POLITECNICO DI TORINO
Repository ISTITUZIONALE

Digital Metrology for Nanoindentation: Synthetic Data Generator for Error Identification

Original
Digital Metrology for Nanoindentation: Synthetic Data Generator for Error Identification / Maculotti, G., Giorio, L., Genta, G., Galetto, M.. - In: MICROMACHINES. - ISSN 2072-666X. - 16:12(2025), pp. 1-20. [10.3390/mi16121394]

Availability:
This version is available at: 11583/3006391 since: 2026-01-09T11:13:59Z

Publisher:
Multidisciplinary Digital Publishing Institute (MDPI)

Published
DOI:10.3390/mi16121394

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

19 November 2026

micromachines

MDPI

Article

Digital Metrology for Nanoindentation: Synthetic Data Generator for Error Identification

Giacomo Maculotti [1,*], Lorenzo Giorio [2], Gianfranco Genta [1] and Maurizio Galetto [1]

[1] Department of Management and Production Engineering, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy; gianfranco.genta@polito.it (G.G.); maurizio.galetto@polito.it (M.G.)
[2] Department of Applied Science and Technology, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy; lorenzo.giorio@polito.it
* Correspondence: giacomo.maculotti@polito.it

Abstract

Digital metrology enables precise, real-time measurement and data analysis using digital tools, which enhances accuracy and efficiency in manufacturing and quality control. Among key enabling technologies, Digital Twins allow continuous control, enabling predictive maintenance, faster error detection, and optimised performance of the measurement system. A current challenge is establishing traceability for the Digital Twins and for the data processing algorithms implemented in digital metrology. Nanoindentation is a challenging measurement technique that may be susceptible to several random and systematic measurement errors. This work presents a parametric synthetic dataset generator for quasi-static, room-temperature nanoindentation that incorporates correlation and covariance among simulated quantities. The method models indentation responses through a power-law formulation fitted via Orthogonal Distance Regression, allowing for traceable and physics-informed datasets. The generator enables the association of uncertainty with simulated results, supporting its use within a metrological framework. Its performance is benchmarked against non-parametric methods such as bootstrapping, showing comparable accuracy with significantly reduced computational cost and improved representativeness. Furthermore, the methodology can simulate main measurement errors for advanced material characterisation and develops a traceable tool based on synthetic data which could be used to train advanced quality control tools for the detection of main measurement errors.

Keywords: nanoindentation; digital twin; synthetic dataset; quality control

check for updates

Academic Editors: Pengyue Zhao and Costas Charitidis

Received: 28 October 2025
Revised: 30 November 2025
Accepted: 9 December 2025
Published: 11 December 2025

Citation: Maculotti, G.; Giorio, L.; Genta, G.; Galetto, M. Digital Metrology for Nanoindentation: Synthetic Data Generator for Error Identification. *Micromachines* **2025**, *16*, 1394. https://doi.org/10.3390/mi16121394

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

By means of Artificial Intelligence (AI) and Machine Learning (ML), digital transformation can achieve higher performances, effectiveness, and increased sustainability by improving manufacturing processes to reduce waste and defects [1,2]. In fact, AI and ML allow enhanced condition monitoring [3], early fault diagnosis, and reliable predictive maintenance [4], fostering goals of zero-defect and zero-waste manufacturing [5,6]. Among the others, Digital Twins (DT) [7,8] have been established as an effective enabling technology for digital transformation to describe, monitor, predict and control [9,10] manufacturing elements, processes, and systems [11,12]. Increasingly, ML approaches and AI are being used to model DTs of physical entities, for such data-driven approaches [13] can overcome expensive modelling efforts required by physics-based and analytical modelling strategies [14].

However, AI and ML severely depend on the quality of the data. Within this framework, data metrology [15] and virtual metrology [16] are gaining a critical and core role to establish trustworthiness of advanced modelling tools. Specifically, data metrology aims at guaranteeing the quality and traceability of data, while providing all relevant information to quantify uncertainty for the particular application [16]. Furthermore, a critical challenge for digital transformation is the availability of big data, which is required for the application of AI and ML to state-of-the-art monitoring approaches, e.g., DTs [17,18]. Also, traceable virtual experiments and digital twins can only be achieved when trustworthy data are available [19].

Accordingly, synthetic datasets are increasingly resorted to augment the representativeness and the numerosity of the training dataset. In fact, rare or extreme conditions can often be underrepresented due to their low occurrence, e.g., due to failures, or extreme costs and risk of related experiments. To overcome such limitations, synthetic data generation has emerged as a strategic solution. In fact, synthetic data generation consists of creating artificial information that accurately reflects the physical and statistical characteristics of real processes. The technique has evolved along a path of increasing sophistication and has found several applications in current industrial and academic applied research.

Synthetic data can be generated mainly by analytical models, data-driven approaches, or by a hybrid method [20]. Analytical models leverage closed-form solutions based on the physics constitutive and descriptive equations of the considered phenomenon. Although highly elegant and computationally effective, they often rely on simplifying assumptions, which might significantly bias the estimated response. Finite element method (FEM) simulations enable overcoming simplifying assumptions by iteratively reaching numerical solutions, albeit at the cost of introducing a large number of parameters, high computational costs, and a cost-efficiency that drives the approximation and bias of the response. Data-driven approaches stand on the opposite side of modelling. In fact, data-driven approaches build models starting from available data without prior knowledge of the system physics. Machine Learning has made data-driven approaches widely used [21,22], e.g., by exploiting generative adversarial networks [23], reinforcement learning, variational autoencoders, Markov chain models, and Gaussian process regressions [21,22]. Inverse measurement problems solutions try to combine data-driven methods, most typically Bayesian metaheuristics, with numerical simulation, to find a set of hyperparameters that minimises the bias of the simulation [24–26]. However, the solution is computationally and experimentally extremely expensive, potentially requiring thousands of indentations in different conditions, and may be not unique [27]. Lastly, hybrid approaches aim to combine the simplicity and potential of ML with the prior knowledge typical of analytical methods. This is typically obtained by constraining some parameters and by defining the analytical form of the model in agreement with the physics of the system [28].

Examples of synthetic data generation can be found both for modelling complex systems and processes and for advanced quality controls.

As far as the complex systems case is concerned, for example, Toro et al. developed a synthetic data generator for smart measurement sensors of electrical quantities integral to advanced grid infrastructures. The generator was based on an analytical model and was required due to the scarcity of real-world data because of privacy, security, and grid accessibility [29]. Urgo et al. exploited synthetic image generation via virtual reality tools to augment the training dataset for a manufacturing system quality control that tracks objects on the production line [3,4]. Lopes et al. exploited synthetic data generated by a Random Graph model to train a DT of a production system, incorporating rare scenarios such as those related to failures, and to tune the response on different bottleneck,

allocation, and productivity scenarios otherwise impossible or extremely expensive to obtain experimentally [30].

Examples of effective use of synthetic datasets to model and control manufacturing processes can be found in Kim et al., who leveraged synthetic data generation to improve the robustness of DT training for a pick-and-place operation by a collaborative robot within a human–machine interaction framework. Such a scenario requires both object detection and gripping, as well as obstacle avoidance, which, for robust training, requires big data collected in a safe environment. Graphic rendering simulations were leveraged to generate synthetic images used to train the object detection and gripping optimisation, while virtual reality was exploited to simulate obstacles and robot response optimisation by means of a reinforcement learning strategy [31]. Loaldi et al. overcame experimental costs to model the complex interaction of process parameters and part quality in micro-injection moulding processes by synthetic data generation through traceable FEM simulations [32]. Similarly, Solis-Rios et al. reduced the cost of investigating the relationship between process parameters for PE-Oxides nanofibres by Neural Network generation of synthetic data [33].

Similarly, measurement and quality controls largely benefited from synthetic data. For example, Nguyen et al. generated synthetic data for automotive wiring to compensate for the cost of acquiring real data by a neural network for geometrical data and by FEM simulation for electrical functionality [34]. Synthetic data generation by generative artificial networks has also been largely exploited to increase the training datasets for Machine Vision applications, e.g., for surveillance [35], and for visual inspection of quality of welds [36] or packaging of microelectronics [37].

Additionally, widespread adoption of synthetic dataset generation can be found for metrological applications to improve the accuracy and measurement uncertainty of measurement techniques. Lafon et al. developed a methodology to generate reference data to benchmark the performance of registration algorithms [38].

Extensive use can also be found in surface science to support nanometrology and increase the informativeness of characterisation techniques. Necas and Klapetek recently reviewed the use of synthetic data for nanometrology of scanning probe microscopy (SPM), highlighting benefits both in terms of accuracy, uncertainty, and measurement duration. Also, synthetic data allows compiling an atlas of measurement error useful both for training and measurement compensations [39]. Advanced deep learning modelling of SPM-based nanoindentation was obtained by synthetic data generation based on contact models [40].

Among the other surface characterisation techniques, nanoindentation [41,42] has largely been the object of data fusion through synthetic data generation. This work aims at developing a metrological framework for synthetic data generation to support error detection and modelling in nanoindentation. The following subsection will briefly introduce the fundamentals of nanoindentation, review state-of-the-art applications of synthetic data to nanoindentation, and define the scope of the work.

1.1. Fundamentals of Nanoindentation

Nanoindentation, i.e., Instrumented Indentation Test in the nano-range [42], is a depth-sensing, non-conventional hardness measuring technique which allows high-resolution characterisation of mechanical properties of surface layers [41,43]. It allows evaluating estimates of Young modulus, creep, and relaxation behaviour of materials [42,43] and coatings [44–46]. It finds extensive applications, e.g., to study grain and phase size, distribution, and mechanical properties [47], to characterise properties' gradient in functional coatings [46,48], directional manufacturing [47], and finishing processes [49], and it can also measure film thickness [50]. Most innovative applications are now focusing on in-

operando characterisation, typically for high-temperature aerospace environments [51] and for biological materials [52].

Nanoindentation is performed by applying a loading–holding–unloading force cycle to a sample by means of an indenter. The (most typically) applied force F and the indenter penetration depth h in the sample are measured during the whole experiment, thus obtaining the Indentation Curve (IC), i.e., $F(h)$. The characterisation at nanoscales is enabled by the calibrated correlation between the area of the contact surface of the indenter with the sample and the penetration depth, i.e., the area shape function $A(h)$.

Traceability is obtained by calibrating the force and displacement scales, and by calibrating the area shape function parameters and the frame compliance C_f, needed to compensate for the elastic deformation of the indentation platform occurring during the load application. In particular, the corrected contact depth h_c is obtained as in Equation (1a), and the maximum corrected contact depth $h_{c,max}$ results in Equation (1b):

$$h_c = h - h_0 - C_f F \tag{1a}$$

$$h_{c,max} = h - h_0 - C_f F_{max} - \epsilon\left(\frac{1}{S_m} - C_f\right)F_{max} \tag{1b}$$

where h_0 is the zero contact point, $C_f F$ accounts for the elastic deformation of the indentation testing machine, and $\epsilon\left(\frac{1}{S_m} - C_f\right)F_{max}$ for the elastic deformation of the sample at the onset of the unloading [42,53]. The latter term includes a correction factor to cater for the indenter geometry, i.e., ϵ, and for the measured contact stiffness, S_m. S_m is evaluated as the derivative of the IC at the onset of the unloading. The derivative evaluation requires modelling the $F(h)$ relationship [42]. Such a task, although other approximations have been suggested [54,55], is best obtained by leveraging Sneddon's solution of Boussinesq's contact problem [56]. This assumes a power-law relationship, see Equation (2), where the parameters depend on the material of the sample and the indenter and on the indenter geometry.

$$F = \beta_0(h_c - \beta_1)^{\beta_2} \tag{2}$$

Two different models, i.e., sets of parameters, can be obtained by studying the loading and unloading segment of the IC, respectively, with parameters $\{\beta_{0,l},\ \beta_{1,l}, \beta_{2,l}\}$, and $\{\beta_{0,u},\ \beta_{1,u}, \beta_{2,u}\}$. The parameters are estimated by nonlinear regression, with the constraint that $1 < \beta_2 < 2$ [56,57]. For the loading segment, the x-axis offset parameter estimates the zero contact point h_0, while for the unloading segment, it estimates the residual indentation depth h_p.

Although the system is calibrated, several measurement errors can still be introduced during the experimental procedure. First, the sample material creep response shall be compensated to avoid any unwanted dynamic contribution. This is obtained by a suitably long primary holding at the maximum force F_{max}. If the holding is insufficient, i.e., in the presence of a significant room temperature creep, a "nose" at the onset of the unloading, see Figure 1a, can be appreciated, hindering proper evaluation of the sample contact stiffness. Second, a thermal drift is generated between the indenter tip and the sample. This generates a trend in h as a function of time t, biasing the results. The thermal drift can be compensated by introducing a secondary holding, conventionally set a 10% of F_{max}. Such secondary holding allows estimating the slope of $h(t)$, which is then used to correct the penetration depth measurements. Figure 1b shows examples of significant thermal drifts. Last, discontinuities in the IC can indicate an abrupt change in the material behaviour, which can be attributed to phase transformation or cracking, as shown in Figure 1c,d.

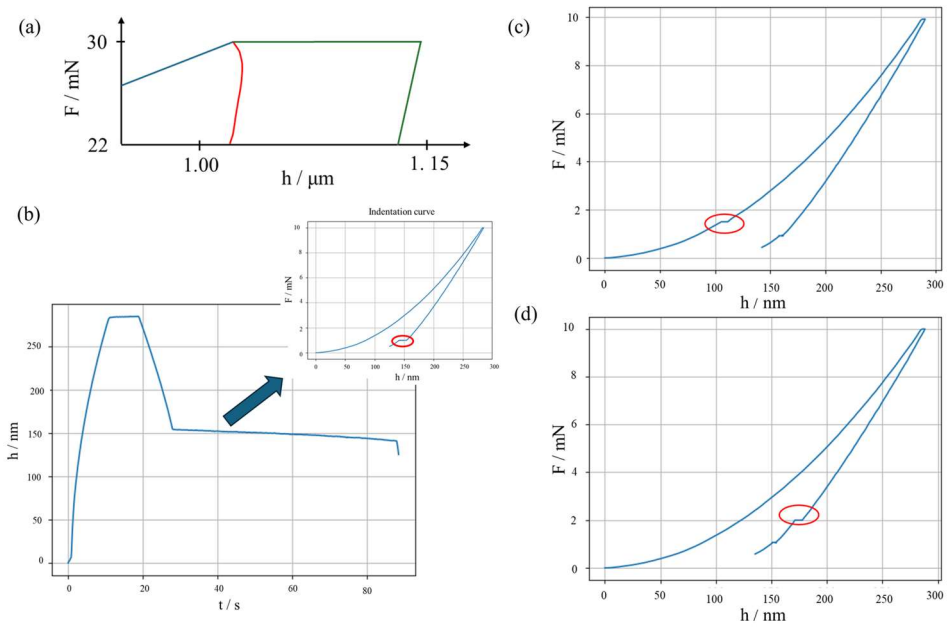


Figure 1. Typical measurement errors in nanoindentation. (**a**) "Nose" due to too short holding: blue, loading; red, unloading with nose; green, primary holding and unloading with correct shape. (**b**) Thermal drift: notice the slope in the secondary holding of $h(t)$; the inset shows the longer secondary holding (circled in red), compared to a typical IC of Figure 2. (**c**) Pop-in event, highlighted by a red circle. (**d**) Pop-out event, highlighted by a red circle.

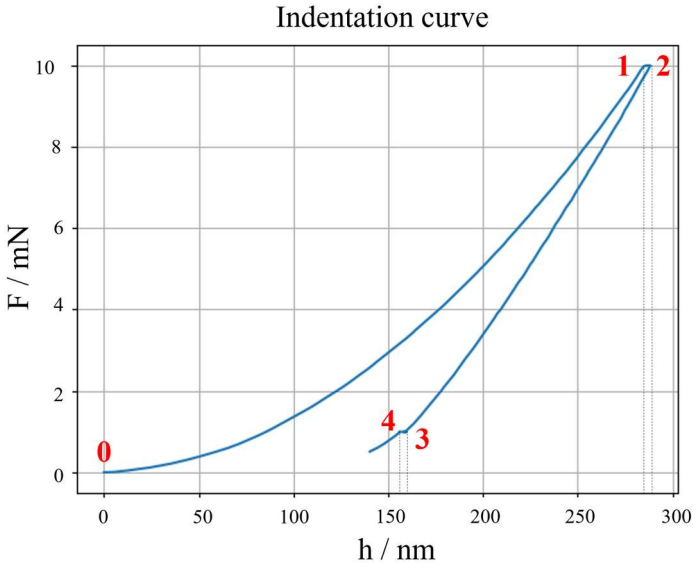


Figure 2. Annotation of critical IC points for continuity: 0: first contact point, 1: end of loading and beginning of primary holding, 2: end of primary holding and start of unloading, 3: end of unloading and start of secondary holding, 4: end of secondary holding.

1.2. Applications of Synthetic Data Generation to Nanoindentation

Nanoindentation has also been the object of synthetic data generation. Several examples can be found in the literature, and, like other measurement techniques—as discussed above—the application involves both advanced characterisation and metrology.

As far as the application for advanced characterisation, these are most often coupled with ML [58]. For example, Koumoulos et al. exploited synthetic data generation by the Synthetic Minority Over-sampling Technique (SMOTE) algorithm to increase the robustness of automatic classification and identification of reinforcement fibres in carbon fibre

reinforced polymers (CFRP) [59]. Giolando et al. exploited a numerical synthetic dataset generator to solve the inverse indentation problem in biological tissues. The synthetic dataset allowed increasing experimental conditions necessary to overcome the lack of uniqueness of the solution [60]. Bruno et al. used k-NN++ synthetic data generation to impute missing data to enhance a correlative microscopy study for transformation induced plasticity (TRIP) steels [61]. Mahmood and Zia trained a generative adversarial network to generate synthetic data to support the prediction of the hardness of diamond-like carbon (DLC) coatings under varying heat treatment processing conditions [62].

Widespread adoption of synthetic dataset generation can also be found for metrological applications. Some target the model optimisation for cutting-edge testing techniques, such as SPM-based nanoindentation. For example, synthetic data were used to train an ML model estimating mechanical parameters [40], which would otherwise require a complex choice of contact model. Other applications can be found for uncertainty estimation and to support calibration methods. In particular, data-driven approaches have been used to investigate the effect of sample size on the accuracy and uncertainty of the calibration of the area shape function parameters and of the frame compliance. They were based on Monte Carlo sampling [63] and on bootstrapping [64]. The generation of synthetic data allowed highlighting a significant contribution of the calibration dataset [64] and a relevant sensitivity of calibration methods to the dataset and experimental conditions [65]. However, although practical, the bootstrapping features a severe limitation in extrapolation and prediction behaviour. Similarly, the Monte Carlo method requires obtaining accurate predictions and avoiding significant overestimation of the measurement uncertainty to model the covariance of simulated quantities, i.e., $F(h)$.

1.3. Scope of the Work

This work aims to develop a traceable synthetic dataset generator for quasi-static room-temperature nanoindentation that can account for the correlation and covariance of simulated quantities. The synthetic dataset generator will be tested for accuracy and will enable associating uncertainty with the simulated results to enable the application within a metrological framework. The developed model will then be benchmarked with other alternatives, e.g., based on bootstrapping, to compare relative performances. Last, the synthetic dataset generator will include the possibility to simulate the most typical measurement errors. Such a feature will allow adopting the developed approach to train and validate advanced quality control tools for automatic measurement error detection in nanoindentation, e.g., a Digital Twin. Innovatively, the proposed method aims at establishing traceability for the synthetic datasets while keeping the experimental and computational cost effective.

The rest of the paper is structured as follows. Section 2 describes the modelling, highlighting methods to establish traceability and to evaluate measurement uncertainty. Section 3 will present results, which are discussed in Section 4. Finally, Section 5 draws conclusions and provides an outlook on future research.

2. Metrological Nanoindentation Synthetic Dataset Generator

This section describes first how a metrological synthetic dataset generator is obtained, and how the uncertainty of the generated data can be estimated. Later, it addresses how the main measurement errors can be simulated. Last, it presents validation methodologies to discuss accuracy with respect to real data, and relative performances with respect to other synthetic nanoindentation dataset generators for metrological applications.

This work proposes to leverage a hybrid approach to model the synthetic data generator. Such a choice, as briefly commented in the Introduction Section, aims to overcome

the simplicity of analytical methods, which requires a strong hypothesis of ideal elastic behaviour and isotropy of the sample; the lack of uniqueness of solutions based on the inverse indentation problem, which challenges metrological applications; and the lack of coherence with the physics of the system of data-driven approaches. Accordingly, a physics-informed modelling is followed.

The statistical methodology is ultimately a parametric simulation approach, where the main simulation parameters, i.e., the distribution shape and related parameters, of measured quantities shall be defined to allow synthetic data generation. The approach proposes to estimate such parameters by exploiting experimental data, thereby allowing for the establishment of traceability for the synthetic dataset. The advantage of this approach lies in the fact that it allows for both modelling correlation between measured quantities and ensuring GUM-compliance for the sake of uncertainty evaluation [66–69].

The synthetic data generation modelling is discussed assuming the most typical choice of a force-controlled loading–holding–unloading force cycle, with a secondary holding for thermal drift compensation. In such a case, both the loading and unloading segments of the IC can be modelled as per Equation (2), while the two holding segments follow the assumption of a constant force as a function of time. Continuity is then constrained, with reference to Figure 2, such that if we consider the i-th indentation, it follows:

$$F_i(h, \boldsymbol{\beta_i}) = \begin{cases} F_{i,l}(h, \boldsymbol{\beta}_{i,l}) = \beta_{i,0,l}(h - \beta_{i,1,l})^{\beta_{i,2,l}}, \;\; F_{0,i} \le F_{i,l} \le F_{1,i} \\ \qquad F_{max,i}, \;\; h_{1,i} \le h_i \le h_{2,i} \\ F_{i,u}(h, \boldsymbol{\beta}_{i,u}) = \beta_{i,0,u}(h - \beta_{i,1,u})^{\beta_{i,2,u}}, \;\; F_{2,i} \ge F_i \ge F_{3,i} \\ \qquad F_{hold2,i}, \;\; h_{3,i} \le h_i \le h_{4,i} \end{cases} \tag{3}$$

where F_{hold2} indicates the average, nominally constant, force of the secondary holding, and $h_{4,i}$ indicates the penetration depth at the end of the unloading.

2.1. Model Training

Model training is based on real data to guarantee traceability. In particular, we can consider a training set of K ICs, each consisting of J data collected at a certain sampling frequency; collected data are typically triplets of $\{F, h, t\}$. Each IC can be modelled with a power-law model for loading and unloading:

$$\hat{F}_{i,l} = F_{i,l}\left(h, \hat{\boldsymbol{\beta}}_{i,l}\right) = \hat{\beta}_{i,0,l}\left(h - \hat{\beta}_{i,1,l}\right)^{\hat{\beta}_{i,2,l}}, \; i: 1, \ldots, K \tag{4a}$$

$$\hat{F}_{i,u} = F_{i,u}\left(h, \hat{\boldsymbol{\beta}}_{i,u}\right) = \hat{\beta}_{i,0,u}\left(h - \hat{\beta}_{i,1,u}\right)^{\beta_{i,2,u}}, \; i: 1, \ldots, K \tag{4b}$$

Parameters $\hat{\boldsymbol{\beta}}_i$ are estimated by nonlinear orthogonal distance regression (ODR) to account for error-in-variables. In fact, ODR assumes error both in the response (ε) and in the regressors (δ), e.g., $F_{i,l}(h, \boldsymbol{\beta}_{i,l}) = f_{i,l}(h + \delta_{i,l}; \boldsymbol{\beta}_{i,l}) + \varepsilon_{i,l}$, and allows obtaining an estimation of model parameters by minimising $\sum_{j=1}^{J_l} \varepsilon_{i,l,j}^2 + \delta_{i,l,j}^2$, which are both expressed for the loading segment of the IC.

Then, it is possible to estimate the average parameters and the mean squared errors of the residuals for both the loading and unloading segments of the IC, e.g., Equations (5a), (5b) and (6) exemplify computations for the loading segment:

$$\overline{MSE_{F,l}} = \frac{1}{K}\sum_{i=1}^{K} MSE_{i,l} \tag{5a}$$

$$MSE_{i,l} = \frac{\sum_{j=1}^{J}\left(F_{i,l,j} - \hat{F}_{i,l,j}\right)^2}{J} \tag{5b}$$

$$\overline{\boldsymbol{\beta}_l}_K = \frac{1}{k}\sum_{i}^{K}\boldsymbol{\beta}_{i,l} \tag{6}$$

According to the modelling of Equation (4a) and (4b) for the loading and unloading segment of the ICs, it is also possible to invert the model to express the estimated penetration depth:

$$\hat{h}_i = h_i\left(F, \hat{\boldsymbol{\beta}}_i\right) = \hat{\beta}_{i,1} + \left(\frac{F}{\hat{\beta}_{i,0}}\right)^{\frac{1}{\hat{\beta}_{i,2}}} \tag{7}$$

2.2. Data Generation

For each synthetic indentation and for each segment of the indentation curve, the number of points J is sampled from a normal distribution with mean and variance evaluated from the K training curves, i.e.,

$$J \sim N\left(\bar{J}_K, s_K^2(J)\right) \tag{8}$$

where the $\bar{J}_K$ indicates the sample average of K observations, and s_K^2 indicates the sample variance of K observations.

Then, for each synthetic indentation and for each segment of the indentation curve, the start time, as in Equation (9a), and end time, as in Equation (9b), are sampled from a normal distribution with mean and variance evaluated from the K curves to train the model. The time vector $\boldsymbol{t}$, see Equation (9d), is then simulated as a linear spacing of J data in the overall duration Δt, as in Equation (9c). The variance of the nanoindentation experiment overall duration is obtained as $s_K^2(\Delta t) = s_K^2(t_{start}) + s_K^2(t_{end})$.

$$t_{start} \sim N\left(\overline{t_{start}}_K, s_K^2(t_{start})\right) \tag{9a}$$

$$t_{end} \sim N\left(\overline{t_{end}}_K, s_K^2(t_{end})\right) \tag{9b}$$

$$\Delta t = t_{end} - t_{start} \sim N\left(\overline{\Delta t}_K, s_K^2(\Delta t)\right) \tag{9c}$$

$$\boldsymbol{t} = \left\{0 : \frac{\Delta t}{J} : \Delta t\right\} \tag{9d}$$

2.2.1. Loading Segment Generation

With reference to Figure 2, the force at point 0, F_0, is sampled from a normal distribution with mean and variance evaluated from the K training ICs, see Equation (10), and the force at point 1, F_1, is sampled from a normal distribution with mean and variance from the primary holding of the K experimental curves, as in Equation (11).

$$F_0 \sim N\left(\overline{F_0}_K, s_K^2(F_0)\right) \tag{10}$$

$$F_1 \sim N\left(\overline{F_1}_K, s_K^2(F_1)\right) \tag{11}$$

Model parameters $\boldsymbol{\beta}_l$ are sampled considering that the estimates by the regression come from a multivariate normal distribution with mean $\overline{\boldsymbol{\beta}_l}_K$ and covariance matrix $\overline{\boldsymbol{\Sigma}_{\boldsymbol{\beta},l}}_K$, as in Equation (12):

$$\boldsymbol{\beta}_l \sim N\left(\overline{\boldsymbol{\beta}_l}_K, \overline{\boldsymbol{\Sigma}_{\boldsymbol{\beta},l}}_K\right) \tag{12}$$

Then, the penetration depth at point 1, h_1, is evaluated as in Equation (13).

$$h_1 = h(F_1, \boldsymbol{\beta_l}) = \beta_{1,l} + \left(\frac{F_1}{\beta_{0,l}}\right)^{\frac{1}{\beta_{2,l}}} \tag{13}$$

Since a force-controlled cycle is considered, the force vector for the loading segment $\boldsymbol{F_l}$ is simulated as a linear spacing vector between F_0 and F_1, as in Equation (14a). A point-wise zero-mean random noise is then added to cater for measurement noise, as in Equation (14b)

$$\boldsymbol{F}_l = \overline{\boldsymbol{F}_l} + e_{F,l} = \left\{F_0 : \frac{F_1 - F_0}{J_l} : F_1\right\} + e_{F,l} \tag{14a}$$

$$e_{F,l} \sim N\left(0, \overline{MSE_{F,l}}\right) \tag{14b}$$

Then, the penetration depth vector is simulated from the regression curve $\boldsymbol{h}_l\left(\overline{\boldsymbol{F}_l}, \boldsymbol{\beta}_l\right)$, as described in Equation (7), constraining $\boldsymbol{h}_l \in [h_0; h_1]$. No further measurement noise, simulating reproducibility, is added as it is already included in the distributions of $\boldsymbol{F}_l$ and $\boldsymbol{\beta}_l$.

2.2.2. Primary Holding Generation

The primary holding is characterised by the maximum applied force F_{max} and by the duration. The primary holding aims to compensate for the room-temperature creep. In fact, the control parameters (F_{max} and duration) induce an increase in the penetration depth Δh computed as the difference between the first and last point of the primary holding, i.e., $\Delta h = h_2 - h_1$, defined with reference to Figure 2. Δh is sampled from a normal distribution with mean and variance evaluated from the K training ICs as in Equation (15), where the variance is obtained by combining individual contributions as $s_K^2(\Delta h) = s_K^2(h_2) + s_K^2(h_1)$.

$$\Delta h \sim N\left(\overline{\Delta h}_K, s_K^2(\Delta h)\right) \tag{15}$$

Then, from the previously sampled h_1, as per Equation (13), the penetration depth at point 2, h_2, is evaluated as in Equation (16a). The penetration depth vector for the primary holding, $\boldsymbol{h}_{holding1}$, is then simulated as a linear spacing vector between h_1 and h_2, as in Equation (16b). A point-wise random noise is added to account for the accuracy, assumed normally distributed, with zero mean, and variance proportional to the measured penetration depth, see Equation (16c) [42]. No measurement reproducibility is further added because it is already included in the distributions of h_1 and Δh.

$$h_2 = h_1 + \Delta h \tag{16a}$$

$$\boldsymbol{h}_{holding1} = \left\{h_1 : J_{holding1} : h_2\right\} + Acc_h \tag{16b}$$

$$Acc_h \sim N\left(0, \left(h \cdot u_{Acc,h,\%}\right)^2\right) \tag{16c}$$

The associated force vector, $\boldsymbol{F}_{holding1}$, is simulated, as in Equation (17a), as a constant force value equal to F_1 (with $F_1 = F_2$), sampled from Equation (11). A point-wise random noise is added to account for the assumed normally distributed, with zero mean, and variance proportional to the measured force, see Equation (17b) [42]. No measurement reproducibility is further added because it is already included in the distributions of F_1.

$$\boldsymbol{F}_{holding1} = \left\{F_1 : J_{holding1} : F_2\right\} + Acc_F \tag{17a}$$

$$Acc_F \sim N\left(0, \left(F \cdot u_{Acc,F,\%}\right)^2\right) \tag{17b}$$

2.2.3. Unloading Generation

The process is highly similar to the generation of the loading, described in Section 2.2.1. The parameters $\boldsymbol{\beta}_u$ of the regression are sampled from a multivariate normal distribution with mean $\overline{\boldsymbol{\beta}_{u\,K}}$ and covariance matrix $\overline{\boldsymbol{\Sigma}_{\boldsymbol{\beta},u\,K}}$, both evaluated from the K training ICs, i.e., similarly to Equation (12) $\boldsymbol{\beta}_u \sim N\left(\overline{\boldsymbol{\beta}_{u\,K}}, \overline{\boldsymbol{\Sigma}_{\boldsymbol{\beta},u\,K}}\right)$. The force at point 3, with reference to Figure 2, is sampled from a normal distribution with mean and variance from the secondary holding, empirically estimated, as in Equation (18), which allows estimating the penetration depth at point 3, h_3, by inverting the regression model, as in Equation (19).

$$F_3 \sim N\left(\overline{F_{3K}}, s_K^2(F_3)\right) \tag{18}$$

$$h_3 = h(F_3, \boldsymbol{\beta}_u) = \beta_{1,u} + \left(\frac{F_3}{\beta_{0,u}}\right)^{\frac{1}{\beta_{2,u}}} \tag{19}$$

Then, the force vector $\boldsymbol{F}_u$ is simulated as a linear spacing vector between $F_2 (= F_1)$ and F_3, as described by Equation (20a). A point-wise random noise is added, assuming a zero mean normal distribution with the variance estimated as in Equation (20b).

$$\boldsymbol{F}_u = \overline{\boldsymbol{F}_u} + e_{F,u} = \left\{F_2 : \frac{F_3 - F_2}{J_u} : F_3\right\} + e_{F,u} \tag{20a}$$

$$e_{F,u} \sim N\left(0, \overline{MSE_{F,u}}\right) \tag{20b}$$

Finally, the penetration depth vector is simulated from the regression curve $\boldsymbol{h}_u\left(\overline{\boldsymbol{F}_u}, \boldsymbol{\beta}_u\right)$, as described in Equation (7), constraining $\boldsymbol{h}_u \in [h_2; h_3]$. No further measurement noise, simulating reproducibility, is added as it is already included in the distributions of $\boldsymbol{F}_u$ and $\boldsymbol{\beta}_u$.

2.2.4. Secondary Holding Generation

Also in this case, the generation process is highly similar to the one introduced in Section 2.2.2 for the primary holding. The secondary holding is introduced to compensate for thermal drifts, which will be simulated in Section 2.4. Conversely, in nominal conditions, a constant force and penetration should be obtained.

Accordingly, the penetration depth vector is created as a constant value equal to h_3 (with $h_4 = h_3$), as described in Equation (21), to which a point-wise random noise is added to account for the measurement accuracy.

$$\boldsymbol{h}_{holding2} = \left\{h_3 : J_{holding2} : h_4\right\} + Acc_h \tag{21}$$

The force vector is created as a constant force value equal to F_3 (with $F_4 = F_3$), to which point-wise random noise is added to account for the accuracy, as in Equation (22)

$$\boldsymbol{F}_{holding2} = \left\{F_3 : J_{holding2} : F_4\right\} + Acc_F \tag{22}$$

2.3. Uncertainty Evaluation

Synthetically generated quantities are obtained by sampling from underlying statistical distributions. Accordingly, it is possible to estimate the uncertainty of the synthetic indentation curve. The loading and unloading segment uncertainty evaluation leverages the law of propagation of uncertainty (LPU) [70] and caters for the fact that model parameters have been estimated by ODR. According to the LPU, the combined variance is obtained as a linear combination of the variance contributions weighted for the squared sensitivity coefficients, i.e., the partial derivatives of the response, in this case F, to the independent

quantities. The model for partial derivative estimation is defined in Equation (23), e.g., for the loading segment.

$$F_l(h, \boldsymbol{\beta}_l) = f_l(h \pm \delta_l; \boldsymbol{\beta}_l) \pm e_{F,l} = \beta_{0,l}(h \pm \delta_l - \beta_{1,l})^{\beta_{2,l}} \pm e_{F,l} \tag{23}$$

In Equation (23), $\delta \sim N(0, ms\delta)$ estimates the error in the regressor variable, and e_F, as introduced in Equations (14b) and (20b), describes the residual error. In particular, on the regression residuals, e.g., for the loading segment as in Equation (23), it is possible to evaluate $ms\delta_l = \frac{1}{J_l}\sum_{i=1}^{J_l} \delta_{i,l}^2$. The *mse* includes, by definition, the error in the response ε, which is also subject to minimization for estimating the parameters $\boldsymbol{\beta}$ by ODR.

Applying the LPU to the metrological model of Equation (23), the variance of the simulated force can be obtained as in Equation (24), which is written, for example, for the loading segment.

$$u^2{}_{F_{syntehtic}} = u^2{}_{F,l} = \begin{bmatrix} \frac{\partial F}{\partial h} \\ \frac{\partial F}{\partial \delta} \\ \frac{\partial F}{\partial \beta} \\ \frac{\partial F}{\partial e_F} \end{bmatrix}^T \begin{bmatrix} u^2{}_h & 0 & 0 & 0 \\ 0 & ms\delta_l & 0 & 0 \\ 0 & 0 & \overline{\boldsymbol{\Sigma}_{\beta_l K}} & 0 \\ 0 & 0 & 0 & u^2{}_F + mse_l \end{bmatrix} \begin{bmatrix} \frac{\partial F}{\partial h} \\ \frac{\partial F}{\partial \delta} \\ \frac{\partial F}{\partial \beta} \\ \frac{\partial F}{\partial e_F} \end{bmatrix} \tag{24}$$

In Equation (24), u_h and u_F are the standard uncertainties of the displacement sensor and the force transducer, typically obtained from a calibration certificate, which shall be added to include contributions from the traceability chain. $\overline{\boldsymbol{\Sigma}_{\boldsymbol{\beta}_K}}$ is the already introduced covariance matrix of the model parameters, estimated by ODR, which caters for the uncertainty in the parameters estimation.

2.4. Error Simulation Approaches

The metrological synthetic dataset generator for nanoindentation described in Section 2.2 also aims at generating the most typical measurement errors. Simulating errors is useful to allow training of error detection models, as reviewed in Section 1.2. In this work, three main errors are modelled: thermal drift, pop-in, and pop-out.

2.4.1. Thermal Drift Simulation

Even in controlled laboratory conditions, a thermal gradient can be present between the indenter tip and the sample. This can occur due to insufficient stabilisation of the sample, because of electronics heating the indenter through conduction, and because of the small amount of heat dissipated through friction during the indentation. To evaluate the thermal drift $q(T)$, the most robust approach, even though not the most time efficient, consists of performing a secondary holding and evaluating the slope of $h(t)$, i.e., $q(T) = \left.\frac{dh}{dt}\right|_{holding2}$, such that: $h_{holding2} = h_a + q(T)t$ [42,71]. In ideal conditions, $q(T) = 0$. However, this is never the case. In optimal experimental conditions, i.e., after a long thermal stabilisation of the sample in the measurement environment, the heating due to electronics plays a major role, typically inducing a thermal flux from the indenter to the sample, inducing $q(T) < 0$ [41,51].

Presence of thermal drift can be simulated by modifying all the penetration depths generated in Section 2.2, such that

$$h_j = h_j + q(T) \cdot (t_j - t_0) \tag{25}$$

Provided that any thermal drift can be generated, if a real-world scenario traceable to experimental data is aimed at, $q(T)$ can be sampled from the K training ICs, assuming a normal distribution $q(T) \sim N\left(\overline{q(T)}_K, s_K^2(q(T))\right)$.

2.4.2. Pop-In Simulation

A pop-in event describes a singularity in the loading segment, such that at a given force level, a discontinuity Δh_{pop-in} in the penetration depth occurs, suddenly increasing the penetration depth. This phenomenon is typically associated with phase changes, e.g., in semiconductors [72,73], or cracking, e.g., for coatings [74]. Accordingly, a pop-in event can be simulated by adding from a certain time instant onwards, during the loading segment, a shift in the penetration depth simulated, according to Section 2.2.1, i.e., $h(t > t^*) = h(t) + \Delta h_{pop-in}$, with $t^* \in [t_0; t_1]$. The specific selection of the t^* that induces the pop-in event shall be modelled depending on the material under study considering the specific loading cycle. In this work, to demonstrate capability of the synthetic dataset generator to include pop-in events, t^* was randomly selected between the instants realising $F \in [F_0; 30\%F_{max}]$ during the loading.

2.4.3. Pop-Out Simulation

Quite in a dual manner, a pop-out event describes a sudden decrease in the penetration depth. Pop-out is typically induced by phase transformation to metastable phases [72,75], or by cracks closing [73], and is induced by the load removal. Thus, it typically takes place during the final part of the unloading.

Accordingly, a pop-out event can be simulated by removing from a certain time instant onwards, during the unloading segment, a shift in the penetration depth simulated, according to Section 2.2.3, i.e., $h(t > t^*) = h(t) - \Delta h_{pop-out}$, with $t^* \in \left[t_3; t_{J_4}\right]$. Also for pop-out, the selection of t^* is strictly material dependent. Thus, to simply demonstrate capability of the synthetic dataset generator to include pop-out events, t^* was randomly selected between the instants realising $F \in [10\%; 30\%]F_{max}$ during the unloading.

2.5. Validation Methodology

The validation of the synthetic dataset generator is performed by testing its capability of simulating data from the real world. In particular, an Anton Paar STeP6 platform equipped with an NHT^3 nanoindentation head is considered. The equipment, hosted in the metrological laboratory of the MInd4Lab of Department of Management and Production Engineering of Politecnico di Torino, was equipped with a Berkovich indenter, calibrated [63] and used to perform a set of 15 indentations of a certified reference material, i.e., a NPL-calibrated sample of SiO_2, having the Young modulus of (73.0 $\pm$ 0.5) GPa and Poisson's ratio of 0.163 $\pm$ 0.002. Indentations implemented a force-controlled indentation cycle with a maximum force of 10 mN, and duration of loading, primary holding, and secondary holding (at 10% of F_{max}), respectively, of 10 s, 8 s, 9 s, and 60 s. The force and displacement sensors have, respectively, a calibrated relative accuracy of 0.061% and 0.058%. Raw force and displacement data, i.e., not automatically corrected for frame compliance, are considered to apply the full analysis pipeline described in previous sections.

The validation aims to test that real-world data cannot be statistically distinguished from the synthetically generated dataset, with a confidence level of 95%. That is, a hypothesis test based on the t-Student distribution, with a null hypothesis $H_0 : F_{synthetic} = F_{experimental}$ is built, having the test statistic $t_{stat} = \frac{F_{synthetic} - F_{experimental}}{u_{F_{synthetic}}}$. The test statistic estimates the standard uncertainty of the synthetically generated force as $u_{F_{synthetic}}$, according to Equation (24). Since Equation (24) includes traceability and reproducibility through u_F

and the *mse*, to avoid overestimation, it is assumed that the dispersion to experimental data is already included.

Further validation is performed by comparing the performances of the proposed metrological synthetic data generator for nanoindentation based on parametric simulation with other approaches available in the literature. In particular, a simulation based on non-parametric bootstrapping is considered. The methodology for bootstrap generation is reported elsewhere [64], and it has been proven effective in various metrological applications [65] overcoming the limits of Monte Carlo approaches. In particular, the presence of significant bias is tested, as well as any effect of systematic underestimation of measurement uncertainty. The former is once more tested by means of a t-Student hypothesis test. The latter by means of a hypothesis test based on the F-Fisher distribution, having the null hypothesis $H_0 : u^2_{F_{synthetic}} = u^2_{F_{Bootstrap}}$, and test statistic $F_{stat} = \frac{u^2_{F_{synthetic}}}{u^2_{F_{Bootstrap}}}$.

3. Results

The collected data, as per Section 2.5, were used to establish a traceable synthetic dataset generator for nanoindentation following the methodology described in Section 2.1.

Figure 3 shows the results of the application of the metrological synthetic data generator according to the methodology outlined in Section 2.2. Insights into relevant and critical regions of the synthetically generated curves are shown in Figure 4. Figures 3 and 4 show the reference model with associated uncertainty with a 95% confidence interval, evaluated as per Section 2.3.

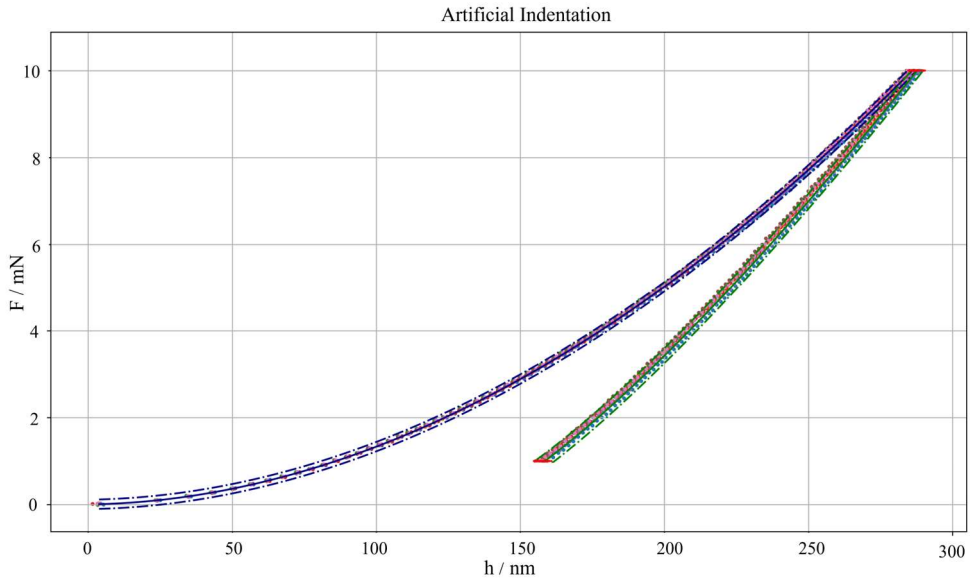


Figure 3. IC of synthetically generated nanoindentations. Points represent synthetically generated curves; lines are the theoretical model from experimental data. Blue: loading segment, green: unloading segment, red: holding segments. Solid lines are average predictions, and dot-dashed lines represent prediction uncertainty intervals with a confidence interval of 95%.

3.1. Generation of Errors

The traceable synthetic dataset generator for nanoindentation is then applied to simulate main measurement errors, to test the methodology described in Section 2.4. Figure 5c–f shows the successful application of the proposed parametric generation approach with examples for a significant thermal drift, pop-in, and pop-out, while Figure 5a,b show, for reference, a typical IC and $h(t)$.

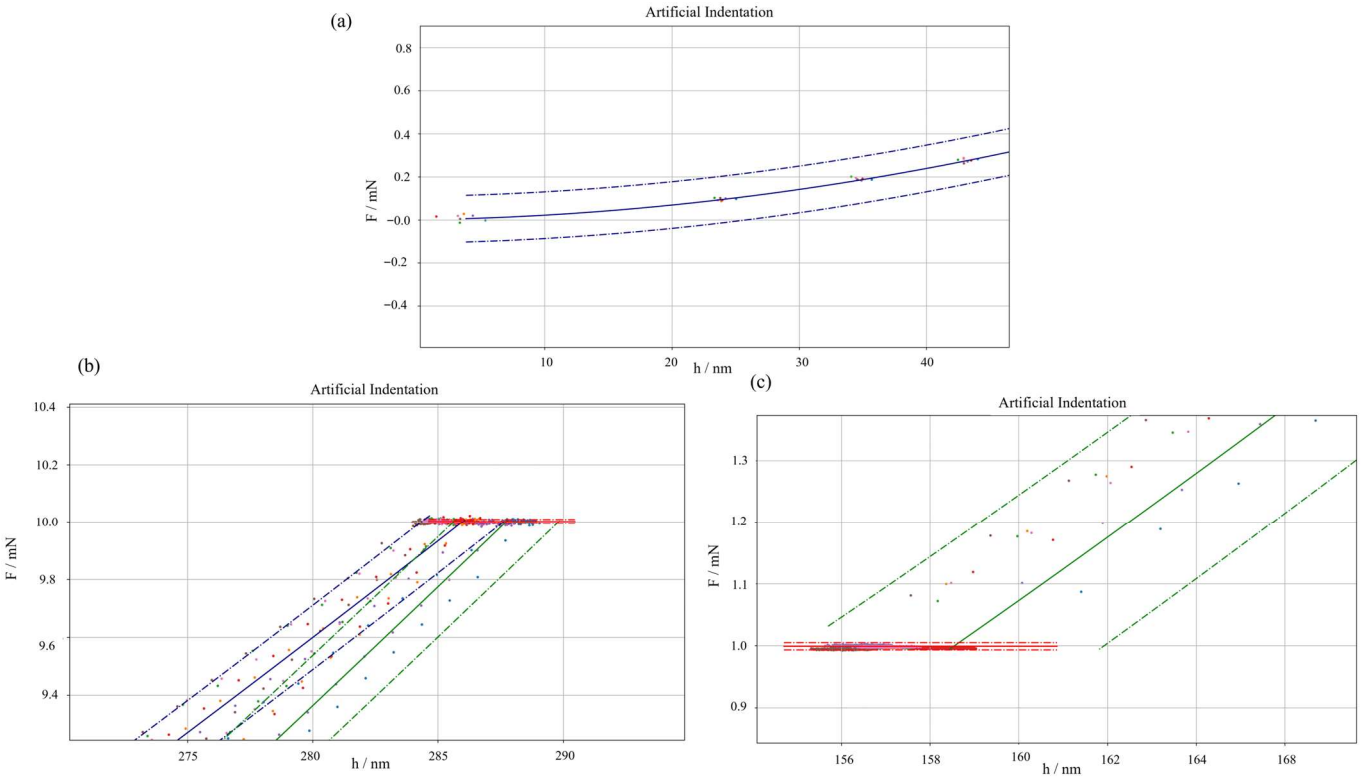


Figure 4. Insights into synthetically generated nanoindentation curves. (**a**) Start of indentation and zero point, (**b**) primary holding, and (**c**) secondary holding. Points represent synthetically generated curves; lines are the theoretical model from experimental data. Solid lines are average predictions, and dot-dashed lines represent prediction uncertainty intervals with a confidence interval of 95%.

As far as the thermal drift is concerned, a relevant slope is introduced, see Figure 5d, which induces an overall distortion of the whole IC, as well as a more apparent secondary holding, see Figure 5c.

Regarding discontinuities, pop-in and pop-out are simulated effectively. Figure 5e shows a pop-in in the loading segment, also clearly visible in the lack of continuity of the related $h(t)$ in Figure 5f. The pop-in offsets the subsequent points of the IC of a constant value, as shown in Figure 5e. Similarly, Figure 5g shows a pop-out in the unloading segment, also clearly visible in the lack of continuity of the related $h(t)$ in Figure 5h.

3.2. Validation

Following the methodology introduced in Section 2.5, the validation of the metrological synthetic dataset generator is performed.

First, the accuracy of the synthetic dataset is assessed by performing a t-Student hypothesis test to compare the synthetic data with experimental data. A graphical representation can be obtained by checking that synthetic datapoints all fall inside the prediction intervals, i.e., the dot-dashed lines in Figure 3. As can be appreciated, the null hypothesis cannot be rejected, with a risk of error of 5%, successfully validating the synthetic dataset generator accuracy. More in detail, insights from Figure 4 show that only a few points (less than 0.1% of generated points), near the transition to one segment to another, are outside the confidence interval limits. This is consistent with how measurement noise was introduced in the synthetic dataset generator to perturb the system further.

Lastly, a comparison with respect to a non-parametric synthetic dataset generation approach based on bootstrap [64] is performed. Figure 6 shows an inset of the loading segment of the IC, highlighting the effect of the synthetic dataset generator. As can be appreciated, the method proposed in this work, i.e., a parametric synthetic dataset generator,

produces a slight overestimation of the measurement uncertainty. However, when the statistical significance of such overestimation is tested by means of a hypothesis test based on the F-Fisher distribution, a *p*-value of 12% results, showing that such overestimation is not statistically significant. As can be appreciated, no systematic shift in the mean prediction can be seen.

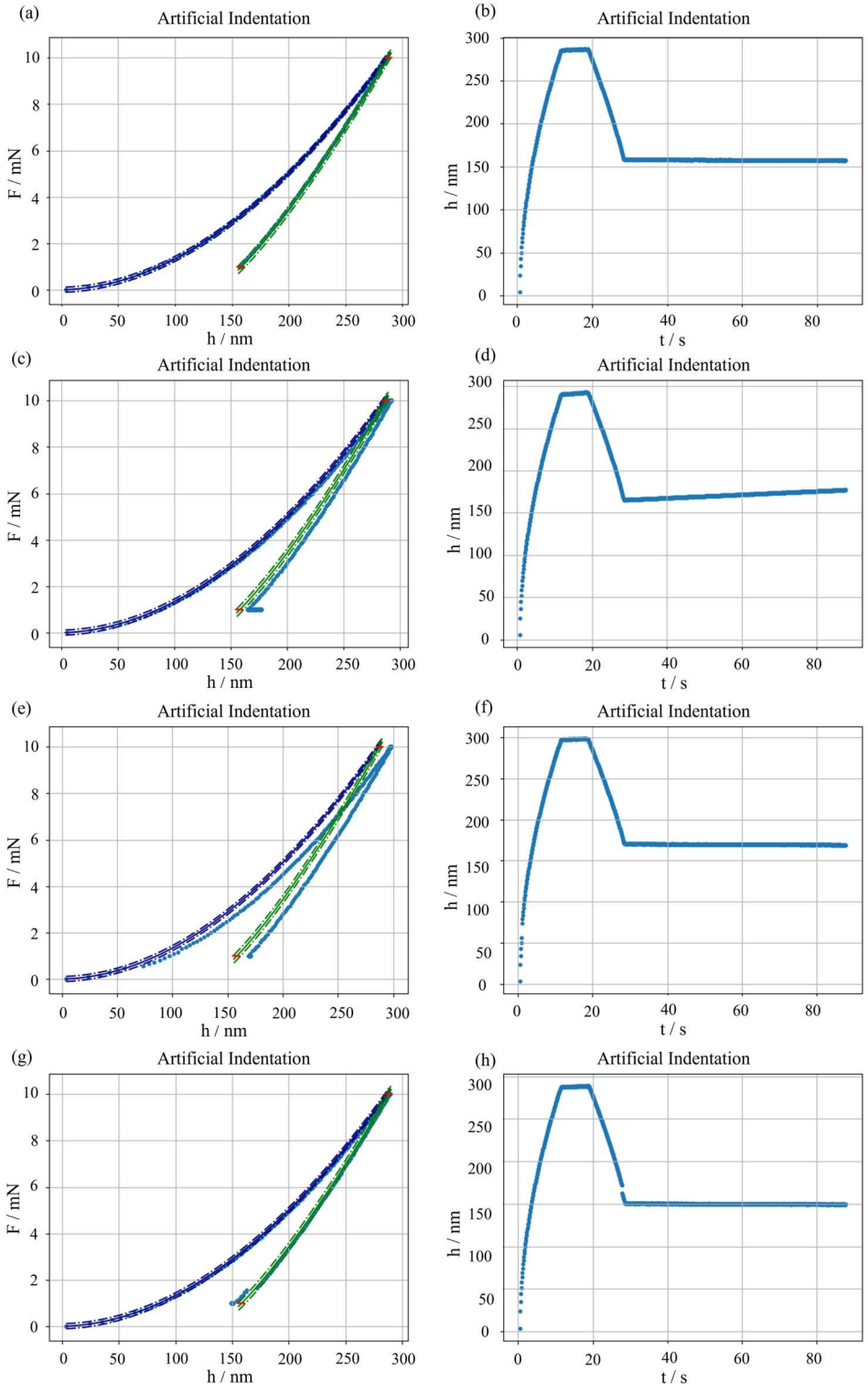


Figure 5. Simulation of main measurement errors in the synthetic dataset. No error (typical measurement): (**a**) IC, (**b**) penetration depth as a function of time. Significant thermal drift: (**c**) IC, (**d**) penetration depth as a function of time. Pop-in: (**e**) IC, (**f**) penetration depth as a function of time. Pop-out: (**g**) IC, (**h**) penetration depth as a function of time.

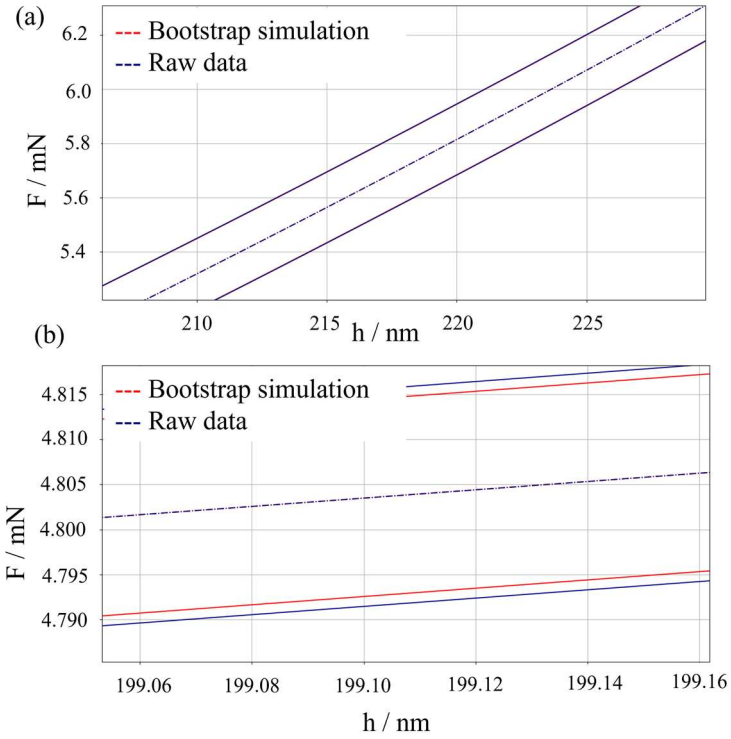


Figure 6. Effect of bootstrapping on measurement uncertainty. Prediction intervals at a 95% confidence level with two different levels of magnification (**a**,**b**) to appreciate the difference. Red: bootstrapping method, blue: parametric synthetic data generator based on raw data.

4. Discussion

The methodology outlined in Section 2 ultimately consists of a parametric synthetic dataset generator for nanoindentation.

With respect to other parametric approaches [63], it allows modelling the correlation between input quantities, thus overcoming possible distortions in the uncertainty evaluation.

Similarly, with respect to non-parametric methods, i.e., based on bootstrapping [64], it has shown no systematic differences, neither in terms of mean nor of dispersion. Compared to bootstrapping, it requires less computational power, is much faster, and allows overcoming issues of sample representativeness needed to perform the non-parametric approach.

The method requires modelling the indentation response of a material by means of a power-law model. Although the mathematical implementation allows for any set of parameters to be used, without the need for an experimental training dataset, the suggested approach in Section 2.1 allows, by means of Orthogonal Distance Regression, to establish a traceable and physics-informed model. Indeed, when the regression is performed to evaluate model parameters, the model representativeness is limited to a specific combination of material and indentation machine. In such a case, the proposed method limits the representativeness of the model to a specific material and specific indentation cycle parameters, i.e., the maximum force and indentation cycle segments' durations. Such limitation is inherent with the proposed methodology and can be considered the cost for having a traceable synthetic dataset generator. To extend the validity of the model to other materials and indentation machines, being the response intertwined due to the indenter geometry and frame compliance, a new training dataset is required.

Further, an underlying assumption of the synthetic dataset generator is the requirement of homogeneous material for training data. This condition is easily met for amorphous materials and for monocrystals. Conversely, for multiphase materials and polycrystalline materials, if the indentation scale of interest allows to distinguish different

phases, a separate synthetic dataset generator per each phase can be trained relying on phase-specific data.

5. Conclusions

This work has proposed a parametric metrological synthetic dataset generator for quasi-static room-temperature nanoindentation. By means of statistical modelling, the proposed approach allows catering for the main uncertainty source, i.e., traceability and reproducibility, in the data generation process, thus enabling a metrological dataset generation. The accuracy of the method has been successfully tested against real-world data.

Although limited to specific experimental conditions, the synthetic dataset generator proposed in this work can be used to generate a traceable dataset. Future work will test the applicability of the method on multiphase materials and coatings. Traceable datasets generated by the synthetic dataset generator here proposed will find application in future works aiming to relieve the experimental effort needed to collect an extensive dataset for training more flexible simulative systems based on finite element simulation, e.g., based on inverse indentation problem, by providing a traceable reference dataset for validating advanced measurement quality control tools based on Digital Twins.

Author Contributions: Conceptualisation, G.M.; methodology, G.M., G.G., and L.G.; software, L.G.; validation, G.M., L.G., and G.G.; formal analysis, G.M.; investigation, G.M. and L.G.; resources, M.G.; data curation, G.M.; writing—original draft preparation, G.M.; writing—review and editing, L.G., G.G. and M.G.; visualisation, L.G. and G.M.; supervision, M.G.; project administration, M.G.; funding acquisition, M.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was carried out within the project 22DIT01 ViDiT, which received funding from the European Partnership on Metrology, cofinanced by the European Union's Horizon Europe Research and Innovation Programme and by the Participating States.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study and the code for the Synthetic dataset generation according to the proposed methodology are available on Zenodo at 10.5281/zenodo.17465545.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
DT	Digital Twin
F	Measured applied force
GUM	Guide to the expression of Uncertainty in Measurement
h	Measured penetration depth
J	Number of data points in each nanoindentation
K	Number of experimental nanoindentations
IC	Indentation Curve
ML	Machine Learning
t	Measured time

References

1. Almada-Lobo, F. The Industry 4.0 Revolution and the Future of Manufacturing Execution Systems (MES). *J. Innov. Manag.* **2016**, *3*, 16–21. [CrossRef]
2. Duflou, J.R.; Sutherland, J.W.; Dornfeld, D.; Herrmann, C.; Jeswiet, J.; Kara, S.; Hauschild, M.; Kellens, K. Towards Energy and Resource Efficient Manufacturing: A Processes and Systems Approach. *CIRP Ann. Manuf. Technol.* **2012**, *61*, 587–609. [CrossRef]

3. Urgo, M.; Terkaj, W.; Simonetti, G. Monitoring Manufacturing Systems Using AI: A Method Based on a Digital Factory Twin to Train CNNs on Synthetic Data. *CIRP J. Manuf. Sci. Technol.* **2024**, *50*, 249–268. [CrossRef]
4. Urgo, M.; Terkaj, W. Integrating Digital Factory Twin and AI for Monitoring Manufacturing Systems through Synthetic Data Generation and Vision Transformers. *CIRP Ann.* **2025**, *74*, 639–643. [CrossRef]
5. Gao, R.X.; Krüger, J.; Merklein, M.; Möhring, H.C.; Váncza, J. Artificial Intelligence in Manufacturing: State of the Art, Perspectives, and Future Directions. *CIRP Ann.* **2024**, *73*, 723–749. [CrossRef]
6. Psarommatis, F.; May, G.; Azamfirei, V. Zero Defect Manufacturing in 2024: A Holistic Literature Review for Bridging the Gaps and Forward Outlook. *Int. J. Prod. Res.* **2024**, 1–37. [CrossRef]
7. *ISO 23247-1:2021*; Automation Systems and Integration-Digital Twin Framework for Manufacturing. Part 1: Overview and General Principles. ISO: Genève, Switzerland, 2021.
8. Shao, G.; Helu, M. Framework for a Digital Twin in Manufacturing: Scope and Requirements. *Manuf. Lett.* **2020**, *24*, 105–107. [CrossRef]
9. Matta, A.; Lugaresi, G. Digital Twins: Features, Models, and Services. In Proceedings of the Winter Simulation Conference, San Antonio, TX, USA, 10–13 December 2023; pp. 46–60.
10. Matta, A.; Lugaresi, G. An Introduction to Digital Twins. In Proceedings of the 2024 Winter Simulation Conference (WSC), Orlando, FL, USA, 15–18 December 2024; pp. 1281–1295.
11. Anwer, N.; Stark, R.; Tao, F.; Erkoyuncu, J.A. Developing and Leveraging Digital Twins in Engineering Design. *CIRP Ann.* **2025**, *74*, 843–868. [CrossRef]
12. de Almeida, S.T.; Mo, J.P.T.; Bil, C.; Ding, S.; Cheng, C.-T. Accurate EDM Calibration of a Digital Twin for a Seven-Axis Robotic EDM System and 3D Offline Cutting Path. *Micromachines* **2025**, *16*, 892. [CrossRef]
13. Maculotti, G.; Khusnuddinov, F.; Kholkhujaev, J.; Genta, G.; Galetto, M. Traceable Digital Twin for Accurate Positioning of Industrial Robot Arms in Human–Robot Collaborative Systems. *Flex. Serv. Manuf. J.* **2025**. [CrossRef]
14. Irino, N.; Kobayashi, A.; Shinba, Y.; Kawai, K.; Spescha, D.; Wegener, K. Digital Twin Based Accuracy Compensation. *CIRP Ann.* **2023**, *72*, 345–348. [CrossRef]
15. Yang, W.T.; Blue, J.; Roussy, A.; Pinaton, J.; Reis, M.S. A Structure Data-Driven Framework for Virtual Metrology Modeling. *IEEE Trans. Autom. Sci. Eng.* **2020**, *17*, 1297–1306. [CrossRef]
16. Eichstädt, S.; Keidel, A.; Tesch, J. Metrology for the Digital Age. *Meas. Sens.* **2021**, *18*, 100232. [CrossRef]
17. Tao, F.; Cheng, J.; Qi, Q.; Zhang, M.; Zhang, H.; Sui, F. Digital Twin-Driven Product Design, Manufacturing and Service with Big Data. *Int. J. Adv. Manuf. Technol.* **2018**, *94*, 3563–3576. [CrossRef]
18. Qi, Q.; Tao, F. Digital Twin and Big Data Towards Smart Manufacturing and Industry 4.0: 360 Degree Comparison. *IEEE Access* **2018**, *6*, 3585–3593. [CrossRef]
19. Maculotti, G.; Marschall, M.; Kok, G.; Chekh, B.A.; van Dijk, M.; Flores, J.; Genta, G.; Puerto, P.; Galetto, M.; Schmelter, S. A Shared Metrological Framework for Trustworthy Virtual Experiments and Digital Twins. *Metrology* **2024**, *4*, 337–363. [CrossRef]
20. Bauer, A.; Trapp, S.; Stenger, M.; Leppich, R.; Kounev, S.; Leznik, M.; Chard, K.; Foster, I. Comprehensive Exploration of Synthetic Data Generation: A Survey. *arXiv* **2024**. [CrossRef]
21. Lin, C.Y.; Tseng, T.L.; Emon, S.H.; Tsai, T.H. Generative AI-Driven Data Augmentation for Robust Virtual Metrology: GANs, VAEs, and Diffusion Models. *IEEE Trans. Semicond. Manuf.* **2025**, *38*, 642–658. [CrossRef]
22. Lu, Y.; Chen, L.; Zhang, Y.; Shen, M.; Wang, H.; Wang, X.; van Rechem, C.; Fu, T.; Wei, W. Machine Learning for Synthetic Data Generation: A Review. *arXiv* **2025**. [CrossRef]
23. Figueira, A.; Vaz, B. Survey on Synthetic Data Generation, Evaluation Methods and GANs. *Mathematics* **2022**, *10*, 2733. [CrossRef]
24. Moy, C.K.S.; Bocciarelli, M.; Ringer, S.P.; Ranzi, G. Identification of the Material Properties of Al 2024 Alloy by Means of Inverse Analysis and Indentation Tests. *Mater. Sci. Eng. A* **2011**, *529*, 119–130. [CrossRef]
25. De Bono, D.M.; London, T.; Baker, M.; Whiting, M.J. A Robust Inverse Analysis Method to Estimate the Local Tensile Properties of Heterogeneous Materials from Nano-Indentation Data. *Int. J. Mech. Sci.* **2017**, *123*, 162–176. [CrossRef]
26. Goto, K.; Watanabe, I.; Ohmura, T. Inverse Estimation Approach for Elastoplastic Properties Using the Load-Displacement Curve and Pile-up Topography of a Single Berkovich Indentation. *Mater. Des.* **2020**, *194*, 108925. [CrossRef]
27. Jiao, Q.; Chen, Y.; Kim, J.; Han, C.-F.; Chang, C.-H.; Vlassak, J.J. A Machine Learning Perspective on the Inverse Indentation Problem: Uniqueness, Surrogate Modeling, and Learning Elasto-Plastic Properties from Pile-Up. *J. Mech. Phys. Solids* **2024**, *185*, 105557. [CrossRef]
28. Karniadakis, G.E.; Kevrekidis, I.G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics-Informed Machine Learning. *Nat. Rev. Phys.* **2021**, *3*, 422–440. [CrossRef]
29. Grasso Toro, F.; Frigo, G. Synthetic PMU Data Generator for Smart Grids Analytics. *Metrology* **2025**, *5*, 12. [CrossRef]
30. Lopes, P.V.; Silveira, L.; Guimaraes Aquino, R.D.; Ribeiro, C.H.; Skoogh, A.; Verri, F.A.N. Synthetic Data Generation for Digital Twins: Enabling Production Systems Analysis in the Absence of Data. *Int. J. Comput. Integr. Manuf.* **2024**, *37*, 1252–1269. [CrossRef]

31. Kim, D.; Choi, M.; Um, J. Digital Twin for Autonomous Collaborative Robot by Using Synthetic Data and Reinforcement Learning. *Robot. Comput.-Integr. Manuf.* **2024**, *85*, 102632. [CrossRef]
32. Loaldi, D.; Quagliotti, D.; Calaon, M.; Parenti, P.; Annoni, M.; Tosello, G. Manufacturing Signatures of Injection Molding and Injection Compression Molding for Micro-Structured Polymer Fresnel Lens Production. *Micromachines* **2018**, *9*, 653. [CrossRef]
33. Solis-Rios, D.; Villarreal-Gómez, L.J.; Goyes, C.E.; Fonthal Rico, F.; Cornejo-Bravo, J.M.; Fong-Mata, M.B.; Calderón Arenas, J.M.; Martínez Rincón, H.A.; Mejía-Medina, D.A. A Neural Network Approach to Reducing the Costs of Parameter-Setting in the Production of Polyethylene Oxide Nanofibers. *Micromachines* **2023**, *14*, 1410. [CrossRef]
34. Nguyen, H.G.; Habiboglu, R.; Franke, J. Enabling Deep Learning Using Synthetic Data: A Case Study for the Automotive Wiring Harness Manufacturing. *Procedia CIRP* **2022**, *107*, 1263–1268. [CrossRef]
35. Delussu, R.; Putzu, L.; Fumera, G. Synthetic Data for Video Surveillance Applications of Computer Vision: A Review. *Int. J. Comput. Vis.* **2024**, *132*, 4473–4509. [CrossRef]
36. Ong, T.Y.; Tan, K.T.; Teoh, P.C.; Haron, M.H. Review of Solder Joint Vision Inspection for Industrial Applications. *Int. J. Adv. Manuf. Technol.* **2025**, *137*, 3257–3272. [CrossRef]
37. Li, T.; Wang, S.; Luo, Y.; Wan, J.; Luo, Z.; Chen, M. 3-D Vision and Intelligent Online Inspection in SMT Microelectronic Packaging: A Review. *IEEE J. Emerg. Sel. Top. Ind. Electron.* **2024**, *5*, 779–789. [CrossRef]
38. Lafon, L.F.; Vissière, A.; Mehdi-Souzani, C.; Anwer, N.; Nouira, H. Reference Data Generation for Evaluating Pairwise Registration Algorithms. *Measurement* **2026**, *257*, 118602. [CrossRef]
39. Nečas, D.; Klapetek, P. Synthetic Data in Quantitative Scanning Probe Microscopy. *Nanomaterials* **2021**, *11*, 1746. [CrossRef]
40. Pacheco, L.R.L.; Ferreira, J.P.S.; Parente, M.P.L. Deep Learning Regressors of Surface Properties from Atomic Force Microscopy Nanoindentations. *Appl. Sci.* **2024**, *14*, 2376. [CrossRef]
41. Lucca, D.A.; Herrmann, K.; Klopfstein, M.J. Nanoindentation: Measuring Methods and Applications. *CIRP Ann. Manuf. Technol.* **2010**, *59*, 803–819. [CrossRef]
42. *ISO 14577-1:2015*; Metallic Materials-Instrumented Indentation Test for Hardness and Materials Parameters—Part 1: Test Method. ISO: Geneva, Switzerland, 2015.
43. Tian, Z.; Xue, W.; Lou, W.; Liu, M.; Feng, H.; Wang, X.; Li, S.; Wu, S. Study on Anisotropic Mechanical Properties of Single-Crystal Silicon at Different Strain Rates. *Micromachines* **2025**, *16*, 744. [CrossRef]
44. *ISO 14577-4:2015*; Metallic Materials—Instrumented Indentation Test for Hardness and Materials Parameters—Part 4: Test Method for Metallic and Non-Metallic Coatings. ISO: Geneva, Switzerland, 2015.
45. Chen, S.; Liu, L.; Wang, T. Investigation of the Mechanical Properties of Thin Films by Nanoindentation, Considering the Effects of Thickness and Different Coating-Substrate Combinations. *Surf. Coat. Technol.* **2005**, *191*, 25–32. [CrossRef]
46. Maculotti, G.; Goti, E.; Genta, G.; Mazza, L.; Galetto, M. Comprehensive Mechanical and Tribological Characterization of Metal-Polymer PTFE+Pb/Bronze Coating by in-Situ Electrical Contact Resistance Measurement Augmented Tribo-Mechanical Tests. *Tribol. Int.* **2024**, *193*, 109397. [CrossRef]
47. Hou, X.; Jennett, N.M.; Parlinska-Wojtan, M. Exploiting Interactions between Structure Size and Indentation Size Effects to Determine the Characteristic Dimension of Nano-Structured Materials by Indentation. *J. Phys. D Appl. Phys.* **2013**, *46*, 265301. [CrossRef]
48. Puchi-Cabrera, E.S.; Staia, M.H.; Iost, A. Modeling the Composite Hardness of Multilayer Coated Systems. *Thin Solid Film.* **2015**, *578*, 53–62. [CrossRef]
49. Maculotti, G.; Genta, G.; Lorusso, M.; Galetto, M. Assessment of Heat Treatment Effect on AlSi10Mg by Selective Laser Melting through Indentation Testing. *Key Eng. Mater.* **2019**, *813*, 171–177. [CrossRef]
50. Genta, G.; Maculotti, G. Thin Coatings Thickness Measurement by Augmented Nanoindentation Data Fusion. *CIRP Ann.* **2024**, *73*, 409–412. [CrossRef]
51. Wheeler, J.M.; Armstrong, D.E.J.; Heinz, W.; Schwaiger, R. High Temperature Nanoindentation: The State of the Art and Future Challenges. *Curr. Opin. Solid State Mater. Sci.* **2015**, *19*, 354–366. [CrossRef]
52. Ezenwafor, T.; Anye, V.; Madukwe, J.; Amin, S.; Obayemi, J.; Odusanya, O.; Soboyejo, W. Nanoindentation Study of the Viscoelastic Properties of Human Triple Negative Breast Cancer Tissues: Implications for Mechanical Biomarkers. *Acta Biomater.* **2023**, *158*, 374–392. [CrossRef]
53. Cagliero, R.; Barbato, G.; Maizza, G.; Genta, G. Measurement of Elastic Modulus by Instrumented Indentation in the Macro-Range: Uncertainty Evaluation. *Int. J. Mech. Sci.* **2015**, *101–102*, 161–169. [CrossRef]
54. Doerner, M.F.; Nix, W.D. A Method for Interpreting the Data from Depth-Sensing Indentation Instruments. *J. Mater. Res.* **1986**, *1*, 601–609. [CrossRef]
55. Barbato, G.; Genta, G.; Cagliero, R.; Galetto, M.; Klopfstein, M.J.; Lucca, D.A.; Levi, R. Uncertainty Evaluation of Indentation Modulus in the Nano-Range: Contact Stiffness Contribution. *CIRP Ann. Manuf. Technol.* **2017**, *66*, 495–498. [CrossRef]
56. Sneddon, I.N. The Relation between Load and Penetration in the Axisymmetric Boussinesq Problem for a Punch of Arbitrary Profile. *Int. J. Eng. Sci.* **1965**, *3*, 47–57. [CrossRef]

57. Oliver, W.C.; Pharr, G.M. Measurement of Hardness and Elastic Modulus by Instrumented Indentation: Advances in Understanding and Refinements to Methodology. *J. Mater. Res.* **2004**, *19*, 3–20. [CrossRef]
58. Puchi-Cabrera, E.S.; Rossi, E.; Sansonetti, G.; Sebastiani, M.; Bemporad, E. Machine Learning Aided Nanoindentation: A Review of the Current State and Future Perspectives. *Curr. Opin. Solid State Mater. Sci.* **2023**, *27*, 101091. [CrossRef]
59. Koumoulos, E.; Konstantopoulos, G.; Charitidis, C. Applying Machine Learning to Nanoindentation Data of (Nano-) Enhanced Composites. *Fibers* **2019**, *8*, 3. [CrossRef]
60. Giolando, P.; Kakaletsis, S.; Zhang, X.; Weickenmeier, J.; Castillo, E.; Dortdivanlioglu, B.; Rausch, M.K. AI-Dente: An Open Machine Learning Based Tool to Interpret Nano-Indentation Data of Soft Tissues and Materials. *Soft Matter* **2023**, *19*, 6710–6720. [CrossRef] [PubMed]
61. Bruno, F.; Konstantopoulos, G.; Fiore, G.; Rossi, E.; Sebastiani, M.; Charitidis, C.; Belforte, L.; Palumbo, M. A Novel Machine Learning Method to Exploit EBSD and Nanoindentation for TRIP Steels Microstructures Analysis. *Mater. Des.* **2024**, *239*, 112774. [CrossRef]
62. Mahmood, T.; Zia, A.W. Predicting the Hardness of Diamond-like Carbon Coatings Using Machine Learning and Generative Adversarial Networks. *J. Manuf. Process.* **2025**, *149*, 129–143. [CrossRef]
63. Galetto, M.; Genta, G.; Maculotti, G. Single-Step Calibration Method for Nano Indentation Testing Machines. *CIRP Ann.* **2020**, *69*, 429–432. [CrossRef]
64. Maculotti, G.; Genta, G.; Galetto, M. An Uncertainty-Based Quality Evaluation Tool for Nanoindentation Systems. *Measurement* **2024**, *225*, 113974. [CrossRef]
65. Maculotti, G.; Giorio, L.; Genta, G.; Galetto, M. Metrological Comparison of Indirect Calibration Methods for Nanoindentation: A Bootstrap-Based Approach. *Materials* **2025**, *18*, 4382. [CrossRef]
66. *JCGM 101:2008*; Evaluation of Measurement Data—Supplement 1 to the "Guide to the Expression of Uncertainty in Measurement"—Propagation of Distributions Using a Monte Carlo Method. JCGM: Sèvres, France, 2008; p. 90.
67. Wübbeler, G.; Marschall, M.; Kniel, K.; Heißelmann, D.; Härtig, F.; Elster, C. GUM-Compliant Uncertainty Evaluation Using Virtual Experiments. *Metrology* **2022**, *2*, 114–127. [CrossRef]
68. Hughes, F.; Marschall, M.; Wübbeler, G.; Elster, C. Uncertainty Evaluation Using Virtual Experiments: Bridging JCGM 101 and a Bayesian Framework. *Tech. Mess.* **2025**, *92*, 130–137. [CrossRef]
69. Marschall, M.; Hughes, F.; Wübbeler, G.; Kok, G.; van Dijk, M.; Elster, C. Using a Multivariate Virtual Experiment for Uncertainty Evaluation with Unknown Variance. *Metrology* **2024**, *4*, 534–546. [CrossRef]
70. *JCGM 100:2008*; Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement (GUM). JCGM: Sèvres, France, 2008.
71. Herrmann, K.; Lucca, D.A.; Klopfstein, M.J.; Menelao, F. CIRP Sponsored International Comparison on Nanoindentation. *Metrologia* **2010**, *47*, S50–S58. [CrossRef]
72. Bradby, J.E.; Williams, J.S.; Wong-Leung, J.; Swain, M.V.; Munroe, P. Mechanical Deformation in Silicon by Micro-Indentation. *J. Mater. Res.* **2001**, *16*, 1500–1507. [CrossRef]
73. Zare, A.; Tunesi, M.; Harriman, T.A.; Troutman, J.R.; Davies, M.A.; Lucca, D.A. Face Turning of Single Crystal (111)Ge: Cutting Mechanics and Surface/Subsurface Characteristics. *J. Manuf. Sci. Eng.* **2023**, *145*, 071007. [CrossRef]
74. Bull, S.J. Nanoindentation of Coatings. *J. Phys. D Appl. Phys.* **2005**, *38*, R393–R413. [CrossRef]
75. Bradby, J.E.; Williams, J.S.; Wong-Leung, J.; Swain, M.V.; Munroe, P. Nanoindentation-Induced Deformation of Ge. *Appl. Phys. Lett.* **2002**, *80*, 2651–2653. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.