J Mater Sci (2024) 59:4863–4881

The Physics of Metal Plasticity: in honor of Professor Hussein Zbib

A Gaussian process autoregressive model capturing microstructure evolution paths in a Ni–Mo–Nb alloy

Andrew Marshall[1], Adam Generale[1], Surya R. Kalidindi[1,*], Bala Radhakrishnan[2], and Jim Belak[3]

[1] *Georgia Institute of Technology, Atlanta, GA, USA*
[2] *Oak Ridge National Laboratory, Oak Ridge, TN, USA*
[3] *Lawrence Livermore National Laboratory, Livermore, CA, USA*

Received: 11 September 2023
Accepted: 3 January 2024
Published online: 20 February 2024

ABSTRACT

Additive manufacturing is increasingly being employed to produce components of complex geometries in structural alloys because of the expected energy savings associated with the near-net-shape capability and the ability to build in novel internal features that are not possible with many conventional manufacturing approaches. However, because of the extreme thermal conditions encountered, the non-equilibrium microstructures produced during powder bed-based additive manufacturing processes must be subjected to custom post-heat treatment processes to recover the target mechanical properties. Phase-field models and simulation techniques have matured to a state where the microstructure evolution paths, and the morphologies of the resulting precipitate phases can be predicted reasonably accurately, considering alloy-specific thermodynamic and kinetic aspects of the nucleation and growth processes. However, phase-field simulations are computationally intensive, which precludes the ability to apply the simulations directly to the length scale of the entire component. Therefore, it is highly desirable to develop low-computational-cost surrogate models that effectively capture the physics at the microstructural length scale, while facilitating the design of optimized processing conditions resulting in location-specific targeted microstructures at the component scale. The work presented here demonstrates the application of the materials knowledge system framework to develop a surrogate model that effectively captures the microstructural path during annealing of a Ni–Mo–Nb alloy containing different Mo and Nb compositions known to segregate during solidification under additive manufacturing conditions. Specifically, the surrogate model built in this work is based on a Gaussian process autoregressive model informed by statistical representation of simulated microstructures using two-point correlations and dimensionality reduction through principal component analysis. This surrogate model is shown to capture the bifurcation of the microstructural path during precipitation, which yields a microstructure

Handling Editor: M. Grant Norton.

Address correspondence to E-mail: surya.kalidindi@me.gatech.edu

https://doi.org/10.1007/s10853-024-09345-6

dominated by the γ'' phase at high Nb concentrations and the δ phase at low Nb concentrations.

Introduction

Post-processing heat treatments are integral to the production of structural alloys and are essential to attaining the desired in-service mechanical properties. It is especially critical for components made using powder bed-based additive manufacturing processes, where the rapid melting and solidification, and repeated thermal excursions in the solid-state result in non-optimum microstructures that need to be modified through customized heat treatment steps [1]. During such heat treatments, phase transformations often occur and form new precipitates of varying sizes and shapes organized in various geometrical patterns; these details are often collectively referred as microstructural details. Traditionally, physics-based modelling approaches such as phase-field methods have been employed to understand the complex microstructures resulting from such transformations [2–6]. However, these approaches are computationally expensive, limiting their ability to explore comprehensively the complete range of microstructures that could be produced through variations in the overall chemical compositions and process histories. Although simpler approaches based on thermodynamics and kinetics-based models are computationally cheap, they often fail to capture the complex microscale interactions involved in such phase transformations and the key features in the resulting microstructures.

Because of the challenges described above, it is highly desirable to mine low-computational-cost surrogate process–microstructure evolution models trained on datasets produced using limited sets of the computationally expensive phase-field simulations. The main advantage of these surrogate models is that they facilitate the practical exploration of inverse solutions identifying specific microstructures and process parameters that result in the desired microstructures [7]. The primary challenge in formulating the desired surrogate models described above lies in the lack of a rigorous framework for the quantitative description of the microstructures. The complete description of a microstructure is inherently high-dimensional. For example, a voxelized two-phase 3-D microstructure with $100 \times 100 \times 100$ voxels would have a dimensionality of 10^6. Such high-dimensional representations pose major challenges to the successful extraction of low-computational-cost surrogates. However, it is also intuitively understood that high-dimensional representations capturing the full details of the microstructure are unwarranted for most practical applications.

Conventionally, microstructures have been quantified using grossly simplified metrics such as phase volume fractions, average precipitate size, and average precipitate shape [8–11]. Such selections of microstructure measures are not necessarily optimized for maximizing the amount of microstructure information captured. In other words, it is entirely possible that the simplified approaches to microstructure quantification can lead to highly correlated microstructure measures that capture partially redundant information. More importantly, these simple measures do not offer a tunable approach for capturing systematically the higher-order microstructural information needed to predict the effective material properties of interest to desired accuracy [7].

In recent years, the materials knowledge systems (MKS) framework [7] has been developed to address the critical need identified above. MKS uses higher-order spatial correlations [12–15] as statistical descriptors of the microstructure together with powerful dimensionality reduction [16–19] algorithms (e.g. principal component analysis (PCA)) to generate practically useful low-order microstructure descriptors (i.e. salient features of the microstructure). These protocols maximize the capture of the variance in a given ensemble of microstructures in very few metrics (i.e. PC scores), and thereby support the efficient development of low-computational-cost surrogates [20–23]. The MKS framework directly addresses the high dimensionality challenge identified earlier with microstructure representation [7, 24, 25]. Most importantly, the MKS framework leads to the representation of a continuous microstructure space. This is because the PC representations of spatial correlations result in a convex space [26, 27], where one can interpolate between known PC representations of microstructures to produce valid representations of new microstructures. The MKS framework has already been demonstrated for a number of applications in materials discovery and development. However, they have largely been focused on establishing predictive models connecting

the material microstructure with its mechanical and functional properties [28–31]. The MKS framework has been used to predict the (quasi-)steady-state microstructures corresponding to a set of input process parameters [32, 33]. However, there have only been limited applications of the MKS framework for capturing the quantitative linkages between process variables and the time-evolving material microstructure [34–36]. This is not only due to the complexities described above in the quantitative representation of the microstructures, but also because of the need for sophisticated surrogate model building approaches for capturing the temporal evolution of the microstructure as a function of the input process variables.

Gaussian process autoregressive (GPAR) models combined with the MKS framework offer tremendous promise for capturing the details of microstructure evolution as a time series model. Specifically, GPAR offers a nonlinear nonparametric modelling technique for producing time series surrogates. Yabansu et al. [36] demonstrated the benefits of GPAR in capturing the microstructure evolution of polycrystalline ceramics during a sintering process simulated by phase-field models. These authors utilized chord length distributions (CLDs) to quantify the material microstructure and employed a local GPAR approach for building the desired surrogate models. Hashemi et al. [34] employed a global GPAR approach for capturing the details of the microstructure evolution during static recrystallization in polycrystalline metals simulated using cellular automaton. These authors used two-point spatial correlations and PCA to obtain the needed reduced-order microstructure representations. The same approach was subsequently demonstrated for capturing the details of microstructure evolution in face-centred cubic (FCC) polycrystalline microstructures subjected to arbitrary stretching tensors [37]. These prior applications explored only uncoupled GPAR models, which assumed that the evolution of each PC score depended only its own history without any influence of the other PC scores. These uncoupled MKS-GPAR surrogates are quite attractive because of their low overall computational cost (both in training and in predictions for new inputs) and high fidelity. Deng et al. [38] used GPAR for predicting the state of charge (SOC) in a lithium-ion battery pack. These authors first computed Pearson correlation coefficients between several key time series signals (e.g. cell voltage) and the SOC. PCA was then used to transform the high-dimensional correlations into low-dimensional features that served as regressors for the GPAR model. The constructed GPAR model displayed low error for the real-time prediction of the state of charge and was successfully employed for self-correcting in cases where an inaccurate initial state of charge value was provided.

Our goal in this paper is to demonstrate the feasibility and benefits of the MKS-GPAR framework for capturing the salient details of microstructure evolution in post-heat treated, multi-phase, AM (additively manufactured)-processed samples. AM processes are being extensively explored for several critical structural components made from advanced alloys, including Ni-based superalloys. In these applications, the as-fabricated components almost always require post-heat treatments to attain the target mechanical properties [39–42]. In this work, the MKS approach is used to obtain efficient low-dimensional representations of phase-field simulated microstructure evolution in heat treatment of a Ni–Mo–Nb alloy containing Nb-enriched regions typical of the compositions of the intercellular regions of the structure produced during AM solidification. The Ni–Nb–Mo ternary alloy was chosen in this study because it has the potential to be used as a surrogate alloy that captures the features of solidification of the nickel base superalloy 625 that involve the segregation Nb and Mo to the liquid. The long-term goal of the research is to capture the cellular solidification microstructure using phase-field simulations and use these as input microstructures to simulate solid-state precipitation characteristics and to relate the precipitation to the predictions of the GPAR model for a complex multi-component alloy such as 625 and other nickel base alloys such as 718. However, this requires generating a vast amount of simulation data involving several microstructural and processing variables and the data analytics using the PCA–GPAR framework will become very complicated. Our goal was to define a simple, yet challenging problem using a limited set of variables to demonstrate the PCA–GPAR approach on a smaller simulation dataset. Since model Ni–Mo–Nb system is also able to capture the competitive precipitation of δ and γ'' upon annealing at 870 °C, similar to the experimental data shown in [43] for 625, the co-precipitation and bifurcation in the precipitation path was used as the test problem for the GPAR model.

In these simulations, a tetragonal γ'' phase and an orthorhombic δ phase are precipitated in the FCC γ matrix phase [44–47]. Each of these phases has a number

of associated crystallographic variants. The γ'' phase has three crystallographic variants, with one variant corresponding to each of the three {100} habit planes [48]. The δ phase has twelve crystallographic variants, with three variants corresponding to each of the four {111} habit planes [46]. The precipitation of γ'' satisfies the cube-on-cube orientation relationship (OR) $(100)^{\gamma''}//(100)^{\gamma}$ and $[100]^{\gamma''}//[100]^{\gamma}$, while the precipitation of δ satisfies the OR $(010)^{\delta}//(111)^{\gamma}$ and $[100]^{\delta}//\left[1\bar{1}0\right]^{\gamma}$ [46]. The transformation strain, gradient coefficient and the elastic constants of the precipitate variants are rotated from the local frame of the precipitate to the computational frame using appropriate rotation matrices based on the lattice correspondence between the precipitate and the matrix as defined in references [46] and [48]. The precipitation of δ phase has been associated with poor mechanical properties [49] in wrought Ni-based superalloys. During powder bed-based metal additive manufacturing processes, it is possible to obtain heterogeneous microstructures with co-precipitation of γ'' and δ inside the FCC γ matrix. Successful capture of high-fidelity microstructure evolution paths [50, 51] in such complex alloys are expected to open new avenues for AM alloy design through optimized post-AM heat treatments [52]. Prior work has already shown that the precipitation of the δ phase relative to the γ'' phase is strongly influenced by the alloy's Nb concentration [45, 49]. In this paper, phase-field simulations are performed to capture solid–solid phase transformations over a range of Nb compositions in Ni–Mo–Nb ternary alloys using the MEUMAPPS code [53–55] developed at Oak Ridge National Laboratory. Specifically, a limited number of datasets were produced using this code and were used successfully to train an MKS-GPAR surrogate model capturing the salient details of the microstructure evolution in the selected material system and process.

Methods

Phase-field simulations

In the MEUMAPPS phase-field model, the total free energy of the system, G, is defined as

$$G = G_{\text{ch}} + G_{\text{int}} + G_{\text{grad}} + G_{\text{el}} \tag{1}$$

where G_{ch} is the chemical energy, G_{int} is the interfacial energy, G_{grad} is the gradient energy, and G_{el} is the elastic energy. The chemical energy is given by

$$G_{\text{ch}} = \left[1 - \sum_{p=1}^{15} h\left(\phi_p\right)\right] G^{\gamma} + \sum_{p=1}^{3} h\left(\phi_p\right) G^{\gamma''} + \sum_{p=4}^{15} h\left(\phi_p\right) G^{\delta} \tag{2}$$

where G^{γ} is the Gibbs free energy of the γ matrix phase, $G^{\gamma''}$ is the Gibbs free energy of the γ'' precipitate phase, and G^{δ} is the Gibbs free energy of the δ precipitate phase. $\phi_p \geq 0$ is the phase-field order parameter for the p^{th} crystallographic variant, where p takes only positive integer values. The values of G^{γ}, $G^{\gamma''}$, and G^{δ} are fitted to quadratic functions of the concentrations of the solute elements using data from CALPHAD (i.e. TCNi8 database in ThermoCalc) [56, 57]. For this work, $1 \leq p \leq 3$ corresponds to the three γ'' crystallographic variants, while $4 \leq p \leq 15$ correspond to the twelve δ crystallographic variants. $h\left(\phi_p\right)$ is an interpolation function expressed as

$$h\left(\phi_p\right) = \phi_p^3\left(6\phi_p^2 - 15\phi_p + 10\right) \tag{3}$$

The interfacial energy is defined as

$$G_{int} = \omega^{\gamma''} \sum_{p=1}^{3} \phi_p^2\left(1-\phi_p\right)^2 + \omega^{\delta} \sum_{p=4}^{15} \phi_p^2\left(1-\phi_p\right)^2 + \omega_p \sum_{i=1}^{15} \sum_{j\neq i}^{15} \phi_i^2 \phi_j^2 \tag{4}$$

where $\omega^{\gamma''}$ and ω^{δ} define the energy well between the matrix and the γ'' and δ phases, respectively, and ω_p is the energy well between any two phase variants. The gradient energy is defined as

$$G_{\text{grad}} = \frac{1}{2} \sum_{p=1}^{15} \nabla \phi_p^T \kappa^p \nabla \phi_p \tag{5}$$

where κ^p represents the second-rank interfacial energy tensor for the p^{th} crystallographic variant. For this work, κ^p is specified by three diagonal values in the crystal frame, along the (100), (010), and (001) directions. The elastic energy calculation is performed using the approach of Steinbach and Apel [58], and is defined as

$$\begin{aligned} G_{el} = & \sum_{p=1}^{15} h\left(\phi_p\right) \frac{1}{2} \epsilon_{ij}^{el,p} : C_{ijkl}^{p} \epsilon_{ij}^{el,p} : \epsilon_{ij}^{el,p} \\ & + \left(1 - \sum_{p=1}^{15} h\left(\phi_p\right)\right) \epsilon_{ij}^{el,m} : C_{ijkl}^{m} : \epsilon_{kl}^{el,m} \end{aligned} \tag{6}$$

where $\epsilon_{ij}^{el,p}$ is a component of the total elastic strain tensor in the precipitate variant p, C_{ijkl}^{p} is the elastic

stiffness tensor for precipitate variant p, $\epsilon_{ij}^{el,m}$ is the elastic strain tensor in the matrix phase, and C_{ijkl}^{m} is the elastic stiffness tensor for the matrix phase. The total elastic strain ϵ_{kl}^{el} is given by,

$$\epsilon_{kl}^{el} = \overline{\epsilon_{kl}} + \frac{1}{2}\left[\frac{\partial u_k}{\partial r_l} + \frac{\partial u_l}{\partial r_k}\right] - \epsilon_{kl}^{*} \tag{7}$$

where $\overline{\epsilon_{kl}}$ is the homogeneous strain, which is a measure of the macroscopic deformation in the system, the second term on the right-hand side is the lattice inhomogeneous strain, and ϵ_{kl}^{*} is the eigenstrain. In the current simulations, $\overline{\epsilon_{kl}}$ is taken to be zero since macroscopic deformation due to externally applied strain is not considered. The elastic strain energy calculation explicitly considers the elastic constants of the matrix and the individual precipitates based on their crystal structure. All calculations are done inside a single matrix grain oriented in the computational frame. The eigenstrain and gradient coefficients for a given phase are usually prescribed on a local frame based on the orientation relationship between the matrix and the phase based on its crystal symmetry. These are rotated to the computational frame using the corresponding rotation matrices.

The evolution equations solved are the time-dependent Ginzburg–Landau (TDGL) equation and the diffusion equation. The TDGL is expressed as

$$\frac{\partial \phi_p}{\partial t} = -\frac{L_\phi}{\tilde{N}} \sum_{p \neq q}^{\tilde{N}} \left(\frac{\delta G}{\delta \phi_p} - \frac{\delta G}{\delta \phi_q} \right) \tag{8}$$

where L_ϕ is the phase-field mobility, $\tilde{N}$ is the total number of variants that coexist at any mesh point (i.e. number of variants with $\phi_p > 0$), and the functional derivatives are defined as $\delta G/\delta\phi = \partial G/\partial\phi - \nabla \cdot \partial G/\partial(\nabla\phi)$. The above formulation of the TDGL assumes that the constraint $\sum_{p=1}^{\tilde{N}+1} \phi_p = 1$ is satisfied at every mesh point, including the matrix. In the current simulations, this is enforced by setting $\sum_{p=1}^{\tilde{N}} \phi_p \leq 1 - \zeta$, where ζ is set to 10^{-4}. This approach allows sufficient soft impingement to occur between precipitates, while a small fraction of the matrix phase is retained between the impinging precipitates so that the local microstructural environment is consistent with the assumptions used in the phase-field model formulation. As already mentioned, the material selected for this study is a ternary Ni–Mo–Nb alloy. The relevant diffusion equation is expressed as

$$\frac{\partial C_{\mathrm{Mo,Nb}}}{\partial t} = \nabla \cdot M_{\mathrm{Mo,Nb}} \nabla \mu_{\mathrm{Mo,Nb}} \tag{9}$$

where M_{Mo} and M_{Nb} are the mobilities of Mo and Nb, and μ_{Mo} and μ_{Nb} represent the chemical potentials of the species in each phase. The Kim–Kim–Suzuki (KKS) formulation requires the calculation of the Mo or Nb compositions at the various diffuse interfaces between phases using the equal chemical potential assumption, $\mu_{\mathrm{Nb,Mo}} = \frac{\partial G^\gamma}{\partial C^\gamma_{\mathrm{Nb,Mo}}}$. This requires solution to $N_S \times N_P$ simultaneous equations where N_S is the number of solutes and N_P is the number of phases in the system. Representing the Gibbs free energies in the quadratic form leads to a system of linear equations. The evolution equations are solved using an established semi-implicit Fourier spectral technique [59]. Parallel implementation of MEUMAPPS-SS on the ORNL supercomputer Summit was accomplished via message passing interface and the use of a parallel three-dimensional fast Fourier transforms (P3DFFT) package [60]. MEUMAPPS-SS shows excellent scaling over thousands of CPU cores in Summit [61].

Nucleation model

Nucleation is a necessary step in carrying out the phase-field simulations and often poses significant challenges. Since the phase-field simulations are based on minimizing the total system energy, they cannot capture the increase in the system energy required to overcome the activation energy required for nucleation. Nucleation occurs at the atomistic level by the random formation of transient clusters that become supercritical nuclei. Often the clusters have a crystal structure that is different from that of the equilibrium solid. While in a pure metal, the clusters are of the same composition as the parent matrix, in the case of alloys the composition as well as the crystal structure within the clusters goes through a complex sequence of changes to transition from a series of metastable phases to nuclei of the equilibrium phase. In addition, the shape of the nucleus at every stage of this sequence is governed by the anisotropy in the interfacial energy and the elastic strain field in the vicinity of the nucleus. These elastic strain fields could be either due to the existence of another phase or defects such as dislocations, point defects, or grain boundaries. Often the problem is not tractable using atomistic simulations such as molecular dynamics or kinetic Monte Carlo

because each has its own limitations in capturing the true physics at the concerned length and timescales. In phase-field simulations, it is customary to implement the nucleation step using either a Langevin noise model, which introduces a stochastic perturbation term in the Ginzburg–Landau and the Cahn–Hilliard equations [62] or an explicit Poisson seeding model [63], which introduces nuclei as critical, diffuse, geometric (i.e. shaped) fluctuations in the order parameter and composition fields. Even though Langevin noise approach works reasonably well at high undercooling where small fluctuations become critical, it is difficult to implement at low undercoolings where large perturbations are required to form the nuclei. It is also somewhat limited in physical insight. The explicit nucleation approach captures better the size and number density of the critical nuclei. However, it becomes very difficult to implement when there are additional complexities due to the presence of multiple crystallographic variants and variant selection based on dynamically varying strain fields. This is particularly relevant to the present study, where there are 15 crystallographic variants in total, and the strain energy associated with nucleation of each variant could be modified by the presence of other pre-existing variants.

In the current simulations, the activation energies for nucleation of the δ and γ'' phases are different, and are dependent on the Nb content in the alloy. The Langevin noise works efficiently for high Nb compositions, which accelerate the precipitation kinetics. At low Nb contents, the amplitude of the noise components for the two phases has to be increased significantly before a critical nucleus of either phase can be initiated. In other words, the system needs to be perturbed to a highly non-equilibrium state before it starts to decay, the non-critical noise components die down, and the supercritical noise components form the required nuclei. This results in an undesired transition period at the start of the phase-field simulation (which starts at the end of the noise step). Because the microstructures in this transition regime are non-physical, they should not be used in the training of the process–microstructure evolution surrogate models. Identifying and removing these transition segments from the phase-field simulation datasets poses a significant challenge. In this work, a critical maximum value of the order parameter introduced by noise was used as the cut-off limit for the noise. After the noise was cut off, the value decayed to a minimum during the transition regime when the critical nuclei were forming, and rose again when the critical nuclei started to grow. However, there was no guarantee that all the noise terms had completely died down when the minimum was reached, because residual non-equilibrium conditions in the system could not be completely ruled out.

Materials knowledge systems (MKS)

The MKS framework [30, 64, 65] starts with a voxelized description of the microstructure, denoted by the array m_s^k, where k indexes the potential distinct material local states (e.g., different thermodynamic phases) that might be encountered in any voxel of the microstructure, s indexes the voxels, and the value of m_s^k reflects the volume fraction of the voxel s occupied by the material local state k. In most practical applications, each voxel is only allowed to be occupied by only one of the allowed material local states; such microstructures are referred as eigen microstructures [66]. In other words, the value of m_s^k is one for the material state present in the voxel and zero for all other material local states.

The representation of the microstructure as the array m_s^k lacks translational invariance, which is automatically implied in the periodic boundary conditions imposed in the phase-field simulations. For this and other reasons [13, 67], the MKS framework utilizes auto and cross-correlations of m_s^k (also referred as two-point spatial correlations) to define statistical measures of the microstructures [33, 68]. Mathematically, the two-point correlations, denoted as $f_t^{kk'}$ are expressed as [69, 70]

$$f_t^{kk'} = \frac{1}{|S|}\sum_{s}^{S} m_s^k m_{s+t}^{k'} \tag{10}$$

where t indexes the set of discrete vectors that are allowed to be placed in the voxelized microstructures, and $|S|$ is the total number of voxels in the microstructure. When k and k' represent the same local state, the spatial correlations are referred as autocorrelations, and as cross-correlations otherwise. For the present case study with three phases, Eq. (10) defines a total of three autocorrelations and six cross-correlations. However, there exists a number of interdependencies between these correlations [71]. In prior work with three material local states [33], it was observed that it is adequate to utilize two autocorrelations and

one cross-correlation for building surrogate models. Equation (10) is implemented [72] utilizing the computationally efficient fast Fourier transform (FFT) algorithm. For each microstructure, the three correlations are flattened and appended into a single row, such that each row represents a complete set of microstructure statistics.

The spatial correlations computed using Eq. (10) result in a very large set of statistical features for each microstructure. As an example, for a microstructure defined using 100^3 voxels, the set of three spatial correlations mentioned above would produce about $3(10)^6$ statistical features. Clearly, this unwieldy representation cannot serve the desired surrogate model building efforts. In the MKS framework, we seek a low-dimensional representation of this large feature vector using PCA [73]. This step involves a rotational (distance-preserving) transformation of the original features into a new frame, whose axes are organized to maximize the capture of variance in the ensemble of microstructures in the least number of orthogonal directions. In other words, the first PC direction explains the largest component of the variance in the original dataset, the second PC direction explains the second largest component of the variance in the original dataset, and so on. The transformed features are referred as PC scores. It is important to scale the different sets of correlations (in the present case we are using three sets of correlations) before performing the PCA to ensure that each set of correlations are equally represented in the resulting PC scores. The details of the scaling protocols have been described in prior work [36].

Gaussian process autoregression (GPAR)

The MKS framework described above can be combined with a SV-MOGPAR to produce surrogate models of microstructure evolution. Let $\boldsymbol{\alpha}^t$ denote the vector of PC scores representing the spatial correlations of an evolving microstructure at a time step indexed by t. In the MKS framework [34], the desired surrogate model can be expressed as

$$\boldsymbol{\alpha}^t \sim GP\left(\boldsymbol{\mu}(\boldsymbol{x}), k\left(\boldsymbol{x}, \boldsymbol{x}'\right)\right) \tag{11}$$

where $\boldsymbol{x}$ is the input feature vector, $\boldsymbol{\mu}(\cdot)$ the mean function, and k the covariance kernel. Frequently, the mean function is taken to be zero without a loss of generality (i.e. $\boldsymbol{\mu}(\boldsymbol{x}) = 0$). An assumption of Gaussian noise is then applied to observed outputs as $\boldsymbol{y} = \mathbf{f}(\boldsymbol{x}) + \boldsymbol{\epsilon}$ (i.e. $\boldsymbol{\epsilon} \sim \mathcal{N}\left(0, \sigma_y^2 \boldsymbol{I}\right)$). More specifically, the input feature vector $\boldsymbol{x}$ contains the microstructure PC scores at T previous time steps (called autoregression order) defined on a uniformly discretized time axis. Prior work [34, 37] has shown that autoregression order of one is adequate for systems governed by first-order time derivatives. Therefore, it was decided to explore only first-order autoregression in this study, where it is assumed that the microstructure PC scores at any specified instant of time are dependent on the values of the microstructure PC scores at only one previous time step. Therefore, for the present study, the feature vector $\boldsymbol{x}$ simply contains the PC scores at the previous time step.

For this work, a Matern 3/2 kernel with automatic relevance determination (ARD) was used [74]. This kernel is expressed as

$$k_{3/2}(\mathbf{x}, \mathbf{x}') = \left(1 + \sqrt{3}\left\|\Delta\left(\mathbf{x} - \mathbf{x}'\right)\right\|_2\right) \exp\left(-\sqrt{3}\left\|\Delta\left(\mathbf{x} - \mathbf{x}'\right)\right\|_2\right) \tag{12}$$

where length scales associated with each input dimension are grouped into $\Delta = \text{diag}\left(l_1^{-1}, l_2^{-1}, \ldots, l_N^{-1}\right)$. The process resulting from this covariance kernel is one mean-squared differentiable, providing sharper transitions in comparison with the squared exponential kernel.

Simultaneous predictions could be made for all values of PC scores considered, through the linear model of co-regionalization (LMC) [75, 76]. This model represents an extension of the standard Gaussian process to learn a multi-output function $\mathbf{f}(\boldsymbol{x}) : \mathcal{X} \to \mathbb{R}^P$ where the input space is $\mathcal{X} \in \mathbb{R}^D$. The p-th output of $\mathbf{f}(\boldsymbol{x})$ is expressed as $\mathrm{f}_p(\boldsymbol{x})$, with its complete representation given as $\mathbf{f} = \left\{\mathrm{f}_p(x_i)\right\}_{i=1}^n$. This is accomplished through constructing a multi-output function from a linear transformation $W \in \mathbb{R}^{P \times L}$ of L independent functions $\boldsymbol{g}(x) = \left\{g_l(x)\right\}_{l=1}^L$. Each function is constructed as an independent $\mathcal{GP}$, $g_l(\boldsymbol{x}) \sim \mathcal{GP}\left(0, k_l\left(\boldsymbol{x}, \boldsymbol{x}'\right)\right)$, each with its own covariance kernel, resulting in the final expression $\mathbf{f}(\boldsymbol{x}) = \boldsymbol{W}\boldsymbol{g}(\boldsymbol{x})$. The multi-output covariance kernel described by this model is then:

$$k\left(\{\boldsymbol{x}, p\}, \{\boldsymbol{x}', p'\}\right) = \sum_{l=1}^{L} W_{pl} k_l(\boldsymbol{x}, \boldsymbol{x}') W_{p'l} \tag{13}$$

This covariance kernel enables improved performance through encoding correlations between output

dimensions and through the linear super positioning of multiple latent $\mathcal{GP}$ s. As a result, this kernel provides for an incredibly expressive model.

Detrimentally, the LMC model greatly expands the dimensionality of the covariance matrix $\boldsymbol{K_{ff}}$ associated with the trained description of the infinite-dimensional stochastic process f. This often results in a prohibitive cost of computing $\left[\boldsymbol{K_{ff}} + \sigma_y^2 \boldsymbol{I}\right]^{-1}$ for even datasets of modest size; this computation needs to be repeatedly performed in optimizing the hyperparameters of Eqs. (12) and (13) through the log marginal likelihood [74]. In order to enable computationally feasible inference, this work relies upon the variational free energy (VFE) approximation [77]. This sparse global approximation to the exact Gaussian process relies upon a collection of M inducing inputs $\boldsymbol{Z} = \{\boldsymbol{z}_m\}_{m=1}^{M}$ independent of $\boldsymbol{X} = \{\boldsymbol{x}_m\}_{n=1}^{N}$, where $M \leq N$ provides a low-rank approximation to the full covariance $\boldsymbol{K_{ff}}$. The positions of these inducing points are then optimized collectively with the model hyperparameters through maximization of the evidence lower bound (ELBO), reflecting a minimization of the Kullback–Leibler (KL) divergence between the variational $\mathcal{GP}$ and the exact posterior $\mathcal{GP}$. Variational inference can then be performed through the expression

$$\mathcal{L} = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log\left|\hat{K}_{\mathbf{ff}}\right| - \frac{1}{2}\boldsymbol{y}^{\top}\hat{K}_{\mathbf{ff}}^{-1}y - \frac{1}{2\sigma_y^2}\mathrm{trace}\left(\boldsymbol{K_{ff}} - \boldsymbol{Q_{ff}}\right) \tag{14}$$

where $\widehat{\boldsymbol{K}}_{\mathbf{ff}} = \boldsymbol{Q_{ff}} + \sigma_y^2 \boldsymbol{I}$, and $\boldsymbol{Q_{ff}} = \boldsymbol{K_{fu}}\boldsymbol{K_{uu}^{-1}}\boldsymbol{K_{uf}}$ provides the low-rank approximation, with $\boldsymbol{u}$ representing the corresponding function values at the inducing points. Elements of the covariance matrices are calculated as $\left[\boldsymbol{K_{uu}}\right]_{mm'} = k(z_m, z_{m'})$ and $\left[\boldsymbol{K_{fu}}\right]_{nm'} = k(\boldsymbol{x}_m, z_{m'})$. The resulting complete set of hyperparameters optimized within Eq. (14) is then $= \{\boldsymbol{W}_1, \ldots, \boldsymbol{W}_L, \boldsymbol{l}, \boldsymbol{Z}\}$.

Results and discussion

Generation of the simulation dataset

A total of 14 MEUMAPPS simulations were performed corresponding to a range of Nb compositions from 0.095 to 0.16 in increments of 0.005. For each Nb composition, the Mo composition was assigned a constant value of 0.06, with Ni making up the balance. The Nb concentrations used in the simulations were selected to reflect those expected at the interdendritic regions during solidification at thermal conditions characteristic of powder bed-based additive manufacturing (AM) processes [78]. Post-AM heat treatment of the alloy at 870 °C shows co-precipitation of the δ and γ'' phases [43], and the preferential formation of the γ'' phase at the expense of the δ phase as Nb concentration is increased. The values of the other input and computational parameters used in these simulations are summarized in Table 1 and are largely taken from prior work. These include the interfacial energy and transformation strains for the γ'' [79] and the transformation strains for the δ phase [47]. The latter assumes that the precipitate is fully coherent with the matrix. However, the use of these values along with the published elastic constants for δ does not result in the observed platelike morphology [45]. It is also known that the broad face of the δ phase contains many interfacial dislocations [46] suggesting loss of coherency. Another unknown in the simulations is the potential variation of the transformation strains with the Nb content in the matrix. Due to the lack of the required data in published literature, a near-incoherent interface was assumed along the broad face of the δ while some coherency was assumed along the thickness of the plate to capture the observed plate morphology for the δ phase.

Prior to performing the phase-field simulations using the nucleation model described in

Table 1 Summary of the input and computational parameters employed in the phase-field simulations performed for this work

Phase-field parameter	Value(s)
Domain size, voxels	126^3
Voxel resolution	1 nm^3
Time step resolution(s)	0.005
Concentration, Mo	0.06
Concentration, Nb	0.095–0.16
Interfacial energy, γ'', (100)	0.12
Interfacial energy, γ'', (010)	0.12
Interfacial energy, γ'', (001)	0.12
Interfacial energy, δ, (100)	0.10
Interfacial energy, δ, (100)	0.10
Interfacial energy, δ, (100)	0.10
Transformation strain, γ'', (100)	0.0286
Transformation strain, γ'', (010)	0.0067
Transformation strain, γ'', (001)	0.0067
Transformation strain, δ, (100)	1.00E−07
Transformation strain, δ, (010)	0.0600
Transformation strain, δ, (001)	1.00E−07

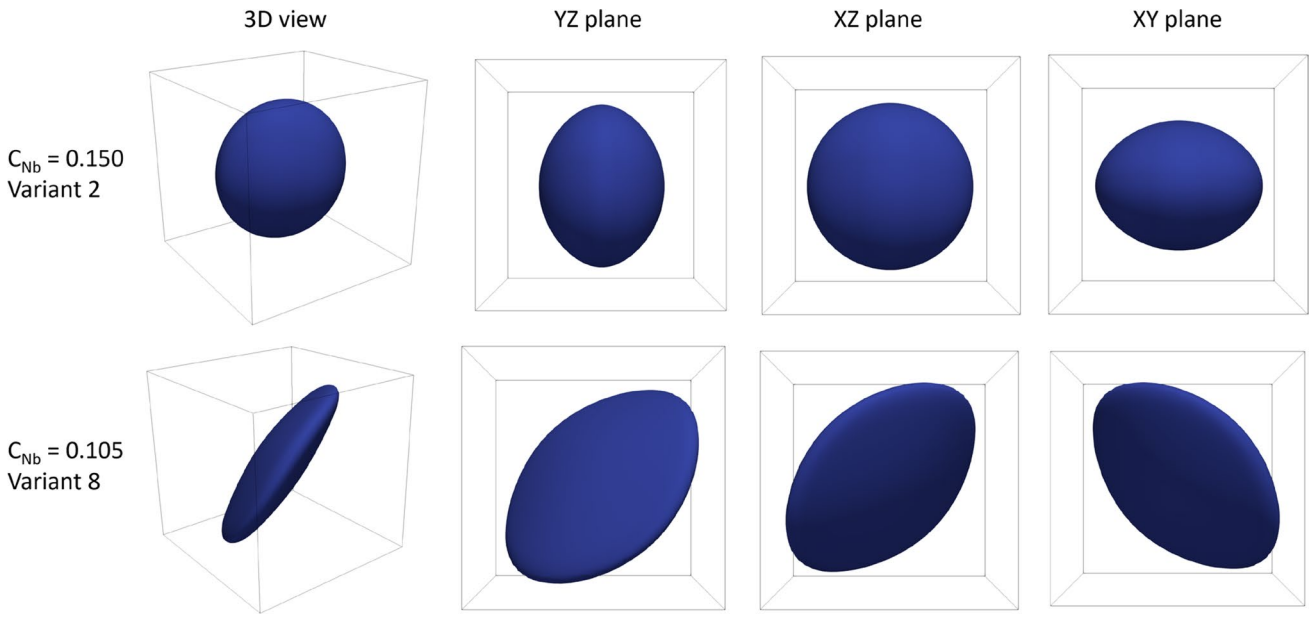

Figure 1 Results from single precipitate simulations. (Top) γ'' precipitate forming on (100) plane. (Bottom) δ precipitate forming on (111) plane.

Sect. "Nucleation model," single precipitate simulations were performed for the γ'' and δ phases as shown in Fig. 1. As expected from the transformation strains and the interfacial energies used in the input parameters, γ'' precipitates grew stably on {100} habit planes, while δ precipitates grew stably on {111} habit planes. The size of the simulation domains used in these simulations was such that the disc shaped morphology was not fully developed for the γ'' phase, although it is clear that the shape is evolving towards a disc parallel to the xz plane for the specific variant shown. However, for the multiple precipitate simulations the domain was not big enough to fully resolve the morphologies of the individual precipitates especially at high Nb contents where significant coalescence of the variants occurred during the simulation.

Each phase-field simulation was performed until the microstructure was observed to have very little change with time: the final time was determined by observing the total free energy curve for each simulation and selecting a point at which the curve displayed asymptotic behaviour. For each MEUMAPPS simulation, model data was collected starting from the time step after which microstructure evolution was largely controlled by the physics-based evolution laws described by the phase-field model and not by the non-physical evolution of the Langevin noise. After the noise was cut off, the maximum value of ϕ introduced by the Langevin noise reached a minimum before it started to rise again as the supercritical nuclei started to grow. The training data was collected from the time step that marked the beginning of the rise in ϕ to the final time step. Furthermore, microstructures were sampled every 15 time steps (i.e. every 0.075 s in simulation time) from the phase-field simulations for training the surrogate model. In other words, the time step for the autoregressive surrogate model corresponded to 15 time steps in the phase-field simulation. This coarser time step for the surrogate model was selected to limit the overall computational cost of training the surrogate model, while ensuring sufficient resolution for capturing the microstructure evolution path for all compositions considered in this study. The protocols described above produced a total of 3,654 microstructures for this work.

Microstructure quantification using MKS approaches

Phase-field simulations employ diffuse interfaces to carry out the computations described in Sects. "Phase-field simulations"–"Nucleation model." Since the precipitates are also in the nanoscale, the cross-sectional dimensions of the δ and γ'' are often comparable to the width of their diffuse interfaces with the matrix. The computations of the predicted volume fractions

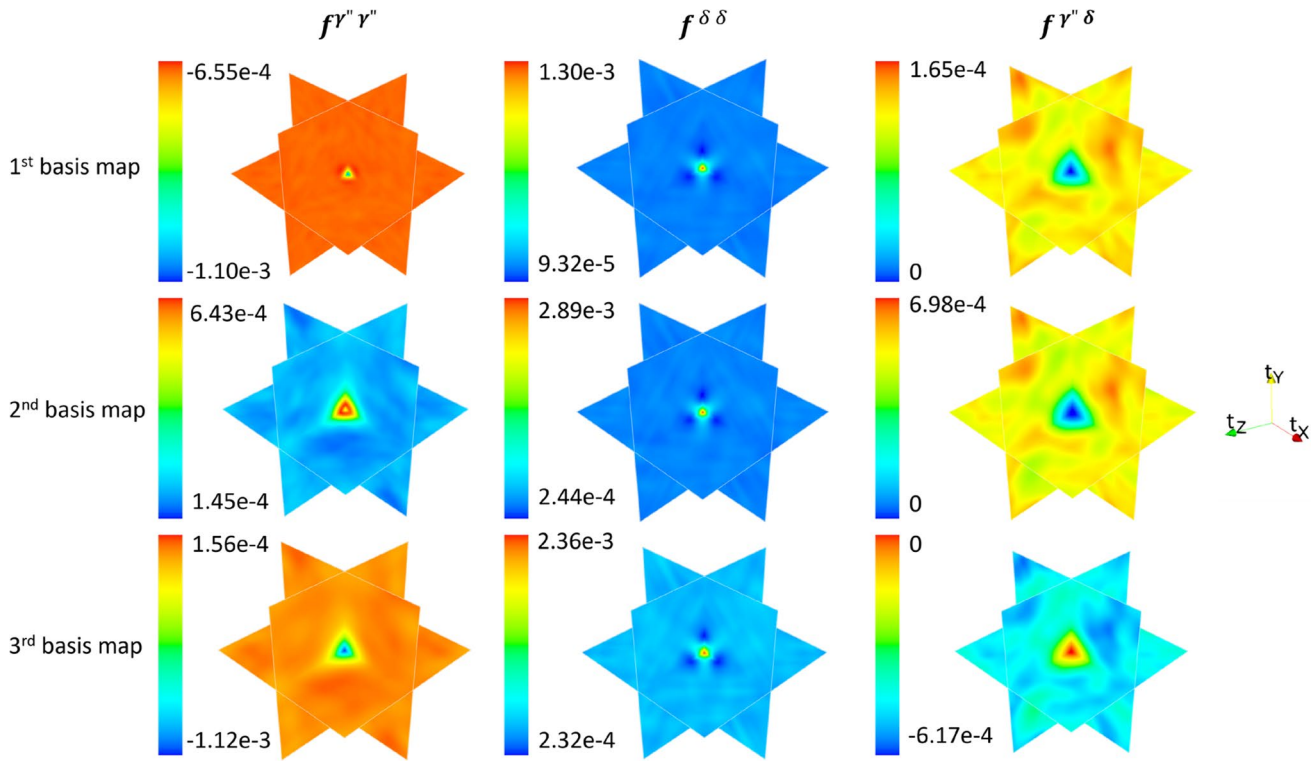

Figure 2 Basis maps for the first 3 PCs computed for the training set aggregated for the present study, shown on three orthogonal sections through the origin.

of the precipitates are significantly influenced by the choice of the threshold value of ϕ (generally referred as segmentation threshold value). In this work, based on an evaluation of precipitate morphologies and the final volume fractions obtained at the end of the phase-field simulations, it was decided to transform the phase-field simulated microstructures to eigen microstructures by applying a threshold value of 0.5. In other words, the microstructure states for the γ'' and δ phases were defined as:

$$m_s^{\gamma''} = \begin{cases} 1, \ \phi_{p,s} \geq 0.5 \forall 1 \leq p \leq 3 \\ 0, \ \text{otherwise} \end{cases} \tag{15}$$

$$m_s^{\delta} = \begin{cases} 1, \ \phi_{p,s} \geq 0.5 \forall 4 \leq p \leq 15 \\ 0, \ \text{otherwise} \end{cases} \tag{16}$$

where $\phi_{p,s}$ is the value of the order parameter for the precipitate variant p at voxel s. Note that these equations ensure that only one of the three material local states (i.e. gamma, delta, or matrix) are allowed in each voxel. The MKS framework described in Sect. "Materials knowledge systems (MKS)" was employed to quantify all the simulated microstructures produced for the present case study. Specifically, three sets of spatial correlations were computed: the γ'' autocorrelations, the δ autocorrelations, and the γ''—δ cross-correlations.

Following the established protocols of the MKS framework, PCA was performed on the computed set of spatial correlations. Consistent with prior applications [30, 32, 33], it was observed that only the first few PC scores are adequate to capture an overwhelming fraction of the overall variance in the collected set of microstructures. Specifically, in the present case, it was observed that the first three PC scores accounted for more than 99% of the explained variance. The basis maps for these scores are presented in Fig. 2.

As noted in prior work [34], the PC bases pose significant challenges in interpretation, because of their extreme dimensionality. In the present case, each PC basis represents 3 sets of weighted spatial correlations. Qualitatively, it is seen the first PC score captures the competing volume fractions between γ'' and δ, indicated by the fact that the highest magnitude points are located in the centre. Thus, as PC1 increases, the δ volume fraction increases while the γ'' volume fraction decreases. PC2 is also strongly correlated with

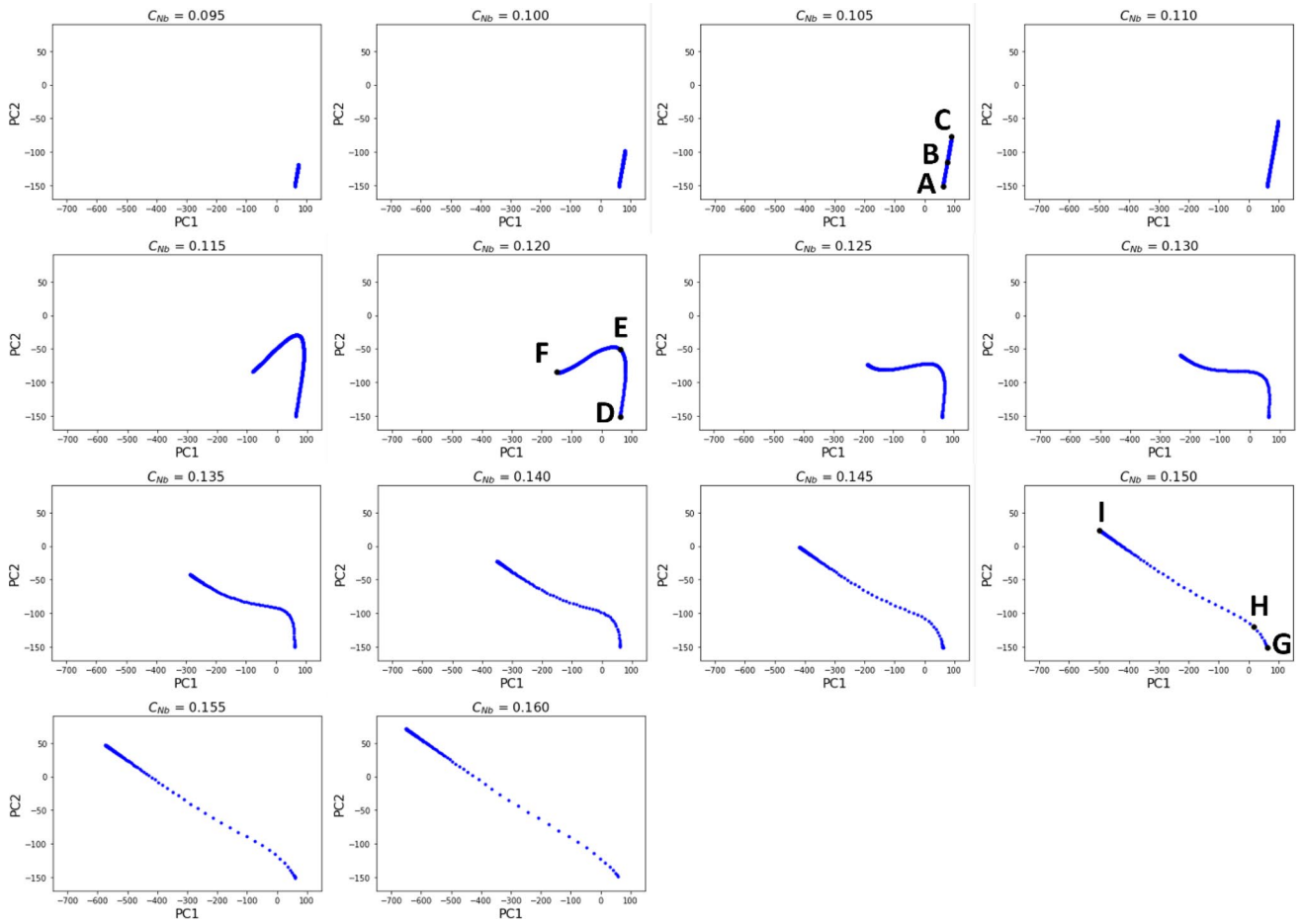

Figure 3 PC1-PC2 projections of MEUMAPPS simulated microstructure evolutions for 14 different compositions. Each point in each figure represents one microstructure. Specific microstructures in these plots are labelled, and their spatial correlations are presented in Figs. 2 and 3.

the δ phase volume fraction, although it may also be observed that it is correlated with the size of γ'' precipitates (indicated by bright bands near but not at the origin).

Since each microstructure is represented as a point in the PC space, the complete details of the microstructure evolution in a phase-field simulation can be represented as a trajectory in the PC space. Indeed, one of the main benefits of employing the MKS framework is that it allows a simple visualization of the extremely high-dimensional information of microstructure evolution in very few dimensions. Figure 3 shows the details of the microstructure evolution for each of the 14 MEUMAPPS simulations projected in the first two PCs. (These two PCs account for 95.7% of the explained variance in the entire ensemble of 3654 phase-field simulated microstructures produced for this study.) For all the trajectories shown, the initial microstructures are quite similar to each other and are found at approximate values of 60 for PC1 and -150 for PC2.

The microstructure evolution trajectories presented in Fig. 3 are smooth and show a systematic trend as the Nb composition is increased. For the low Nb compositions, the trajectories are found to be essentially linear segments (e.g. see the trajectory labelled by points A, B, and C, for $C_{Nb} = 0.105$; the corresponding microstructures are shown in Fig. 4). These trajectories represent the microstructure evolutions dominated by the formation of δ precipitates, with the longer segments representing the production of higher δ volume fractions. For the high Nb compositions, the microstructure trajectories are seen to be slightly nonlinear, but their direction is

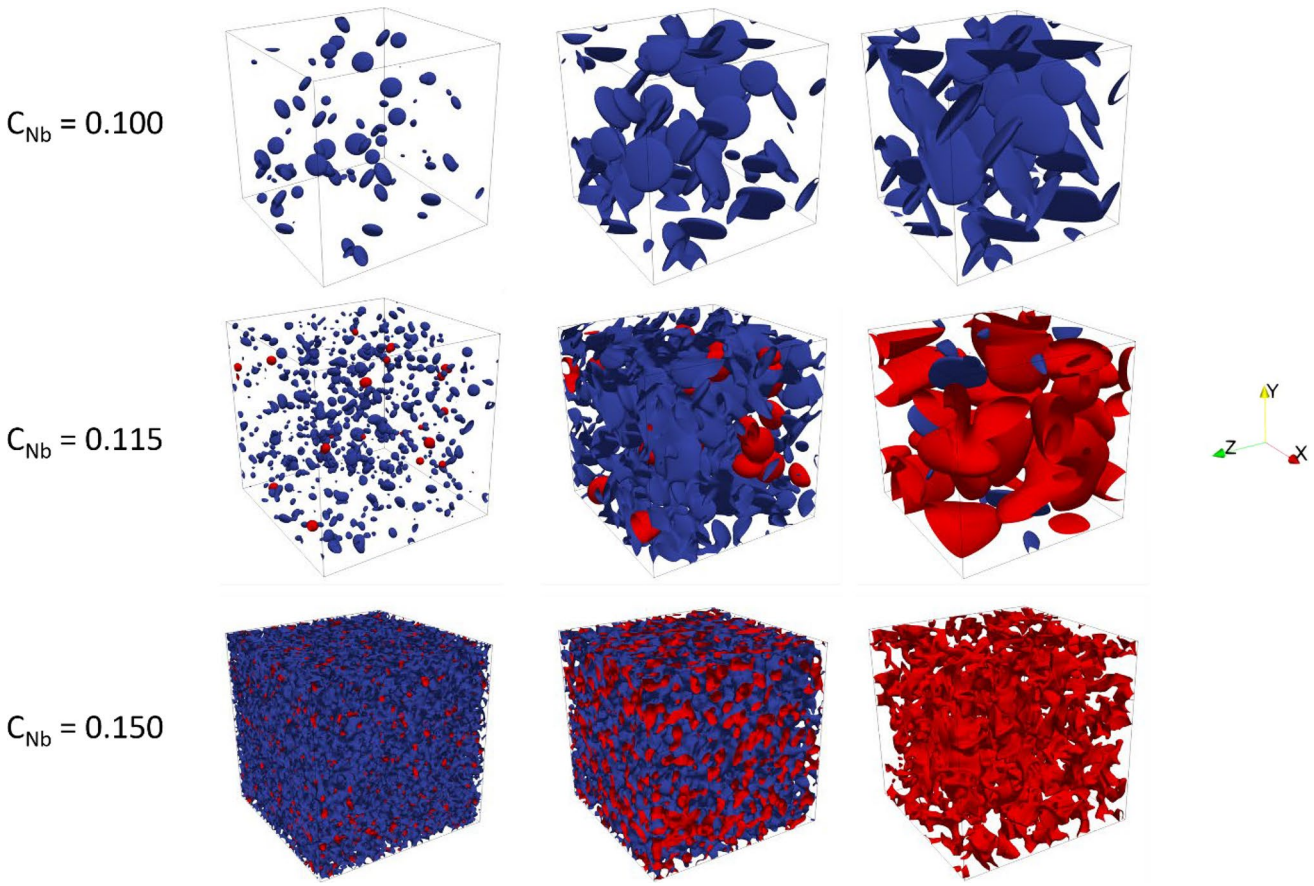

Figure 4 Phase-field simulated microstructures corresponding to the ones labelled in Fig. 1. The three rows in the figure depict the microstructure evolution for three different Nb compositions. γ'' precipitates are shown in red, while δ precipitates are shown in blue. The evolution paths for these three different Nb compositions are clearly distinct from each other.

distinctly different from the trajectories corresponding to the low Nb compositions (e.g. see the trajectory labelled by points G, H, and I, for $C_{Nb} = 0.15$ and compare it with the ABC trajectory; the corresponding microstructures are shown in Fig. 4). These trajectories represent microstructure evolutions dominated by the formation of γ'' precipitates, with the longer segments once again reflecting the production of higher γ'' volume fractions. For the intermediate Nb composition, the trajectories smoothly transition between the two types of distinct trajectories described above (compare the trajectory DEF with the other trajectories described above; the corresponding microstructures are shown in Fig. 4). This is because the microstructure evolution in these intermediate Nb compositions appears to be initially dominated by the formation of δ, followed by dissolution of δ, and eventual formation of γ''. The fact that PCA captured all of these details in very simple visualizations is quite impressive, especially since PCA is completely unsupervised. In other words, none of the microstructures in the collected ensemble were labelled in any manner before the application of PCA.

Figure 4 shows that the phase-field simulations are able to capture the nucleation and growth of the δ phase giving rise to the characteristic plate like morphology. However, the shape of the γ'' phase deviates significantly from the expected disc morphology on (100) planes, especially at high Nb concentrations. This is mainly due to the high density of nucleation and significant coalescence of the three crystallographic variants of γ''. However, the simulations do capture the preferential growth of γ'' at the expense of δ at intermediate (0.115 Nb) and high (0.15 Nb) concentrations.

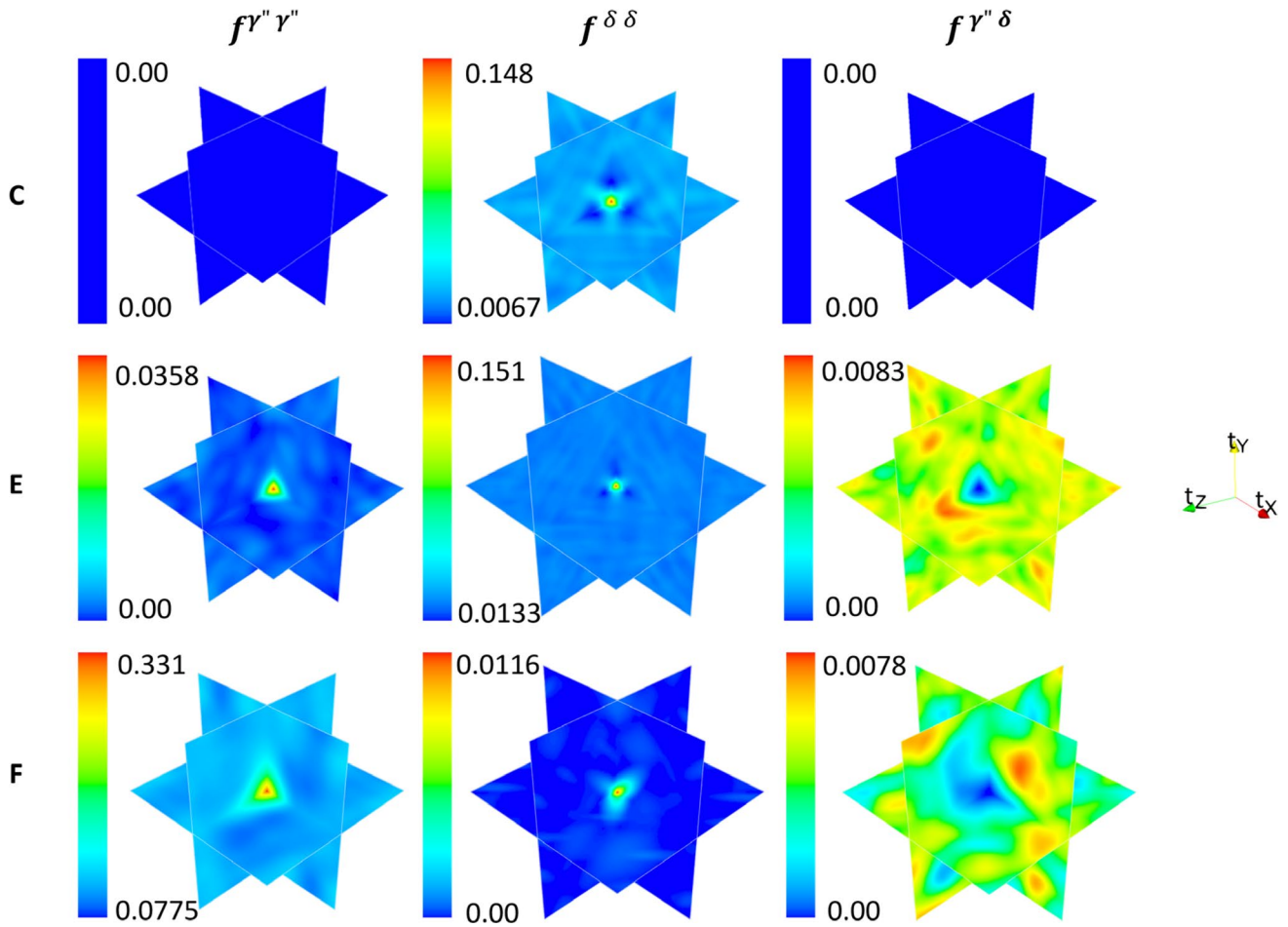

Figure 5 Computed spatial correlations for microstructures C, E, and F. Only the three mid-planes passing through the origin are shown in these plots for clarity.

Figure 5 presents the computed spatial correlations for three example microstructures (one in each row). The peaks at the origin in the autocorrelations depict the respective phase volume fractions. Therefore, it can be seen that microstructure C is composed only of δ, while microstructure E shows a high volume fraction of δ combined with a low volume fraction of γ''. Conversely, microstructure F shows a high volume fraction of γ'' and a low volume fraction of δ. The cross-correlations provide information on the relative placement of the two phases in the microstructures. The spatial correlations shown in the figure also carry significant additional information about precipitate size, orientation, and morphology distributions. For example, although microstructures C and E exhibit similar volume fractions of δ phase, the precipitate size is noticeably larger for microstructure C compared to microstructure E, which is reflected through the much broader central peak in the autocorrelation maps.

SV-MOGPAR model building

In contrast to prior applications of the MKS-GPAR framework, the evolution of the low-dimensional microstructure measures (i.e. PC scores) exhibited significant coupling. In other words, the evolution of a PC score at any given time step was found to be significantly influenced by the values of other PC scores at the previous time step. In this work, the SV-MOGPAR model systematically accounts for this interdependence through the off-diagonal components of the autoregressive covariance matrix,

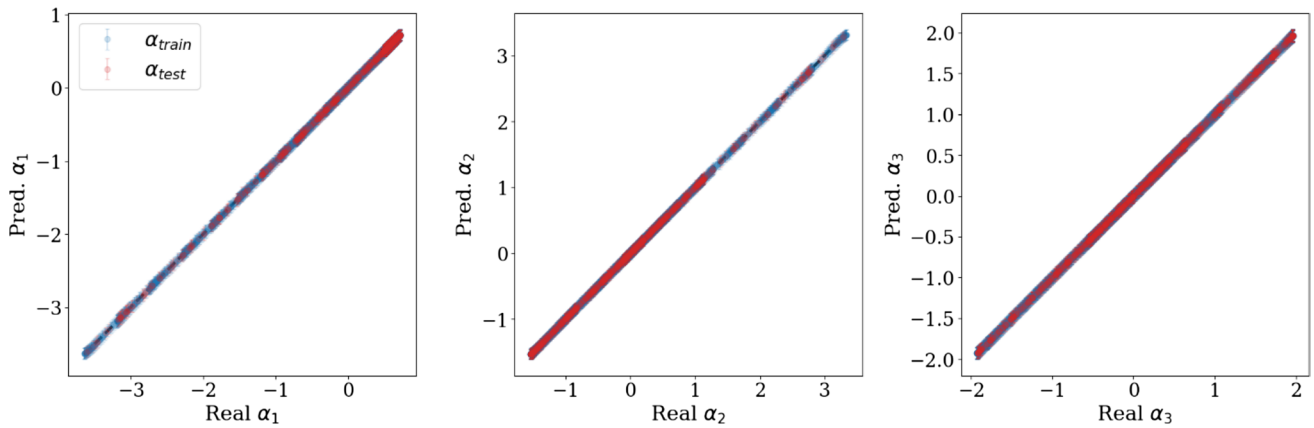

Figure 6 SV-MOGPAR model fits for (left) PC1, (middle) PC2 and (right) PC3. Blue points indicate training points, while red points indicate test points.

naturally accounting for correlations between the PC scores considered.

This model was constructed with the time evolution of standardized[1] PC scores for the entire ensemble, which were then broken into train and test partitions with an 80/20 split. The number of latent $\mathcal{GP}$ s (L) considered in the LMC model was 6, with 116 summarizing inducing points ($\boldsymbol{Z}$). Training was performed over 4000 epochs with the Adam optimizer [80] alongside cosine annealing of the initial 1e−2 learning rate in order to maximize the expression in Eq. (14). Once trained, the SV-MOGPAR model was then capable of providing autoregressive time series predictions for the evolution of the PC scores considered (taken to be equal to 15 time steps of the phase-field simulation). The performance of the model can be seen in the parity plots of Fig. 6 for the first three target PC scores, with error metrics summarized in Table 2. The validation performance of the model ensures high numerical stability, since the predictive accuracy varies little with the input parameter values (i.e. PC scores) or the MEUMAPPS simulation (i.e. the Nb composition value).

SV-MOGPAR predictions of microstructure evolution

Equipped with the trained SV-MOGPAR model, predictions were performed for each of the MEUMAPPS simulations (i.e. the compositions used to train the model). The initial set of PC scores for each simulation was taken as an input to the surrogate model, which subsequently was capable of producing predictions for the complete evolution pathways in PC space. Figure 7 summarizes the SV-MOGPAR surrogate model performance across the range of Nb concentrations, where it is apparent that the model is capable of producing accurate predictions across varied Nb concentration levels. The highest errors were observed in the simulations where there was a transition from the microstructures dominated by δ to those dominated by γ''. Since this transition occurs over a very narrow range of Nb compositions (from 11 to 12% Nb), there are very few of these microstructures in the training set used in this study. It is therefore quite reasonable that the highest errors in the surrogate model were observed for these transition microstructures. Even in these cases, it is seen that the final values of the PC scores were well reflected by the surrogate model predictions.

Table 2 Mean absolute error (MAE) and normalized mean absolute error (NMAE) of the SV-MOGPAR model fitting errors for the present work

	MAE		NMAE (%)	
	Train	Test	Train	Test
PC1	7.61E−04	7.30E−04	0.100	0.098
PC2	1.05E−03	9.26E−04	0.139	0.125
PC3	7.88E−04	8.36E−04	0.098	0.104

[1] Features were standardized by subtracting their mean values and dividing by their standard deviations.

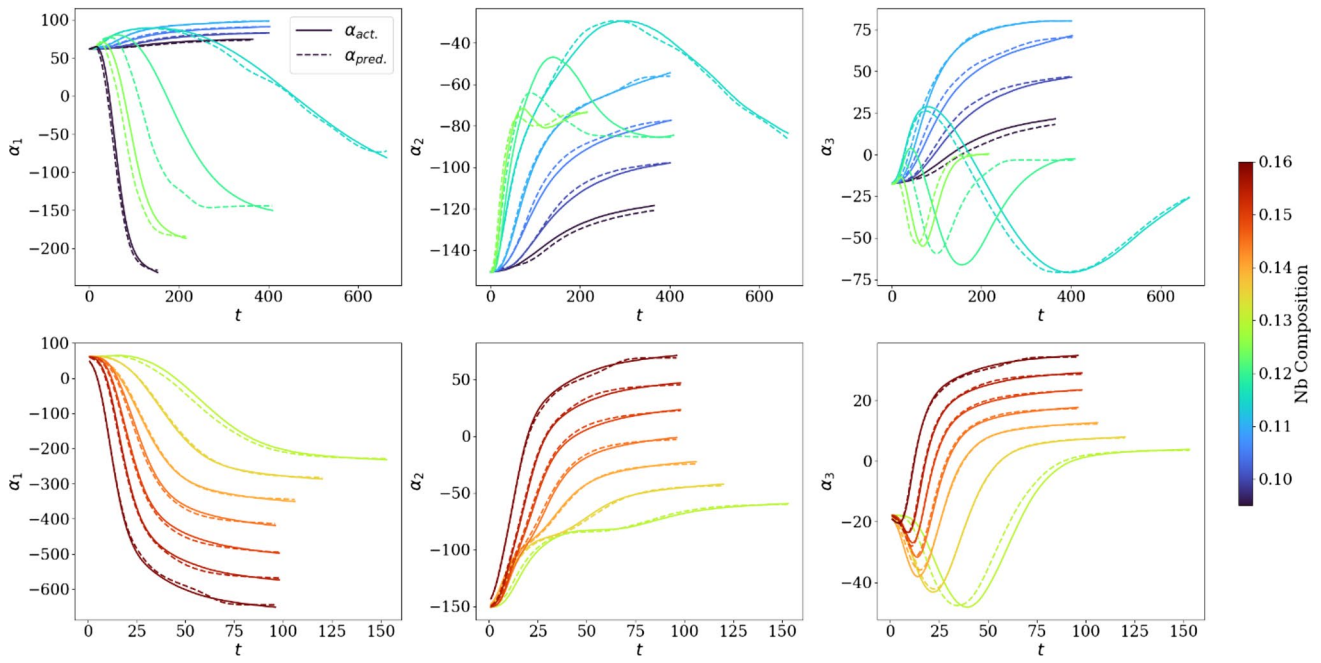

Figure 7 Time evolution of PC1-3 scores predicted by the phase-field simulations and the SV-MOGPAR model for a range of Nb compositions. The phase-field results are shown as solid black lines, while the corresponding SV-MOGPAR predictions are shown as dashed lines coloured by initial Nb concentration. The top row displays time evolutions for Nb compositions from 0.095 to 0.125, while the bottom row displays time evolutions for Nb compositions from 0.13 to 0.16.

Table 3 Mean absolute error (MAE) mean and standard deviation for reconstructions of SV-MOGPAR-predicted correlations compared to original correlations

	γ''–γ''	δ–δ	γ''–δ
Mean	0.00383	0.01354	0.00927
Standard deviation	0.00204	0.00029	0.00009

Table 4 Mean absolute error (MAE) mean and standard deviation for SV-MOGPAR-predicted δ and γ'' volume fractions compared to original δ and γ'' volume fractions

	γ''	δ
Mean	0.064	0.097
Standard deviation	0.039	0.018

Reconstructions of microstructure statistics

Once the microstructure evolution paths had been predicted, it was possible to reconstruct the truncated two-point correlations (i.e. reconstructions obtained from 3 PC scores and basis vectors). The reconstructions are obtained by performing matrix multiplication of the PC scores and basis vectors; this procedure has been covered in previous work [32]. Reconstructions of the two-point correlations for both were computed for all 14 MEUMAPPS simulations at the final recorded time step. The reconstructions were computed for both target PC values as well as PC values predicted by the SV-MOGPAR model. Mean absolute error (MAE) was computed between the SV-MOGPAR statistics and the original statistics for each of the 14 final time steps, with these results being presented in Table 3.

Additionally, volume fraction errors may be computed since volume fraction information is contained within the autocorrelations. Volume fraction values were extracted from the SV-MOGPAR-predicted correlations by selecting the origin values of the reconstructed autocorrelations and comparing them to the original volume fraction values. Since reconstructed microstructure statistics can be obtained from PC scores via simple matrix multiplication, going from a set of PC scores to volume fraction values may be considered to be a unique mapping. Additionally, since γ'' and δ are stoichiometric compounds, the average composition of the matrix can be predicted directly from mass balance once the γ'' and δ volume fractions are known.

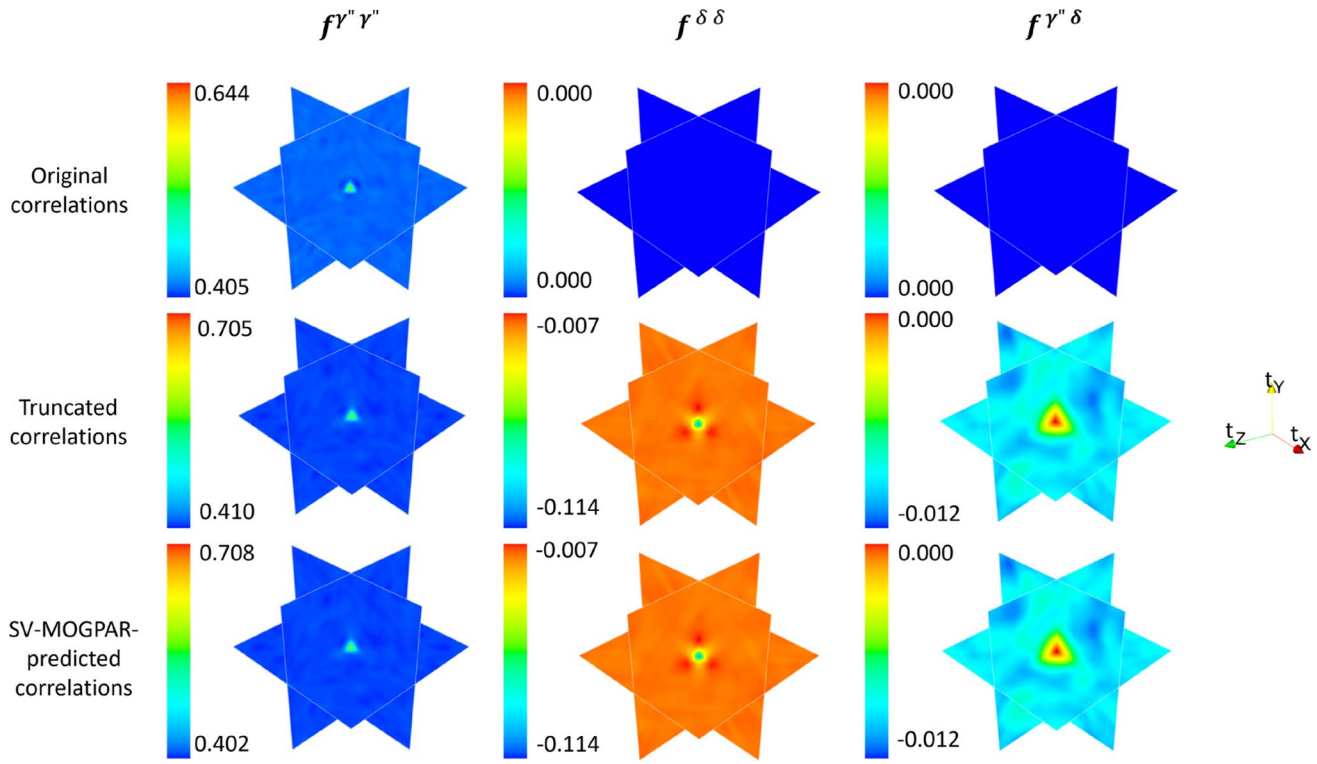

Figure 8 Best-case scenario of reconstructed two-point correlations compared to original two-point correlations. These results were obtained at the final simulation time step for a Nb composition of 0.15.

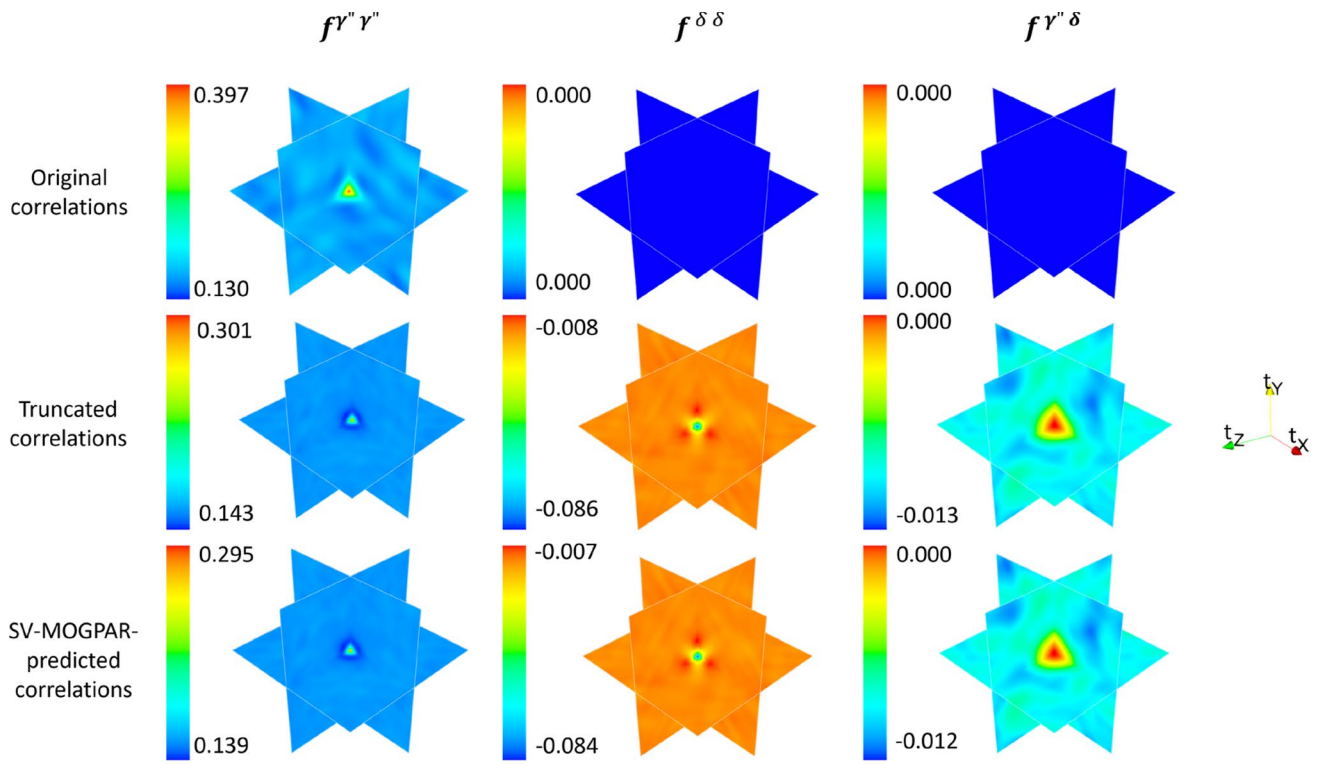

Figure 9 Worst-case scenario of reconstructed two-point correlations compared to original two-point correlations. These results were obtained at the final simulation time step for a Nb composition of 0.12.

MAE was computed for both δ and γ'' volume fractions at each of the final 14 simulation time steps, with the results being presented in Table 4.

By examining the MAE performance for each of the 14 simulations, high- and low-error cases may be identified. Figures 8 and 9 show examples of best- and worst-case scenarios, respectively, for reconstructed correlations in this work.

For both the best- and worst-case scenarios, the majority of the error is derived from the truncation to 3 PC scores and basis vectors. Therefore, performance may potentially be improved by extending the SV-MOGPAR to a larger number of PC scores. Though it was outside the scope of this work, there has been promising work in recovering original microstructure information from microstructure statistics via multi-output Gaussian random fields [70]. Future work could involve applying this approach to the reconstructed microstructure statistics to obtain predictions of the thresholded phase-fields.

Conclusions

A surrogate model using the toolkits in the materials knowledge system consisting of two-point correlations, principal component analysis and SV-MOGPAR was developed by using training data generated by phase-field simulations to capture the microstructural evolution paths during annealing of Ni–Mo–Nb alloys with varying levels of Nb that are likely to be present in the microsegregated regions of the solidification microstructures produced during powder bed-based AM processes. The model was shown to capture reasonably well the highly complex microstructure evolution paths during annealing that is characterized by a bifurcation from the δ phase dominated structure at low Nb contents to a γ'' structure at high Nb contents. The Langevin noise-based nucleation model used in the phase-field simulation posed significant issues due to the uncertainty involved in separating the artificially introduced noise from the physics-based evolution of the supercritical nuclei and their growth.

Importantly, this work provides several notable extensions on prior efforts in modelling the evolving spatial correlations of heterogeneous microstructures. Most notably, the microstructure evolution pathways demonstrated coupled evolution of their low-dimensional PC representation that provided accurate predictions across PC scores. The use of a multi-output Gaussian process, incorporating such interdependencies facilitated the production of accurate trajectory predictions. These interdependencies between low-dimensional microstructure features are a natural consequence of considering multiple precipitate phases and variants, and adjoining their corresponding two-point correlations prior to performing PCA. This work establishes a foundational framework for future efforts aimed at the modelling of high-dimensional microstructure evolution pathways in complex material systems.

Acknowledgements

This research was partially supported by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. DOE Office of Science and the National Nuclear Security Administration at the Oak Ridge National Laboratory. This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under contract DE- AC05-00OR22725. SK acknowledges support from NSF 2119640. The authors acknowledge drawing inspiration from the many research works of Prof. Zbib, and are grateful for the opportunity to submit our work to the special issue honouring him.

Author contributions

M conducted phase-field simulations, performed feature engineering tasks, and contributed to manuscript. Generale built SV-MOGPAR model and contributed to manuscript. K provided guidance and feedback throughout the process, and contributed to manuscript. R provided expertise on phase-field simulations and the MEUMAPPS code, and contributed to manuscript. B provided guidance and feedback throughout the process.

Data availability

The dataset of PC scores used to train the SV-MOGPAR model for this work as well as the Python code

used for generating predictions may be provided upon request.

Declarations

Conflict of interest The authors have no conflicts of interest or competing interests to report for this work.

Ethical approval Not applicable.

References

[1] Laleh M, Sadeghi E, Revilla RI, Chao Q, Haghdadi N, Hughes AE, Xu W, De Graeve I, Qian M, Gibson I (2022) Progr Mater Sci 133:101051

[2] Baganis A, Bouzouni M, Papaefthymiou S (2021) Metals 11:241

[3] Fleck M, Schleifer F, Holzinger M, Glatzel U (2018) Metall Mater Trans A 49:4146

[4] Schleifer F, Fleck M, Holzinger M, Lin Y-Y, Glatzel U (2020) Phase-field modeling of γ′ and γ″precipitate size evolution during heat treatment of Ni-based superalloys. Superalloys. Springer, p 500

[5] Wen Y, Wang B, Simmons J, Wang Y (2006) Acta Mater 54:2087

[6] Yenusah CO, Ji Y, Liu Y, Stone TW, Horstemeyer MF, Chen L-Q, Chen L (2021) Comput Mater Sci 187:110123

[7] Kalidindi SR (2015) Hierarchical materials informatics: novel analytics for materials data. Elsevier

[8] Alpas A, Zhang J (1994) Metall Mater Trans A 25:969

[9] Huang L, Geng L, Wang B, Wu L (2013) Mater Des 45:532

[10] Nazarian A, Stauber M, Zurakowski D, Snyder BD, Müller R (2006) Bone 39:1196

[11] Oñorbe E, Garcés G, Pérez P, Adeva P (2012) J Mater Sci 47:1085

[12] Baniassadi M, Ahzi S, Garmestani H, Ruch D, Remond Y (2012) J Mech Phys Solids 60:104

[13] Cecen A, Yabansu YC, Kalidindi SR (2018) Acta Mater 158:53

[14] Moore AW, Connolly AJ, Genovese C, Gray A, Grone L, Kanidoris N II, Nichol RC, Schneider J, Szalay AS, Szapudi I (2001) Fast algorithms and efficient statistics: N-point correlation functions, in Mining the Sky. Springer, p 71

[15] Sun Y, Cecen A, Gibbs JW, Kalidindi SR, Voorhees PW (2017) Acta Mater 132:374

[16] Jolliffe I (2005) Encyclopedia of statistics in behavioral science

[17] Rajan K, Suh C, Mendez PF (2009) Stat Anal Data Min ASA Data Sci J 1:361

[18] Suh C, Rajagopalan A, Li X, Rajan K (2002) Data Sci J 1:19

[19] Wold S, Esbensen K, Geladi P (1987) Chemom Intell Lab Syst 2:37

[20] Iskakov A, Yabansu YC, Rajagopalan S, Kapustina A, Kalidindi SR (2018) Acta Mater 144:758

[21] Latypov MI, Kühbach M, Beyerlein IJ, Stinville J-C, Toth LS, Pollock TM, Kalidindi SR (2018) Mater Charact 145:671

[22] Latypov MI, Toth LS, Kalidindi SR (2019) Comput Methods Appl Mech Eng 346:180

[23] Yabansu YC, Iskakov A, Kapustina A, Rajagopalan S, Kalidindi SR (2019) Acta Mater 178:45

[24] Bro R, Smilde AK (2014) Anal Methods 6:2812

[25] Ringnér M (2008) Nat Biotechnol 26:303

[26] Adams BL, Henrie A, Henrie B, Lyon M, Kalidindi S, Garmestani H (2001) J Mech Phys Solids 49:1639

[27] Niezgoda SR, Kanjarla AK, Kalidindi SR (2013) Integr Mater Manuf Innov 2:54

[28] Cecen A, Dai H, Yabansu YC, Kalidindi SR, Song L (2018) Acta Mater 146:76

[29] Gupta A, Cecen A, Goyal S, Singh AK, Kalidindi SR (2015) Acta Mater 91:239

[30] Marshall A, Kalidindi SR (2021) JOM 73:2085

[31] Yang Z, Yabansu YC, Al-Bahrani R, Liao W-K, Choudhary AN, Kalidindi SR, Agrawal A (2018) Comput Mater Sci 151:278

[32] Harrington GH, Kelly C, Attari V, Arroyave R, Kalidindi SR (2022) Integr Mater Manuf Innov 11:433

[33] Yabansu YC, Steinmetz P, Hötzer J, Kalidindi SR, Nestler B (2017) Acta Mater 124:182

[34] Hashemi S, Kalidindi SR (2021) Comput Mater Sci 188:110132

[35] de Montes Oca Zapiain D, Stewart JA, Dingreville R (2021) Comput Mater 7:3

[36] Yabansu YC, Rehn V, Hötzer J, Nestler B, Kalidindi SR (2019) Modell Simul Mater Sci Eng 27:084006

[37] Hashemi S, Kalidindi SR (2023) Int J Plast 162:103532

[38] Deng Z, Hu X, Lin X, Che Y, Xu L, Guo W (2020) Energy 205:118000

[39] Cheruvathur S, Lass EA, Campbell CE (2016) Jom 68:930

[40] Khan MA, Gupta K (2020) Trans Indian Inst Met 73:429

[41] Sangid MD, Book TA, Naragani D, Rotella J, Ravi P, Finch A, Kenesei P, Park J-S, Sharma H, Almer J (2018) Addit Manuf 22:479

[42] Teixeira Ó, Silva FJ, Ferreira LP, Atzeni E (2020) Metals 10:1006

[43] Lass EA, Stoudt MR, Katz MB, Williams ME (2018) Scripta Mater 154:83
[44] Azadian S, Wei L-Y, Warren R (2004) Mater Charact 53:7
[45] Huang Y, Langdon TG (2007) J Mater Sci 42:421
[46] Mahadevan S, Nalawade S, Singh JB, Verma A, Paul B, Ramaswamy K (2010) Evolution of δ phase microstructure in alloy 718. In: Proceedings of the 7th International Symposium on Superalloy
[47] Sundararaman M, Mukhopadhyay P, Banerjee S (1988) Metall Trans A 19:453
[48] Hong S, Chen W, Wang T (2001) Metall Mater Trans A 32:1887
[49] Du J, Lu X, Deng Q, Qu J, Zhuang J, Zhong Z (2007) Mater Sci Eng A 452:584
[50] Gallmeyer TG, Moorthy S, Kappes BB, Mills MJ, Amin-Ahmadi B, Stebner AP (2020) Addit Manuf 31:100977
[51] Teng Q, Li S, Wei Q, Shi Y (2021) J Manuf Process 61:35
[52] Radhakrishnan B, Gorti S, Acharya R, Song Y (2022) Predictive tools for customizing heat treatment of additively manufactured aerospace components. ORNL/TM-2022/2474. https://doi.org/10.2172/1883850
[53] Radhakrishnan B, Fatebert J-l, Gorti SB, Haxhimali T, El-Wardany T, Acharya R, Staroselsky A (2017) Integrated predictive tools for customizing microstructure and material properties of additively manufactured aerospace components, ORNL/TM-2017/718. https://doi.org/10.2172/1414688
[54] Radhakrishnan, B., S.B. Gorti, J.A. Turner, and R. Acharya.
[55] Song Y, Radhakrishnan B, Gorti S, Acharya R (2021) Phys Rev Mater 5:053401
[56] Andersson J-O, Helander T, Höglund L, Shi P, Sundman B (2002) Calphad 26:273
[57] Saunders N, Miodownik AP (1998) CALPHAD (calculation of phase diagrams): a comprehensive guide. Elsevier
[58] Steinbach I, Apel M (2006) Physica D 217:153
[59] Chen LQ, Shen J (1998) Comput Phys Commun 108:147
[60] Pekurovsky D (2012) SIAM J Sci Comput 34:C192
[61] Turner JA, Belak J, Barton N, Bement M, Carlson N, Carson R, DeWitt S, Fattebert J-L, Hodge N, Jibben Z (2022) Int J High Perform Comput Appl 36:13
[62] Shi R, Wang Y (2013) Acta Mater 61:6006
[63] Heo TW, Chen L-Q (2014) Jom 66:1520
[64] Cecen A, Fast T, Kalidindi SR (2016) Integr Mater Manuf Innov 5:1
[65] Fullwood DT, Niezgoda SR, Kalidindi SR (2008) Acta Mater 56:942
[66] Gao X, Przybyla C, Adams B (2006) Metall Mater Trans A 37:2379
[67] Fullwood DT, Niezgoda SR, Adams BL, Kalidindi SR (2010) Prog Mater Sci 55:477
[68] Choudhury A, Yabansu YC, Kalidindi SR, Dennstedt A (2016) Acta Mater 110:131
[69] Latypov MI, Kalidindi SR (2017) J Comput Phys 346:242
[70] Robertson AE, Kalidindi SR (2022) Acta Mater 232:117927
[71] Niezgoda S, Fullwood D, Kalidindi S (2008) Acta Mater 56:5285
[72] Fullwood D, Kalidindi S, Niezgoda S, Fast A, Hampson N (2008) Mater Sci Eng A 494:68
[73] Niezgoda SR, Yabansu YC, Kalidindi SR (2011) Acta Mater 59:6387
[74] Rasmussen CE, Williams CK (2006) Gaussian processes for machine learning. Springer
[75] Alvarez MA, Rosasco L, Lawrence ND (2012) Found Trends Mach Learn 4(3):195–266. https://doi.org/10.1561/2200000036
[76] Journel AG, Huijbregts CJ (2003) Mining geostatistics. Blackburn Press
[77] Titsias M (2009) Proceedings of the Twelth International-Conference on Artificial Intelligence and Statistics, PMLR 5:567–574
[78] Mohammadpour P, Yuan H, Phillion A (2022) Addit Manuf 55:102824
[79] Zhou N, Lv D, Zhang H, McAllister D, Zhang F, Mills M, Wang Y (2014) Acta Mater 65:270
[80] Kingma DP, Ba J (2014) arXiv:1412.6980