



Machine Learning–Based Reduce Order Crystal Plasticity Modeling for ICME Applications

Mengfei Yuan¹ · Sean Paradiso² · Bryce Meredig³ · Stephen R. Niezgoda^{1,4}

Received: 8 October 2018 / Accepted: 3 December 2018 / Published online: 18 December 2018
© The Minerals, Metals & Materials Society 2018

Abstract

Crystal plasticity simulation is a widely used technique for studying the deformation processing of polycrystalline materials. However, inclusion of crystal plasticity simulation into design paradigms such as integrated computational materials engineering (ICME) is hindered by the computational cost of large-scale simulations. In this work, we present a machine learning (ML) framework using the material information platform, Open Citrination, to develop and calibrate a reduced order crystal plasticity model for face-centered cubic (FCC) polycrystalline materials, which can be both rapidly exercised and easily inverted. The reduced order model takes crystallographic texture, constitutive model parameters, and loading condition as inputs and returns the stress-strain curve and final texture. The model can also be inverted and take a stress-strain curve, loading condition, and final texture as inputs and return the initial texture and constitutive model parameters as outputs. Principal component analysis (PCA) is used to develop an efficient description of the crystallographic texture. A viscoplastic self-consistent (VPSC) crystal plasticity solver is used to create the training data by modeling the stress-strain behavior and evolution of texture during deformation processing.

Keywords Reduced order model · Microstructure quantification · Machine learning · Crystal plasticity · Dimensionality reduction · Parameter optimization

Introduction

In recent decades, the development of computationally aided design methodologies has revolutionized the way products

are manufactured, from first prototype to final processing, assembly, and testing. The next frontier is the extension of computational design to include material structure, properties, and processing as critical design variables, which can be optimized to deliver superior material and component-scale performance, rather than constraints on the design process. This is the fundamental goal of Integrated Computational Materials Engineering (ICME) [1]. Leveraging the exponential advances in physics-based materials simulation tools, ICME is predicated on replacing expensive, repeated experimentation and testing with optimization through simulation. While offering unprecedented fidelity and predictive power, spatially and temporally resolved material simulation tools can be too computationally expensive or require too much calibration data to be effectively used within the ICME framework [2, 3]. In order to properly utilize these models in design, we must address the problem of developing reduced order models which replicate the fundamental physical behavior of the material system but, which can be rapidly exercised and easily inverted for integration within an optimization framework. Further, given that the statistical confidence in the final design (previously obtained through

✉ Stephen R. Niezgoda
niezgoda.6@osu.edu

Mengfei Yuan
yuan.358@osu.edu

Sean Paradiso
sean@citrine.io

Bryce Meredig
bryce@citrine.io

¹ Department of Materials Science and Engineering, The Ohio State University, Columbus, OH 43210, USA

² Scientific Software Engineer, Citrine Informatics, Redwood City, CA 94063, USA

³ CSO, Citrine Informatics, Redwood City, CA 94063, USA

⁴ Department of Mechanical and Aerospace Engineering, The Ohio State University, Columbus, OH 43210, USA

repeated material and component-scale testing) must now be obtained through simulation, it is critical that the developed reduced order models are integrated with formal uncertainty quantification as part of the validation process [4].

In this work, a reduced order framework for crystal plasticity, developed through machine learning (ML), is presented. Crystal plasticity was chosen as an initial case study as it is a relatively mature and widely utilized simulation area [5–7]. Also, crystal plasticity simulations are often computationally expensive and underdetermined with respect to the parameters, as the single crystal constitutive theory requires more model parameters than there are experimental observations for calibration [8–10]. Even relatively simple theories require a large amount of parameters to capture the non-linearity in the stress-strain response, particularly with respect to strain rate, temperature, multiple slip and twinning modes, and strain path effects [8]. Thus, demonstration of the ML framework to identify correlations between model parameters during the training of the reduced order model and then correctly predict constitutive theory model parameters from limited input data serves as a good test of the robustness of the proposed framework.

Traditionally, the calibration of crystal plasticity models was performed by manually adjusting model parameters to best match the available experimentally measured stress-strain and crystallographic texture data. This can be a time-consuming process, with large amounts of human intervention due to the large parameter space. This process needs to be repeated anytime the single crystal mechanical properties are changed [6, 8]. The need for expensive manual calibration renders exploration across many materials and processing paths difficult if not intractable. Therefore, it is important to have an automated process to rapidly estimate model parameters at the preliminary stages based on limited stress-strain response and texture data. Previous attempts at the automated optimization of fitting parameters have not been entirely successful due to strongly non-linear and potentially non-unique relationships between the model parameters and the stress-strain response of the polycrystal [11–13]. Another contributing factor is the degree of sample-to-sample variability in both initial microstructure and measured properties. Any automated approach to optimization of model parameters must be robust to uncertainty in the single crystal properties due to noise or variance in: (i) the initial texture and microstructure of the sample, (ii) measurement noise due to variability in the mechanical test, and (iii) limitations of the constitutive theory (e.g., missing physics or model assumptions).

The development of reduced order models has the potential for wide application beyond the design process. One of the tenets of the Materials Genome Initiative (MGI) is the development of an infrastructure to accelerate the development of new materials [14]. The construction of large-scale

materials structure-processing-properties databases is a critical aspect of this MGI infrastructure [15]. Significant progress towards population of these databases can potentially be made from the automated mining of legacy material data from the existing materials science and engineering literature in the form of manuscript, text, tables, graphics, and figures [6–8, 14]. However, it is also understood that significant gaps will be left in the databases as legacy sources may not have collected or reported complete microstructure data. For example, initial or final texture data is frequently missing from reports on mechanical testing, due to the historic difficulty of texture measurement before the advent of commercial electron back-scatter diffraction (EBSD) systems. Fully calibrated reduced order models, such as the one presented here, can be invaluable tools to fill these missing gaps. To illustrate this idea, a proof-of-concept example is demonstrated where the initial state of the microstructure (texture) is predicted from the final observed structure in this work.

In this study, we apply Open Citrination, a materials informatics platform [15], to establish a “fast-acting reduced order crystal plasticity model” for polycrystalline materials. A viscoplastic self-consistent (VPSC) crystal plasticity formulation with Voce hardening is used as a case study model. Rather than calibrating to isolated mechanical test results either from the lab or literature, the training data is generated by repeatedly exercising VPSC over a range of initial textures and hardening parameters. The trained reduced order model is demonstrated by predicting the stress-strain response and final texture of a sample given the initial texture, strain path, and hardening parameters. The model is inverted to predict the initial texture and optimized hardening parameters given the stress-strain curve and final texture. This process is schematically shown in Fig. 1. The proposed reduced order model can help to quickly predict the behaviors of products in manufacturing design, while providing additional material information according to limited experimental data. Additionally, in order to quantify the high microstructural variance, the dimensionality reduction technique, principal component analysis (PCA), is applied to capture the characteristic microstructural information using a low-dimension description. The correlation between microscale texture/deformation evolution, macroscale stress-strain behavior, and optimal crystal plasticity model will be calibrated through Open Citrination as a case study in this paper.

Methodology

The deformation response of a single-phase, face-centered cubic (FCC) copper, with 12 independent slip systems of type $\{111\} \langle 110 \rangle$, is used as the initial case study. The

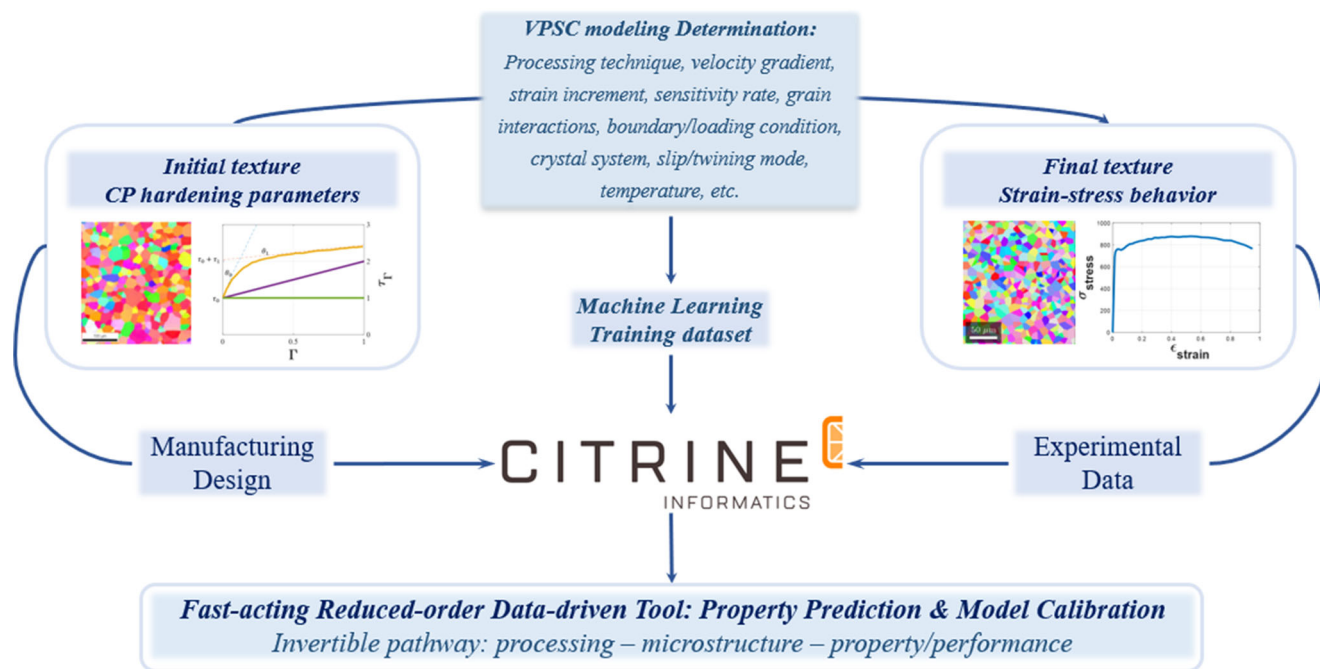


Fig. 1 Schematic description of the proposed machine learning framework

well-known VPSC framework will be used as the target crystal plasticity model for the machine learning reduced order model [8]. Voce hardening is selected to simulate the plastic hardening response during processing. Representative FCC textures, selected loading conditions, and strain rates are considered as variables in the learning progress. The Open Citrination platform is adopted to learn the complex correlation between the microstructure features and crystal plasticity in this study. Using the computational network established in this study, the deformation behavior and microstructure evolution of FCC copper under arbitrary conditions can be quickly predicted. Additionally, the experimental stress-strain data can also be fitted to identify optimal constitutive model parameters. In order to better focus the discussion on the details of the framework, our initial case study does not extend to multiple materials or multiple deformation mechanisms. Future work will consider more complicated material systems, such as body-centered cubic (BCC) and hexagonal closest packed (HCP) alloys with multiple non-degenerate slip and twinning modes with strong temperature dependence and multi-phase materials.

Viscoplastic Self-Consistent Crystal Plasticity Modeling

There are many well-known constitutive theories and numerical approaches to model the deformation of crystalline materials at the polycrystalline level [5–7, 16–18]. For brevity, we will only discuss details relevant to this study, readers who require a review of the field are referred

to the following references [6–10, 17, 19–22]. For this work, we will use VPSC as the physics-based model that we wish to replicate. In contrast to crystal plasticity finite elements (CPFE) and the elasto-viscoplastic fast Fourier transform approaches, which are full-field three dimensional spatially resolved solvers, VPSC is a mean field or homogenized model that does not consider the spatial arrangement of grains but computes the effective mechanical response of the polycrystal using the orientation, size, and shape of the individual crystallites [8, 23–25].

Self-consistent models are commonly used to estimate the homogenized response of heterogeneous systems [17]. VPSC is routinely used to model the mechanical response of polycrystal aggregates during plastic deformation [8]. Within VPSC, the polycrystal is approximated as an ensemble of weighted statistically representative (SR) grains each with a specific crystallographic orientation. The individual SR grains are modeled as Eshelby inclusions within a homogeneous effective medium [26, 27]. The properties of the effective medium are calculated as an average over the ensemble of SR grains. As the response of the SR grains and the properties of the effective medium are mutually dependent, they must be iteratively updated (starting with an initial estimate of the effective properties) until convergence [6].

The constitutive model applied in this case study is selected as the non-linear rate-dependent power law formulation which relates the viscoplastic strain rate, $\dot{\epsilon}_p^r$, to the stress σ as

$$\dot{\epsilon}_p^r = \mathbf{M}_{ijkl}^r : \sigma_{kl}^r = \gamma_0 \sum_s \mathbf{m}_{ij}^s \left(\frac{\mathbf{m}_{kl}^s : \sigma_{kl}^r}{\tau_0^s} \right)^n \quad (1)$$

where $\dot{\epsilon}_p$ and σ are represented as the average viscoplastic strain rate and stress in each SR grain. \mathbf{M}^r is the viscoplastic compliance. τ_0^s is the critical resolved shear stress (CRSS) for slip system s , and \mathbf{m}^s is the symmetric Schmid tensor associated with the activated slip/twinning system, s , in homogenized SR grain r . $n = 20$ is assigned in this case study as a reasonable rate-sensitivity exponent constant. The reference shear rate, γ_0 , served as a normalization factor, and the viscoplastic strain rate, $\dot{\epsilon}_p$, can be non-linearly scaled by the magnitude of applied velocity gradient or applied stress components during deforming. Detailed explanation will be provided in “Parameterization of Applied Deformation”.

The Voce hardening model is applied to evolve the slip resistances for each slip system with accumulated deformation [8]. Equation 2 describes the evolution of the threshold stress, $\hat{\tau}^s$, that is produced by the accumulated shear strain, $\Gamma = \sum_s \Delta\gamma^s$, in each respective grain of a polycrystalline material, where γ^s is the shear activity in a grain. τ_0^s and $\tau_0^s + \tau_1^s$ are the initial and the back-extrapolated critical shear stress (CRSS) in slip system s , where θ_0^s and θ_1^s are the initial and the asymptotic hardening rate. The evolution of plastic hardening behavior for polycrystalline aggregates can be defined using following equations.

$$\hat{\tau}^s = \tau_0^s + (\tau_1^s + \theta_1^s \Gamma) \left(1 - \exp \left(-\Gamma \left| \frac{\theta_0^s}{\tau_1^s} \right| \right) \right) \quad (2)$$

$$\begin{aligned} \frac{d\hat{\tau}^s}{d\Gamma} &= \theta_1 + \left(\left| \frac{\theta_0}{\tau_1} \right| \tau_1 - \theta_1 \right) \exp \left(-\Gamma \left| \frac{\theta_0}{\tau_1} \right| \right) \\ &+ \left| \frac{\theta_0}{\tau_1} \right| \theta_1 \Gamma \exp \left(-\Gamma \left| \frac{\theta_0}{\tau_1} \right| \right) \end{aligned} \quad (3)$$

As shown in Fig. 2, the standard hardening behavior (defined as “Kosher” Voce hardening in the VPSC manual) demonstrates where the corresponding flow stress increases and hardening rate decreases with increasing strain, requires the condition, $\theta_0 \geq \theta_1 \geq 0$ [8]. The limiting case of standard is the right-perfectly-plastic hardening, which maintains the relation, $\theta_0 = \theta_1 = 0$, where other non-standard situations (defined as “Non-kosher” law in VPSC) are shown in Fig. 2. As an initial case study, we randomly generated data grids for hardening parameters according to the reasonable fitting parameters from existing literature [7, 8].

The increment in threshold stress, $\Delta\tau^s$, is relative to coupled dislocation interactions when other slip systems, s' are activated in a same grain (Eqs. 3 and 4) as,

$$\Delta\tau^s = \frac{d\hat{\tau}^s}{d\Gamma} \sum_{s'} h^{ss'} \Delta\gamma^{s'} \quad (4)$$

The coupling coefficient or hardening matrix, $h_{ss'}$, defines the effect of slip on system s' on the slip resistance for s . For

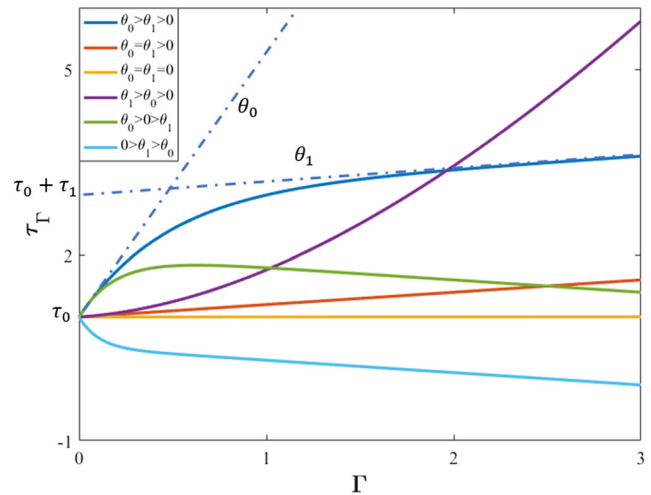


Fig. 2 Physical interpretation of Voce hardening parameters (Eq. 2)

this case study, we consider both self and latent hardening and take the limiting case of $h_{ss'} = 1 \quad \forall (s, s')$.

Parameterization of Applied Deformation

As the stress-strain response and texture evolution are strong functions of the imposed boundary condition, the complete range of physically meaningful effective velocity gradient tensors for the polycrystal must be parameterized for the machine learning schema described below (see “Machine Learning Algorithm”).¹ Table 1 shows four basic loading conditions with a unit loading rate. Each condition can be fully described in the deformation reference coordinate frame by a nine-component velocity gradient tensor. However, the identical loading/deformation path imposed upon a polycrystalline metal can be more compactly expressed in a principal deformation coordinate frame [28, 29]. Equation 5 details how the loading matrix in reference frame (L_R) is transformed to the one in principal frame (L_P) through multiplication of a stretching tensor, $f_{P \rightarrow R}$ (Table 2).

$$L_P = f_{P \rightarrow R}^T * L_R * f_{P \rightarrow R} \quad (5)$$

In the principal frame, the necessary parameters used to describe a loading condition are reduced from an arbitrary nine-component matrix to a three-component diagonal matrix containing a reference strain rate, $\dot{\epsilon}_0^*$ (Eq. 6). Varying

¹For convenience, this work restricts the space of boundary conditions to imposed deformation rate. The extension to stress control and mixed loading paths would require an additional parameterization of the boundary condition space.

Table 1 Loading condition in the reference frame (L_R)

Compression/tension	Shear	Rolling
$\begin{bmatrix} \pm 0.5 & 0 & 0 \\ 0 & \pm 0.5 & 0 \\ 0 & 0 & \mp 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}$

the effective applied deformation rate can be achieved through multiplication with a reference strain rate, $\dot{\epsilon}_0^*$, as

$$L_R = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix} \iff L_P = \begin{bmatrix} \cos(\beta^* - \frac{\pi}{3}) & 0 & 0 \\ 0 & \cos(\beta^* + \frac{\pi}{3}) & 0 \\ 0 & 0 & -\cos(\beta^*) \end{bmatrix} \dot{\epsilon}_0^* \quad (6)$$

The above parameterization reduces the dimension of the description of any applied deformation/loading paths with effective deformation rates into five parameters, where $\beta_1, \beta_2, \beta_3$, and β^* define the transformation stretching tensor between the reference and the principal frame. As an example, Table 2 displays the values of parameterized loading parameters in the principal frame for four corresponding conditions corresponding to Table 1.

Reduced Order Representation of Crystallographic Texture

For this work, crystallographic texture, described by an orientation distribution function (ODF), is assumed to be the controlling microstructure feature for mechanical response. The body of literature on crystal plasticity simulation largely shows that effective stress-stain curves can be well-captured by crystallographic texture while the specific grain structure/geometry affects the stress localization and other fluctuations of the mechanical fields [6–8, 30–32]. In order to properly calibrate the desired reduced order crystal plasticity model via machine learning, the range of possible ODFs must be sampled. This poses a difficulty as the space

of possible crystallographic textures is vast. For example, discretizing the orientation space for cubic materials into 5° bins produces 15,552 ODF components, each of which can be varied individually. Fortunately, these components do not evolve independently during deformation and a more efficient sampling scheme based on experimentally observed texture components can be identified.

Previous work from Raabe and Roters demonstrated that the representation of an arbitrary texture by the weighted mixture of standard texture components was an effective method to capture the effective mechanical response and texture evolution of polycrystals within a crystal plasticity framework [19, 30]. In this study, we follow a similar approach and consider the most important texture components observed to be produced by deformation processing in FCC metals. These include (i) the uniform or random ODF, which has identical values for all orientations, (ii) important unimodal texture components that contain a strong concentration around a preferred orientation including the cube, goss, brass, copper, S1, S2, S3, and Taylor textures often detected in FCC material, and (iii) fiber ODFs which represent all orientations formed by a rotation between two specific orientations including the α, β , and τ fiber textures [31, 33–35]. These texture components are defined explicitly in Table 3 and their positions in the Bunge-Euler space are shown in Fig. 3. The figure also highlights the relationship between the unimodal texture components and the fiber textures in cubic crystal structures. Numerical representation of the orientation distribution functions is accomplished via a Fourier series representation utilizing Bunge’s generalized spherical harmonics as a basis set [36]. This choice was simply one of the authors’ preference, as other representations such as direct binning in either the Euler angle or Rodrigues space would also work equally well. The free MATLAB toolbox MTEX is used in this study to carry out all ODF calculations and visualizations in this work [33–35].

The texture components described above form a convenient framework for describing the space of ODFs that need to be considered to adequately train a reduced order model via machine learning. However, each ODF is still expressed

Table 2 Loading parameters in the principal frame (L_P)

Loading in principal frame (L_P)	Compression/tension	Shear	Rolling
$L_P(\beta_1, \beta_2, \beta_3)$	(0, 0, 0)	$(\frac{3\pi}{2}, \frac{\pi}{4}, \frac{\pi}{2})$	(0, 0, 0)
$f_{P \rightarrow L}(\beta^*, \dot{\epsilon}_0^*)$	(0, ± 1)	$(\frac{\pi}{6}, \frac{-2}{\sqrt{3}})$	$(\frac{\pi}{6}, \frac{2}{\sqrt{3}})$
Transformation matrix ($f_{P \rightarrow R}$)			
$\begin{bmatrix} \cos\beta_1\cos\beta_3 - \sin\beta_1\sin\beta_3\cos\beta_2 & \sin\beta_1\cos\beta_3 + \cos\beta_1\sin\beta_3\cos\beta_1 & \sin\beta_3\sin\beta_2 \\ -\cos\beta_1\sin\beta_3 - \sin\beta_3\cos\beta_3\cos\beta_2 & -\sin\beta_1\sin\beta_3 + \cos\beta_1\cos\beta_3\cos\beta_2 & \cos\beta_3\sin\beta_2 \\ \sin\beta_1\sin\beta_2 & -\cos\beta_1\sin\beta_2 & \cos\beta_2 \end{bmatrix}$			

Table 3 Standard FCC texture components: Miller indices and Euler angles

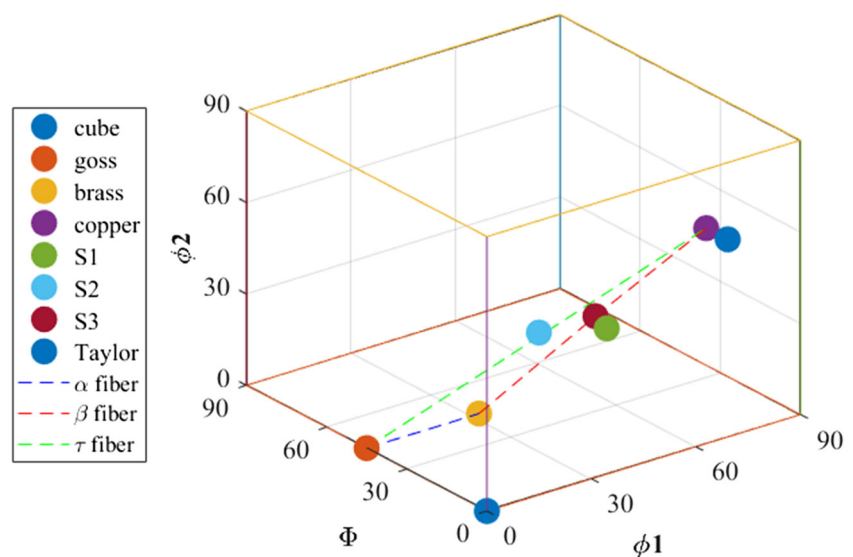
Texture name	Miller indices	Bunge-Euler angle
Cube	{1 0 0}<0 0 1>	(0 0 0)
Goss	{1 1 0}<0 0 1>	(0 45 0)
brass	{1 1 0}<1 $\bar{1}$ 2>	(35 45 0)
Copper	{1 1 2}<1 1 1>	(90 35 45)
S1	{1 2 4}<2 1 $\bar{1}$ >	(59 29 63)
S2	{1 2 3}<4 1 $\bar{2}$ >	(47 37 63)
S3	{1 2 3}<6 3 $\bar{4}$ >	(59 37 63)
Taylor	{4 4 11}<11 11 $\bar{8}$ >	(90 27 45)
α fiber	extend from goss to brass	
β fiber	extend from copper to S3 to brass	
τ fiber	extend from goss to copper	
Uniform	identical orientation everywhere	

as a Fourier series containing thousands of spherical harmonic coefficients [36]. Therefore, individual ODFs still need to be represented in a small number of parameters suitable for the training model. The naïve approach would be used to simply represent any ODF as a vector containing the weights of each of the texture components. However, this imposes the limitation that the set of final deformation textures and the set of input textures used for model calibration both span the same subspace of the ODF space. As an alternate approach, we could also use the strategy of microstructure-sensitive design (MSD) and consider the set of all possible deformation textures given the set of initial texture components; however, this approach leads to a large parameter set [37–46]. In order to retain flexibility and reduce the number of parameters, we have adopted principal component analysis (PCA) to identify an appropriate low-dimensional representation. PCA is a dimensionality

reduction method that conducts orthogonal decomposition on the high-dimensional raw data and maps to a new low-dimensional orthogonal coordinate frame, i.e., the principal components (PCs) represent as the new coordinate systems [47–49].

PCA can be simply explained as projection of the high-dimensional data onto a low-dimensional subspace, defined by a basis set (the principal components), that captures the critical features in the dataset. Each datapoint can be written as a weighted linear combination of the principal components. The new variables, interpreted through PCA, are the weights of the principal components. The basis is defined such that the first PC is the direction of highest variability through the original dataset. Next PC is defined by taking the direction of highest variability orthogonal to the first, and so on. In this way, differences between individual datapoints in the set can be represented by a small number of principal components. Since PCA and the generalized spherical harmonic representation are both linear transformations of the orientation distribution data, the final low-dimensional representation should be identical for both the binned orientation data and the spherical harmonic coefficients of the ODF. Therefore, in this case study, the dimension of each material texture (originally containing thousands of coefficients) is interpreted by PCA since the first couple of PCs can effectively capture the main characteristic of texture.

Similarly, the reconstruction of specific texture can also be completed by picking the corresponding coordinate in PC space [50, 51]. The accuracy and efficiency of the texture reconstruction using PCA has been measured with the L2-norm between the original ODF and the reconstructed one that computed as a function of the number of principal components. Figure 4 shows the PCA reduction based on thousands of initial and final deformation textures, and

Fig. 3 Standard texture component representation in Bunge-Euler angle space

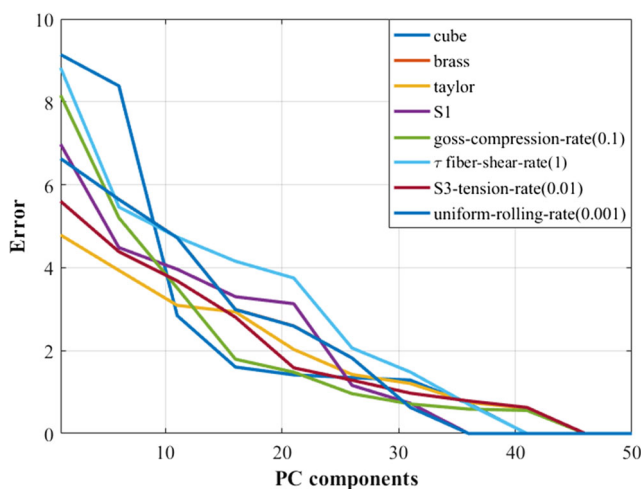


Fig. 4 Accuracy of texture reconstruction using PCA

how the error reduces when using more PCs to reconstruct textures. Both the MTEX synthesized texture and VPSC simulated texture are reconstructed using the same PC basis. The error between the actual and the reconstructed texture converges to zero by picking enough PCs when adding both the linear combinations of the texture components (from Table 3) and deformation textures predicted by VPSC under a range of imposed boundary conditions. Most textures can be represented by using more than 30 PCs with minimum error (Fig. 4). However, as the number of relevant texture components is increased, the number of necessary

PCs required to accurately represent the raw data will also increase. In that case, more advanced dimensionality reduction techniques should be recommended, such as kernel PCA, which provides more accuracy when raw data structure is non-linear [52–56]

In order to simplify the machine learning framework, we would like to use as efficient a representation as possible. Figure 5 shows examples of the reconstruction of both the synthesized and deformation texture using weighted PCs. In Fig. 5b, the ideal simulated ODF (left) shows the orientation distribution resulting from the shearing of an initial cube texture. The main characteristics of the ODF can be captured with acceptable accuracy by reducing the number of PCs to 16. The number of principal components considered is also motivated by the current limitations on the number of degrees of freedom on the Open Citrination platform. Therefore, for this work, we will consider 16 PCs as representative inputs to the machine learning framework, with the understanding that this may not be the optimal choice, but rather a compromise between accuracy and computational resources.

Machine Learning Algorithm

In the field of material science, ML is usually applied to predict various functional properties/behaviors and calibrate reduced order models from limited experimental data [57–61]. Numerous ML techniques have been applied across the engineering fields (such as *k*-nearest neighbors (*k*-NN),

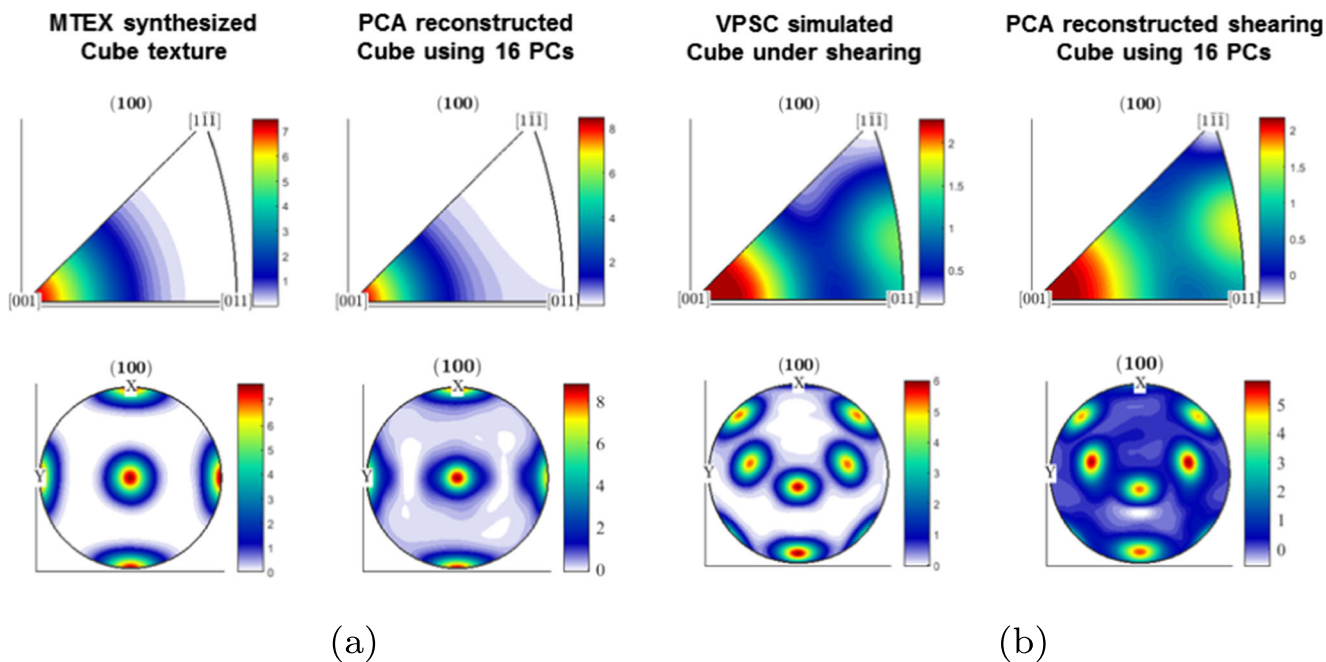


Fig. 5 Example of texture reconstruction using 16 PCs for **a** MTEX Synthesized unimodal cube texture **b** VPSC simulated cube under shearing. The initial texture was subsequently deformed to a strain of $\epsilon_0^* = 1$ under shear loading

decision trees (DT), random forest (RF), support vector machine (SVC), artificial neural networks (ANN), Bayesian networks, and deep learning) [62–64]. Additionally, many convenient toolboxes are developed to analyze and model data on commonly used Integrated Development Environments (IDEs), such as scikit-learn in python, statistics and machine learning toolbox, and neural network toolbox in MATLAB [62, 65, 66]. In this study, we use the Open Citrination platform [15, 67, 68] to construct the ML-based reduced order model for crystal plasticity. The default estimator on Open Citrination is based on random forest [68], which is an ensemble method wherein individual learners are decision trees [69, 70]. More complete description of the Citrination platform and the machine learning model on random forest can be learned from the open-source Lolo scala library [71]. Specific details related to the training of our model such as the number of estimators, minimum samples per leaf, and maximum tree depth used for the reduced order model in this paper are demonstrated on the Open Citrination platform and shown in Appendix [72].

Calibration and Validation of Reduced Order Model

The set of crystal plasticity model parameters and initial conditions (including imposed velocity gradient tensor, material texture, and hardening parameters) were systematically varied in order to build a comprehensive training dataset using VPSC. The details are described in Table 4. Four typical loading conditions are included, pure shear, rolling, uniaxial compression, and tension, while the loading rate varies between $1 \times 10^{-3} s^{-1}$ and $1 s^{-1}$. In order to consider the natural randomness at the simulation volume element level, for each ODF, ten different VPSC input grain files were created by discretely sampling 1000 grains. Additionally, ten random combinations of τ_1^s , θ_0^s , and θ_1^s are picked from the predefined range of normalized Voce hardening parameters for each simulation [8].

The simulation dataset from VPSC is randomly separated into two groups, the training set and the validation set. The training dataset is used to discover the correlation between the input and output of the predefined configuration, while the validation set is used to check the accuracy of

the corresponding mathematical relationship that was built by the training set. In total, approximately two thousand computational results (by picking different parameter sets in Table 4) were used for training and calibration. The reduced order model was trained using one thousand VPSC simulation results out of the complete data set. The remainder of the VPSC simulations was used as validation. Additional validation examples were also created by creating “new” textures, that were not included in the calibration data, as linear combinations of the standard texture components. Results from these additional validation simulations are described below.

Two configurations are defined in this case study as “Property prediction” and “Model calibration” corresponding to the invertible pathway shown in Fig. 6. Property prediction is used as a reduced order simulation tool that quickly generates the stress-strain curve and the processed texture given any initial sample texture and condition. The invertible configuration is used to quickly calibrate the optimal hardening parameters given any strain-stain curve and providing the initial state of a sample based on the EBSD of the processed sample.

The prediction of each property is calculated through all inputs; however, each property is predicted mainly through several important features among the training dataset. Open Citrination provides the important score of each feature corresponding to the contribution to model performance of each feature (see Appendix). As an example of model calibration, Voce hardening parameters are mainly determined by features produced from stress-strain curves, while the initial ODF is mainly determined by the input of final ODFs. The accuracy and the uncertainty of the reduced order model model are evaluated during training and validation through different error metrics, such as root mean squared error, and non-dimensional model error. Additionally, the standardized residual measures the difference between ideal and predicted values on Open Citrination.

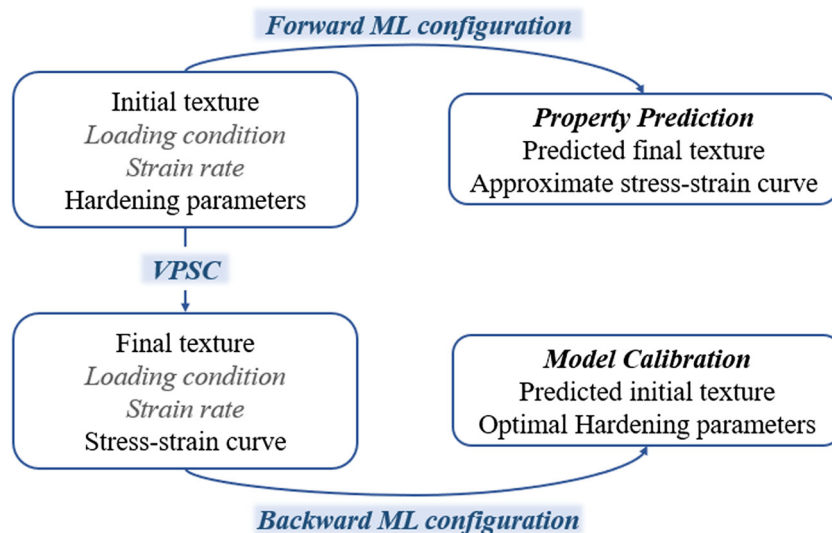
Results and Discussion

Several examples are presented below to demonstrate the accuracy of predictions from the reduced order model. All of the results presented are from the validation dataset and

Table 4 VPSC model parameters sampling for calibration data

Parameters	Number	Boundary condition
Loading condition	4	Compression, tension, shear, rolling
Normalized Voce hardening parameters	10	$\frac{\tau_1^s}{\tau_0^s} \in [0,1]$, $\frac{\theta_0^s}{\tau_0^s} \in [0,5]$, $\frac{\theta_1^s}{\tau_0^s} \in [0,2]$,
Initial texture	12	Cube, goss, brass, copper, S1, S2, S3, Taylor, uniform, α fiber, β fiber, τ fiber
Loading rate	4	1, 0.1, 0.01, 0.001

Fig. 6 Training configurations of “Property prediction” and “Model calibration”



were not used to calibrate the reduced order model. As mentioned in “[Reduced Order Representation of Crystallographic Texture](#)”, both initial and final ODF are represented by the first 16 PCs and the stress-strain behaviors are represented by ten selected points on stress-strain curves. This effectively means that this limited set of 26 variables can efficiently describe the critical microstructural features and the effective mechanical behavior of a FCC sample.

Examples in Fig. 7 show the accuracy of the reduced-order model prediction for the final texture and the stress-strain behavior for a range of different initial textures and loading conditions. The left ODF shown in Fig. 7a is an example of input crystallographic texture ODF, τ -fiber, used to verify the accuracy of the reduced order model. The middle pole figure shows the VPSC predicted texture resulting from the τ -fiber ODF under a shear loading condition, while the right one shows the predicted ODF from the reduced order model. The intensity of both pole figures are mostly captured when comparing the VPSC simulated results and the reduced order prediction.

In Fig. 7b, the red solid line shows the plastic stress-strain curve computed from VPSC for τ -fiber ODF under shear loading while ten blue dots represent the reduced order prediction of stress-strain behavior under the same condition. Most of the predicted points are located very close to the actual stress-strain curve for this case. Figure 7c–f show another two examples of the predicted ODF and plastic stress-strain curve for τ -fiber under tension and uniform ODF under compression. Additionally, Fig. 7g, h shows an example of the prediction of the final texture and the stress-strain curve when input a mixed FCC texture, 50% cube + 50% brass into the reduced order model. These results indicate the calibrated machine learning reduced order model can accurately capture the evolution of crystallographic texture during deformation

processing, while the predicted stress-strain curve shows a good correspondence with the VPSC simulation.

The overall expected level of error for the prediction of the stress-strain curve from Open Citrination is quantified in Fig. 8. Both the specific cases for “ τ -fiber ODF under shearing (Fig. 7b)” and “uniform ODF under compression (Fig. 7f)” are shown in Fig. 8. The box plot shows the distribution of the percentage error among the training dataset. The percentage error is defined as $\frac{\sigma_{true} - \sigma_{predicted}}{\sigma_{true}}$ and the average percentage error is limited in 2%. As a good prediction example, “ τ -fiber ODF under shearing” has a prediction error within the first and the third quartile (Q1 and Q3), while the percentage is less than 5%. While the percentage error of the worse case (“uniform ODF under compression”) locates around the extreme outlier. Additionally, the standard deviation (SD) of the percentage error of predictions are also shown as a function of the stain, and the overall uncertainty increase along with the strain.

Similarly, the examples in Fig. 9 shows the accuracy of the inverse case: model calibration for the initial texture, and identification of the optimal Voce hardening model parameters. In Fig. 9a, the ground truth initial condition texture, Taylor ODF (middle), is compared with the prediction from the inverse reduced order model. The deformed texture provided as input the to inverse reduced order model is shown on the left. The pole figures demonstrate excellent agreement between the ground truth initial texture and the inverse reduced order model prediction, with both qualitative features and trends in intensities being captured. We believe that the quantitative differences are largely due to truncating the PCA expansion to 16 components, and better agreement can be realized if a more terms could have been used. However, for the future application of identifying missing data in material database the semi-quantitative predictions are likely sufficient.

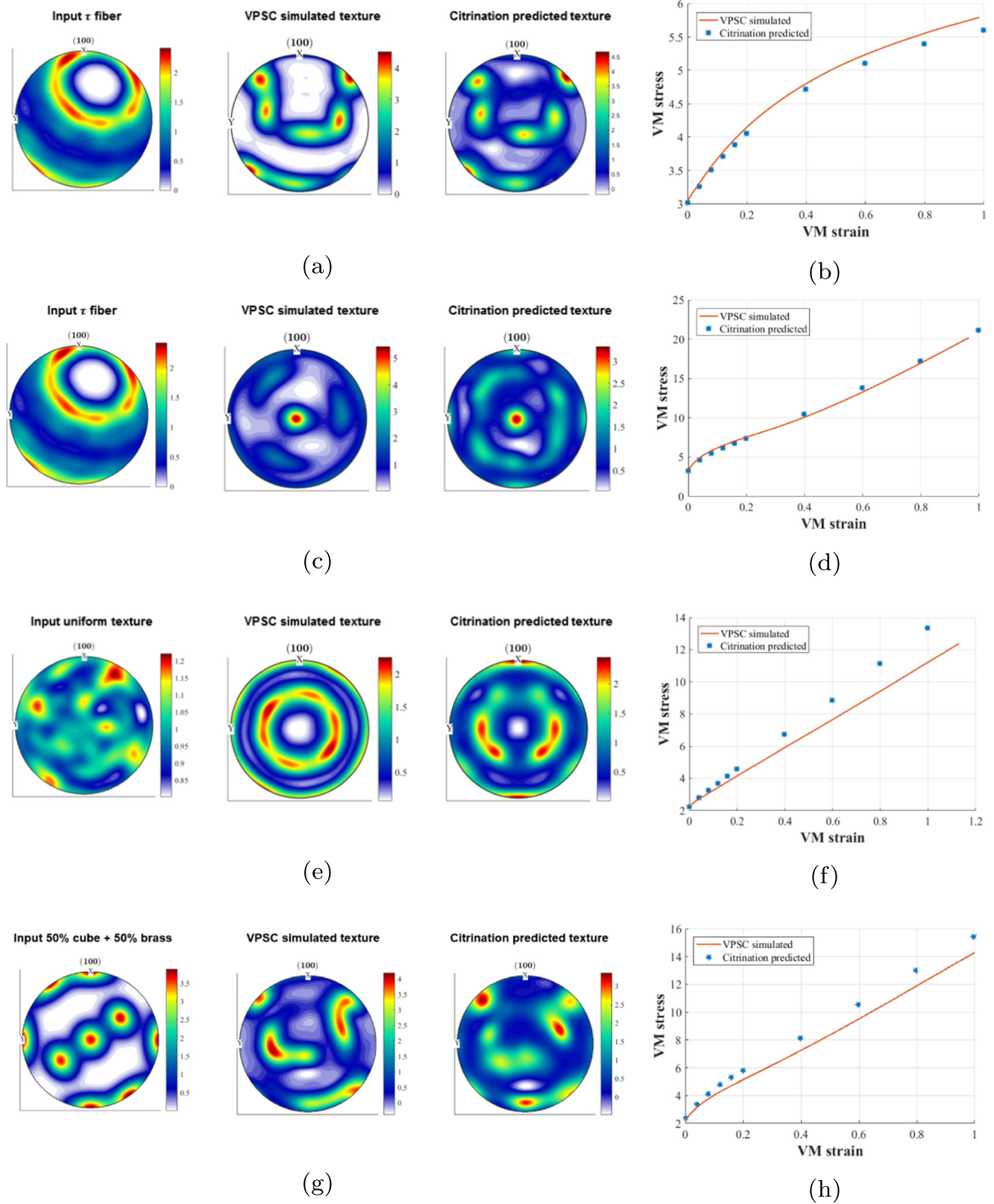


Fig. 7 Pole figure of the final texture generated by VPSC simulation (left) and Open Citration’s prediction (right) for the following: **a** τ -fiber ODF under shearing; **c** τ -fiber ODF under tension; **e** uniform ODF under compression; **g** 50% cube + 50% brass under shearing; The

Von-Mises stress-strain curve generated by VPSC simulation (solid line) and Open Citration’s prediction (scattering points) for: **b** τ -fiber ODF under shearing; **d** τ -fiber ODF under tension; **f** uniform ODF under compression; **(h)** 50% cube + 50% brass under shearing;

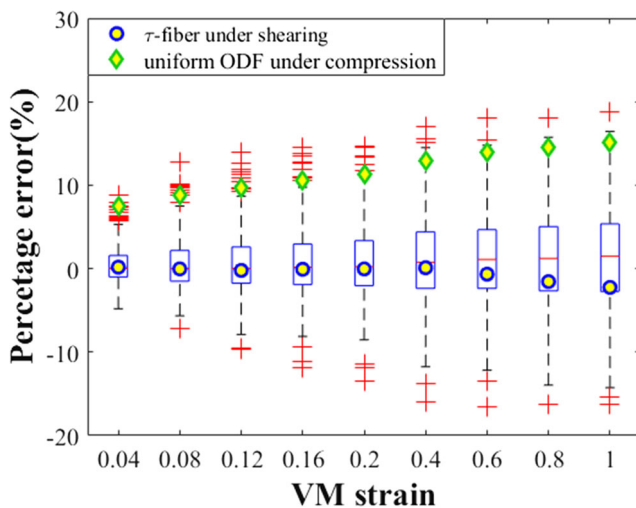


Fig. 8 Error assessment of stress-strain predictions using the reduced order model (The overall percentage error for validation set and the corresponding cases sourced from Fig. 7 are demonstrated). The red line indicates the mean error level and the boxes contain the middle 50% (2nd and 3rd quartile). The error bars indicate 10% and 90% of all samples. Individual extreme values are marked with +

Additional examples are also shown in Fig. 9c, e. Figure 9g shows the inverse reduced order model can also be utilized when the initial texture is not one of the standard ODF components. In this case, the initial texture of 50% goss + 50% brass was subject to rolling.

The inverse reduced order model can also be utilized to determine optimal Voce hardening parameters from limited stress-strain data. Figure 9b, d, f, and h compares the ground truth VPSC stress-strain curves resulting from the input Voce parameters with stress-strain curves resulting from the inverse reduced order model prediction of the hardening parameters. Stress-strain curves are generated covering the full confidence interval for the Voce parameters to better visualize the expected accuracy resulting from using the reduced order model for parameter estimation.

The accuracy of the validation set for Voce hardening parameters are compared with the exact VPSC simulated value in error bar plot in Fig. 10a. The asymptotic hardening rate is perfectly predicted by the proposed tool. Most of the back-extrapolated CRSS and the initial hardening rates are predicted with some uncertainty. Figure 10b, c shows the cumulative density function (CDF) and the probability density function (PDF) curve of the validation set of normalized hardening parameters from multiple cases. Figure 10b shows the upper predicted boundary (UPB) and the lower predicted boundary (LPB) of Open Citration's predicted values that corresponded to the positive and negative variance. Figure 10c shows the probability distribution of the positive variance over the entire training and validation set.

For practical applications, crystal plasticity calculations are exercised to predict material response under different conditions (typically either initial texture or strain path) than the calibration data. In order to check whether the uncertainty of the stress-strain behavior that was computed using the estimated hardening parameters is allowable and acceptable for this purpose, the optimal hardening parameters were estimated and used to simulate the stress-strain response under multiple loading conditions. Figures 11 and 12 show the reproduced stress-strain curve of selected texture under different loading conditions. In Figs. 11 and 12, the blue dash line shows the stress-strain data used to estimate the Voce parameters from the inverse reduced order model. The calibrated Voce hardening parameters are then used to predict the stress-strain response under differing loading conditions for the same textures. As can be seen in the figures, the reproduced stress-strain curves are very close to the ground truth response (solid red line).

The case studies above demonstrate that the Open Citration-trained model successfully captures the underlying physics represented in the VPSC model with Voce hardening, and can be used as a surrogate model for the more computationally expensive crystal plasticity simulation. The model can also be inverted quickly to learn material properties, in this case single crystal yield and hardening parameters, from mechanical tests avoiding manual fitting. This can effectively reduce the cost of large amounts of computational and experimental results at the preliminary design stage with the adoption of the machine learning model.

However, it should be noted that the uncertainty of the calibrated parameters is neither uniform across the three normalized hardening parameters nor constant for a given parameter. As shown in Fig. 10, the calibrated values of the initial hardening rate, θ_0 , and the asymptotic hardening rate, θ_1 , have smaller variance than the back-extrapolated CRSS, τ_1 . This is expected from the form of the Voce hardening law, both θ_0 and θ_1 define slopes of the stress-strain curve which describe the trend of the curve, which means that training data points at multiple strain values contribute to these parameters. However, τ_1 is the interception of back-extrapolation asymptotic stress as very large strains with the stress axis. This value is not independent of the other parameters and that the calibrated uncertainty of τ_1 will always be greater than others. The uncertainty is also not constant in that θ_1 can be determined more accurately when it is numerically small than when it is the parameter takes a large value. Or equivalently the hardening parameters can be more accurately determined in systems that display a smaller degree of work hardening.

The quality of the machine learning fit was also found to be sensitive to how the input stress-strain curves were

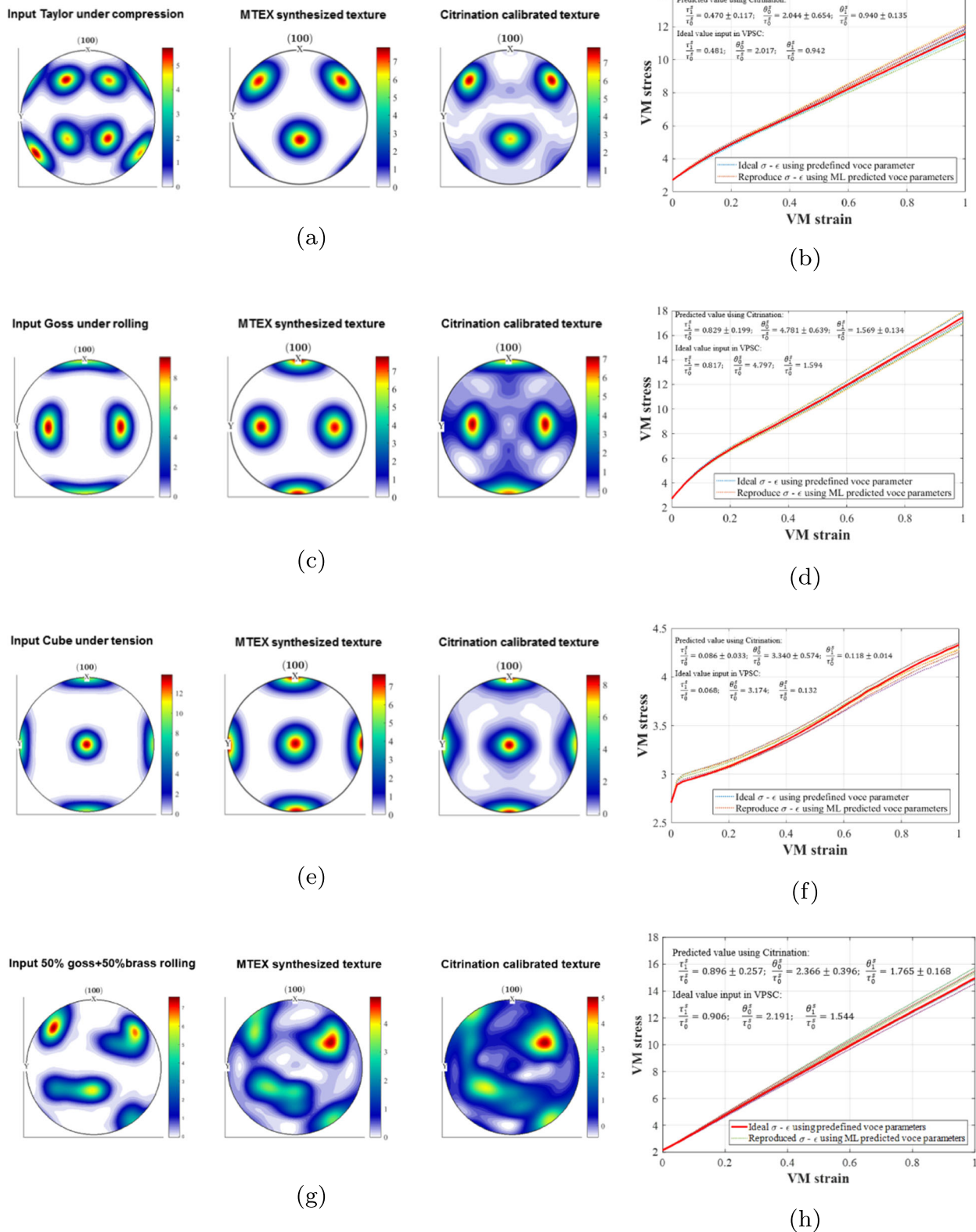


Fig. 9 Pole figure of the final texture generated by MTEX synthesized (left) and Open Citrination’s calibration (right) for: **a** Taylor texture; **c** Goss texture; **e** Cube texture; **g** 50% goss+ 50% brass texture; The Von-Mises stress-strain curve generated by VPSC simulation using predefined Voce hardening parameters (solid red line)

and Open Citrination’s predicted curves using couple calibrated Voce hardening parameters with variance (dotted lines) for: **b** Taylor texture under compression; **d** Goss texture under rolling; **f** Cube texture under tension; **h** 50% goss+ 50% brass texture under rolling

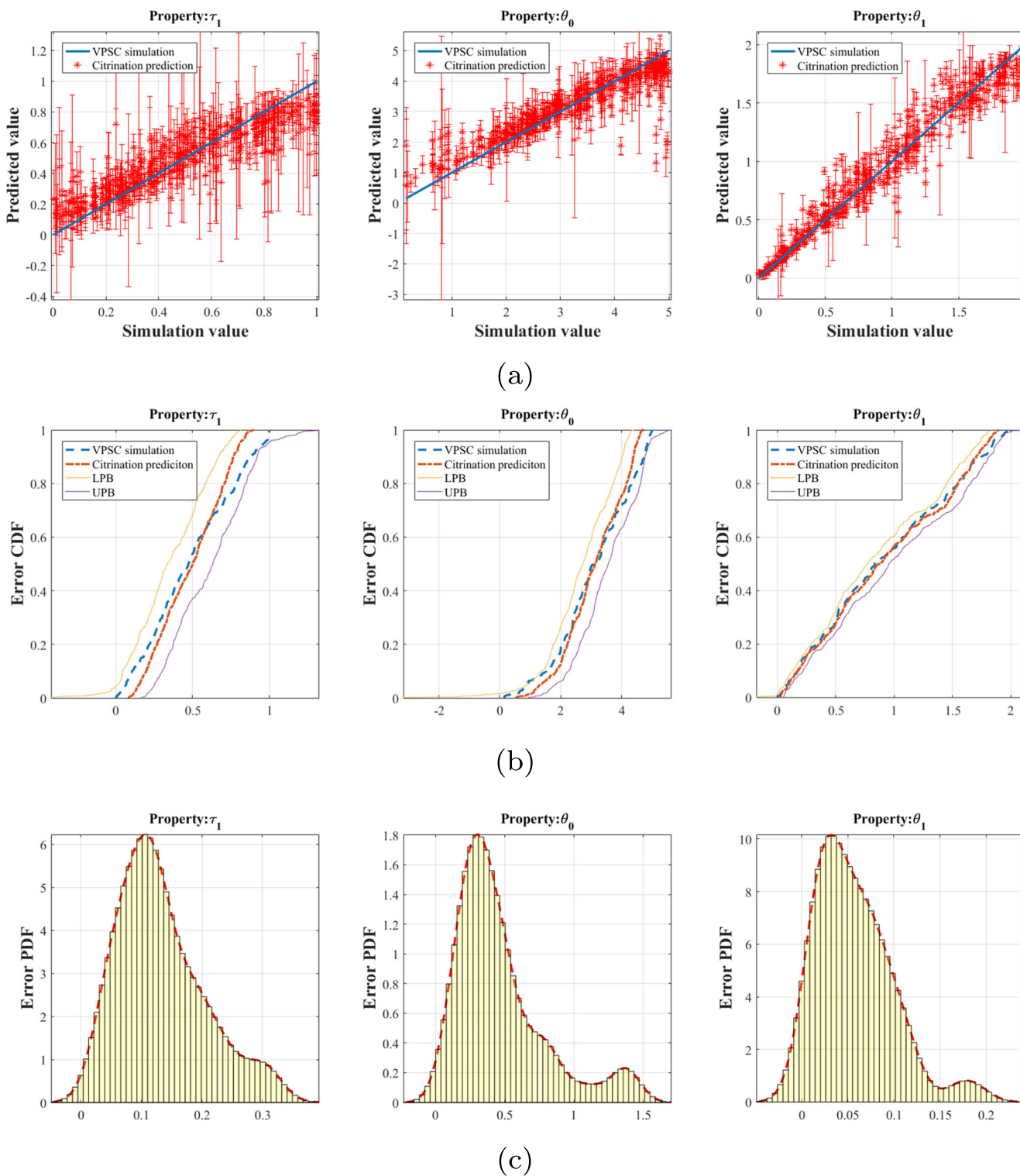


Fig. 10 **a** The comparison between the ideal and predicted Voce hardening parameters; **b** Cumulative distribution function (CDF) of predicted Voce hardening parameters using Open Citrination; **c** Probability distribution function (PDF) of predicted Voce hardening parameters using Open Citrination

sampled. For initial calibration attempts, we used a uniform (with respect to strain) sampling. However, this found to provide poor fits for parameter sets corresponding to rapid

hardening or rapid saturation of stress with increasing strain, i.e., cases where $\frac{d\sigma}{d\varepsilon}$ was large. Instead, a non-uniform sampling scheme was applied where a higher density of

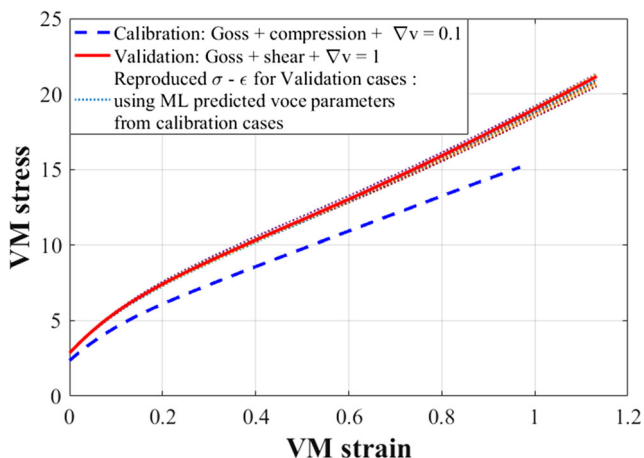


Fig. 11 Validation of optimal hardening parameters using goss texture under different loading conditions

stress-strain values were taken early in the curve where hardening rates were highest. This scheme is likely to fail for the case of non-kosher Voce parameters (see Fig. 2) as the effective stress-strain response would exhibit curves or inflections in regions not heavily sampled. While a seemingly trivial step in the process, the discretization of the input data can have a large impact on the accuracy and efficacy of the fit machine learning tool.

One potential limitation of this initial implementation is that the texture predictions from the model are all taken at large strains ($\epsilon = 1$). As shown above, the calibrated machine learning model does a good job of predicting the final texture as a function of initial texture and strain path at large strain, but as it currently stands, has not been calibrated to predict the texture as a function of stress-strain curve for a fixed strain path. Intermediate textures were not used in the calibration process. However, this is a feature that would be reasonably straightforward to add in the future.

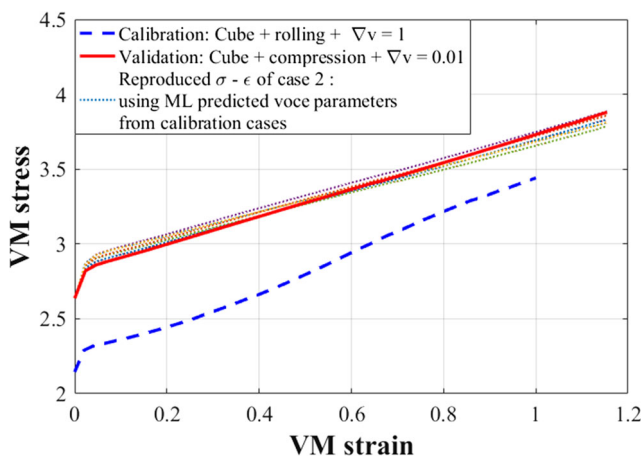


Fig. 12 Validation of optimal hardening parameters using cube texture under different loading conditions

Furthermore, this work shows a good prediction of deformation textures resulting from commonly used loading conditions. However, in real-world applications, material may be manufactured by complex loading path, such as bi-axial loading, sequential combination of different paths, or even superposition of strain paths. Analogously to the interim texture predictions, the extension to a broader range of strain paths should also be straightforward.

Overall, as an initial case study, this work only considered the deformation of FCC materials. FCC was chosen for practical reasons, as the plastic response is easier to predict as only one slip mode, containing 12 independent and degenerate $\{111\} \langle 110 \rangle$ slip systems, needs to be considered. The “texture-processing-properties” map for materials with BCC is significantly more complex to mine as multiple slip modes and up to 48 slip systems have a opportunity to be activated depending on the deformation path. For other non-cubic structures, such as HCP materials, the deformation behavior is harder to capture because twinning can be an important mechanism for accommodation of plasticity. The inclusion of non-degenerate slip modes drastically increases the number of parameters that need to be estimated in order to predict the stress-strain response and texture evolution. Also for simplicity, the limiting case of uniform latent hardening was considered for the case studies presented above. With non-degenerate slip modes, the form of the hardening matrix (see Eq. 4) also needs to be learned in addition to the hardening parameters for each slip mode. Deformation twinning also usually produces remarkable effect on texture evolution due to grain reorientation. Additionally, the hardening response will be affected since twin lamellae are potent barriers to dislocation motion. This will manifest as a strong degree of latent hardening associated with twin systems. However, in order to become a practical design tool, the framework presented here needs to be extend to both BCC and HCP initially (followed by a wider range of lower symmetry structures).

Conclusion and Future Work

This work demonstrates the feasibility and utility of using a machine learning approach to develop reduced order or fast-acting models to capture the behavior of more complex “physics-based” materials simulations. These reduced order models can then be used in ICME to efficiently explore the relationship between processing, microstructure, and property. The development of the proposed tool can effectively reduce the time and cost of property and performance prediction, and improve the quality of the real-world manufacturing design. Specifically, in this case study, instead of using a crystal plasticity solver to calculate the evolved microstructure and the mechanical properties under

different conditions, the proposed tool can predict the result directly. Additionally, the existing material properties can be supplemented without repeatable experiments or simulations. According to the invertible linkage established in this study, the proposed tool can computationally optimize a processing route for a given starting material to meet a set of designer specified criteria.

The initial case study stated in this paper shows that both “Property prediction” and “Model calibration” pathways can be established by training a comprehensive computational database of FCC copper on Open Citrination. This will become a preliminary fast-acting tool that finds the processing path to the desired structure and the macroscale functional properties. The overarching goal of this project can be extended to the “data-driven material design” for expected microstructures with desired crystal plasticity properties depending upon given element composition, crystal structure, possible activated slip/twinning systems, initial texture, and processing techniques and conditions. In the future, we can continue to expand our database by adding experimental data and other textures in various crystal systems and consider other microstructural features as important factors. Moreover, other conditions and model parameters, such as arbitrary loading types, temperature, constitutive sensitive rate, and different crystal plasticity laws can be gradually added as variance in the existing database to complement the current machine learning model.

Funding Information The authors received financial support from the U.S. Department of Energy National Energy Technology Laboratory award DE-FE0027776 and DARPA Young Faculty Award grant D15AP00103.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Appendix: Machine Learnign Details and Hyperparameters

The proposed crystal plasticity reduced order model is built on a random forest model that includes uncertainty estimation. Random forests can be understand as featured bagging of decision trees, where decision trees are basic models that define a piece-wise function that analyzing the input space recursively. Random forests have much better ability than single decision trees when predicting data with noise, and overcome the non-linearity by random draw, and featured bagging of the complete training dataset. The final decision/prediction of random forest is aggregated by voting/averaging of each tree so that limiting the overfitting on dataset without substantially increasing error due to the variance and bias.

Table 5 Hyperparameter and important features for training session

Configuration	Number of estimators	Minimum samples per leaf	Maximum tree depth
Forward	189	1	30
Backward	210	1	30
Property	Important features		
Yield stress	loading rate, loading condition, τ_1		
Hardening curve	$\tau_1, \theta_0, \theta_1$, PCs of initial textures		
Final texture	loading condition, PCs of initial textures		
Voce parameters	Stress-strain discrete points, PCs of final textures		
Initial texture	PCs of final textures		

Table 5 demonstrates the machine learning algorithm and hyperparameters used for the machine learning training session on Open Citrination, while the estimator is sourced from the open-source machine learning Lolo scala library [71]. The main important features used to train the model are also shown in Table 5. The complete description of the fraction for each important feature used to train the model can be found on Open Citrination [72]. (Importance scores of each feature sums to one and are determined by a given feature’s contribution to the model’s performance.) Moreover, the detailed model reports and summary of the training session for this case study can be reached at Dataviews 5506 (forward configuration) and 5507 (backward configuration) on Open Citrination.

References

- Allison J (2011) Integrated computational materials engineering: a perspective on progress and future steps. *JOM* 63(4):15
- Agrawal A, Choudhary A (2016) Perspective: Materials informatics and big data: realization of the “fourth paradigm” of science in materials science. *Apl Materials* 4(5):053208
- Olson GB (1997) Computational design of hierarchically structured materials. *Science* 277(5330):1237
- Panchal JH, Kalidindi SR, McDowell DL (2013) Key computational modeling issues in integrated computational materials engineering. *Comput Aided Des* 45(1):4
- Asaro RJ (1983) Crystal plasticity. *J Appl Mech* 50(4b):921
- Lebensohn R, Tomé C (1994) A self-consistent viscoplastic model: prediction of rolling textures of anisotropic polycrystals. *Mater Sci Eng A* 175(1-2):71
- Lebensohn R, Tomé C, Castaneda PP (2007) Self-consistent modelling of the mechanical behaviour of viscoplastic polycrystals incorporating intragranular field fluctuations. *Philos Mag* 87(28):4287
- Tomé C, Lebensohn R (2007) Visco-plastic self-consistent (vpssc), Los Alamos National Laboratory (USA) and Universidad Nacional de Rosario (Argentina) 6
- Jia N, Peng RL, Wang Y, Johansson S, Liaw P (2008) Micromechanical behavior and texture evolution of duplex stainless steel studied by neutron diffraction and self-consistent modeling. *Acta Mater* 56(4):782

10. Wang H, Wu P, Tomé C, Huang Y (2010) A finite strain elastic–viscoplastic self-consistent model for polycrystalline materials. *J Mech Phys Solids* 58(4):594
11. Anglin B, Gockel B, Rollett A (2016) Developing constitutive model parameters via a multi-scale approach. *Integr Mater Manuf Innov* 5(1):11
12. Aguir H, BelHadjSalah H, Hambli R (2011) Parameter identification of an elasto-plastic behaviour using artificial neural networks–genetic algorithm method. *Mater Des* 32(1):48
13. Aguir H, Chamekh A, BelHadjSalah H, Dogui A, Hambli R (2008) Identification of constitutive parameters using hybrid ann multi-objective optimization procedure. *Int J Mater Form* 1(1):1
14. White A (2012) The materials genome initiative: one year on. *MRS Bull* 37(08):715
15. O'Mara J, Meredig B, Michel K (2016) Materials data infrastructure: a case study of the citrination platform to examine data import, storage, and access. *JOM* 68(8):2031
16. Taylor GI (1938) Plastic strain in metals, our. *Inst Metals* 62:307
17. Hill R (1965) A self-consistent mechanics of composite materials. *Journal of the Mechanics and Physics of Solids* 13(4):213
18. Anand L, Kothari M (1996) A computational procedure for rate-independent crystal plasticity. *J Mech Phys Solids* 44(4):525
19. Roters F, Eisenlohr P, Hantcherli L, Tjahjanto DD, Bieler TR, Raabe D (2010) Overview of constitutive laws, kinematics, homogenization and multiscale methods in crystal plasticity finite-element modeling: theory, experiments, applications. *Acta Mater* 58(4):1152
20. Molinari A, Canova G, Ahzi S (1987) A self consistent approach of the large deformation polycrystal viscoplasticity. *Acta Metall* 35(12):2983
21. Segurado J, Lebensohn RA, LLorca J, Tomé CN (2012) Multiscale modeling of plasticity based on embedding the viscoplastic self-consistent formulation in implicit finite elements. *Int J Plast* 28(1):124
22. Proust G, Kalidindi SR (2006) Procedures for construction of anisotropic elastic–plastic property closures for face-centered cubic polycrystals using first-order bounding relations. *J Mech Phys Solids* 54(8):1744
23. Liu B, Raabe D, Roters F, Eisenlohr P, Lebensohn R (2010) Comparison of finite element and fast fourier transform crystal plasticity solvers for texture prediction. *Model Simul Mater Sci Eng* 18(8):085005
24. Lebensohn RA (2001) N-site modeling of a 3d viscoplastic polycrystal using fast fourier transform. *Acta Mater* 49(14):2723
25. Lebensohn RA, Rollett AD, Suquet P (2011) Fast fourier transform-based modeling for the determination of micromechanical fields in polycrystals. *JOM* 63(3):13
26. Eshelby JD (1957) The determination of the elastic field of an ellipsoidal inclusion, and related problems. *Proc R Soc Lond A* 241(1226):376
27. Eshelby JD (1959) The elastic field outside an ellipsoidal inclusion. *Proc R Soc Lond A* 252(1271):561
28. Van Houtte P (1994) Application of plastic potentials to strain rate sensitive and insensitive anisotropic materials. *Int J Plast* 10(7):719
29. Kalidindi SR, Duvvuru HK, Knezevic M (2006) Spectral calibration of crystal plasticity models. *Acta Mater* 54(7):1795
30. Raabe D, Roters F (2004) Using texture components in crystal plasticity finite element simulations. *Int J Plast* 20(3):339
31. Kocks UF, Tomé CN (2000) *HR Wenk Texture and anisotropy: preferred orientations in polycrystals and their effect on materials properties*. Cambridge University Press, Cambridge
32. Kalidindi SR, Bronkhorst CA, Anand L (1992) Crystallographic texture evolution in bulk deformation processing of fcc metals. *J Mech Phys Solids* 40(3):537
33. Hielscher R, Schaeben H (2008) A novel pole figure inversion method: specification of the mtex algorithm. *J Appl Crystallogr* 41(6):1024
34. Bachmann F, Hielscher R, Schaeben H (2010) In solid state phenomena. *Trans Tech Publ* 160:63–68
35. Mainprice D, Hielscher R, Schaeben H (2011) Calculating anisotropic physical properties from texture data using the mtex open-source package. *Geol Soc Lond Spec Publ* 360(1):175
36. Bunge HJ (2013) *Texture analysis in materials science: mathematical methods*. Elsevier, Amsterdam
37. Adams BL, Henrie A, Henrie B, Lyon M, Kalidindi S, Garmestani H (2001) Microstructure-sensitive design of a compliant beam. *J Mech Phys Solids* 49(8):1639
38. Adams BL, Gao XC, Kalidindi SR (2005) Finite approximations to the second-order properties closure in single phase polycrystals. *Acta Mater* 53(13):3563
39. Sundararaghavan V, Zabarar N (2005) Classification and reconstruction of three-dimensional microstructures using support vector machines. *Comput Mater Sci* 32(2):223
40. Sundararaghavan V, Zabarar N (2004) A dynamic material library for the representation of single-phase polyhedral microstructures. *Acta Mater* 52(14):4111
41. Sundararaghavan V, Zabarar N (2008) A multi-length scale sensitivity analysis for the control of texture-dependent properties in deformation processing. *Int J Plast* 24(9):1581
42. Fullwood DT, Niezgoda SR, Adams BL, Kalidindi SR (2010) Microstructure sensitive design for performance optimization. *Prog Mater Sci* 55(6):477
43. Kalidindi SR, Houskamp JR, Lyons M, Adams BL (2004) Microstructure sensitive design of an orthotropic plate subjected to tensile load. *Int J Plast* 20(8):1561
44. Niezgoda SR, Yabansu YC, Kalidindi SR (2011) Understanding and visualizing microstructure and microstructure variance as a stochastic process. *Acta Mater* 59(16):6387
45. Niezgoda SR, Kanjarla AK, Kalidindi SR (2013) Novel microstructure quantification framework for databasing, visualization, and analysis of microstructure data. *Integr Mater Manuf Innov* 2(1):1
46. Galbincea ND, Yuan M, Niezgoda SR (2017) Computational design tools for quantifying uncertainty due to material variability. In: 19th AIAA non-deterministic approaches conference, p 0816
47. Fukunaga K (2013) *Introduction to statistical pattern recognition*. Academic Press, Cambridge
48. Johnson RA, Wichern D (2002) *Multivariate analysis*. Wiley Online Library, New York
49. Ghodsi A (2006) Dimensionality reduction a short tutorial. In: Department of Statistics and Actuarial Science, Univ of Waterloo, Ontario, Canada, vol 37, p 38
50. Maaten Lvd, Hinton G (2008) Visualizing data using t-sne. *J Mach Learn Res* 9:2579
51. Wattenberg M, Viégas F, Johnson I (2016) How to use t-sne effectively. *Distill* 1(10):e2
52. Schölkopf B, Smola A, Müller KR (1998) Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput* 10(5):1299
53. Ma X, Zabarar N (2011) Kernel principal component analysis for stochastic input model generation. *J Comput Phys* 230(19):7311
54. Ringné M (2008) What is principal component analysis? *Nat Biotechnol* 26(3):303
55. Abdi H, Williams LJ (2010) *Principal component analysis*. Wiley Interdiscip Rev Comput Stat 2(4):433
56. Shlens J (2014) A tutorial on principal component analysis. arXiv:1404.1100
57. Kalidindi SR, Niezgoda SR, Salem AA (2011) Microstructure informatics using higher-order statistics and efficient data-mining protocols. *JOM* 63(4):34

58. Sundararaghavan V, Zabarav N (2009) A statistical learning approach for the design of polycrystalline materials. *Statistical Analysis and Data Mining* 1(5):306
59. Pilania G, Wang C, Jiang X, Rajasekaran S, Ramprasad R (2013) Accelerating materials property predictions using machine learning. *Sci Rep* 3:2810
60. Meredig B, Agrawal A, Kirklin S, Saal JE, Doak J, Thompson A, Zhang K, Choudhary A, Wolverton C (2014) Combinatorial screening for new materials in unconstrained composition space with machine learning. *Phys Rev B* 89(9):094104
61. Schütt K, Glawe H, Brockherde F, Sanna A, Müller K, Gross E (2014) How to represent crystal structures for machine learning: towards fast prediction of electronic properties. *Phys Rev B* 89(20):205118
62. Robert C (2014) Machine learning a probabilistic perspective
63. Nasrabadi NM (2007) Pattern recognition and machine learning. *J Electron Imaging* 16(4):049901
64. Goldberg DE, Holland JH (1988) Genetic algorithms and machine learning. *Mach Learn* 3(2):95
65. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V et al (2011) Scikit-learn: machine learning in python. *J Mach Learn Res* 12(Oct):2825
66. Demuth H, Beale M (1993) Neural network toolbox for use with matlab – user’s guide verion 3.0
67. Hill J, Mannodi-Kanakithodi A, Ramprasad R, Meredig B (2018) Materials science with large-scale data and informatics: unlocking new opportunities. In: *Computational materials system design*. Springer, pp 193–225
68. Ling J, Hutchinson M, Antono E, Paradiso S, Meredig B (2017) High-dimensional materials and process optimization using data-driven experimental design with well-calibrated uncertainty estimates. *Integr Mater Manuf Innov* 6(3):207
69. Breiman L (2001) Random forests. *Mach Learn* 45(1):5
70. Liaw A, Wiener M et al (2002) Classification and regression by randomforest. *R news* 2(3):18
71. Hutchinson M, Paradiso S lololo <https://githubcom/CitrineInformatics/lolo>
72. Citrine informatics <https://citrinationcom> Accessed: 2018-11-01