indenter with $R = 50$ μm, by applying a maximum load of 5 N: (a) and (b) indentation curves and (c) and (d) residual imprints. Only two materials are displayed per graph for clarity. The model is capable of accurately representing the indentation behaviour for ASTM A516 and SAE 1080 steels, but the errors are higher for AISI 304L and AISI 415 steels.

Table 6

Average number of grains (SAE 1080, ASTM A516, AISI 304L) or martensite blocks (AISI 415) contained within a plastic zone approximated from the residual imprint topography on the undeformed surface prior to indentation. These values refer to the number of grains visible on the surface, not the number of grains in the three-dimensional plastic zone.

Material	Plastic zone radius (μm)	Average number of sampled grains or blocks on the surface
SAE 1080	30	12
ASTM A516	50	80
AISI 415	30	8
AISI 304L	60	4

indentation data for ASTM A516 and SAE 1080 steels are shown in Fig. 15, as well as the corresponding macroscopic experimental tensile curves for comparison. In the case of SAE 1080, the extrapolation procedure used in Fig. 10 was used to improve the estimation of ε_{pl}. Extrapolations of the experimental true stress–true strain curves up to values of ε_{max} are also shown since data is not available after the onset of necking.

The parameters extracted from the surrogate step are shown in Table 7. The values of E and σ_y, found in the optimization step, as well as their relative errors as compared to experimental values obtained by

tensile tests and the average error over the hardening region, $|\overline{\Delta\sigma_p}|$, are shown in Table 8.

The results demonstrate that the proposed inverse method captures the presence of a plastic plateau for both materials. For SAE 1080, an overestimation is observed for all but the last pre-defined true strain value, leading to an average error in the hardening region of 3.9%. This overestimation is maximum at the yield stress, where it reaches 20.3%. These differences can partly be explained by the modelling errors for this steel (see Fig. 14). Indeed, the modelled indentation curve using the experimental tensile curve shows a lower reaction force for the same indentation depth as compared to the indentation experiment. To minimize the error between the numerical and experimental indentation curves, the optimized true stress–true strain curve consequently corresponds to a material which is more resistant to plastic deformation. Finally, a significant increase in the error is observed in the estimation of E, when compared to that obtained in the numerical study (−6% vs. 0.2%).

In the case of ASTM A516 steel, the overall shape of the curve is better approximated as compared to SAE 1080. The average error over the hardening region is 0.89%. The yield stress is also better estimated for this material, with an error of −4.8%. This was expected since the modelling errors are smaller for ASTM A516 (see Fig. 14). Finally, the elastic modulus, E, is also underestimated for this steel. Table 5 shows that modelling errors for the contact stiffness, S, are positive for
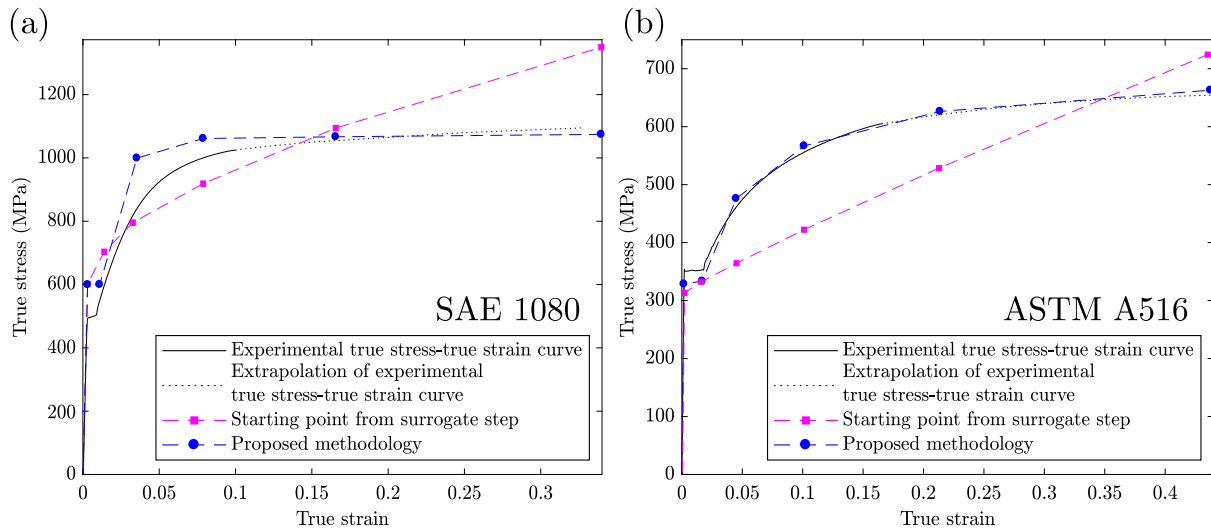
Fig. 15. Comparison of the experimental true stress–true strain curves to those estimated by the surrogate step and upon full completion of the proposed methodology using experimentally obtained target indentation data for ASTM A516 and SAE 1080 steels.

Table 7
Optimization starting points defined by elasto-plastic parameters of a Ramberg–Osgood hardening model obtained by the surrogate step of the optimization procedure for SAE 1080 and ASTM A516 steels. The starting points were obtained using experimental indentation data.

SAE 1080				ASTM A516			
E (GPa)	σ_y (MPa)	K (MPa)	n	E (GPa)	σ_y (MPa)	K (MPa)	n
193	600	1421	0.57	208	339	917	0.89

SAE 1080 and ASTM A516 steels, meaning that the model is stiffer than the experiments and thus explaining the underestimation of E for both materials.

The proposed methodology is thus capable of estimating true stress–true strain curves with satisfying precision even in the presence of experimental and moderate modelling errors. Most important is the capability to approach the general shape of the experimental tensile curve, without the use of a hardening model.

7. Conclusion

In this study, a methodology was developed to estimate the true stress–true strain curves by spherical indentation while avoiding the assumption of a hardening model. The numerical and experimental studies of the method brought to light the following conclusions:

- The proposed method, when used with a spherical indenter ($R = 50$ μm), an applied maximal force of 5 N and numerically generated indentation data, is capable of generating a true stress–true strain curve which fits the shape of the target macroscopic tensile curves of four different steels. The average error between the estimated and experimental tensile curves ranged from 0.96% to 2.42%. It was shown that the estimated tensile curves were either closer to the targets when using the proposed method rather than assuming a hardening model or similar results were obtained. *The proposed method is therefore more versatile than other common methods that assume a hardening model.*
- The computation time required for the convergence of the proposed method can be substantially longer than when using a hardening model. The user must weigh if the added precision is worth the significant increase in required time for their application.
- The choice of the optimization algorithm, OrthoMADS, is justified for this complex problem because of its capacity of performing

well under constrained problems. Such an algorithm is not necessary for common methods assuming a hardening model since no constraints are present. The performance profiles of the proposed methodology for four materials, show convergence is efficient and an example of the algorithm escaping a local minimum is shown in the case of the application to SAE 1080 steel. *This algorithm is thus concluded to behave satisfactorily within the proposed methodology.*
- The experimental application of the method showed that the presence of a plastic plateau for both studied carbon steels, SAE 1080 and ASTM A516, was captured by the proposed method. The average differences between the estimated and experimental tensile curves over their hardening regions are 3.9% for SAE 1080 and 0.9% for ASTM A516. The general overestimation of the extracted true stress–true strain curve for SAE 1080 steel can be explained by the errors induced by the finite element modelling hypotheses. *The proposed method is shown to perform well even when using experimental indentation data.*
- The limitations of the method include the pre-defined strain values, which can lead to significant errors in the estimation of the length of a plastic plateau, if present (*e.g.* $\Delta\varepsilon_{pl} = -19.1\%$ for ASTM A516 in the numerical study). Another important limitation is the capability of the studied material to be modelled by classical plasticity using von Mises yield criterion and isotropic hardening. Materials exhibiting, for example, strain-induced phase transformations will not be accurately modelled since these effects are not taken into account in the current finite element model and the method will yield erroneous results.

The method developed in this work demonstrates that it is possible to estimate the elasto-plastic behaviour of a metal without assuming a hardening model. Many applications can benefit from the proposed method, especially when heterogeneous materials are present, with evolving microstructures, like in welds. It is important to bear in mind that welded joints contain thermal residual stresses, which are known to have significant effects on indentation data (Clyne et al., 2021). These stresses could be measured through another non-destructive method, such as X-ray diffraction, and be modelled by finite element within the inverse methodology.

The ability to obtain reliable local tensile properties in these cases can lead to increasingly precise models which can predict the behaviour of these materials when in service. This can then lead to better design, increased maintenance efficiency and reduction of repair costs.

Table 8
Values of the extracted elastic modulus, E, and yield stress, σ_y, as well as their relative errors with respect to the experimental tensile curves, ΔE and $\Delta \sigma_y$, and the average of the absolute value of the error over the hardening region of the estimated true stress–true strain curve, $|\overline{\Delta \sigma_p}|$, obtained by the material parameters identified in the optimization step of the method for ASTM A16 and SAE 1080 steels. Experimental indentation data were used in the application of the method.

SAE 1080					ASTM A516				
E (GPa)	ΔE (%)	σ_y (MPa)	$\Delta \sigma_y$ (%)	$|\overline{\Delta \sigma_p}|$	E (GPa)	ΔE (%)	σ_y (MPa)	$\Delta \sigma_y$ (%)	$|\overline{\Delta \sigma_p}|$
192.8	−6.0	599.4	20.3	3.9	187.0	−10.1	328.5	−4.8	0.89

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

References

Abramson, M., Audet, C., Couture, G., Dennis, Jr., J., Le Digabel, S., Tribes, C., 2018. The NOMAD project. Software available at https://www.gerad.ca/nomad/. URL https://www.gerad.ca/nomad/.

Abramson, M.A., Audet, C., Dennis, Jr., J.E., Digabel, S.L., 2009. OrthoMADS: A deterministic MADS instance with orthogonal directions. SIAM J. Optim. 20 (2), 948–966.

Anon, 2011. Standard Test Method for Tension Testing of Metallic Materials. ASTM, PA, United-States, Standard.

Anon, 2018. ANSYS Academic Research Mechanical, Release 19.2. Help System : Contact Technology Guide. ANSYS, Inc.

Audet, C., Dennis, Jr., J.E., 2009. A progressive barrier for derivative-free nonlinear programming. SIAM J. Optim. 20 (1), 445–472.

Audet, C., Digbel, S.L., tribes, C., 2009. NOMAD User Guide. Technical Report G-2009-37, Les Cahiers du GERAD, Montreal, Canada.

Audet, C., Hare, W., 2017. Derivative-Free and Blackbox Optimization. Springer.

Beghini, M., Bertini, L., Fontanari, V., 2006. Evaluation of the stress–strain curve of metallic materials by spherical indentation. Int. J. Solids Struct. 43 (7–8), 2441–2459.

Bocciarelli, M., Bolzon, G., Maier, G., 2005. Parameter identification in anisotropic elastoplasticity by indentation and imprint mapping. Mech. Mater. 37 (8), 855–868.

Bocciarelli, M., Maier, G., 2007. Indentation and imprint mapping method for identification of residual stresses. Comput. Mater. Sci. 39 (2), 381–392.

Bolzon, G., Maier, G., Panico, M., 2004. Material model calibration by indentation, imprint mapping and inverse analysis. Int. J. Solids Struct. 41 (11–12), 2957–2975.

Bouzakis, K.-D., Michailidis, N., 2004. Coating elastic–plastic properties determined by means of nanoindentations and FEM-supported evaluation algorithms. Thin Solid Films 469, 227–232.

Bouzakis, K.-D., Michailidis, N., 2006. An accurate and fast approach for determining materials stress–strain curves by nanoindentation and its FEM-based simulation. Mater. Charact. 56 (2), 147–157.

Bouzakis, K.-D., Michailidis, N., Erkens, G., 2001. Thin hard coatings stress–strain curve determination through a FEM supported evaluation of nanoindentation test results. Surf. Coat. Technol. 142, 102–109.

Campbell, J., Thompson, R., Dean, J., Clyne, T.W., 2018. Experimental and computational issues for automated extraction of plasticity parameters from spherical indentation. Mech. Mater. 124, 118–131.

Campbell, J., Thompson, R., Dean, J., Clyne, T., 2019. Comparison between stress-strain plots obtained from indentation plastometry, based on residual indent profiles, and from uniaxial testing. Acta Mater. 168, 87–99.

Chakraborty, A., Eisenlohr, P., 2017. Evaluation of an inverse methodology for estimating constitutive parameters in face-centered cubic materials from single crystal indentations. Eur. J. Mech. A Solids 66, 114–124.

Chen, X., Ogasawara, N., Zhao, M., Chiba, N., 2007. On the uniqueness of measuring elastoplastic properties from indentation: the indistinguishable mystical materials. J. Mech. Phys. Solids 55 (8), 1618–1660.

Clyne, T.W., Campbell, J., Burley, M., Dean, J., 2021. Profilometry-based inverse fem indentation plastometry (pip). Adv. Engrg. Mater..

Dean, J., Clyne, T.W., 2017. Extraction of plasticity parameters from a single test using a spherical indenter and FEM modelling. Mech. Mater. 105, 112–122.

Drucker, D.C., 1957. A Definition of Stable Inelastic Material. Tech. Rep., Brown Univ. Providence RI.

Fréchette, G., 2018. Étude du comportement au fluage de l'alliage 13Cr-4Ni en vue de simuler la redistribution des contraintes résiduelles lors du traitement thermique post-soudage (Master's thesis). École de technologie supérieure.

Heinrich, C., Waas, A., Wineman, A., 2009. Determination of material properties using nanoindentation and multiple indenter tips. Int. J. Solids Struct. 46 (2), 364–376.

Kang, J., Becker, A., Sun, W., 2012. Determining elastic–plastic properties from indentation data obtained from finite element simulations and experimental results. Int. J. Mech. Sci. 62 (1), 34–46.

Liu, Z., Harsono, E., Swaddiwudhipong, S., 2009. Material characterization based on instrumented and simulated indentation tests. Int. J. Appl. Mech. 1 (1), 61–84.

Meng, L., Breitkopf, P., Raghavan, B., Mauvoisin, G., Bartier, O., Hernot, X., 2019. On the study of mystical materials identified by indentation on power law and Voce hardening solids. Int. J. Mater. Form. 12 (4), 587–602.

Pelletier, H., 2006. Predictive model to estimate the stress–strain curves of bulk metals using nanoindentation. Tribol. Int. 39 (7), 593–606.

Perdahcıoğlu, E., Geijselaers, H.J., 2012. A macroscopic model to simulate the mechanically induced martensitic transformation in metastable austenitic stainless steels. Acta Mater. 60 (11), 4409–4419.

Pham, T.-H., Kim, S.-E., 2015. Determination of mechanical properties in SM490 steel weld zone using nanoindentation and FE analysis. J. Construct. Steel Res. 114, 314–324.

Pham, T.-H., Kim, J.J., Kim, S.-E., 2015. Estimating constitutive equation of structural steel using indentation. Int. J. Mech. Sci. 90, 151–161.

Renner, E., Bourceret, A., Gaillard, Y., Amiot, F., Delobelle, P., Richard, F., 2020. Identifiability of single crystal plasticity parameters from residual topographies in Berkovich nanoindentation on FCC nickel. J. Mech. Phys. Solids 138, 103916.

Tabor, D., 1951. The Hardness of Metals: By D. Tabor. Clarendon Press, Oxford.

Wang, M., Wu, J., 2018. Study on the identifiability of material properties using solely the residual imprint in instrumented indentation experiment. In: Multidisciplinary Digital Publishing Institute Proceedings, Vol. 2, p. 443.

Wang, M., Wu, J., Hui, Y., Zhang, Z., Zhan, X., Guo, R., 2017. Identification of elastic-plastic properties of metal materials by using the residual imprint of spherical indentation. Mater. Sci. Eng. A 679, 143–154.

Wang, M., Wu, J., Zhan, X., Guo, R., Hui, Y., Fan, H., 2016. On the determination of the anisotropic plasticity of metal materials by using instrumented indentation. Mater. Des. 111, 98–107.

Zhao, M., Ogasawara, N., Chiba, N., Chen, X., 2006. A new approach to measure the elastic–plastic properties of bulk materials using spherical indentation. Acta Mater. 54 (1), 23–32.