



Gaussian process autoregression models for the evolution of polycrystalline microstructures subjected to arbitrary stretching tensors

Sepideh Hashemi^a, Surya R. Kalidindi^{a,b,*}

^a George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, 30332, GA, USA

^b School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, 30332, GA, USA

ARTICLE INFO

Keywords:

Reduced-order models
Design of Experiment
Gaussian process autoregressive model
Process-structure evolution linkage
Crystal plasticity
Polycrystalline materials

ABSTRACT

Crystal plasticity finite element models (CPFEM) have shown tremendous potential for simulating the microstructure evolution paths in polycrystalline aggregates subjected to large plastic strains. However, their high computational cost has hindered their broader deployment in design efforts where large process design spaces need to be explored. In this work, a novel machine learning framework is developed to establish low-computational cost reduced-order models for predicting the details of microstructure evolution in face-centered cubic (FCC) polycrystalline microstructures subjected to arbitrary stretching tensors. Within this framework, the previously established feature engineering of polycrystalline materials within the material knowledge system (MKS) framework is extended such that it is applicable to the highly deformed microstructures obtained during large deformations. Gaussian process autoregression (GPARG) approaches combined with Bayesian design of experiment strategies are employed for building the desired surrogate models to optimize the generation of the computationally expensive training data (produced using CPFEM). It is demonstrated that a relatively small training set of 1400 datapoints is adequate to produce a high-fidelity reduced-order model for predicting the details of the microstructure evolution in a very broad set of FCC polycrystalline aggregates subjected to arbitrary macroscopically imposed stretching tensors.

1. Introduction

It is well known that certain salient details of the polycrystalline microstructures of metals strongly influence their overall mechanical properties (McDowell, 2018; Ghosh and Dimiduk, 2011; Jones and Ashby, 2012). Since the metallic microstructures can be modulated easily with a broad variety of thermo-mechanical processing routes, there is a strong interest in developing low-computational cost reduced-order models relating the process parameters to the salient details of the microstructure evolution (Ibragimova et al., 2021; Fernandez-Zelaia and Melkote, 2019; Khandelwal et al., 2021; Fernandez-Zelaia and Melkote, 2019; Khosravani et al., 2017; Hashemi and Kalidindi, 2021). An important parameter characterizing metal manufacturing processes is the

* Corresponding author.

E-mail address: surya.kalidindi@me.gatech.edu (S.R. Kalidindi).

imposed deformation history, which is best captured as a time-dependent stretching tensor. It should be noted that the complete domain of isochoric¹ stretching tensors (i.e., the space of all traceless symmetric second-rank tensors) serves as the formal design space for developing metals with improved effective properties. This design space is rather large, and a rigorous exploration of such large design spaces requires the development of validated reduced-order models connecting the large input domains and output ranges involved. The central challenge in this task comes from the lack of a suitable mathematical framework for establishing practically useful, but universally applicable, low-dimensional representations of the inherently high-dimensional polycrystalline microstructures.

Crystal plasticity finite element models (CPFEM) represent a class of established physics-based tools for simulating the textural and morphological details of microstructure evolution during room-temperature plastic deformations in single-phase cubic polycrystalline metals deforming by crystallographic slip (Kalidindi et al., 1992; Khan et al., 2015; Ali et al., 2017; Luccarelli et al., 2017). CPFEM models essentially apply a crystal plasticity constitutive model at each integration point of the meshed polycrystalline aggregate. Although CPFEM have enjoyed many successes in predicting the evolution of heterogeneous microstructures in various metal-forming applications (Gupta et al., 2018; Tasan et al., 2014; Sedighiani et al., 2022; Knezevic et al., 2014; Chen et al., 2018; Wang et al., 2019; Sedighiani et al., 2021), their broader adoption in material design has been hampered by the extremely high computational cost of these tools (Iftikhar et al., 2021; Farooq et al., 2020; Bonatti et al., 2022). As already described earlier, a validated low-computational cost reduced-order model could open an attractive avenue. Since the phenomena of interest involves microstructure evolution resulting from an imposed deformation history, it is most logical to formulate the desired reduced-order model as an auto-regressive (AR) model (Box et al., 2015). In these approaches, the inputs to the model would include the salient descriptors of the microstructure at the beginning of a fixed time step and the imposed stretching tensor during the time step, while the output would constitute the salient features of the microstructure at the end of the time step. This type of formulation allows for recursive usage of the model by taking the output from one time step as a part of the input for the next time step. It should also be noted that although AR models are generally formulated to mainly capture linear relationships between the inputs and outputs, they can be extended to include nonlinear relationships (Zhou and Ding, 2020; Chen et al., 2018; Waheeb et al., 2019). The central challenges in establishing reduced-order AR models for capturing the salient features of microstructure evolution in metal polycrystalline aggregates subjected to arbitrary deformation modes include: (i) the earlier mentioned lack of established formalisms for capturing the salient features of polycrystalline microstructures in low dimensional representations, and (ii) lack of demonstrated workflows capable of efficiently building the desired reduced-order models with as few training datapoints as possible (this is important because of the high computational cost of CPFEM).

Majority of prior studies addressing the problem described above have incorporated highly simplified feature engineering, i.e., they employed grossly simplified representations of the complex polycrystalline microstructures. For example, some studies have focused only on the morphological features and disregarded completely the information on the crystallographic texture (Fernandez-Zelaia and Melkote, 2019; Yabansu et al., 2019; Muhammad et al., 2021). Alternately, other studies have focused exclusively on the averaged texture information in the polycrystalline aggregate (Knezevic and Kalidindi, 2007; Kalidindi et al., 2009; Knezevic and Bhattacharyya, 2017; Sankaran and Zabarar, 2007). More recent studies have exploited the implicit feature engineering capabilities of neural networks (including both the simple feed-forward neural networks as well as the convolution neural networks). These studies have mainly incorporated the texture information, and have successfully predicted the flow behavior, stress-strain response and texture evolution of various FCC polycrystals undergoing plastic deformation (Ibragimova et al., 2021; Ali et al., 2019; Saidi et al., 2022; Dai et al., 2021; Acar, 2020; Yuan et al., 2018; Gorji et al., 2020; Ali, 2021). Only a few recent studies incorporated both textural and morphological information to predict localized stress and strain fields (Ibragimova et al., 2022; Mangal and Holm, 2018). It should be noted that the neural network-based approaches require a large training dataset, posing significant challenges to problems where data generation is computationally expensive (such as the CPFEM). Since both textural and grain morphology information in the representative volume element (RVE) play an important role in controlling its effective response, it is desirable to develop feature engineering approaches that rigorously capture all of the salient features of the polycrystalline microstructures.

The recently developed Material Knowledge System (MKS) framework (Kalidindi, 2015; Iskakov et al., 2018; Latypov et al., 2019; Yabansu et al., 2020) leverages 2-point spatial correlations (Torquato and Haslach Jr, 2002; Niezgodna et al., 2011; Niezgodna et al., 2013) and principal component analyses (PCA) (Hastie et al., 2005) to obtain low dimensional representations of polycrystalline microstructures. Prior studies have demonstrated that the 2-point spatial correlations of lattice orientations in polycrystalline aggregates can be computed efficiently utilizing generalized spherical harmonics (GSH) (Hashemi and Kalidindi, 2021; Kalidindi, 2015; Paulson et al., 2017; Bunge, 1993). The main advantage of using GSH representations lies in its ability to reflect compactly all relevant crystal and/or sample symmetries. However, the complete set of the 2-point spatial correlations of orientations in polycrystalline aggregates contains an extremely large number of microstructure descriptors (often as many as $10^6 - 10^9$ (Hashemi and Kalidindi, 2021; Paulson et al., 2017; Paulson et al., 2018)), making it intractable for the formulation of low-cost reduced-order models. Prior work has addressed this challenge by employing PCA to obtain a reduced-order representation of each microstructure by using only the first 10-15 principal component (PC) scores (Khosravani et al., 2017; Iskakov et al., 2018; Yabansu et al., 2020; Paulson et al., 2017; Fernandez-Zelaia et al., 2019). The MKS feature engineering approach is distinctly different from the other feature engineering approaches mentioned earlier. First, it rigorously captures most of the textural and morphological information of the microstructure. Second, it employs an unsupervised learning approach in that the identified features are unaware of the targets (i.e., model outputs). In other words, they are agnostic to the targets, and can potentially be employed for a very broad set of user-selected targets. Third, highly

¹ The hydrostatic components of imposed stretching tensors in most metals are very small and produce only elastic deformations; these are ignored in this work.

efficient computational protocols for evaluating these features have already been developed in prior work (Cecen et al., 2016; Fullwood et al., 2010). Indeed, it has been shown that high-fidelity structure-property (S-P) linkages can be established for polycrystalline materials using these low-dimensional representations (Paulson et al., 2017; Paulson et al., 2018; Priddy et al., 2017; Paulson et al., 2019; Tallman et al., 2019). In this work, our goal is to demonstrate that the same protocols can be used to formulate reduced-order process-structure evolution (P-S) linkages as well, thereby leading to a consistent framework for exploring the complete set of process-structure-property (PSP) linkages needed for materials design (Kalidindi, 2015; Kalidindi, 2019; McDowell and Olson, 2008; Ramakrishna et al., 2019). It should be noted that the establishment of P-S linkages is generally more challenging compared to the formulation of the S-P linkages. This is because of the need to track the time evolution of the material microstructure for a specified process history. Although a number of different model building strategies exist, the nonparametric Gaussian process autoregressive (GPAP) models (Requeima et al., 2019) are particularly attractive because of their high expressivity with relatively small training datasets. This approach allows the formulation of P-S linkages without a priori knowledge of the expected model form, while providing a natural way for the quantification of the prediction uncertainties. In prior work, GPAP has been explored in problems involving relatively simpler microstructures (e.g., two-phase microstructures) (Brough et al., 2017) or simplified descriptions of complex microstructures (e.g., grain size distribution) (Fernandez-Zelaia and Melkote, 2019; Fernandez-Zelaia and Melkote, 2019; Yabansu et al., 2019) or relatively smaller design spaces (e.g., including only morphological parameters such as grain shape and size) (Hashemi and Kalidindi, 2021).

The main goal of this work is to explore the feasibility of formulating GPAP-based microstructure evolution linkages for polycrystalline materials subjected to an arbitrary stretching tensor. This goal encounters two main challenges. The first challenge arises from the fact that the MKS framework uses a voxelized description of the microstructures, where the spatial domain is uniformly tessellated. However, upon deformation, the uniformly tessellated microstructures become distorted (i.e., the deformed microstructure is presented on a non-uniform grid). The second challenge comes from the need to explore a vast input domain (i.e., the complete space of the initial microstructures input to the simulation of a time step of the microstructure evolution) spanning a large number of morphologies and textures. Additionally, these input microstructures can be subjected to an arbitrary macroscopically imposed stretching tensor in the time step being simulated. Since computationally expensive CPFEM simulations need to be performed to generate the training data required for establishing the desired P-S linkages of interest, it is essential to devise a computational scheme that efficiently explores such expansive input domains while minimizing the size of training data needed. The challenges mentioned above are addressed in this work through suitable extensions to the MKS framework. Specifically, by taking advantage of the inherent periodicity of the CPFEM simulated RVEs, an efficient computational scheme is devised to map the deformed polycrystalline microstructures (presented on a non-uniform spatial grid) onto a uniform spatial grid. The second challenge described above is addressed by designing and implementing a Bayesian design of experiment (DOE) strategy that utilizes the uncertainty quantification implicit in the GPAP. The Bayesian DOE strategy sequentially updates the training set until the prediction uncertainty over the entire input domain of interest is reduced to an acceptable level. This study establishes that the GPAP-MKS framework can be effectively combined with the novel computational protocols described above (i.e., mapping fields from a nonuniform spatial grids to a uniform spatial grid, Bayesian DOE) to establish a high fidelity reduced-order microstructure evolution model for predicting the temporal evolution of face-centered cubic (FCC) polycrystalline microstructures subjected to an arbitrary macroscopically imposed stretching tensor. Most importantly, it is demonstrated that the use of the Bayesian DOE strategy dramatically reduces the need for large training datasets.

2. Background

2.1. Feature engineering of polycrystalline materials via MKS

The MKS framework starts with an unsupervised feature engineering of the microstructures being studied. For this purpose, we first compute translationally invariant features of the polycrystalline RVEs via 2-point spatial correlations. The 2-point spatial correlations $f(g, g' | r)$ capture the conditional probability density of finding lattice orientations g and g' located at the head and tail of a vector r randomly placed within the microstructure (Torquato and Haslach Jr, 2002; Niezgoda et al., 2011; Niezgoda et al., 2013). The digitized Fourier representation of the 2-point spatial correlations for the polycrystalline material are expressed as (Kalidindi, 2015; Paulson et al., 2017),

$$f(g, g' | r) = \sum_Q \sum_L \sum_u F_u^{LQ} T_L(g) T_Q(g') \chi_u(r) \quad (1)$$

where F_u^{LQ} are the Fourier coefficients, and $\chi_u(r)$ is a primitive basis function which equates to one if the vector r lies in the vector bin u and is zero otherwise. $T_L(g)$ represent the symmetrized GSH basis functions that already reflect the desired crystal and sample symmetries for the problem. Suitable truncation of the GSH representations and the discretized r -space (resulting from the use of voxelated microstructure domains) allows for the identification of a finite set of F_u^{LQ} for a comprehensive representation of polycrystalline microstructures capturing many of their salient features (Hashemi and Kalidindi, 2021; Paulson et al., 2017; Paulson et al., 2018). Most importantly, it has been shown (Cecen et al., 2016; Fullwood et al., 2010) that this extremely large feature vector can be computed by leveraging the fast Fourier transform (FFT) algorithm (Brigham, 1988).

As the next step, the MKS framework employs the principal component analysis (PCA) (Hastie et al., 2005) to reduce the dimensionality of the complete set of 2-point spatial correlations selected for model building. PCA rotationally transforms the original data onto the orthogonal directions of highest variance (called PC basis) ordered descendingly based on their captured variance in the

original dataset. Prior studies have shown that one can achieve reasonably accurate low-dimensional representation of highly complex polycrystalline microstructures with only a handful of PC scores (Khosravani et al., 2017; Iskakov et al., 2018; Yabansu et al., 2020; Paulson et al., 2017; Fernandez-Zelaia et al., 2019). In the present study, the PC scores representing the polycrystalline microstructures are denoted by the vector α .

2.2. Gaussian Process Autoregression Models

Gaussian Processes (GPs) can be combined with time series autoregression (AR) models; the resultant GPAR models (Requeima et al., 2019) are capable of capturing nonlinear system dynamics. Mathematically, the GPAR model is expressed as a zero-mean GP (Williams and Rasmussen, 2006):

$$\alpha_i^t \sim \mathcal{GP}(0, k(\gamma, \gamma')) \quad (2)$$

where α_i^t denotes an individual targeted PC score representing the polycrystalline microstructure at the current time t , $\gamma = \{\alpha_i^{t-1}, \alpha_i^{t-2}, \dots, \alpha_i^{t-\beta}, \theta\}$ denotes the input vector comprising normalized variables capturing the known prior time history (over the previous β timesteps, with β referred as the autoregression order) as well as the external loading parameter θ (captures details of the imposed deformation in the current time step), and k denotes the covariance kernel. The automatic relevance determination squared exponential (ARD-SE) kernel (Williams and Rasmussen, 2006) often serves as a good choice for the kernel function due to its expressivity and interpretability. The ARD-SE kernel can be mathematically expressed as (Williams and Rasmussen, 2006)

$$k(\gamma, \gamma') = \sigma_f^2 \exp \left[-\frac{1}{2} \sum_{l=1}^{\beta+1} \frac{(\gamma_l - \gamma'_l)^2}{\sigma_l^2} \right] + \sigma_n^2 \delta_{\gamma\gamma'} \quad (3)$$

where $\delta_{\gamma\gamma'}$ is the Kronecker delta. The scaling factor σ_f , the length scale parameters σ_l , and the noise factor σ_n represent hyperparameters that need to be optimized during the model building process. Relatively higher/smaller values of length scale σ_l results in smooth/noisy predictions, providing important insights into the influence of each input feature on the output predictions.

The joint distribution between the observed train set and unobserved test set is expressed by a zero-mean Gaussian process (Williams and Rasmussen, 2006) as

$$\begin{bmatrix} \alpha_i^t \\ \alpha_{i*}^t \end{bmatrix} \sim \mathcal{GP} \left(\mathbf{0}, \begin{bmatrix} K(\mathbf{A}_i, \mathbf{A}_i) & K_*(\mathbf{A}_i, \mathbf{A}_{i*}) \\ K_*^\dagger(\mathbf{A}_i, \mathbf{A}_{i*}) & K_{**}(\mathbf{A}_{i*}, \mathbf{A}_{i*}) \end{bmatrix} \right) \quad (4)$$

where $\mathbf{A}_i$ and $\mathbf{A}_{i*}$ are data matrices of inputs corresponding to the training and test sets of the i^{th} PC score, respectively, and the components of the covariance matrices are computed using the kernel described in Eq. (3). One can obtain the posterior distribution of the GPAR by conditioning the joint distribution in Eq. (4). The posterior distribution is a Gaussian, fully described by its mean μ_{i*} and covariance function Σ_{i*} (Williams and Rasmussen, 2006) computed as

$$\begin{aligned} \mu_{i*} &= K_*^\dagger K^{-1} \alpha_i^t \\ \Sigma_{i*} &= K_{**} - K_*^\dagger K^{-1} K_* \end{aligned} \quad (5)$$

Although the GPAR framework is presented above using a single input, it is easily extendable to multiple inputs. However, when GPAR models are formulated with the dependence of the output on the time histories of all the inputs, the larger data matrix sizes and need to optimize the large number of hyperparameters increase the computational cost significantly. In prior work (Hashemi and Kalidindi, 2021) using PC representations of the high-dimensional inputs, it was seen that the history dependence of the PC scores can be decoupled (possibly because of the orthogonal decomposition). This observation raises the possibility that GPAR models can be established independently for each PC score at significantly lower computational cost. In this work, we will specifically explore the feasibility of establishing such uncoupled GPAR models for the evolution of polycrystalline microstructures subjected to plastic deformation.

2.3. Bayesian DOE strategies

Typically, in problems involving vast design spaces and/or high cost of data generation it is desirable to devise a Bayesian design of experiment (DOE) strategy to minimize the number of training data points needed for the extraction of the high-fidelity surrogate model of interest. Bayesian DOE facilitates this process by (1) indicating which data point (here, a combination of polycrystalline microstructure and the arbitrary stretching tensor) should be chosen next to maximize the expected information gain, and (2) providing a stopping criterion when additional training data points are not expected to further improve the model performance.

Within the Bayesian DOE framework, an initial GP is built using a small initial training data (possibly covering the design space uniformly). Then, the GP prediction for all candidate data points targeted for training are evaluated. From the candidate set, the data point that maximizes the expected information gain on model performance is selected as the next training point. The computational expensive physics-based model is executed for the selected training point, and the results are used to update the GP model. The process is then repeated until one obtains an acceptable model performance. Prior studies have largely used either active learning Mackay

(ALM) (MacKay, 1992; Montagna and Tokdar, 2016; He, 2015; Seo et al., 2000) or active learning Cohn (ALC) (He, 2015; Seo et al., 2000; Cohn, 1996; Sauer et al., 2022) to select additional data points. While ALM selects the data point that has the greatest standard deviation in the GP prediction, ALC selects the data point that maximizes the reduction in the overall prediction variance. In this study, we used ALC since it results in global reduction in the variance of the GP predictions instead of focusing on local data points of highest variance.

3. Data generation

3.1. Synthetic polycrystalline microstructures

For the present study, a rich ensemble of synthetic FCC polycrystalline microstructures was generated to serve as the candidate set for training the desired surrogate models. The generation was designed to account for various morphologies and textures commonly encountered in cubic polycrystalline materials, and comprised of three steps:

- (i) Generation of a space-filling 3-D grain structure with targeted grain morphology (specified by an average grain size and an average aspect ratio),
- (ii) Generation of orientation distribution functions (ODFs) from various commonly encountered textures and
- (iii) Assignment of grain orientations sampled from generated ODFs to the grain structures produced in step (i).

Specifically, the open-source Python package KanaPy (Prasad et al., 2019) was used for addressing step (i). The volumes generated corresponded to a $51 \mu\text{m} \times 51 \mu\text{m} \times 51 \mu\text{m}$ cube domain with a $1 \mu\text{m} \times 1 \mu\text{m} \times 1 \mu\text{m}$ voxel size. Since there is no explicit length scale in the constitutive model used in this study, the results are independent of the voxel size. The main specifications in generation of each microstructure are grain aspect ratios parameterized as $1 : 1 : c$ with $c \in [1, 2, 3, 4, 5]$ (elongating grains in the z-direction) and the total number of grains chosen to range from 40 to 1500 grains. A total of 600 distinct grain structures were generated by uniform sampling of combinations of aspect ratio and number of grains from the ranges described above. For step (ii), ODFs are defined using weighted linear combinations of commonly encountered texture components (Suwas and Ray, 2014), which can be described using Miller indices as $\{hkl\}\langle uvw \rangle$ specifying the crystallographic planes and directions, respectively. The different texture components utilized in this study to generate different ODFs are shown in Fig. 1, and include (a) texture components typically observed in deformed samples of cubic metals: Goss $\{110\}\langle 001 \rangle$, Brass $\{110\}\langle 112 \rangle$, S $\{123\}\langle 112 \rangle$, Copper $\{112\}\langle 111 \rangle$ and Cube $\{100\}\langle 001 \rangle$; (b) rolling textures: brass (with strong Brass and weak Goss components) and copper (equal proportions of Copper, Brass and S components); (c) fiber textures: $\langle 110 \rangle // \text{RD}$ and $\langle 111 \rangle // \text{ND}$; and (d) uniform texture (Suwas and Ray, 2014; Toth et al., 1990; Venkattraman et al., 2021). Each ODF is generated via MTEX (Mainprice et al., 2015) by randomly sampling 10 different weights (w_i) for each of the aforementioned texture components while ensuring $\sum_{i=1}^{10} w_i = 1$. Following this procedure, a total of 500 distinct ODFs were generated. In step (iii), for each grain structure produced in step (i), the desired number of grain orientations are sampled from each generated ODF, and randomly assigned to the grains. The process described above produced a total of 300,000 ($= 600$ grain structures times 500 ODFs) distinct polycrystalline microstructures, serving as the candidate pool for the planned DOE efforts. This collection will be referred to as the ensemble of microstructure volume elements (MVEs) (Paulson et al., 2017; Yang et al., 2018; Yabansu et al., 2014; Latypov and Kalidindi, 2017). Prior work (Paulson et al., 2017; Yang et al., 2018; Yabansu et al., 2014; Latypov and Kalidindi, 2017) has demonstrated that it is advantageous to use the smaller MVEs (instead of RVEs) to train the surrogate models of interest. This is

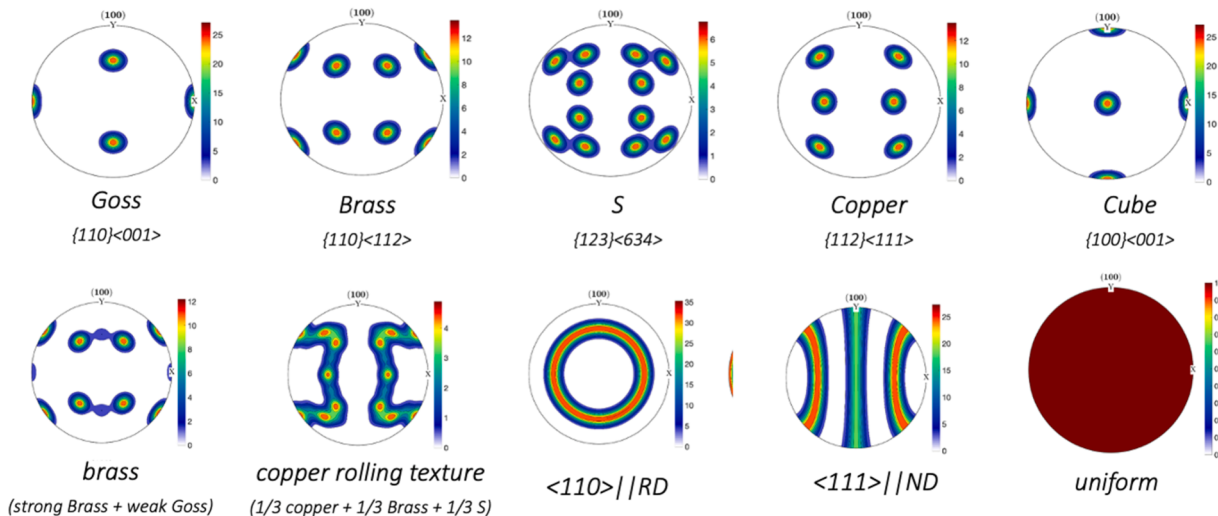


Fig. 1. Texture components utilized in the present study to generate an ensemble of ODFs.

because these generated MVEs are large enough to allow the surrogate models to learn the grain interactions within the polycrystalline aggregates, while reducing the computational cost of data generation using the computationally expensive physics-based models.

Following the MKS feature engineering framework described in Section 2.1, 2-point spatial correlations were computed for the complete set of 300,000 candidate polycrystalline MVEs. Following prior work (de Oca Zapiain and Kalidindi, 2019), a total of 5,173,428 features were computed for each microstructure. These included 39 GSH coefficients representing the orientation distribution function in the MVE (correspond to the volume averaged values of each of the 39 GSH coefficients computed for each voxel in the MVE) and 39 sets of 2-point spatial correlations (results in $39 \times 51^3 = 5,173,389$ features for each MVE; details can be found in prior work (Hashemi and Kalidindi, 2021)). PCA was performed on the full ensemble of the computed microstructural statistics. Prior to the application of PCA, each set of spatial correlations is scaled in such a way that it exhibits a unit variance across the entire dataset. This ensures all 39 sets of spatial correlations are given equal consideration in the PC representation. Incremental principal component analysis (IPCA) (Li et al., 2016) was employed in this study to avoid the memory issues arising from the large dataset of the size $300,000 \times 5,173,428$ collected for this study. IPCA (implemented in Scikit package in python programming language (Hao and Ho, 2019)) retrieves the first k principal components by considering small batches of data at a time, and sequentially updating the PC representation as more data is taken into consideration. In the present study, the GPAR-based MKS model for polycrystalline aggregates subjected to arbitrary stretching tensors is obtained using the first 10 PC scores ($k = 10$). This decision was made primarily to limit the computational cost involved. Furthermore, since the reduced-order models are being built in a completely de-coupled manner (i.e., separately for each PC score), one can augment the models produced here as needed by simply extending the protocols to the additional PC scores.

It is emphasized that PCA is performed only once in this work, and the established PC basis is used consistently throughout this work. In other words, as new microstructures are encountered (e.g., deformed microstructures are predicted by CPFEM), they are projected onto the already established PC basis. Since the candidate pool established for this study comprised a diverse set of microstructures, the established PC basis was found to be comprehensive in capturing the changes in the polycrystalline microstructures arising from the imposed plastic deformation (discussed in more detail in Section 3.3).

3.2. CPFEM simulations

All crystal plasticity simulations for this study were performed using a previously validated CPFEM model with a computationally efficient fully-implicit time-integration scheme (Kalidindi et al., 1992). Each voxel of the MVE is converted to a 8-noded 3-D solid element (C3D8) of a finite element (FE) mesh. Because of our interest in capturing the role of plastic deformation on microstructure evolution, high values of elastic stiffness parameters were assigned to the elements to limit the elastic strains to negligibly small numbers (~ 0.0001). Although, it is in principle possible to treat the evolution of the slip resistance as an additional material local state, it was decided to limit the present study to the non-hardening case for simplicity. Other parameters used in the CPFEM simulations include the reference slip rate and the slip rate sensitivity, which were set to 0.001/s and 0.01, respectively. In the CPFEM simulations, each MVE was subjected to periodic boundary conditions corresponding to a selected macroscopically imposed stretching tensor. Following strategies developed in earlier work (Van Houtte, 1994; de Oca Zapiain et al., 2017; de Oca Zapiain et al., 2022; Alharbi and Kalidindi, 2015; Kalidindi et al., 2020), the surrogate model developed in this work is formulated in the principal frame of the traceless symmetric stretching tensor, $\mathbf{D}$, which can be conveniently parametrized as

$$\mathbf{D} = \sqrt{\frac{2}{3}} D_0 \begin{bmatrix} \cos(\theta - \pi/3) & 0 & 0 \\ 0 & \cos(\theta + \pi/3) & 0 \\ 0 & 0 & -\cos(\theta) \end{bmatrix} \quad (6)$$

where $D_0 = |\mathbf{D}|$ denotes the magnitude of the macroscopically imposed strain rate, and $\theta \in [0, \frac{\pi}{3}]$ captures the deformation mode. As specific examples, $\theta = 0$ and $\theta = \frac{\pi}{6}$ correspond to uniaxial compression and plane strain compression, respectively.

For generating the training data, CPFEM simulations corresponding to a selected θ are performed on MVEs selected from the generated ensemble using a DOE strategy. The deformed microstructures predicted by CPFEM after an imposed strain of $\varepsilon = D_0 \times t = 0.02$ (i.e., an imposed equivalent strain of 0.02) were recorded as the final microstructures (this includes the spatial positions as well as crystal lattice orientations at each integration point of the deformed elements). Note that the imposed strain rate is not expected to have a significant influence on the predicted final microstructure, since the model employed in this study exhibits an almost rate-independent behavior (the rate-sensitivity parameter was assigned a value of 0.01 (Wu et al., 2008)). We therefore formulate the GPAR model in this work such that the time steps correspond to a fixed strain step of 0.02, independent of the imposed strain rate (i.e., the value of D_0). The approach outlined here therefore allows us to parameterize the deformation mode with a single variable, θ . Further, it should be noted that for a rigid-viscoplastic non-hardening crystal plasticity constitutive model employing the rate-dependent slip power law, the relationship between the imposed strain rate tensor $\mathbf{D}$ and the deviatoric stress tensor $\boldsymbol{\sigma}'$ is one-to-one. In other words, a single value of $\mathbf{D}$ corresponds to a single value of $\boldsymbol{\sigma}'$, and vice-versa. This is because the crystal plasticity relationship between these two tensors produces 5 equations with 5 unknowns with a unique solution, independent of the number of potential slip systems. Consequently, there is no need for any other history-tracking state variables. The computational simplicity of the updated Lagrangian framework described above is also the main reason for its selection. Although a total Lagrangian scheme (Kalidindi et al., 1992; Kalidindi and Anand, 1992; Kouchmeshky and Zabarar, 2009; Sundararaghavan and Kumar, 2012) offers many advantages in the conventional approaches to the crystal plasticity computations, it actually makes it much more difficult to establish the machine learning model desired here as it requires us to model the time history of much more complicated variables such as the

plastic deformation gradient tensor.

3.3. Extension of MKS feature engineering to deformed polycrystalline microstructures

The voxels of the CPFEM predicted deformed microstructures do not lie on uniform grids, inhibiting the direct use of the fast Fourier transform (FFT) algorithms (Cecen et al., 2016; Fullwood et al., 2010) in the computation of 2-point spatial correlations. Without the use of the FFT algorithms, the cost of computing the 2-point correlations for polycrystalline microstructures becomes extremely high. For this work, this challenge was addressed by devising new protocols. Following prior work (Hashemi and Kalidindi, 2021; Paulson et al., 2017), the first step in this protocol involves the representation of the material local state in each voxel by a local orientation distribution function (i.e., voxel-level ODF), expressed as GSH coefficients. Since the CPFEM predictions produce a crystal orientation at each of the integration points, the voxel-level ODF is simply defined by an average over the GSH coefficients corresponding to the orientations at each of the integration points in the deformed mesh (Bunge, 1993). Taking advantage of the fact that GSH coefficients of real-valued functions (such as the spatial correlations defined in Eq. (1)) come in complex-conjugate pairs, they are reshaped to equivalent real-valued representations (Bunge, 1993). Note that this representation is on a non-uniform spatial grid for the deformed microstructures, and needs to be re-sampled onto the uniform grids (see Fig. 2(a)-(b)).

Fig. 2(a) shows schematically a 2-dimensional section of a 3-D deformed microstructure, where the grains are shown in different colors. A bounding rectangular parallelepiped box of dimensions $L_x^{Box} \times L_y^{Box} \times L_z^{Box}$ is identified based on the extremes of the nodal coordinates in the deformed microstructure $l_x \times l_y \times l_z$, padded by r_{max} in all directions (see Fig. 2(a)). r_{max} identifies the maximum vector component lengths to which the 2-point spatial correlations are to be computed. For the present study, considering that the undeformed MVEs are $51 \times 51 \times 51$ voxels, the value of r_{max} was chosen as 25 voxels to ensure adequate sampling of the vectors of the selected lengths in computing the spatial correlations (Cecen et al., 2016; Sun et al., 2017). Each dimension of the bounding box is also rounded-up to correspond to an integer number of the voxels. The bounding box is then placed around the deformed microstructure, as shown in Fig. 2(a). Note that the bounding box can be placed with any amount of gap between the deformed microstructure and the bounding box as long as the deformed microstructure is contained fully within the box. The voxels that are not associated with the deformed microstructure in Fig. 2(a) are all assigned zero values (also called padding). The padded microstructure shown in Fig. 2(a) is denoted as $\tilde{m}_s^g$. Next, we create another microstructure denoted $\hat{m}_s^g$ by extending the deformed microstructure in all directions to completely fill out completely the same bounding box $L_x^{Box} \times L_y^{Box} \times L_z^{Box}$. Note that the second microstructure has no padded voxels.

The real-valued representations of the voxel-level ODF computed on the non-uniform deformed microstructure are then re-sampled, using a nearest neighbor interpolation algorithm (Veldman, 2016), onto the uniform grids identified in Fig. 2 for both microstructures $\tilde{m}_s^g$ and $\hat{m}_s^g$. Convolving these re-sampled microstructures with each other using the FFT algorithm, and normalizing by the number of voxels in the re-mapped domain of the deformed microstructure in $\tilde{m}_s^g$ results in the efficient and correct computation of the 2-point spatial correlations up to vector $\tilde{r} = (|r_x| \leq r_{max}, |r_y| \leq r_{max}, |r_z| \leq r_{max})$ (Kalidindi, 2015; Cecen et al., 2016). The computed 2-point spatial correlations are projected onto the already established PC basis to obtain the PC scores for the deformed microstructure; these serve as the targets for the training of the GPAR model, while the PC scores of the initial microstructures and the deformation mode serve as regressors.

4. GPAR-MKS framework

As previously stated, our goal in the present study is to produce high-fidelity low-computational cost process-microstructure evolution surrogates for polycrystalline aggregates subjected to any arbitrary stretching tensor in the form of a GPAR model with autoregression order of one. In other words, these models would take as input the PC scores of a microstructure at the beginning of a strain step² (set equal to 0.02 for the present study) and the imposed deformation mode (quantified by θ), and predict the values of PC scores at the end of the strain step. A suitable workflow was designed for this training and is schematically shown in Fig. 3. The workflow is comprised of three main steps (shown using three differently colored borders in Fig. 3 for the steps involved): (1) the initial design, (2) the Bayesian DOE loop to identify the next training point to be added and the corresponding update of the GPAR model, and (3) the GPAR model validation.

4.1. Initial design

The workflow starts by selecting an initial training set of datapoints. Each datapoint is obtained by selecting a microstructure from the pool of candidate microstructures (generated in Section 3.1) and applying a strain step of 0.02 corresponding to a selected θ using the CPFEM tool described earlier. An Euclidean distance-based K-centroid clustering algorithm (Hastie et al., 2005) was used to identify the k ($= 25$ for the present study) well-separated microstructures from the full ensemble of 300,000 candidate microstructures, each of which was then subjected to six deformation modes corresponding to $\theta = \{0^\circ, 10^\circ, 20^\circ, 30^\circ, 40^\circ, 50^\circ\}$ to generate the

² Due to the fact that the constitutive model considered in the present study is almost rate-independent, it is natural to formulate the GPAR model here in terms of the imposed strain step. This is because the imposed strain step is much more relevant to tracking the microstructure evolution, compared to the time step over which the strain step is imposed.

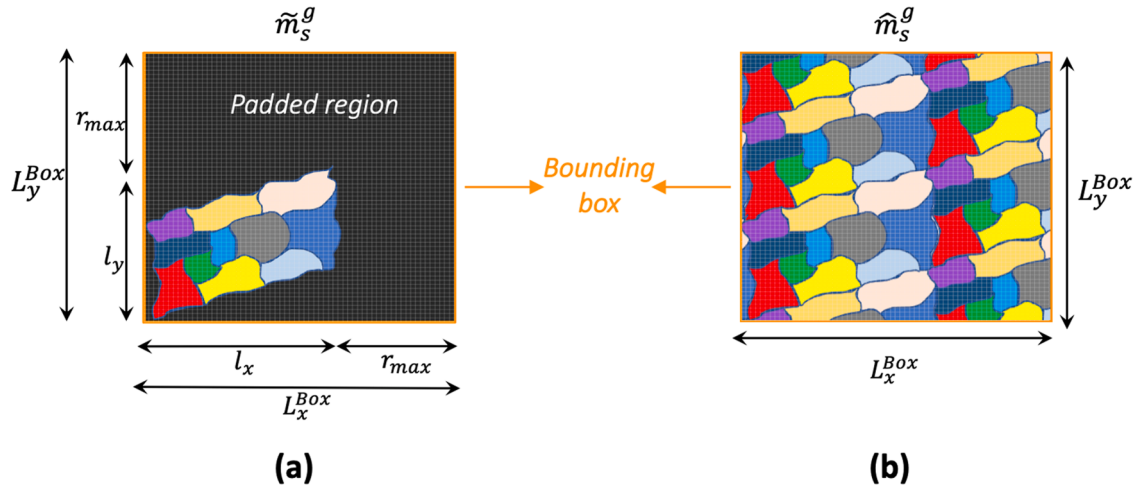


Fig. 2. Protocols used for the efficient computation of 2-point spatial correlations in the deformed microstructures. (a) A schematic of the 2-D cross-section of a 3-D deformed microstructure padded within a large bounding box and mapped to a uniform grid. (b) Extended deformed microstructure within the same bounding box mapped to a uniform grid.

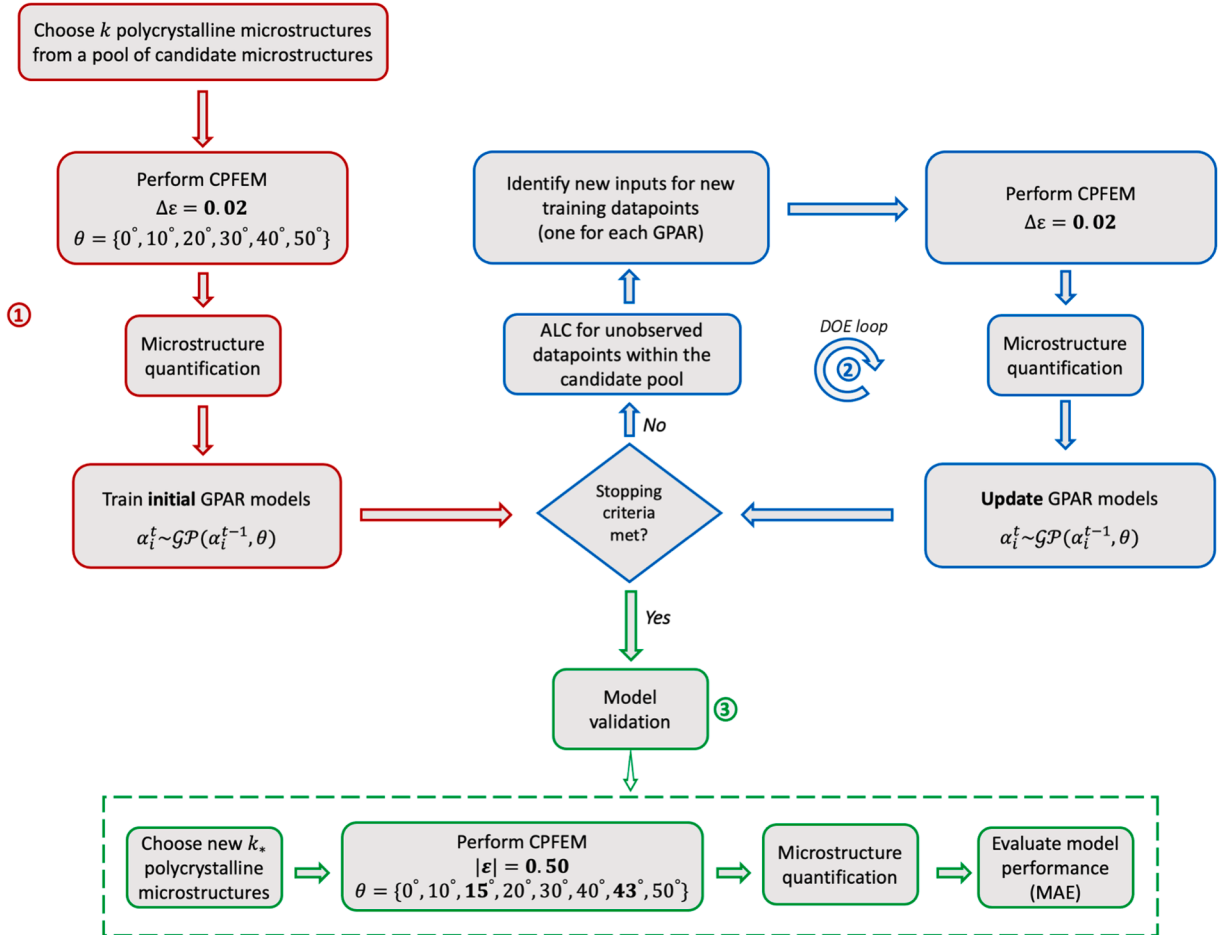


Fig. 3. Schematic description of the workflow used in this study to train the GPAR model.

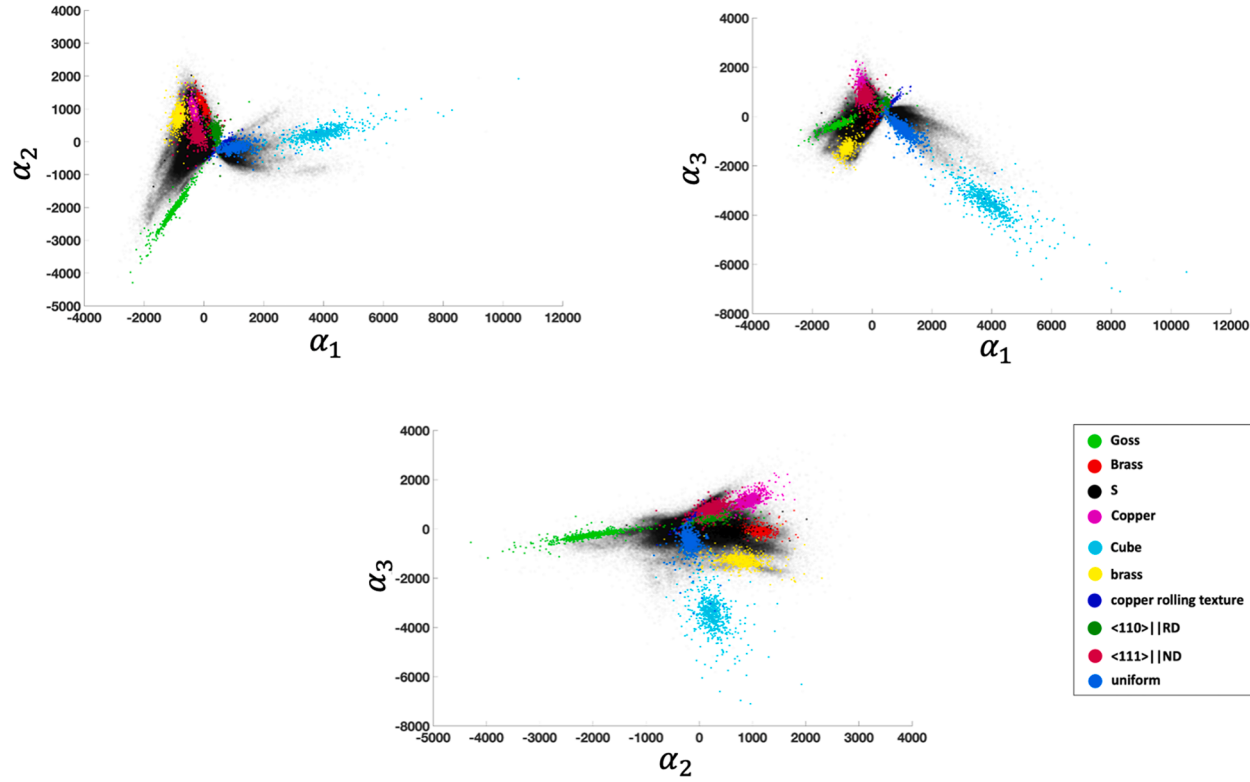


Fig. 4. Low-dimensional PC projections of the full ensemble of candidate polycrystalline microstructures considered in this study. In these plots, microstructures whose texture was predominantly from a single texture component are colored differently.

initial training set. The 150 (25 microstructures X 6 deformation modes) CPFEM predicted deformed microstructures were quantified using the extended MKS feature engineering scheme presented in Section 3.3 to produce the initial training dataset.

Initial GPAR models with autoregression order of one were trained. Following protocols developed in prior studies (Hashemi and Kalidindi, 2021), the dependence of the evolution of each PC on the other PCs was evaluated by exploring trial GPAR models. These trials revealed that the influence of the time history of a PC score on its own evolution was significantly higher than its influence on the evolution of the other PC scores. This observation confirms our expectation that the GPAR models can be developed in a fully de-coupled manner, i.e., the model input for α_i^t can be taken simply as $\gamma = \{\alpha_i^{t-1}, \theta\}$. Note that this greatly simplifies the model building effort.

4.2. Bayesian DOE loop

One loop of the Bayesian DOE entails selection of new inputs (one new input is selected for each of the 10 GPAR models, one for each PC score, being trained) where the new training points should be generated, performing the necessary CPFEM simulations, adding the new training point to the database of available training datapoints, and updating the GPAR models. A stopping criterion is implemented to decide when to terminate the loop. In the present study, the 10-fold cross-validation mean absolute error associated with each model was evaluated using the entire training dataset. These errors were then normalized by the mean absolute value of the target PC scores. The normalized cross-validation mean absolute errors (nCVMAE) were required to be lower than a specified value, chosen as 0.1 here. If this stopping criterion is not met, the GP-based uncertainty associated with the ensemble of the candidate microstructures considering all values of θ used in the training is evaluated for each trained GPAR model. Active Learning Cahn (ALC) (Cohn, 1996; Gramacy, 2016; Gramacy and Apley, 2015) is then employed to choose new inputs (microstructure and deformation mode), whose addition to the training set would result in maximum expected information gained for each GPAR model by maximizing the reduction in overall prediction uncertainties. CPFEM simulations are performed for the set of the identified microstructures and deformation modes, and the GPAR models are updated using these new data points. The DOE loop described above is repeated, until the stopping criterion is met.

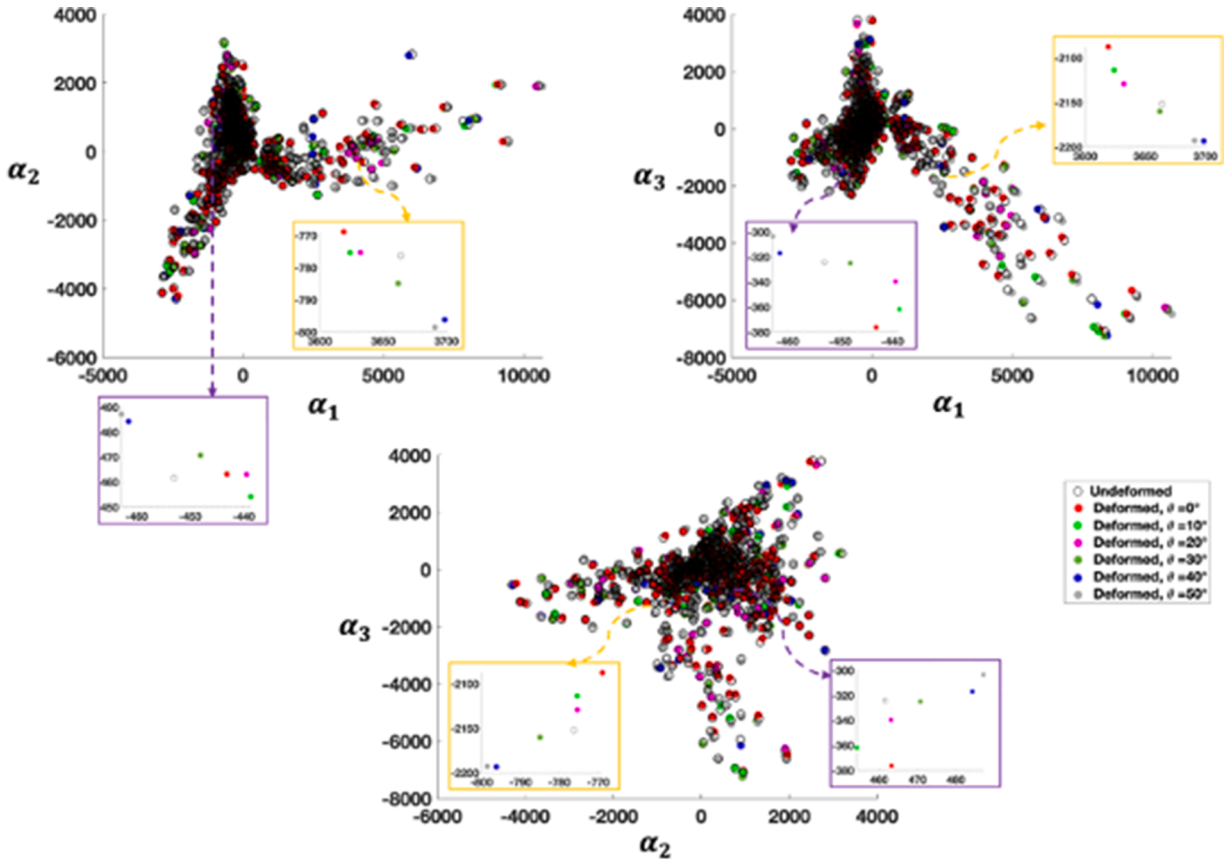


Fig. 5. The sets of undeformed and deformed microstructures generated as the training set by the DOE strategy employed in this work. The datapoints corresponding to the deformed microstructures are colored differently, based on the applied value of θ .

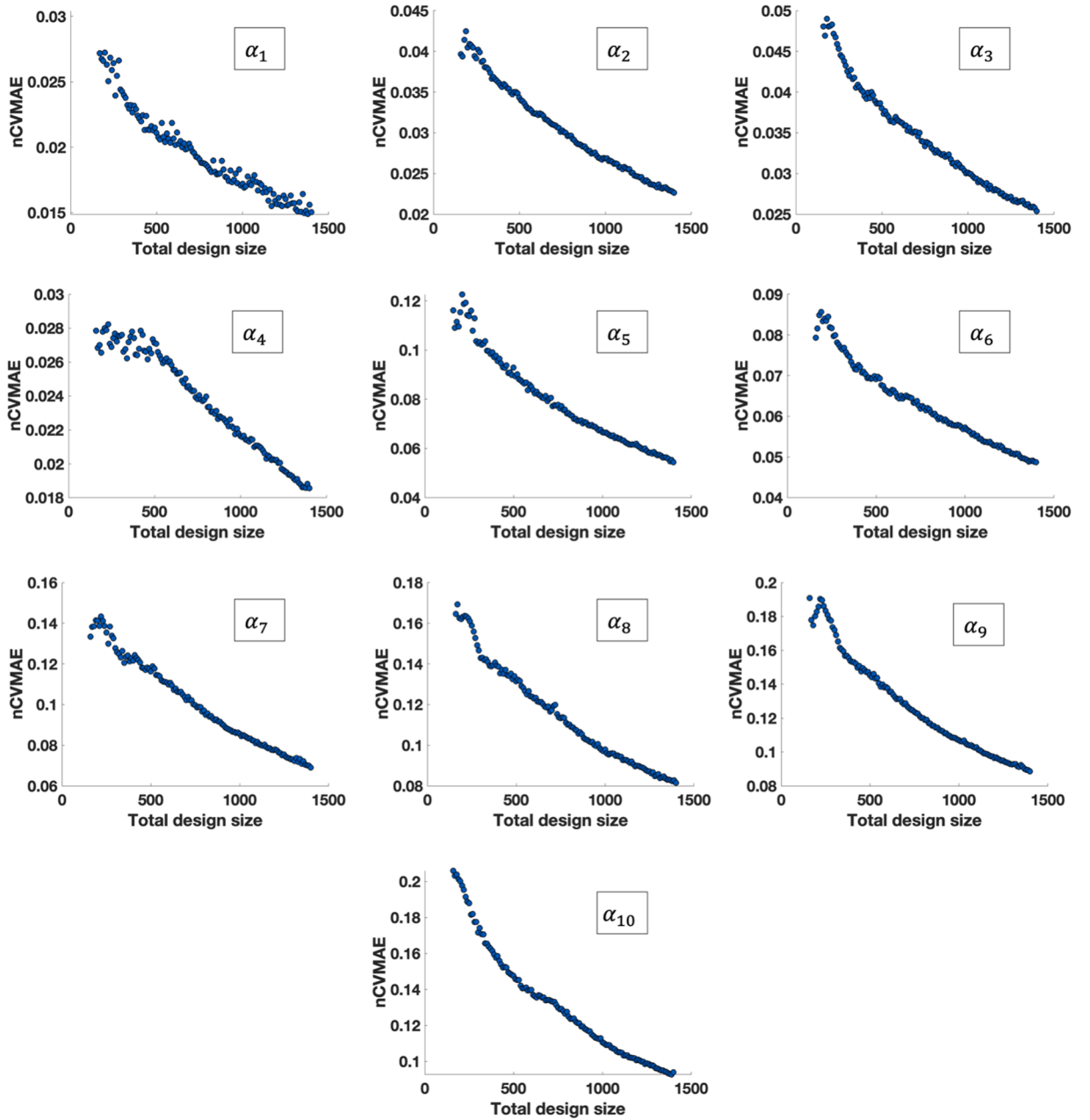


Fig. 6. The decay in the values of nCVMAE with the design size for all GPAR models built in this study.

4.3. Model validation

A test dataset comprised of $k_* = 8$ microstructures is chosen from the candidate pool of microstructures not used in the training. Each microstructure within the test set is randomly subjected to one of the six deformation modes used in the training process as well as two additional previously unseen values of $\theta = 15^\circ$ and $\theta = 43^\circ$. These new values of the deformation mode θ enable assessing the GPAR model's predictive capabilities for other loading conditions not considered in the training step. The CPFEM simulations of test cases were conducted to a total strain of $|\epsilon| = 0.50$. This allows us to evaluate the predictive capability of the GPAR to large strains through recurrent use of the model (recall that all training simulations were only performed to a total strain of 0.02). The low-dimensional representation of the CPFEM-predicted deformed microstructures are computed using the extended MKS feature engineering scheme described in Section 3.3. It should be noted that since the model is built in the principal frame of the stretching tensor, the MVE for the test case needs to be described in this frame to use the models correctly. For monotonic paths, this means that the initial MVE is defined in this frame. In other words, the crystal orientations are defined using the principal frame of the stretching tensor as

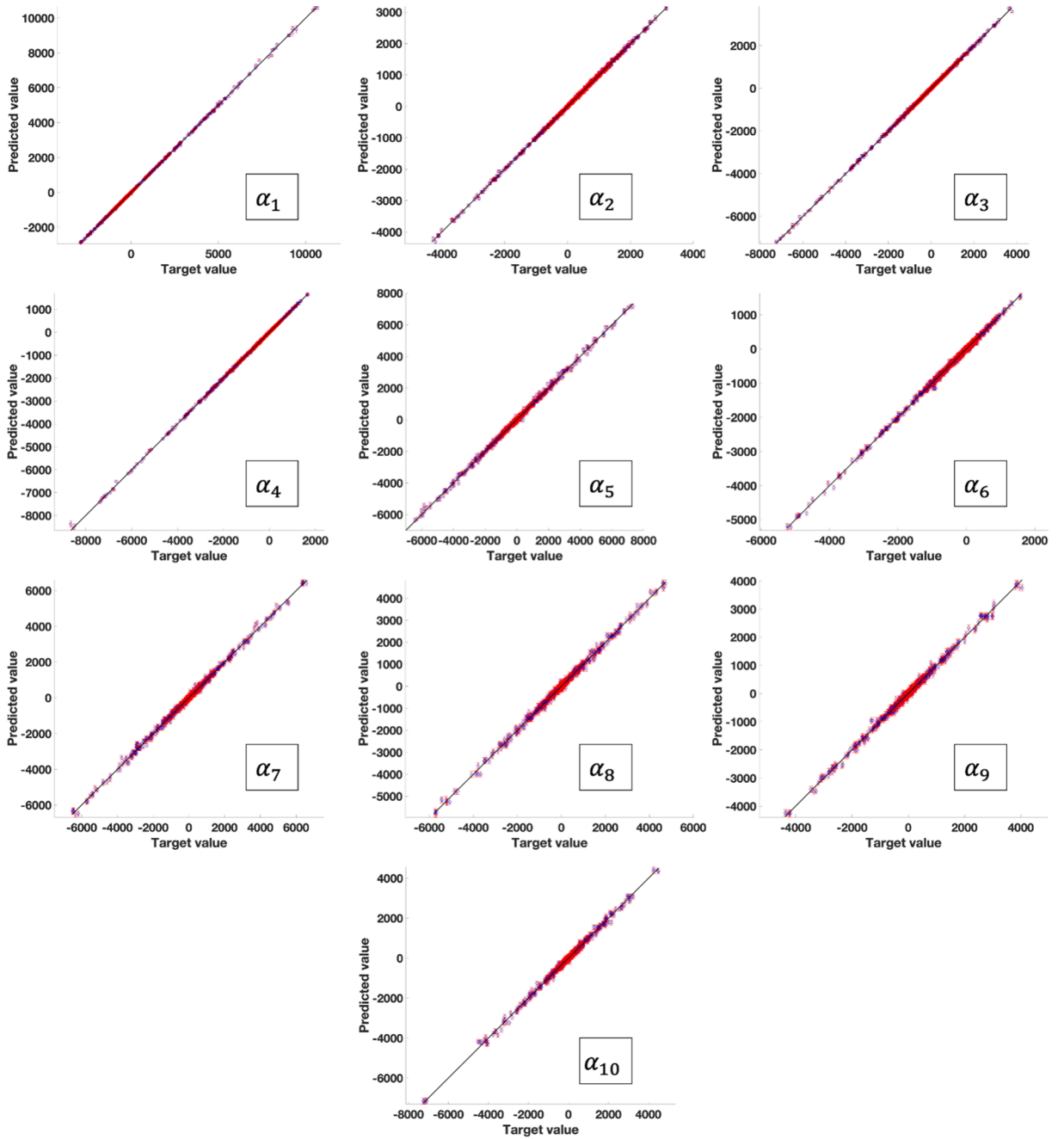


Fig. 7. Out of fold CV parity plots showing the accuracies of all ten GPAR models built in this work. The circles represent the predicted means, while the error-bars identify the one standard deviation from the mean value.

the sample reference frame; the voxelization and the spatial computations are also computed using this frame. For non-monotonic paths, this means that the spatial correlations need to be suitably rotated between time increments before applying the model recurrently. Finally, the performance of the trained models is evaluated using the mean absolute error (MAE) between the predicted and actual values of PC scores at the end of the imposed total strain of 0.5.

5. Results and Discussion

As mentioned earlier, in this study, only the first 10 PC scores are used in the low-dimensional representation of the microstructures. This was done mainly to limit the computational cost of the training involved. This truncation level resulted in a representation

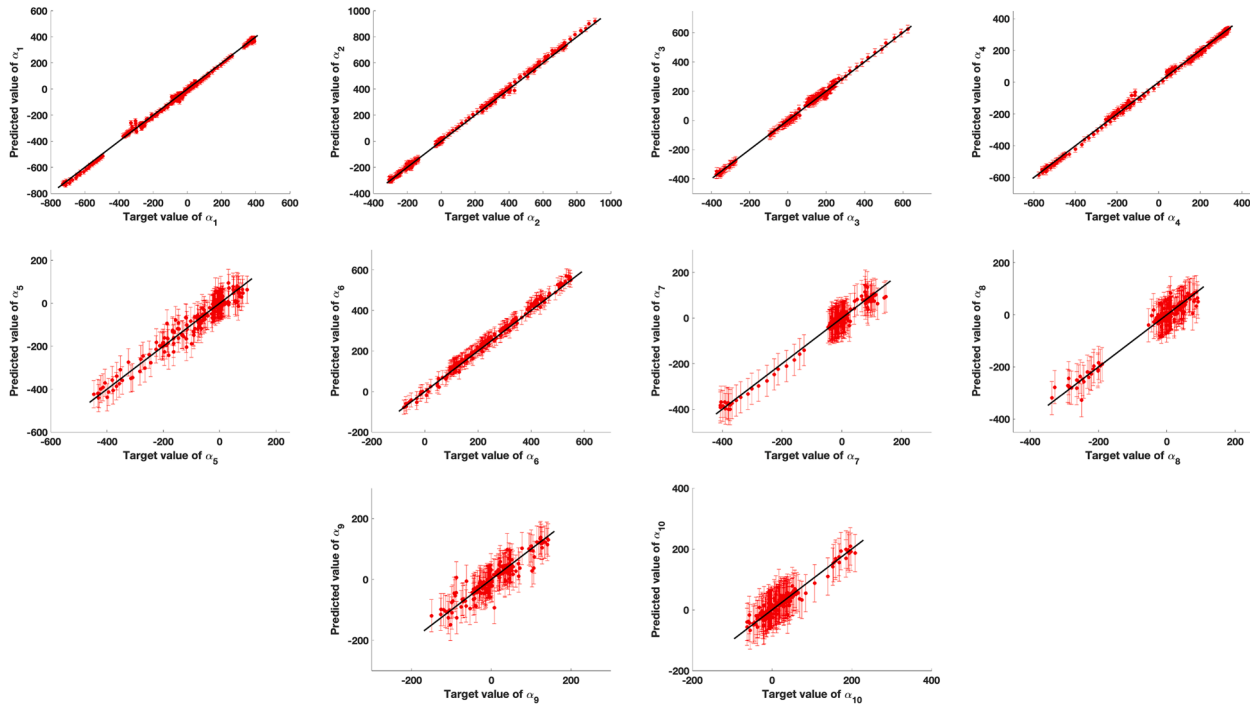


Fig. 8. Parity plots showing the accuracies of all ten GPAR models built in this work applied to unseen test data corresponding to imposed total plastic strain of $|\epsilon| = 0.5$. The circles represent the predicted means, while the error-bars identify the one standard deviation from the mean value.

error of $err = \sum \frac{|F_u^{LQ} - \tilde{F}_u^{LQ}|}{\sum |F_u^{LQ}|} = 0.3$, where $\tilde{F}_u^{LQ}$ denotes the GSH representations of the 2-point spatial correlations recovered from using 10 first PCs. Limited availability of computational resources did not allow us to consider a larger number of PC scores, which will undoubtedly reduce this representation error. Fig. 4 presents the low-dimensional representation of the entire ensemble of 300,000 polycrystalline microstructures considered in this study, projected onto the first 3 PC scores. In this figure, microstructures whose texture was predominantly from a single texture component (see Section 3.1) are colored differently. It is clearly seen that within the first 3 PCs, microstructures are automatically clustered based on their textures. The intra class variabilities are mainly attributed to differences in the microstructure morphologies (i.e., grain shape, size and their spatial distributions). Although, we are visualizing only 2-D projections of the very large PC space, it is re-assuring that there is natural (unsupervised) organization in the PC space. This is impressive because the actual microstructure space for the polycrystals, as already discussed, is extremely high-dimensional. We reiterate here our earlier argument that the unsupervised feature engineering of the MKS framework has the potential to be used with any user-specified set of targets. Essentially, the low dimensional representations obtained here are chosen to efficiently span a very large input domain.

Adopting the Bayesian DOE scheme presented in the Section 4, a total of 1400 microstructures were identified and simulated to obtain the training dataset used in this study. Fig. 5 presents low-dimensional projections of the entire training set, including both the undeformed microstructures used as inputs as well as the CPFEM-predicted deformed microstructures (to an imposed total strain of 0.02, color-coded based on the value of θ). It is observed that the deformed microstructures lie close to the initial microstructures, but do depict the expected large influence of the deformation mode, θ . These observations further confirm that the MKS feature engineering framework is producing, practically useful, low-dimensional representations of the inherently high-dimensional polycrystalline microstructures.

Fig. 6 demonstrates the steady decay in the values of nCVMAE with the addition of the training data points identified by the Bayesian DOE strategy (the total number of training data points is also referred as the design size). Note that in each Bayesian DOE loop, 10 new training datapoints are identified (i.e., the design size is increased in steps of 10 in our application). The steady decrease in nCVMAE for all ten GPARs built in this work indicates strongly that the DOE strategy employed in this study is working well for our application. In the present study, it is observed that after 1400 training data points, the nCVMAE values corresponding to all PCs are well below the stipulated accuracy of 0.1. Thus, the Bayesian DOE loop was terminated and the final GPAR models were established using the complete set of the identified 1400 training datapoints. Fig. 7 demonstrates a parity plot showing a comparison between the out of fold predictions in a 10-fold cross-validation performed for each trained GPAR model and the target values. It is seen that leveraging the devised DOE scheme, with only 1400 data points, we have obtained successfully a reasonably accurate and robust set of GPAR models.

Next, we evaluate the accuracy of the GPAR models established above for their ability to march forward in time and predict the complete trajectory of unseen microstructures subjected to a larger imposed strain of 0.5. Given only the PCs of initial microstructures as well as the imposed deformation modes, the trained GPAR models were used recursively to predict the evolution of each PC score to the imposed large strains. Fig. 8 demonstrates a parity plot serving as a comparison between the predicted values and the target values of PC scores of the microstructures within the test set (recall this set comprised of 8 microstructures each subjected to a distinct deformation mode). The normalized MAE across all PCs at the end of the last strain step was found to be smaller than 3.5%. The GP-based uncertainties with respect to model predictions are found to be within 4% of the mean absolute target value. A summary of the normalized mean absolute error for out of fold cross validations (nCVMAE) as well as the test normalized mean absolute error (nMAE) are shown in Table 1. Although the parity plot showed in Fig. 8 provides valuable insight concerning the model prediction accuracy and its associated uncertainty, it is incapable of demonstrating the evolutionary path of the PC score. Fig. 9 shows the trajectory of two selected PC scores of all test microstructures, each colored differently based on their associated applied value of the deformation mode θ . It is observed the GPAR predictions closely followed the CPFEM predicted microstructure evolution trajectories, even for unseen values of θ .

As a final confirmation of the accuracy of the GPAR models built in this work, Fig. 10 compares 2-D cross-sections of the reconstructed 2-point spatial autocorrelations from the actual (target) and predicted PC scores for a selected test microstructure subjected to the unseen value of $\theta = 15^\circ$. These cross-sections are depicted for 3 different values of imposed total strain of 0.02 (associated with the training data), 0.24 and 0.50. It is demonstrated that microstructural statistics obtained from the GPAR predictions are quite close to the target values. In fact, the mean absolute error between the reconstructed target and predicted 2-point spatial autocorrelations is within 1.25% of the mean target value. Similar observations were made for the remaining sets of 2-point spatial correlations.

In this paper, we have treated the 2-point spatial correlations of the microstructure as the essential quantitative representation of the microstructure, as opposed to the more commonly employed 3D volumes (such as those used as inputs or obtained as outputs in CPFEM simulations). This is mainly because we treat any specific 3D volume as only one of many possible instantiations of the microstructure in a given physical sample. In other words, we subscribe to the paradigm that a statistically homogeneous microstructure in a given physical sample is most accurately characterized by the n-point spatial correlations (Torquato and Haslach Jr,

Table 1

The mean absolute cross-validation and test errors for each of the ten independent GPAR models built in this work.

PC score	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9	α_{10}
out of fold nCVMAE	0.015	0.023	0.025	0.018	0.055	0.048	0.069	0.081	0.089	0.091
Test nMAE	0.014	0.015	0.014	0.013	0.026	0.027	0.018	0.022	0.033	0.025

2002; Niezgoda et al., 2011; Niezgoda et al., 2013), and one can sample a large number of stochastic instantiations (3D volumes) for any specified incomplete set of spatial correlations (Niezgoda et al., 2011; Niezgoda et al., 2013; Robertson and Kalidindi, 2022). The set of stochastic instantiations can then inform the uncertainty in the predicted properties associated with the incomplete set of spatial correlations. This fundamental hypothesis is supported by advanced composite theories (Kalidindi et al., 2006; Milhans et al., 2011; Nemat-Nasser and Hori, 2013; Matous et al., 2017) as well as our prior work (Paulson et al., 2017; Paulson et al., 2018; Paulson et al., 2019) where the 2-point spatial correlations were successfully correlated with effective properties of the polycrystalline aggregates. It is acknowledged that the mapping of the specified spatial correlations to stochastic 3D instantiations can be quite challenging, especially for polycrystalline microstructures.

As another limitation of the current work, we acknowledge that future work should consider a larger number of PCs and an even richer set of candidate microstructures. Note specifically that the use of 10 PCs in the current work resulted in a representation error of ~ 0.3 . Similarly, there is a need to further extend the candidate set of microstructures to additional textures and grain morphologies. Although this would incur a very high one-time computational expense (mainly for carrying out PCA), establishing such a large library of 3D microstructures and their PC representations would serve as a valuable resource for future efforts. Because the PC representations are obtained in unsupervised learning, such candidate sets can be used for any property of interest associated with the candidate microstructures (not restricted to the plastic properties considered in this work).

As a final discussion point, it is emphasized that the reduced-order GPAR models built in this work will significantly reduce the computational cost involved in polycrystalline microstructure or process design efforts. On an average, each simulation of the training set corresponding to an imposed total strain of 0.02 took ~ 50 minutes using 16 CPU cores and one V100 NVIDIA GPU core on Georgia Tech's HIVE computing clusters. On the other hand, the average run time of each test simulation corresponding to an imposed total strain of 0.50 was ~ 26 hours using 16 CPU core and four V100 NVIDIA GPU cores on Georgia Tech's HIVE computing clusters. The GPAR models trained in this study took less than 5 seconds on a single core 2.8 GHz processor with 16 GB RAM desktop computer to predict the complete trajectory of microstructure evolution for new (not seen in the training set) microstructures subjected to an imposed total strain of 0.50. It is therefore clear that the high one-time computational cost involved in generating the GPAR models covering an extremely large space of microstructures and deformation modes will be very valuable to future microstructure and process design efforts.

6. Conclusions

A novel framework and associated computational strategies have been devised and demonstrated for establishing a high-fidelity reduced-order model capable of predicting the temporal evolution of an extremely broad range of single-phase polycrystalline FCC microstructures undergoing any arbitrary plastic stretching tensor. These new protocols leveraged the previously established MKS feature engineering schemes and extended their application to deformed polycrystalline microstructures described on a non-uniform spatial grid (as typically output from a CPFEM simulation). Most importantly, the proposed protocols successfully implemented a Bayesian design of experiment (DOE) strategy scheme to optimally generate the computationally expensive training data set needed. This was critical for the success of the effort described in this work because of the extremely large input domain involved, which in itself represents a product space of two large spaces (the polycrystalline microstructure space and the deformation mode space). Combining all of the elements described above with a Gaussian process autoregression (GPAR) model building strategy, wherein the evolution of the individual microstructure features (PC scores) was decoupled, allowed for the successful formulation of the desired reduced-order models. The modeling strategies described in this work open new avenues of research because of their broad

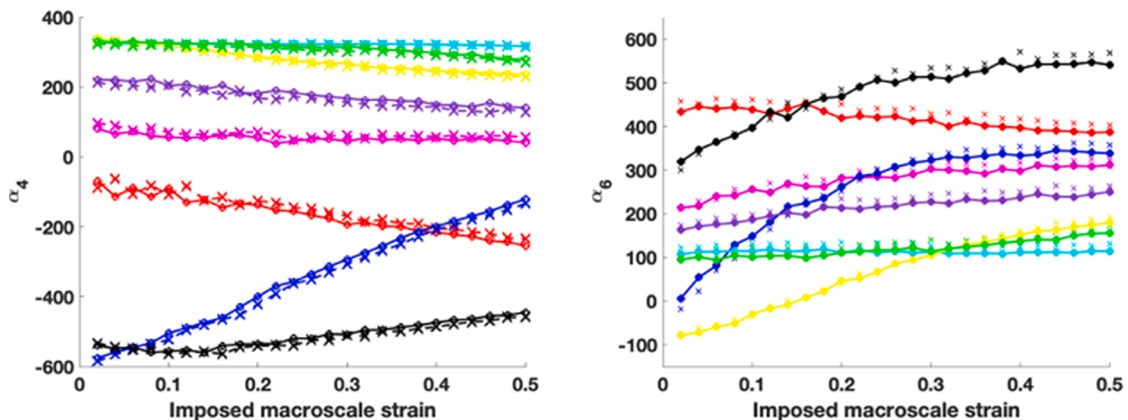


Fig. 9. The predictions of two selected GPAR models built in this work. The circles/solid curves and the cross/dashed curve represent the actual and predicted trajectories, respectively. The GPAR predictions are seen to be reasonably accurate for the entire time evolution for all the test microstructures.

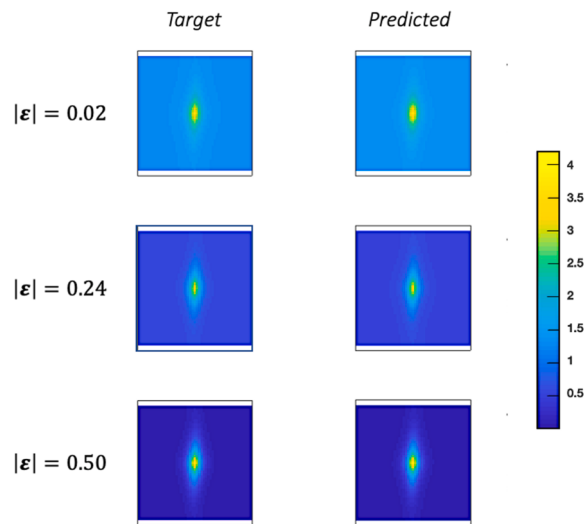


Fig. 10. A comparison between the reconstructed 2-point spatial correlations from the actual and predicted PC scores for a selected microstructure from the test set ($\theta = 15^\circ$).

applicability to similarly intractable problems because of their exceedingly large input domains and output ranges.

CRediT authorship contribution statement

Sepideh Hashemi: Conceptualization, Methodology, Data curation, Formal analysis, Writing – original draft. **Surya R. Kalidindi:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Surya R. Kalidindi reports financial support was provided by National Science Foundation.

Data availability

Data will be made available on request.

Acknowledgement

The authors acknowledge funding from NSF 2027105. This work utilized the Hive computing cluster supported by NSF 1828187 and managed by PACE at Georgia Institute of Technology, United States.

References

- Acar, P., 2020. Machine learning reinforced crystal plasticity modeling under experimental uncertainty. *AIAA J.* 58 (8), 3569–3576.
- Alharbi, H.F., Kalidindi, S.R., 2015. Crystal plasticity finite element simulations using a database of discrete Fourier transforms. *Int. J. Plast.* 66, 71–84.
- Ali, U., Odoh, D., Muhammad, W., Brahme, A., Mishra, R.K., Wells, M., Inal, K., 2017. Experimental investigation and through process crystal plasticity-static recrystallization modeling of temperature and strain rate effects during hot compression of AA6063. *Mater. Sci. Eng. A Struct. Mater.* 700, 374–386.
- Ali, U., Muhammad, W., Brahme, A., Skiba, O., Inal, K., 2019. Application of artificial neural networks in micromechanics for polycrystalline metals. *Int. J. Plast.* 120, 205–219.
- Ali, U., 2021. Artificial Neural Network Based Crystal Plasticity Framework to Simulate Flow Behavior In Aluminum Alloys. Charlottetown, PE, Canada.
- Bonatti, C., Berisha, B., Mohr, D., 2022. From CP-FFT to CP-RNN: recurrent neural network surrogate model of crystal plasticity. *Int. J. Plast.*, 103430.
- G.E. Box, G.M. Jenkins, G.C. Reinsel, G.M. Ljung, *Time Series Analysis: Forecasting and Control*, John Wiley & Sons 2015.
- Brigham, E.O., 1988. *The Fast Fourier Transform and its Applications*. Prentice-Hall, Inc.
- Brough, D.B., Kannan, A., Haaland, B., Bucknall, D.G., Kalidindi, S.R., 2017. Extraction of process-structure evolution linkages from x-ray scattering measurements using dimensionality reduction and time series analysis. *Integr. Mater. Manuf. Innov.* 6 (2), 147–159.
- Bunge, H.-J., 1993. *Texture Analysis in Materials Science: Mathematical Methods*. Elsevier.
- Cecen, A., Fast, T., Kalidindi, S.R., 2016. Versatile algorithms for the computation of 2-point spatial correlations in quantifying material structure. *Integr. Mater. Manuf. Innov.* 5 (1), 1–15.

- Chen, L., Xiao, X., Yu, L., Chu, H., Duan, H., 2018. Texture evolution and mechanical behaviour of irradiated face-centred cubic metals. *Proc. R. Soc. A* 474 (2210), 20170604.
- Chen, G.-y., Gan, M., Chen, G.-l., 2018. Generalized exponential autoregressive models for nonlinear time series: stationarity, estimation and applications. *Inf. Sci.* 438, 46–57.
- Cohn, D.A., 1996. Neural network exploration using optimal experiment design. *Neural Netw.* 9 (6), 1071–1083.
- Dai, W., Wang, H., Guan, Q., Li, D., Peng, Y., Tomé, C.N., 2021. Studying the micromechanical behaviors of a polycrystalline metal by artificial neural networks. *Acta Mater.* 214, 117006.
- de Oca Zapiain, D.M., Kalidindi, S.R., 2019. Localization models for the plastic response of polycrystalline materials using the material knowledge systems framework. *Modell. Simul. Mater. Sci. Eng. R. Rep.* 27 (7), 074008.
- de Oca Zapiain, D.M., Popova, E., Kalidindi, S.R., 2017. Prediction of microscale plastic strain rate fields in two-phase composites subjected to an arbitrary macroscale strain rate using the materials knowledge system framework. *Acta Mater.* 141, 230–240.
- de Oca Zapiain, D.M., Shanker, A., Kalidindi, S.R., 2022. Convolutional neural networks for the localization of plastic velocity gradient tensor in polycrystalline microstructures. *J. Eng. Mater. Technol.* 144 (1).
- Farooq, H., Cailletaud, G., Forest, S., Ryckelynck, D., 2020. Crystal plasticity modeling of the cyclic behavior of polycrystalline aggregates under non-symmetric uniaxial loading: global and local analyses. *Int. J. Plast.* 126, 102619.
- Fernandez-Zelaia, P., Melkote, S.N., 2019. Process-structure-property modeling for severe plastic deformation processes using orientation imaging microscopy and data-driven techniques. *Integrat. Mater. Manuf. Innovat.* 1–20.
- Fernandez-Zelaia, P., Melkote, S.N., 2019. Process–structure–property relationships in bimodal machined microstructures using robust structure descriptors. *J. Mater. Process. Technol.* 273, 116251.
- Fernandez-Zelaia, P., Yabansu, Y.C., Kalidindi, S.R., 2019. A comparative study of the efficacy of local/global and parametric/nonparametric machine learning methods for establishing structure–property linkages in high-contrast 3D elastic composites. *Integr. Mater. Manuf. Innov.* 8 (2), 67–81.
- Fullwood, D.T., Niezgoda, S.R., Adams, B.L., Kalidindi, S.R., 2010. Microstructure sensitive design for performance optimization. *Prog. Mater. Sci.* 55 (6), 477–562.
- Ghosh, S., Dimiduk, D., 2011. *Computational Methods for Microstructure-Property Relationships*. Springer.
- Gorji, M.B., Mozaffar, M., Heidenreich, J.N., Cao, J., Mohr, D., 2020. On the potential of recurrent neural networks for modeling path dependent plasticity. *J. Mech. Phys. Solids* 143, 103972.
- Gramacy, R.B., Apley, D.W., 2015. Local Gaussian process approximation for large computer experiments. *J. Comput. Graph. Statist.* 24 (2), 561–578.
- Gramacy, R.B., 2016. laGP: large-scale spatial modeling via local approximate Gaussian processes in R. *J. Stat. Softw.* 72, 1–46.
- Gupta, A., Bettaieb, M.B., Abed-Meraim, F., Kalidindi, S.R., 2018. Computationally efficient predictions of crystal plasticity based forming limit diagrams using a spectral database. *Int. J. Plast.* 103, 168–187.
- Hao, J., Ho, T.K., 2019. Machine learning made easy: a review of scikit-learn package in python programming language. *J. Educ. Behav. Stat.* 44 (3), 348–361.
- Hashemi, S., Kalidindi, S.R., 2021. A machine learning framework for the temporal evolution of microstructure during static recrystallization of polycrystalline materials simulated by cellular automaton. *Comput. Mater. Sci.* 188, 110132.
- Hastie, T., Tibshirani, R., Friedman, J., Franklin, J., 2005. The elements of statistical learning: data mining, inference and prediction. *Math. Intell.* 27 (2), 83–85.
- He, Y., 2015. Online detection and modeling of safety boundaries for aerospace applications using active learning and bayesian statistics. In: 2015 International Joint Conference on Neural Networks (IJCNN). IEEE, pp. 1–8.
- Ibragimova, O., Brahme, A., Muhammad, W., Lévesque, J., Inal, K., 2021. A new ANN based crystal plasticity model for FCC materials and its application to non-monotonic strain paths. *Int. J. Plast.* 144, 103059.
- Ibragimova, O., Brahme, A., Muhammad, W., Connolly, D., Lévesque, J., Inal, K., 2022. A convolutional neural network based crystal plasticity finite element framework to predict localised deformation in metals. *Int. J. Plast.* 157, 103374.
- Ifthikhar, C.M.A., Li, Y.L., Kohar, C.P., Inal, K., Khan, A.S., 2021. Evolution of subsequent yield surfaces with plastic deformation along proportional and non-proportional loading paths on annealed AA6061 alloy: experiments and crystal plasticity finite element modeling. *Int. J. Plast.* 143, 102956.
- Iskakov, A., Yabansu, Y.C., Rajagopalan, S., Kapustina, A., Kalidindi, S.R., 2018. Application of spherical indentation and the materials knowledge system framework to establishing microstructure–yield strength linkages from carbon steel scoops excised from high-temperature exposed components. *Acta Mater.* 144, 758–767.
- Jones, D.R., Ashby, M.F., 2012. *Engineering Materials 2: An Introduction to Microstructures and Processing*. Butterworth-Heinemann.
- Kalidindi, S.R., Anand, L., 1992. An approximate procedure for predicting the evolution of crystallographic texture in bulk deformation processing of fcc metals. *Int. J. Mech. Sci.* 34 (4), 309–329.
- Kalidindi, S.R., Bronkhorst, C.A., Anand, L., 1992. Crystallographic texture evolution in bulk deformation processing of FCC metals. *J. Mech. Phys. Solids* 40 (3), 537–569.
- Kalidindi, S.R., Binci, M., Fullwood, D., Adams, B.L., 2006. Elastic properties closures using second-order homogenization theories: case studies in composites of two isotropic constituents. *Acta Mater.* 54 (11), 3117–3126.
- Kalidindi, S.R., Knezevic, M., Niezgoda, S., Shaffer, J., 2009. Representation of the orientation distribution function and computation of first-order elastic properties closures using discrete Fourier transforms. *Acta Mater.* 57 (13), 3916–3923.
- Kalidindi, S.R., Gupta, A., Popova, E., 2020. Computationally efficient crystal plasticity simulations using spectral databases. *Handbook of Materials Modeling: Methods: Theory and Modeling*, pp. 1685–1710.
- Kalidindi, S.R., 2015. *Hierarchical Materials Informatics: Novel Analytics for Materials Data*. Elsevier.
- Kalidindi, S.R., 2019. A Bayesian framework for materials knowledge systems. *MRS Commun.* 9 (2), 518–531.
- Khan, A.S., Liu, J., Yoon, J.W., Nambori, R., 2015. Strain rate effect of high purity aluminum single crystals: experiments and simulations. *Int. J. Plast.* 67, 39–52.
- Khandelwal, S., Basu, S., Patra, A., 2021. A machine learning-based surrogate modeling framework for predicting the history-dependent deformation of dual phase microstructures. *Mater. Today Commun.* 29, 102914.
- Khosravani, A., Cecen, A., Kalidindi, S.R., 2017. Development of high throughput assays for establishing process-structure-property linkages in multiphase polycrystalline metals: application to dual-phase steels. *Acta Mater.* 123, 55–69.
- Knezevic, M., Bhattacharyya, A., 2017. Characterization of microstructure in Nb rods processed by rolling: effect of grooved rolling die geometry on texture uniformity. *Int. J. Refract. Met. Hard Mater.* 66, 44–51.
- Knezevic, M., Kalidindi, S.R., 2007. Fast computation of first-order elastic–plastic closures for polycrystalline cubic-orthorhombic microstructures. *Comput. Mater. Sci.* 39 (3), 643–648.
- Knezevic, M., Drach, B., Ardeljan, M., Beyerlein, I.J., 2014. Three dimensional predictions of grain scale plasticity and grain boundaries using crystal plasticity finite element models. *Comput. Meth. Appl. Mech. Eng.* 277, 239–259.
- Kouchmeshky, B., Zabarás, N., 2009. Modeling the response of HCP polycrystals deforming by slip and twinning using a finite element representation of the orientation space. *Comput. Mater. Sci.* 45 (4), 1043–1051.
- Latypov, M.I., Kalidindi, S.R., 2017. Data-driven reduced order models for effective yield strength and partitioning of strain in multiphase materials. *J. Comput. Phys.* 346, 242–261.
- Latypov, M.I., Toth, L.S., Kalidindi, S.R., 2019. Materials knowledge system for nonlinear composites. *Comput. Meth. Appl. Mech. Eng.* 346, 180–196.
- Li, L., Liu, S., Peng, Y., Sun, Z., 2016. Overview of principal component analysis algorithm. *Optik* 127 (9), 3935–3944.
- Luccarelli, P.G., Pataky, G., Sehitoglu, H., Foletti, S., 2017. Finite element simulation of single crystal and polycrystalline Haynes 230 specimens. *Int. J. Solids Struct.* 115, 270–278.
- MacKay, D.J., 1992. Information-based objective functions for active data selection. *Neural Comput.* 4 (4), 590–604.
- Mainprize, D., Bachmann, F., Hielscher, R., Schaeben, H., 2015. Descriptive tools for the analysis of texture projects with large datasets using MTEX: strength, symmetry and components. *Geol. Soc. Lond. Spec. Publ.* 409 (1), 251–271.
- Mangal, A., Holm, E.A., 2018. Applied machine learning to predict stress hotspots I: face centered cubic materials. *Int. J. Plast.* 111, 122–134.

- Matouš, K., Geers, M.G., Kouznetsova, V.G., Gillman, A., 2017. A review of predictive nonlinear theories for multiscale modeling of heterogeneous materials. *J. Comput. Phys.* 330, 192–220.
- McDowell, D.L., Olson, G.B., 2008. Concurrent Design of Hierarchical Materials and Structures, Scientific Modeling and Simulations. Springer, pp. 207–240.
- McDowell, D.L., 2018. Microstructure-Sensitive Computational Structure-Property Relations in Materials design, Computational Materials System Design. Springer, pp. 1–25.
- Milhans, J., Li, D., Khaleel, M., Sun, X., Garmestani, H., 2011. Prediction of the effective coefficient of thermal expansion of heterogeneous media using two-point correlation functions. *J. Power Sources* 196 (8), 3846–3850.
- Montagna, S., Tokdar, S.T., 2016. Computer emulation with nonstationary Gaussian processes. *SIAM/ASA J. Uncertain. Quantificat.* 4 (1), 26–47.
- Muhammad, W., Brahme, A.P., Ibragimova, O., Kang, J., Inal, K., 2021. A machine learning framework to predict local strain distribution and the evolution of plastic anisotropy & fracture in additively manufactured alloys. *Int. J. Plast.* 136, 102867.
- Nemat-Nasser, S., Hori, M., 2013. Micromechanics: Overall Properties of Heterogeneous Materials. Elsevier.
- Niezgoda, S.R., Yabansu, Y.C., Kalidindi, S.R., 2011. Understanding and visualizing microstructure and microstructure variance as a stochastic process. *Acta Mater.* 59 (16), 6387–6400.
- Niezgoda, S.R., Kanjarla, A.K., Kalidindi, S.R., 2013. Novel microstructure quantification framework for databasing, visualization, and analysis of microstructure data. *Integr. Mater. Manuf. Innov.* 2 (1), 54–80.
- Paulson, N.H., Priddy, M.W., McDowell, D.L., Kalidindi, S.R., 2017. Reduced-order structure-property linkages for polycrystalline microstructures based on 2-point statistics. *Acta Mater.* 129, 428–438.
- Paulson, N.H., Priddy, M.W., McDowell, D.L., Kalidindi, S.R., 2018. Data-driven reduced-order models for rank-ordering the high cycle fatigue performance of polycrystalline microstructures. *Mater. Des.* 154, 170–183.
- Paulson, N.H., Priddy, M.W., McDowell, D.L., Kalidindi, S.R., 2019. Reduced-order microstructure-sensitive protocols to rank-order the transition fatigue resistance of polycrystalline microstructures. *Int. J. Fatigue* 119, 1–10.
- Prasad, M.R., Vajragupta, N., Hartmaier, A., 2019. Kanapy: A Python package for generating complex synthetic polycrystalline microstructures. *J. Open Source Softw.* 4 (43), 1732.
- Priddy, M.W., Paulson, N.H., Kalidindi, S.R., McDowell, D.L., 2017. Strategies for rapid parametric assessment of microstructure-sensitive fatigue for HCP polycrystals. *Int. J. Fatigue* 104, 231–242.
- Ramakrishna, S., Zhang, T.-Y., Lu, W.-C., Qian, Q., Low, J.S.C., Yune, J.H.R., Tan, D.Z.L., Bressan, S., Sanvito, S., Kalidindi, S.R., 2019. Materials informatics. *J. Intell. Manuf.* 30 (6), 2307–2326.
- Requeima, J., Tebbutt, W., Bruinsma, W., Turner, R.E., 2019. The gaussian process autoregressive regression model (gpar). In: The 22nd International Conference on Artificial Intelligence and Statistics. PMLR, pp. 1860–1869.
- Robertson, A.E., Kalidindi, S.R., 2022. Efficient generation of anisotropic N-field microstructures from 2-point statistics using multi-output Gaussian random fields. *Acta Mater.* 232, 117927.
- Saidi, P., Pirgazi, H., Sanjari, M., Tamimi, S., Mohammadi, M., Bèland, L.K., Daymond, M.R., Tamlyn, I., 2022. Deep learning and crystal plasticity: a preconditioning approach for accurate orientation evolution prediction. *Comput. Meth. Appl. Mech. Eng.* 389, 114392.
- Sankaran, S., Zabar, N., 2007. Computing property variability of polycrystals induced by grain size and orientation uncertainties. *Acta Mater.* 55 (7), 2279–2290.
- Sauer, A., Gramacy, R.B., Higdon, D., 2022. Active learning for deep Gaussian process surrogates. *Technometrics* 1–15.
- Sedighiani, K., Shah, V., Traka, K., Diehl, M., Roters, F., Sietsma, J., Raabe, D., 2021. Large-deformation crystal plasticity simulation of microstructure and microtexture evolution through adaptive remeshing. *Int. J. Plast.* 146, 103078.
- Sedighiani, K., Traka, K., Roters, F., Sietsma, J., Raabe, D., Diehl, M., 2022. Crystal plasticity simulation of in-grain microstructural evolution during large deformation of IF-steel. *Acta Mater.* 237, 118167.
- Seo, S., Wallat, M., Graepel, T., Obermayer, K., 2000. Gaussian Process Regression: Active Data Selection and Test Point Rejection, Mustererkennung 2000. Springer, pp. 27–34.
- Sun, Y., Cecen, A., Gibbs, J.W., Kalidindi, S.R., Voorhees, P.W., 2017. Analytics on large microstructure datasets using two-point spatial correlations: coarsening of dendritic structures. *Acta Mater.* 132, 374–388.
- Sundararaghavan, V., Kumar, A., 2012. Probabilistic modeling of microstructure evolution using finite element representation of statistical correlation functions. *Int. J. Plast.* 30, 62–80.
- Suwas, S., Ray, R.K., 2014. Crystallographic Texture of Materials. Springer.
- Tallman, A.E., Stopka, K.S., Swiler, L.P., Wang, Y., Kalidindi, S.R., McDowell, D.L., 2019. Gaussian-process-driven adaptive sampling for reduced-order modeling of texture effects in polycrystalline α -Ti. *JOM* 71 (8), 2646–2656.
- Tasan, C.C., Hoefnagels, J.P., Diehl, M., Yan, D., Roters, F., Raabe, D., 2014. Strain localization and damage in dual phase steels investigated by coupled in-situ deformation experiments and crystal plasticity simulations. *Int. J. Plast.* 63, 198–210.
- Torquato, S., Haslach Jr, H., 2002. Random heterogeneous materials: microstructure and macroscopic properties. *Appl. Mech. Rev.* 55 (4), B62–B63.
- Toth, L., Jonas, J., Daniel, D., Ray, R., 1990. Development of ferrite rolling textures in low-and extra low-carbon steels. *Metall. Trans. A* 21 (11), 2985–3000.
- Valdman, J., 2016. Applications from Engineering with MATLAB Concepts. BoD-Books on Demand.
- Van Houtte, P., 1994. Application of plastic potentials to strain rate sensitive and insensitive anisotropic materials. *Int. J. Plast.* 10 (7), 719–748.
- Venkatraman, A., de Oca Zapiani, D.M., Lim, H., Kalidindi, S.R., 2021. Texture-sensitive prediction of micro-spring performance using Gaussian process models calibrated to finite element simulations. *Mater. Design* 197, 109198.
- Waheeb, W., Ghazali, R., Shah, H., 2019. Nonlinear Autoregressive Moving-average (NARMA) time series forecasting using neural networks. In: 2019 International Conference on Computer and Information Sciences (ICCIS). IEEE, pp. 1–5.
- Wang, H., Lu, C., Tieu, A.K., Su, L., Deng, G., 2019. Coupled effects of initial orientation scatter and grain-interaction to texture evolution: a crystal plasticity FE study. *Int. J. Mater. Form.* 12 (1), 161–171.
- Williams, C.K., Rasmussen, C.E., 2006. Gaussian Processes for Machine Learning, 2. MIT press Cambridge, MA.
- Wu, X., Kalidindi, S.R., Necker, C., Salem, A.A., 2008. Modeling anisotropic stress-strain response and crystallographic texture evolution in α -titanium during large plastic deformation using Taylor-type models: influence of initial texture and purity. *Metall. Mater. Trans. A* 39 (12), 3046–3054.
- Yabansu, Y.C., Patel, D.K., Kalidindi, S.R., 2014. Calibrated localization relationships for elastic response of polycrystalline aggregates. *Acta Mater.* 81, 151–160.
- Yabansu, Y.C., Rehn, V., Hötzer, J., Nestler, B., Kalidindi, S.R., 2019. Application of Gaussian process autoregressive models for capturing the time evolution of microstructure statistics from phase-field simulations for sintering of polycrystalline ceramics. *Modell. Simul. Mater. Sci. Eng.* 27 (8), 084006.
- Yabansu, Y.C., Altschuh, P., Hötzer, J., Selzer, M., Nestler, B., Kalidindi, S.R., 2020. A digital workflow for learning the reduced-order structure-property linkages for permeability of porous membranes. *Acta Mater.* 195, 668–680.
- Yang, Z., Yabansu, Y.C., Al-Bahrani, R., Liao, W.-k., Choudhary, A.N., Kalidindi, S.R., Agrawal, A., 2018. Deep learning approaches for mining structure-property linkages in high contrast composites from simulation datasets. *Comput. Mater. Sci.* 151, 278–287.
- Yuan, M., Paradiso, S., Meredig, B., Niezgoda, S.R., 2018. Machine learning-based reduce order crystal plasticity modeling for ICME applications. *Integr. Mater. Manuf. Innov.* 7 (4), 214–230.
- Zhou, Y., Ding, F., 2020. Modeling nonlinear processes using the radial basis function-based state-dependent autoregressive models. *IEEE Signal Process. Lett.* 27, 1600–1604.