

ARTICLE OPEN



Data-driven Bayesian model-based prediction of fatigue crack nucleation in Ni-based superalloys

Maxwell Pinz¹, George Weber¹, Jean Charles Stinville², Tresa Pollock² and Somnath Ghosh³✉

This paper develops a Bayesian inference-based probabilistic crack nucleation model for the Ni-based superalloy René 88DT under fatigue loading. A data-driven, machine learning approach is developed, identifying underlying mechanisms driving crack nucleation. An experimental set of fatigue-loaded microstructures is characterized near crack nucleation sites using scanning electron microscopy and electron backscatter diffraction images for correlating the grain morphology and crystallography to the location of crack nucleation sites. A concurrent multiscale model, embedding experimental polycrystalline microstructural representative volume elements (RVEs) in a homogenized material, is developed for fatigue simulations. The RVE domain is modeled by a crystal plasticity finite element model. An anisotropic continuum plasticity model, obtained by homogenization of the crystal plasticity model, is used for the exterior domain. A Bayesian classification method is introduced to optimally select informative state variable predictors of crack nucleation. From this principal set of state variables, a simple scalar crack nucleation indicator is formulated.

npj Computational Materials (2022)8:39; <https://doi.org/10.1038/s41524-022-00727-5>

INTRODUCTION

Ni-based superalloys, such as René 88DT, are widely used in high-temperature applications especially in the aerospace propulsion industry, on account of their excellent thermomechanical and chemical properties. Despite significant advantages, the fatigue life of these materials is compromised due to the presence of heterogeneities in the microstructure. While very few metallurgical defects like inclusions or voids are found in their polycrystalline microstructures, typically manufactured by powder metallurgy techniques, heterogeneities manifest as spatial variations in the crystallographic and grain morphology distributions, as well as intragranular compositions¹. Additionally, annealing twins are critical heterogeneities that form in the polycrystalline microstructures during processing. These microstructural variations are major contributing factors to their fatigue failure under operating conditions².

A number of studies, e.g. in^{3–7} have observed nucleating cracks near coherent $\Sigma 3$ twin boundaries under fatigue loading conditions. For the polycrystalline René 88DT, fatigue cracks are often found to nucleate at these twin boundaries, especially where the adjacent grains also have a high Schmid factor^{1,8}. Crack nucleation is preceded by the formation of slip bands that are parallel to the coherent twin boundaries, yet slightly offset from the boundary itself^{9,10}. In addition to the high Schmid factor, high elastic modulus mismatch across the boundary and long twin boundaries have also been associated with the probable locations of crack nucleation in^{11–13}. A hypothesis behind such observations of crack nucleation is that these geometric and crystallographic metrics can be correlated with high values of the resolved shear stress on shared slip systems between twin and parent grains, parallel to the twin boundary^{14–19}. The studies conjecture that rather than high Schmid factors, the high resolved shear stresses promote dislocation glide along the shared plane leading to nucleation. Out-of-plane forest dislocations do not disrupt slip on the shared plane, and hence dislocation glide is unimpeded along the boundary. With the progression of cyclic loading, highly localized slip parallel to the

twin boundary trigger the formation of persistent slip bands (PSB), which precede crack nucleation and propagation²⁰. Experiments on Ni-based superalloys have shown that PSBs can form at strains below 0.35% within a single cycle of deformation⁹.

These observations necessitate the development of robust mechanism-driven predictive models that can relate evolving state variables in the underlying polycrystalline microstructure to the nucleation of cracks. The last two decades have witnessed significant advances in the understanding and prediction of the connection between microstructure and local mechanical fields, along with critical failure indicators under fatigue loading. This has been enabled by a combination of advanced experimental characterization techniques like electron backscatter diffraction (EBSD), high-energy diffraction microscopy (HEDM), and X-ray tomography/diffraction techniques, and computational models like adaptive finite element methods (FEM) and fast Fourier transform (FFT) methods integrated with crystal plasticity constitutive models. In the computational modeling approaches, virtual 2D and 3D polycrystalline microstructures are generated from materials characterization data, and subsequently image-based micromechanical simulations are conducted to interrogate the locations of crack evolution^{21–23}. Comprehension of material states near the critical cracked regions has guided researchers in constructing metrics of fatigue and failure indicators. Both physics-based and data-driven approaches have been pursued to create the appropriate fatigue metrics, such as dislocation density and plastic slip-based criteria combined with crack opening stresses for susceptible grain configurations^{24,25}, energetics and stability of PSBs in²⁶, accumulated plastic slip and elastic anisotropy in¹⁴, and cyclic stored plastic energy density^{21,23,27,28}. Important fatigue indicator parameters have thereby been derived through the use of correlation metrics, variance-based sensitivity, or physics-derived arguments in combination with machine learning and Bayesian techniques^{21,22,29–31}.

¹Civil & Systems Engineering, Johns Hopkins University, Baltimore, MD 21218, USA. ²Materials Department, University of California, Santa Barbara, CA 93106, USA. ³Departments of Civil & Systems Engineering, Mechanical Engineering, and Materials Science & Engineering, Johns Hopkins University, Baltimore, MD 21218, USA. ✉email: sghosh20@jhu.edu

Despite significant advances, shortcomings, e.g. in experimental throughput rates and computational efficacy, have in some cases mitigated the effectiveness of these models. Limitations in the collection of experimental data sets linking microstructural configurations to crack nucleation sites have often led to the development of models from very limited microstructural data sets, typically encompassing a single grain configuration. Alternatively, the microstructural geometry has been abandoned entirely in favor of higher-scale statistical analysis. The lack of realistic experimental data leads to inherent uncertainty in the model output. The resulting fatigue indicators are not likely to retain fidelity with the introduction of new data. Furthermore, probabilistic methods, required to handle multiple experiments in a high-dimensional input data space, suffer from various issues resulting from the problem dimensionality. From a modeling point of view, the computational methods must ensure accurate determination of state variables like stresses, plastic strains, dislocation densities, etc. at potential crack nucleation sites for effective development of mechanics-based models of crack nucleation. Boundary conditions play a very important role in the prediction of localized variables in the image-based micromechanical models^{32,33}. Conventionally used periodicity boundary conditions can lead to gross inaccuracies when modeling fatigue behavior.

This paper introduces a data-driven Bayesian model framework for predicting fatigue crack nucleation in polycrystalline microstructures of the Ni-based superalloy René 88DT, overcoming some of the limitations mentioned above. The overall approach leverages large data sets obtained from crystal plasticity FEM-based micromechanical simulations with statistics and machine learning algorithms. Within this framework, the probability of observing a crack nucleation event is derived based on computationally generated micromechanical data on state variables at a potential nucleation point. High fidelity, image-based crystal plasticity finite element (CPFE) models developed in^{34–37} are adopted for simulating evolving state variables in microstructural representative volume elements (RVE) of polycrystalline Ni-based superalloys. The deficiencies with improper boundary conditions on the RVE is alleviated in this work through a concurrent multiscale model, where the microstructural RVE is

embedded in a homogenized exterior material domain. The exterior domain is modeled using an upscaled, anisotropic continuum plasticity constitutive model that is obtained from homogenization of crystal plasticity RVE simulations^{38–44}. This model is discussed in the Supplementary Methods Section.

The paper introduces a robust Bayesian model framework for extracting fatigue nucleation criterion using a combination of simulation and experiments. The primary contributions of this framework are delineated below.

- The probabilistic crack nucleation model accounts for multiple experimental microstructures that are simulated using crystal plasticity FEM with one-to-one virtual representations of EBSD scans. A major factor for model calibration and validation is the generation of statistically significant results over a set of multiple experimental microstructures, covering a range of microstructural variabilities. By incorporating multiple one-to-one experimental data sets directly in the probabilistic modeling framework, the approach mitigates uncertainties arising from a single microstructure dependency and insufficient validation data. An important additional aspect is the accounting for material states of both cracked and uncracked regions in the development, through efficient algorithms for handling large data sets.
- The Bayesian crack nucleation model is formulated using a customized Nataf transformation to represent multivariate joint posteriors. The resulting probabilistic framework yields several advantages over other comparable methods, viz. (i) semi-analytical expressions for high-dimensional joint probability distributions, (ii) easy sampling of arbitrary points in the data-space with no reliance on specialized sampling techniques, (iii) simple calculation of moments and measures on the joint probability distribution, and (iv) straightforward regularization of discrete data that reduces a generalized model error. These advantages enable efficient use of the final posterior joint distribution in predicting crack nucleation.
- The automated simulation framework creates pipelines, which take a set of EBSD scans delineating grain configurations and crack nucleation sites as input, and returns a reduced order state-variable-based Bayesian model for probabilistic nucleation prediction as shown in Fig. 1. The framework integrates

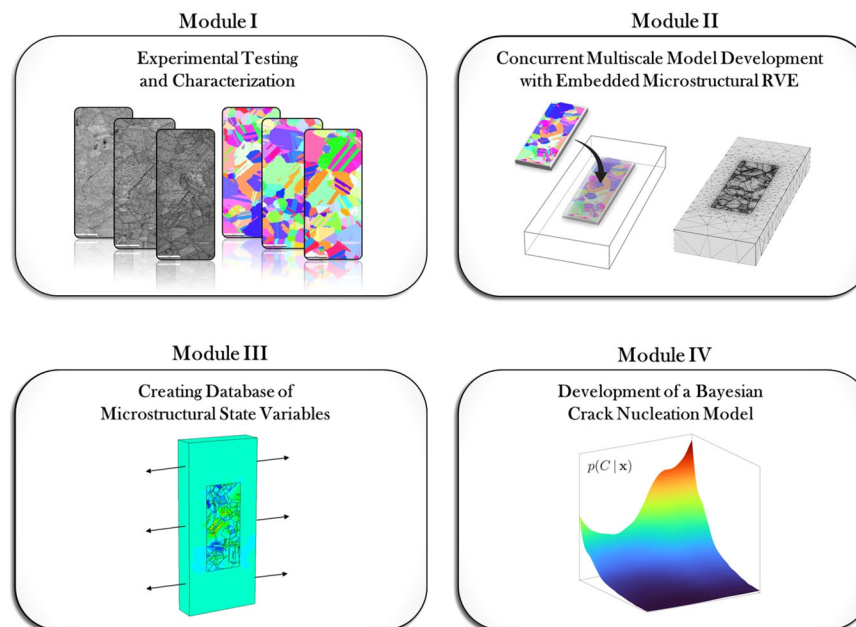


Fig. 1 Four modules comprising the pipeline for developing the data-driven crack nucleation model. (I) experimental testing and characterization data, (II) development of the concurrent multiscale model with embedded microstructural RVE, (III) creation of a database of microstructural state variables, and (IV) development of the Bayesian crack nucleation model.

- multiple computational procedures into a continuous workflow.
- The model introduces an algorithm for selecting important state variables and isolating critical metrics that indicate locations of crack nucleation. This algorithm assesses successive multi-dimensional joint distributions and accounts for highly correlated sets of variables to be incrementally optimal in its predictions.
- The automated procedure with minimal intervention uses experimental data to inform the probabilistic model of the most informative parameters. These parameters yield mechanistic insights into the drivers of crack nucleation for this material. The resulting probabilistic model demonstrates robust performance across variations in the microstructure and local grain configuration.
- A scalar crack nucleation indicator is derived from the simulation platform.

The process of developing a probabilistic crack nucleation modeling framework integrates experimental testing and characterization with a pipeline of simulation tools for mechanical modeling and data-driven, supervised learning techniques. The sequence of four major modules in this pipeline is schematically depicted in Fig. 1.

Experimental testing and characterization

In this module, specimens of the Ni-based superalloy René 88DT are subjected to interrupted fatigue loading at various stages of deformation, to track the precursors of crack nucleation. Scanning electron microscopy (SEM) maps are acquired and utilized for identifying microstructural locations of crack nucleation. Simultaneously, EBSD maps are acquired to determine the crystallographic orientations and grain morphologies at locations of interest in the microstructure.

Concurrent multiscale model development with embedded microstructural RVE

This module builds a two-scale concurrent model, consisting of an embedded polycrystalline microstructural RVE in a homogenized outer domain. Important microstructural deformation variables are extracted from simulations of the concurrent model of the experimental specimen. The two-scale concurrent representation is constructed to avoid direct application of boundary conditions on the RVE, thereby averting boundary condition-related inaccuracies in micromechanical fields. The embedded microstructural region is generated by extruding 2D EBSD scans in the out-of-plane direction. It is modeled using a crystal plasticity constitutive model for individual grains in the polycrystalline microstructure, while the outer domain is modeled by a homogenized, rate-dependent elasto-plastic constitutive model that is calibrated to the average mechanical response of the embedded RVE.

Creating a database of microstructural state variables

The concurrent multiscale model is simulated with fatigue loading conditions applied on the external boundary, from which the evolving state variables in the embedded microstructural RVE region are extracted. The state variables are then processed and archived in a crystal plasticity simulation database for subsequent stages of the data-driven modeling.

Development of a Bayesian crack nucleation model

The crystal plasticity simulation database provides the input data for the modeling framework to determine the probability of crack nucleation corresponding to a mechanical state in the material volume. The Bayesian model selects the most informative state variables for predicting crack nucleation by navigating the covariance structure of candidate predictor variables. The results

are used to create a simple model for identifying potential locations of crack nucleation in the microstructure.

Each module is standardized and automated so that the model building process can be rapidly and systematically repeated for all experimental microstructures. Furthermore, this approach allows for the extension to other critical events as well as the inclusion of new data sets.

RESULTS

Material description

The material investigated in this study is a polycrystalline Ni-based superalloy René 88DT. Generally produced by powder metallurgy techniques^{45,46}, René 88DT has a nominal chemical composition (weight percent) of 13% Co, 16% Cr, 4% Mo, 4% W, 3.7% Ti, 0.7% Nb, 0.03% C, 0.015% B, and the balance in Ni⁴⁷. The intragranular microstructure of René 88DT consists of an FCC-structured Ni matrix (γ phase) and an embedded population of L1₂-structured Ni₃Al precipitates (γ' phase) with a bimodal size distribution of large (~100 nm) secondary and small (<10 nm) tertiary γ' precipitates. At the polycrystalline scale, the material has an average grain size of about 26 μ m, very weak crystallographic texture, and a significant population of Σ 3 twin boundaries. Details on the René 88DT microstructure are given in^{8,9}.

Experiments on fatigue tests and data acquisition

An experimental data set consisting of 12 specimens with independent grain structure configurations is utilized for developing the Bayesian model in the present study. Room temperature, low-cycle fatigue tests are conducted to evaluate the fatigue life of the Ni-based superalloy René 88DT, for which the mechanical testing procedures are detailed in⁹. The tests are performed in air on an electromechanical machine in a uniaxial, symmetric push-pull mode. Fully reversed loading tests with an R-ratio equal to -1 at a frequency of 1 Hz, are carried out in a stress-controlled mode at a maximum stress of 758 MPa. Cylindrical specimens with a gauge diameter of 5 mm and a gauge length of 16 mm are used in these tests. The strains in the specimen are measured by a mechanical extensometer that is positioned on the specimen gauge. The average fatigue lifetime of the specimens is observed to be 65340 cycles with a scatter (standard deviation) of 5530 cycles. A maximum and minimum total strain of $0.35\% \pm 0.04\%$ and $-0.35\% \pm 0.03\%$ are measured for the first cycle. The material exhibits cyclic evolution of the strain, with the maximum and minimum strain stabilizing at $0.40\% \pm 0.05\%$ and $-0.30\% \pm 0.04\%$ respectively, after half of the lifetime i.e. around 30–40,000 cycles. The strain range remains constant at $0.70\% \pm 0.03\%$ during the entire fatigue test. Interrupted tests are performed to enable specimen surface examination using SEM and EBSD measurements at about 45% of the average fatigue lifetime (~30,000 cycles). The SEM scans identify regions exhibiting evidence of micro-crack formation, while the EBSD measurements interrogate the configuration of the surface grains surrounding the micro-cracks. The interrupted tests capture two types of cracks, viz. those confined to one grain⁴⁸ (type I) and those extended between one and three-grain diameters, crossing one or two grain boundaries on the surface (type II). This approach facilitates the identification of the low-cycle crack nucleation sites that are characteristic for each microstructure. A few typical microstructures, showing nucleated cracks under the fatigue loading conditions, are shown in Fig. 2.

Experimentally generated polycrystalline microstructures

Subsequent to EBSD acquisition, an automated data processing pipeline is applied to each of the 12 investigated grain structures in the microstructure generator software DREAM.3D⁴⁹ to remove

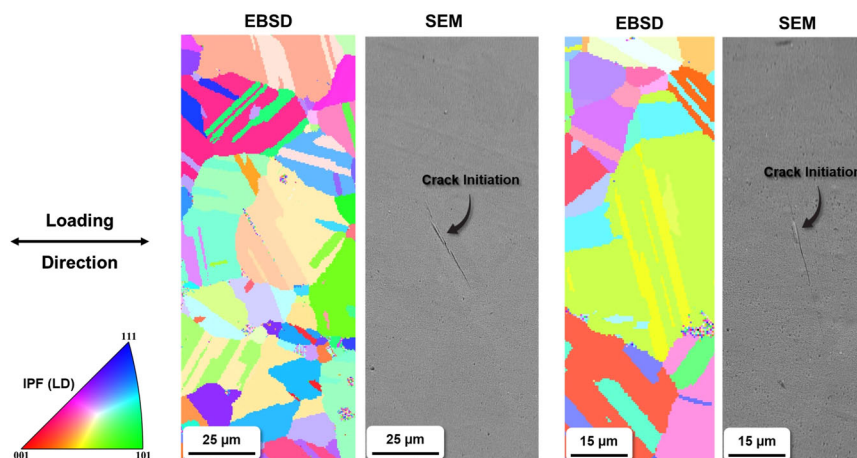


Fig. 2 Two sample experimental data sets used in the development of the Bayesian crack nucleation model. The SEM images display fatigue crack nucleation near and parallel to the twin boundaries. The associated EBSD maps present the grain structure around the crack nucleation sites.

any measurement defects. Low image quality voxels are identified and automatically filled. Grains are determined by grouping voxels within a 5° misorientation angle threshold. Grains containing an area $< 20 \mu\text{m}^2$ are merged with their surrounding grains using a nearest-neighbor identification method. These processing stages result in a collection of 2D segmented images, free of artifacts that can cause difficulties with generating robust finite element meshes.

Key experimental observations on fatigue crack nucleation

The primary sources of heterogeneity in René 88DT are derived from variations in microstructural features since minimal metallurgical defects like inclusions and voids are generated with powder metallurgy processing. The most prominent microstructural features are annealing twins, whose boundaries are uniformly dispersed throughout the grain structure during heat treatment. These boundaries have unique crystallographic and morphological properties and thereby affect dislocation motion during material deformation.

Prior research, e.g. in^{1,9,50,51}, suggests a number of experimental correlations between the material grain structure and physical location of crack nucleation sites. A few key observations are delineated below:

1. Fatigue crack nucleation and the highest degree of strain localization are systematically reported to occur near and parallel to $\Sigma 3$ twin boundaries in the low, high and very high cycle regimes⁹;
2. Crack nucleation always occurs between pairs of grains, which are relatively oriented in a parallel slip configuration;
3. Additional geometric, crystallographic, and mechanical factors, such as high Schmid factor, large difference in elastic modulus between adjacent grains, and long straight twin boundary length, have been shown to be associated with crack nucleation sites;
4. Slip bands at crack nucleation sites can form as early as the first cycle, suggesting state variables at early stages of deformation are predictive of the crack nucleation locations;
5. For both high and low-cycle fatigue regimes, the previous studies have reported that all crack nucleation sites are associated with twin boundaries that display parallel slip configurations¹⁵.

The above observations are taken into account in this study for predicting the microstructure and grain configurations that

generate crack nucleation. Next the pipeline requires the development of a concurrent multiscale model for fatigue analysis, followed by the creation of a microstructural state variable database related to fatigue nucleation. These are respectively discussed in Sections I and II of the Methods Section.

A Bayesian framework for predicting crack nucleation

A stochastic machine learning (ML) method is developed for assessing the probability of crack nucleation in the René 88DT superalloy microstructure. The Bayesian framework, outlined in the Methods Section III, is chosen for generating the probability of observing a crack nucleation event, as a function of the micromechanical state in the microstructure. This is a supervised learning classification approach that is effective with a limited set of high-dimensional data. It has model parameters that are physically interpretable functions of the micromechanical state variables and offers a level of uncertainty in its predictions. The Bayesian modeling framework naturally yields a probabilistic outcome as opposed to a binary (0 or 1) response, which is advantageous over other machine learning classification techniques. The finite element simulation data, generated by the concurrent model in the Methods Section I, is aggregated over all microstructures and used as training data for this stochastic model. The results of these simulations are collected into a segmented database containing over 20 potential crack nucleation predictors in Methods Section II.

A Bayesian framework for predicting crack nucleation

Constructing a Bayesian model using all of these parameters would result in over-fitting, providing little useful information. To remedy this, a model selection technique is developed to systematically reduce the number of state variables to only those that most influence the posterior distribution $p(C|\mathbf{x})$ as a combined set. The objective is to select a reduced order model with only the most significant state variables from the possible $k = 20$ state variables in the data set \mathbf{X} listed in Table 2 as well as their regularized difference terms.

A recursive build-up approach is executed to add state variables one by one to the model and assess the incremental value of each new addition. In order to compare across multiple models, an objective function is designed to quantify the overall effectiveness of each posterior distribution. Let the objective function ρ_C be

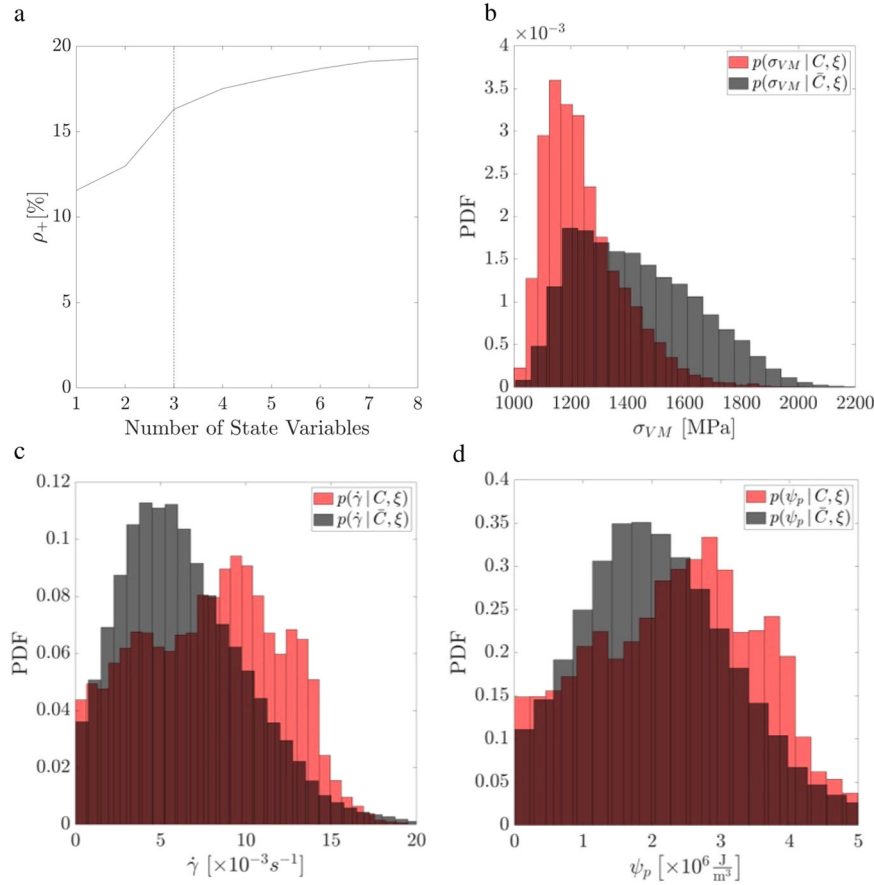


Fig. 3 Most predictive state variables for crack nucleation locations. **a** Increasing geometric mean probability of prediction for known cracked material volumes with incremental addition of state variables into the posterior PDF; Cracked and uncracked marginal likelihood distributions for **b** the von Mises stress σ_{VM} , **c** the maximum plastic slip-rate $\dot{\gamma}$, and **d** the plastic defect energy ψ_p .

defined as:

$$\rho_C = \exp\left(\sum_i^{N_C} w_C^{(i)} \log(p(C|\mathbf{x}^{(i)}))\right) \quad (1)$$

where N_C is the total number of cracked material volumes across all microstructures and $w_C^{(i)}$ are the normalized volumes for each, i.e. $\sum_i^{N_C} w_C^{(i)} = 1$. This objective function ρ_C represents the geometric mean of the probability that the model correctly labels each known cracked volume of the observed data set. In practice, $\log \rho_C$ is used for numerical computations.

The reduced order posterior model $p(C|\mathbf{x}_*)$ is determined by choosing a subset of state variables $\mathbf{X}_* \subset \mathbf{X}$ that incrementally maximizes ρ_C for a fixed number of state variables $k_* \leq k$. This optimization procedure is described in Algorithm 1 in the Supplementary Methods Section. The fundamental idea is to train the posterior PDF with each of the state variables individually, and assess the effectiveness of each model by comparing their respective ρ_C . After determining the most informative state variable with the maximal objective value, the model is trained again with each of the remaining state variables individually, as well as the previously chosen most informative state variable, in order to select the second most informative state variable. This incremental process is repeated until a priori chosen $k_* \leq k$ state variables are established. The incremental selection process algorithm is given in the Supplementary Methods Section under Algorithm 1.

The incremental selection process is executed with the present data set \mathbf{X} comprised of $k = 20$ state variables, using $k_* = 8$. In each

iteration, the method optimally adds the most informative state variable. With each additional state variable included in the probabilistic model, the overall performance characterized by ρ_C increases. Figure 3a plots the incremental effectiveness of the posterior with additional state variables. The slope of the plot gradually saturates as less new and independent crack-delineating information is provided to the joint posterior PDF from the additional simulation output data.

The above procedure yields a systematic ranking of state variables from the data set in terms of effectiveness, into a set \mathbf{X}_* . It provides a clear path to reduce the degrees of freedom in the final model. The top three state variables determined by this process are:

1. The von Mises stress σ_{VM} .
2. The maximum plastic slip-rate $\dot{\gamma}$.
3. The plastic defect energy ψ_p .

These three quantities contain the most combined information that is capable of distinguishing between cracked and uncracked material volumes within the regions undergoing parallel slip. The marginal distributions for these state variables are shown in Fig. 3b–c, illustrating the overall trends for the crack data separation. Generally, low σ_{VM} , high $\dot{\gamma}$, and high ψ_p all correlate with an increased likelihood of cracked nucleation at the material volumes in a parallel slip configuration.

A reduced-order posterior is generated from these three state variables by defining a set $\mathbf{x}_r = [\sigma_{VM}, \dot{\gamma}, \psi_p]$. The three-dimensional likelihood PDFs $p(\mathbf{x}_r|C, \xi)$ and $p(\mathbf{x}_r|\bar{C}, \xi)$ are

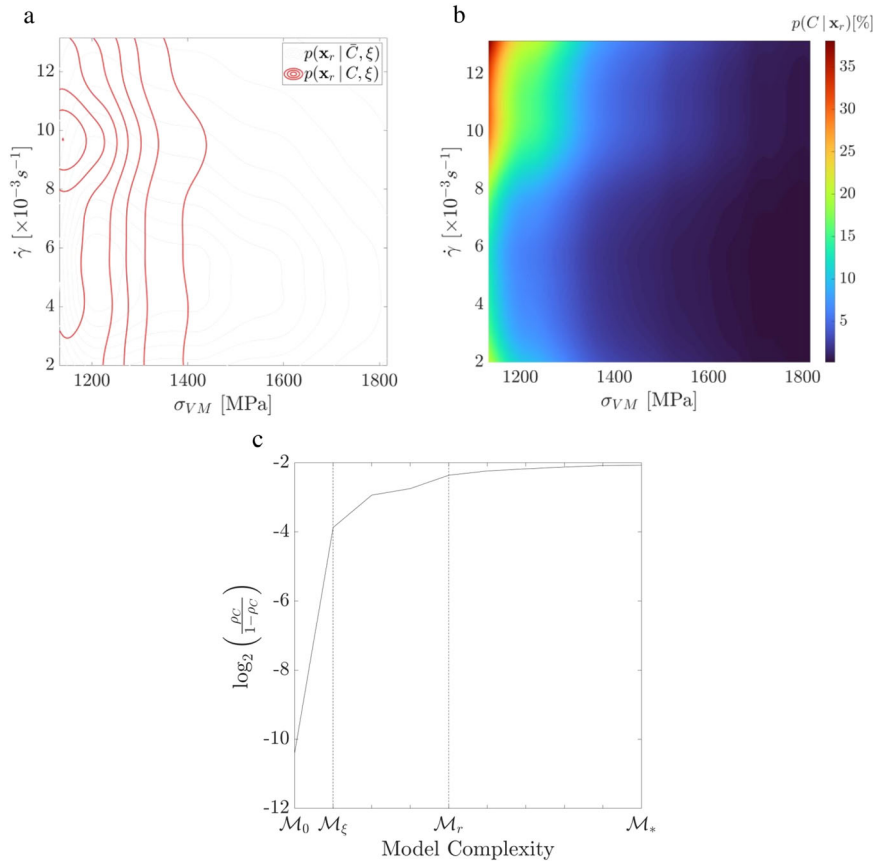


Fig. 4 Illustration of crack nucleation model performance. **a** 2D projection of the iso-probabilistic contours of the final uncracked and cracked likelihood PDFs, and **b** 2D projection of the final posterior joint PDF indicating high crack probability at low σ_{VM} and high $\dot{\gamma}$ values. **c** Log-odds of the geometric mean of the probability of correct crack nucleation prediction for increasingly model complexity. The included models are $\mathcal{M}_0 = p(C)$, $\mathcal{M}_\xi = p(C|\xi)$, $\mathcal{M}_r = p(C|\xi, \mathbf{x}_r)$, and $\mathcal{M}_* = p(C|\xi, \mathbf{x}_*)$.

constructed, and subsequently the resultant posterior PDF $p(C|\mathbf{x}_r)$ is calculated. Two-dimensional projections of the likelihoods are visualized in Fig. 4a, which shows the tendency of the cracked data to separate according to the values of σ_{VM} and $\dot{\gamma}$. Incorporating the priors yields the final projected posterior as shown in Fig. 4b, where the highest probability of crack nucleation commences for low σ_{VM} and high $\dot{\gamma}$ values.

The posterior PDF, when compared with other existing crack nucleation models of varying complexity, demonstrates excellent improvement. The original prior $\mathcal{M}_0 = p(C)$, the prior after conditioning on parallel slip configuration regions $\mathcal{M}_\xi = p(C|\xi)$, and additional posteriors including an increasing number of state variables, are ranked according to their predictive capacity on the training data set. In Fig. 4c, the log-odds of each PDF's geometric mean probability ρ_C (in bits) is used as the measure of comparison and is calculated as $\log_2\left(\frac{\rho_C}{1-\rho_C}\right)$. The log-odds gives a natural interpretation of the strength of evidence gained by adding model complexity to a probabilistic model⁵². A unit increase in log-odds corresponds to an additional bit of evidence. This is roughly interpreted as observing 2:1 evidence in support of the model's hypothesis, i.e. that it correctly predicts crack nucleation sites. In this context, $p(C|\xi)$ adds about 6.5 bits of evidence over the naive $p(C)$, by requiring that crack nucleation occurs at material volumes in a parallel slip configuration. Furthermore, $\mathcal{M}_r = p(C|\xi, \mathbf{x}_r)$ adds about 1.5 bits of evidence on top of $p(C|\xi)$, by incorporating the top three state variables in the posterior PDF. The model therefore greatly improves the probability of locating a

rare crack nucleation event, by about 8 bits relative to the original prior from the data.

From a mechanistic point of view, an interpretation is offered for the model's emphasis on σ_{VM} , $\dot{\gamma}$, and ψ_p in terms of the experimental evidence supporting long, thin cracks in PSB regions at parallel slip configurations. The von Mises stress, which is a norm of the deviatoric stress, typically represents a driving force for the volume-preserving shear deformation such as plasticity and dislocation motion. However, since the material volume is already in a parallel slip configuration, a high directionally independent measure of shear stress, such as σ_{VM} , indicates that dislocation slip is likely to occur on multiple slip systems, rather than focusing on only the parallel slip plane. This importance of directionality is further evidenced by a large $\dot{\gamma}$ measure, also acting as a crack-delineating state variable. The combination of the low σ_{VM} and high $\dot{\gamma}$ is an environment that promotes dislocation motion along the primary slip plane, which is the parallel slip plane in a parallel slip configuration. It represses slip along with other secondary slip systems that interferes with the main deformation mode. The plastic defect energy as the third major indicative variable of crack nucleation is linked to the build-up of dislocation substructures that are typical of persistent slip band (PSB) regions. Regions of large ψ_p store more external work into complex and organized arrays of tangled dislocations, rather than dissipating it entirely as plastic work.

After constructing the full posterior distribution $p(C|\mathbf{x}_r)$, probability fields are overlaid on the microstructures to determine the most likely crack nucleation sites. A sample polycrystalline

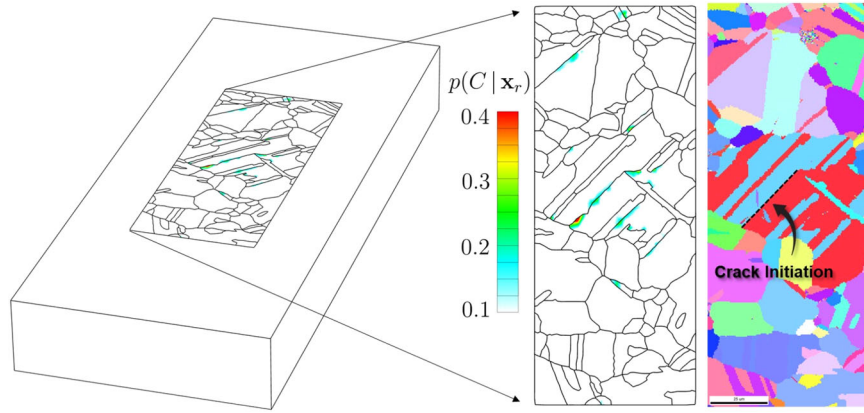


Fig. 5 Comparison of the crack nucleation model predictions with the observed crack locations of an experimental microstructure. The probability of crack nucleation predicted from the posterior distribution in an embedded microstructure of the concurrent multiscale model (red corresponding to the most probable), compared with the corresponding EBSD map and experimental nucleation site.

microstructure is shown in Fig. 5 with the corresponding orientation map from the EBSD scan. The probability contours by the Bayesian model concur with the experimental crack nucleation sites, as probable locations for observing a crack.

Cross-validation

The predictive capacity of the probabilistic model is assessed in this section. Delineation of the training and test data allows for a comparison of the performance of the methodology in terms of both training error and generalization error. Model over-fitting is a concern when the only efficacy measure of a model is on the data it is trained on. Therefore, a “leave-two-out” cross-validation approach is used to ensure the crack nucleation model generalizes onto unseen data. The leave-two-out method is a standard, exhaustive cross-validation approach, in which the model is trained on every available set of data except for two microstructures. The model is then tested/validated on the two microstructures that are left out from the training set. This process is repeated, leaving two different data sets out each time, and retraining the model on the remaining ten microstructures. This approach is one of the best for small data sets and for training a computationally expensive model, as it increases the amount of training data available for each validation evaluation.

For the case of $M = 12$ microstructures, there are $M(M - 1)/2 = 66$ ways to leave two unique data sets out and train on the remaining ten. Therefore, the model is trained and tested 66 times for all combinations. The geometric mean of correct crack nucleation prediction ρ_C is used as a measure of the effectiveness of the model in both training and validation. An additional related measure is defined as:

$$\rho_{\bar{C}} = \exp \left(\frac{1}{N_{\bar{C}}} \sum_i w_{\bar{C}}^{(i)} \log(1 - p(C | \mathbf{x}^{(i)})) \right) \quad (2)$$

where $N_{\bar{C}}$ is the total number of uncracked material volumes across all microstructures of interest, and $w_{\bar{C}}^{(i)}$ are the normalized volumes for each. This measure is the geometric mean of the probability that the model correctly labels every known uncracked material volume of the observed data set.

The arithmetic mean and geometric mean of ρ_C and $\rho_{\bar{C}}$ over all 66 cases are given in Table 1 for both the training and testing data sets, by employing the leave-two-out method. The resulting evaluation measures indicate that the probabilistic model yields similar abilities for both the training and testing sets. Only the geometric mean for the validation cracked cases is slightly lower than that of the test case. This deviation is due to the geometric

Table 1. Mean probabilities of nucleation prediction for all “leave-two-out” cases.

	Training		Validation	
	ρ_C	$\rho_{\bar{C}}$	ρ_C	$\rho_{\bar{C}}$
Arithmetic mean	16.6%	88.8%	16.9%	88.7%
Geometric mean	16.3%	88.8%	13.0%	88.6%

mean’s sensitivity to variance and bias towards low values, which are higher in the testing examples. Overall, these results demonstrate that the model generalizes well, and is predictive on new data.

The predictive behavior of the model is further emphasized by plotting the posterior distribution over a grain structure for one of the 66 leave-two-out cases. Fig. 1 in the Supplementary Methods Section shows the probability field and the corresponding EBSD map over one of the two cases that are not included in the training. The probability field illustrates likely locations for crack nucleation, and in particular, highlights the locations near the experimental crack nucleation site. These regions are comprised of material volumes in a parallel slip configuration with high propensity for primary slip and have already undergone substantial defect build-up.

DISCUSSION

The probabilistic crack nucleation model consisting of a validated three-dimensional posterior PDF can be used to create a simple, heuristic scalar indicator model of crack nucleation probability. A one-dimensional functional form is constructed to become a separator of the cracked and uncracked regions. The dimensionless crack nucleation indicator is defined at material volumes in a parallel slip configuration ($\Xi = \xi$) as:

$$\Psi_C = \kappa_1 \left(\frac{\log(\frac{\dot{\gamma}}{\dot{\gamma}_0})}{\sigma_{VM}^2} \right) + \kappa_2 \psi_p \quad (3)$$

where $\dot{\gamma}$ is the slip-rate, $\dot{\gamma}_0 = 5 \times 10^7$ is the reference slip-rate, ψ_p is the plastic defect energy, and κ_1 , κ_2 are parameters to be determined. The crack nucleation indicator incorporates the basic mechanistic ideas discussed in the previous section. The first term of Ψ_C corresponds to a state of material which heavily favors parallel slip along a boundary rather than homogeneous plastic

deformation across multiple slip systems. The numerator $\log\left(\frac{\dot{\gamma}}{\dot{\gamma}_0}\right)$ is approximately proportional to the effective RSS driving slip through thermal activation, and the denominator σ_{VM}^2 is proportional to the elastic distortion energy. The second term of Ψ , often used in the Helmholtz energy representation of phase-field crack models⁵³, indicates that the material has a tendency to build up dislocation substructures to promote the onset of persistent slip band formation. The proposed crack nucleation indicator encompasses and combines features of fatigue indicator parameters proposed in the literature, such as the emphasis on the plane of maximum plastic slip in²⁴ and the significance of the stored plastic defect energy in²³.

The two parameters κ_1 and κ_2 of Ψ_c are linear coefficients with respect to the inputs of the function. Correspondingly they can be determined using linear discriminant analysis (LDA). LDA is a classification method that finds the linear combination of inputs that best separates two distinct distributions, to find a reduced order model. The application of LDA to the crack nucleation data yields explicit values for the linear coefficients of Ψ_c , i.e. $\kappa_1 = -0.28 \cdot 10^{-18} \text{ Pa}^{-2}$ and $\kappa_2 = 0.26 \cdot 10^{-6} \left(\frac{\text{J}}{\text{m}^3}\right)^{-1}$. This process results in a single scalar value for crack nucleation prediction and is used as a sole indicator in the probabilistic model $p(C|\mathbf{x})$, where $\mathbf{x} = [\Psi_c]$.

Visual analysis of the likelihood distributions of Ψ_c in Fig. 6a establishes it as a useful separator of the cracked and uncracked experimental data. Furthermore, the model in Eq. 3 is assessed similar to the previous posteriors by calculating p_c . Applying the cracked data to the model $p(C|\Psi_c)$ yields $p_c = 11.8\%$, which is superior to building a model with any of the other previously evaluated state variables individually. Consequently, the crack nucleation indicator, which embodies the mechanistic motivation of parallel slip within a persistent slip band, is a consistent measure for identifying cracked regions of the microstructures.

The contour plots of the crack nucleation indicator field are shown for four polycrystalline microstructures in Fig. 2 of the Supplementary Methods Section. This enables a visual comparison to assess its performance with respect to predicting crack nucleation sites. Qualitatively, large values of Ψ_c highlight areas associated with the experimentally identified nucleation sites for each case.

A final study is performed to quantify the sensitivity of the scalar-valued probabilistic model development to the inclusion of new data. The 66 cross-validation cases are revisited for this analysis. For each case, the posterior model is constructed by training with data from ten microstructures, using only the data from the crack nucleation indicator Ψ_c . The posterior $p(C|\Psi_c)$ is generated and plotted as a function of Ψ_c for each case. In Fig. 6b, the 66 models are aggregated into an average model shown with the solid dark line, and standard deviation bounds depicted with dotted lines. This result demonstrates that most models lie within about 5–10% of the average model, and therefore are robust and relatively insensitive to the choice of training data set.

In summary, a robust Bayesian inference-based probabilistic predictive model for crack nucleation in the Ni-based superalloy René 88DT under fatigue loading is developed in this paper. The process fuses several statistical and data-driven methods operating on micromechanical and experimental data with important mechanistic observations. The model systematically reduces a large set of experimentally acquired microstructural data and corresponding crystal plasticity-based micromechanical simulation results to a 3D posterior probability density function (PDF) for crack nucleation. Three hierarchically important state variables in the microstructural material volume, relevant to the crack nucleation process, are identified by the framework. They are: (i) the von Mises stress, (ii) the maximum plastic slip rate, and (ii) the plastic defect energy. This data-driven selection process is performed in an automated fashion

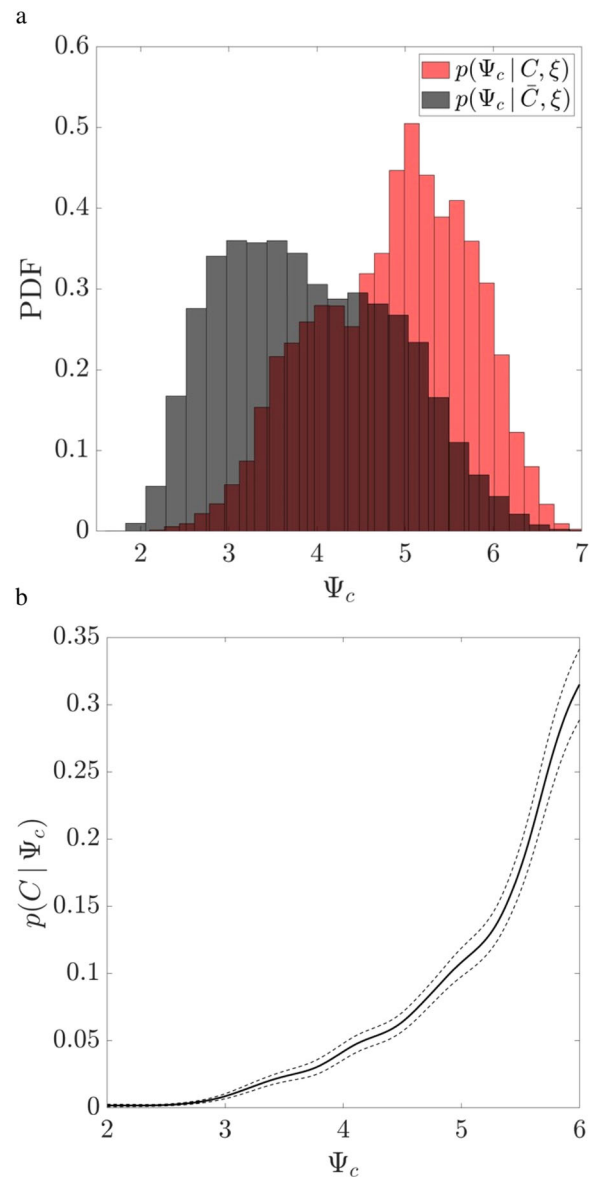


Fig. 6 Performance and robustness of the scalar crack nucleation indicator. **a** The likelihood distributions of the scalar crack nucleation indicator for cracked and uncracked states, **b** aggregated posterior distribution of the 66 models, conditioned on only the scalar crack nucleation indicator, for different combinations of ten microstructure in the training data sets. The average of all models is depicted by the dark solid line and the standard deviation bounds are depicted by the dotted lines.

without the need for external intervention. It follows a statistically unbiased approach that makes an informed decision while accounting for various experimental microstructures and mechanical simulations.

Given a set of EBSD scans along with the associated pre-identified crack nucleation sites as input, the procedure ranging from the processing of experimental EBSD and SEM data, multiscale simulations for creating a reference database, to the development of the probabilistic crack nucleation model is implemented through an automated computational pipeline. It comprises four sequential modules given below.

1. Collection of post-fatigue loading local microstructural data, and subsequent cleaning and processing of EBSD data;

2. Development of a concurrent multiscale model with the image-based 3D microstructure embedded in a homogenized material modeled by an upscaled, anisotropic continuum plasticity constitutive model;
3. Multiscale simulations with different experimental microstructures to create a database of micromechanical state variables relevant to the crack nucleation process;
4. Formulation of a Bayesian crack nucleation model that identifies crack nucleation sites from the database of local state variable fields.

The probabilistic models for indicating crack nucleation are based on the core concept of parallel slip at susceptible coherent twin boundaries. The material near these boundaries possesses particular crystallographic configurations that promote an environment for recurring dislocation slip parallel to the boundary. State variables, such as the von Mises stress, the maximum plastic slip rate, and the plastic defect energy in this region, have a strong effect on this mechanism. Slip is driven on the primary plane and builds up dislocation substructures like PSBs that further promote future slip along the same path. Simultaneously, little driving force exists for forest dislocations to impede obstacles on the primary plane. The combination of these events creates a probable setting for the incidence of crack nucleation.

The multi-dimensional Bayesian inference model is constructed to capture this sequence of mechanistic events. It is validated for various microstructures and probabilistic measures, using the “leave-two-out” cross-validation process. Finally, a simple crack nucleation indicator is formulated to integrate the parallel slip concepts into a single scalar model. This composite indicator outperforms other individual state variables in differentiating between cracked and uncracked material volumes. In conclusion, this paper provides a robust framework for employing sophisticated data analysis techniques in conjunction with multiscale simulations to investigate the mechanisms underlying complex physical phenomena, and produces interpretable results. The Bayesian crack nucleation model should be valid for different microstructural compositions and loading conditions, as long as the main mechanisms of deformation such as plastic slip, dislocation glide, etc., remain the same.

The overall modeling framework developed in this paper is not limited to only superalloys considered in this study. Most aspects of the framework are transferable to other polycrystalline alloys for which, state variable evolution corresponding to parallel slip and PSB formation govern their fatigue behavior. A majority of steps in the automated framework, including the embedding methodology, finite element meshing, state variable calculation, probabilistic model procedure, and the reduced model selection technique, are material agnostic within the realm of metallic materials. Transferring the microstructure to deformation behavior and fatigue platform to other metallic materials will require the corresponding experimental data, viz. EBSD scans and crack nucleation sites, a calibrated crystal plasticity constitutive model, and the associated upscaled anisotropic continuum plasticity model. Additionally, the probabilistic classification framework has many other potential applications in data-rich and knowledge-poor circumstances.

For component-scale analysis, it is necessary to introduce a detailed parametrically upscaled constitutive model (PUCM) and associated parametrically upscaled crack nucleation model (PUCNM) developed in^{39–43}. The thermodynamically consistent PUCMs, incorporate a parametric representation of lower-scale microstructural descriptors in higher-scale constitutive coefficients. The PUCM coefficients are expressed as functions of representative aggregated microstructural parameters (RAMPs), representing lower-scale descriptors of microstructural morphology and crystallography. While the upscaled anisotropic continuum plasticity model developed for the concurrent model is a

step in this direction, it does not have the parametric forms in terms of the RAMPs. The PUCM, in conjunction with the PUCNM, readily performs component-scale analysis with direct connection to the microstructure. The models developed in this study can be used for SERVE analysis to develop the PUCM and PUCNM for the superalloy considered for this study.

METHODS

Concurrent multiscale model for fatigue analysis

The concurrent multiscale model in Fig. 7 consists of a polycrystalline microstructural RVE that is embedded in an exterior computational domain. The RVE and the exterior domain are modeled by a crystal plasticity constitutive model^{34–36} and a rate-dependent homogenized elasto-plasticity constitutive model respectively. Cyclic loads representing fatigue loading are applied on the outer surface of the homogenized exterior domain. The constitutive relations used in the concurrent multiscale model are summarized in the Supplementary Methods Section of this paper.

Concurrent multiscale model with embedded microstructural RVE. The application of periodicity or other boundary conditions directly on the microstructural RVE domain can result in significant inaccuracies due to improper representation of the domain and consequent boundary effects^{32,33,54}. The concept of concurrent multiscale domains, in which the RVE is embedded in a homogenized exterior domain, has been successfully introduced in^{43,55,56} to alleviate this discrepancy. Boundary conditions are applied on the exterior boundary of the concurrent domain, while the RVE shares a common interface with the homogenized domain. Similar concurrent approaches have also been proposed for coupled continuum-polycrystalline domains in^{29,57}.

The embedded RVE in Fig. 7 is generated by directly mapping the 2D EBSD micrograph of a polycrystalline microstructure shown in Fig. 2, and extruding it in the orthogonal direction to a suitable depth. Each section of the RVE contains about 50–200 grains. Previous studies on Ni-based superalloys in^{34,54,58} have demonstrated the effectiveness of microstructural RVEs of this size. Some of the EBSD micrographs containing cracked regions are not fully supported by first and second nearest neighbor grains. While this may lead to some error in the simulation results due to limited grain representation in RVE models, the relatively lower strains levels may not significantly alter the response fields as noted in⁵⁹. The homogenized exterior medium in the concurrent model may also mitigate the surrounding boundary effect due to influence from neighboring grains. Furthermore, the final posterior model is constructed using multiple experimental microstructures and testing. With increasing data, the Bayesian model improves at distinguishing between all the sources of model noise and the meaningful relation existing between crack nucleation sites and state variable fields.

For fatigue crack modeling, it is necessary that the microstructural RVEs be constructed directly from EBSD scans of the region containing cracks for one-to-one correspondence with experimental observations. Statistically equivalent RVE (SERVEs), with matching morphological and crystallographic distributions, are not ideal for this one-to-one correspondence. 3D microstructures may consequently be created by focused ion beam (FIB)-based serial-sectioning, or synchrotron-based CT methods. However, it is challenging to obtain interrupted 3D experimental microstructures in the context of fatigue experiments. The decision to create 3D microstructural RVEs by extruding 2D EBSD scans was made since such types of data were not available for this study. In⁶⁰, the authors have inferred that sub-surface crystallography and grain morphology, are needed to accurately predict surface strain distributions. In the present study, it is assumed the error due to the extrusion will not be significant since the thickness of the RVE is smaller than the average grain size. Additionally, the use of multiple microstructural instantiations may have a mitigating effect on the crack nucleation model.

The concurrent multiscale model can mitigate other spurious boundary effects on the evolving RVE variables. A sample microstructural RVE, shown in Fig. 7, has dimensions of $168\ \mu\text{m} \times 69\ \mu\text{m} \times 7.5\ \mu\text{m}$, while the homogenized exterior rectangular domain has dimensions of $336\ \mu\text{m} \times 138\ \mu\text{m} \times 42\ \mu\text{m}$. This corresponds to twice the length and width of the RVE and a depth of $7.5\ \mu\text{m}$ plus half of the RVE width. Five of the six faces of the embedded RVE share an interface with the

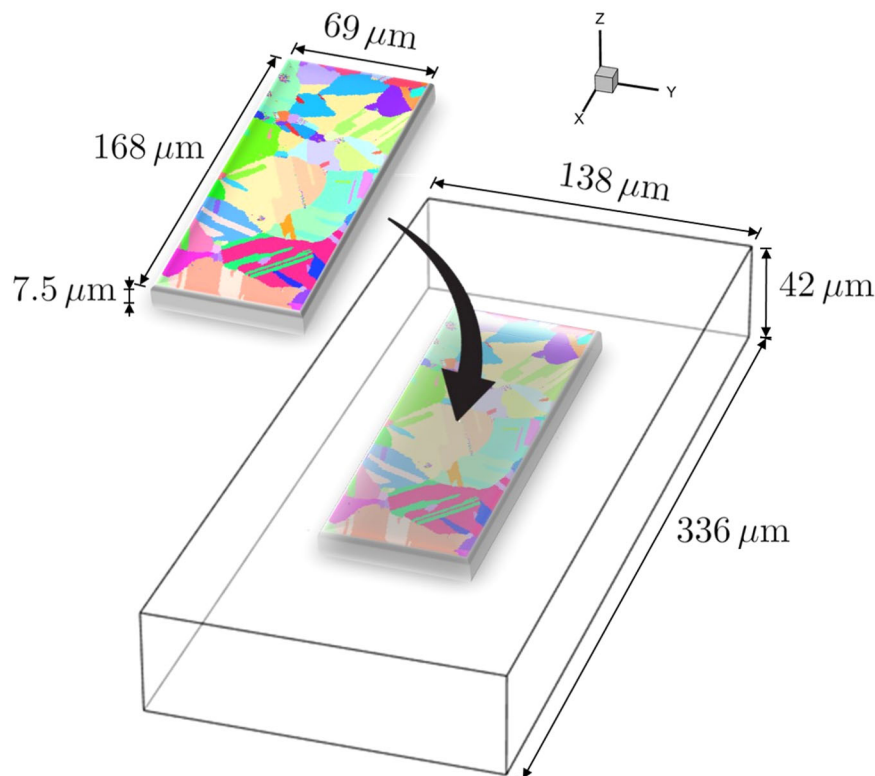


Fig. 7 Schematic of the concurrent computational domain. Microstructural RVEs are embedded in homogenized exterior domains.

exterior homogenized domain, enforcing displacement compatibility across these interfaces. The remaining RVE face represents the free surface of the fatigue experiment and is unconstrained during the cyclic simulations.

The concurrent FE model is discretized into TET4 elements, generated using an automated meshing pipeline built within the Simmetrix code⁶¹. This minimizes the computational and memory requirements and provides a robust discretization in terms of numerical accuracy and geometric conformity. Since the state variables near twin boundaries are key to the crack model, a graded mesh with highly refined elements near the grain/twin boundaries, and coarser elements away from these critical regions, is developed as seen in Module II of Fig. 1. The exterior domain mesh matches the RVE mesh at the interface, but coarsens away from the interface.

Creating a microstructural state variable database related to fatigue nucleation

The following steps are executed for creating the database.

Experimentally obtained training data sets for supervised learning. The Bayesian modeling approach to fatigue crack nucleation requires a data set for supervised learning, consisting of experimentally observed points for which the binary cracked/uncracked states are known. To generate this data set, the locations of crack nucleation sites are recorded for the experimental micrographs, e.g. in Fig. 2. Each TET4 element in the concurrent model, for which three nodes are on a cracked interface is marked as belonging to the set of cracked elements. All other elements in the ensemble are designated as uncracked. This labeling method classifies the data set into two distinct states for training the probabilistic model.

Extracting core microstructural state variables for the crack nucleation model. The concurrent multiscale model in Methods Section I is now developed for each of the 12 experimental microstructures and subsequently loaded under fatigue loading conditions. Analogous to the experiments in⁵¹, a stress-controlled fully reversed ($R = -1$) cyclic loading with a triangular wave-form and 1 Hz frequency is applied on the computational model. The peak stress in this loading cycle corresponds to

Table 2. List of core state variables used in the probabilistic crack nucleation analysis.

State variable	Symbol	Mathematical expression
von Mises stress	σ_{VM}	$\sqrt{\frac{3}{2}} \ \mathbf{dev}(\boldsymbol{\sigma})\ $
Principal Cauchy stress components (3)	σ_i	$\sigma_i \mathbf{n}_i = \boldsymbol{\sigma} \mathbf{n}_i$
Maximum resolved shear stress	τ_*	$\max_a (\mathbf{m}_0^a \cdot \mathbf{Mn}_0^a)$
Effective plastic strain	ϵ_p	$\epsilon_p = \int_0^t \ \mathbf{D}_p\ \mathbf{dev} dt'$
Elastic Helmholtz free energy	ϕ_e	$\frac{1}{2} \mathbf{E}_e : (\mathbf{C}_e \mathbf{E}_e)$
Plastic defect energy	ψ_p	$\frac{\beta}{2} \sum_a^{N_{slip}} (s_a^a + s_{th}^a) \gamma_{acc}^a$ ($\beta = 1$ and $\gamma_{acc} = \int_0^t \dot{\gamma} dt'$)
Maximum plastic slip-rate	$\dot{\gamma}$	$\max_a (\dot{\gamma}^a)$
Maximum slip system normal stress	τ_n	$\max_a (\mathbf{n}_0^a \cdot \mathbf{Mn}_0^a)$

a 1% total strain. All state variables are recorded at the peak tensile load of each cycle. This process generates a complete database of mechanical fields for all the 12 experimental microstructures, which is used for training the Bayesian crack nucleation model.

Time-histories of state variables are calculated and recorded as quantities of interest in the development of the fatigue crack nucleation model. From a comprehensive set of mechanistic variables, including stress and strain measures, dislocation evolution indicators, fatigue indicator parameters, etc., a set of ten core state variables are deduced. Most of the other indicators are functions of these core parameters. The core variables are tabulated in Table 2.

The set of variables in Table 2 is designated as \mathbf{X} . Additionally, a regularized difference operator (similar to unsharp masking) acting on each state variable is appended to the set \mathbf{X} . Correspondingly the dimension of the set is $k = \dim(\mathbf{X}) = 20$. The values of \mathbf{X} are assembled into a state variable database for use in the training and testing of the probabilistic model.

Observations on parallel slip configuration from experiments. Observations from fatigue experiments in⁹ suggest that dislocation slip parallel to the coherent twin boundaries results in persistent slip band (PSB) and thin shear band formation due to strain localization. Adjacent to the twin boundary, plasticity occurs due to dislocation glide on slip planes parallel to the twin boundary. The repeated passing of these dislocations and migration of dislocations to this region, shears the γ' precipitates, promoting additional plasticity by reducing the barrier to subsequent dislocation motion. It has been observed that these highly localized regions of plasticity nucleate cracks, several nanometers offset from the actual twin boundary.

Some twin boundaries are identified by grains that are oriented such that their slip systems favor dislocation motion along their shared boundary plane. This special grain-pair configuration is called a parallel slip configuration. It is characterized by a grain pair separated by a coherent twin boundary, where the resolved shear stress on one of the shared slip systems is significantly higher than that of the other slip systems. This configuration results in primary slip along that single shared plane and no new dislocation migration occurs from other planes to the slipped region. The twin boundary lengths are also relatively long and straight compared to other grain boundaries. Consequently, this geometric effect and the lack of out-of-plane dislocation motion leaves the shared slip system free of obstacles, which allows dislocations to glide unimpeded for long distances. In cyclic deformation, uninterrupted localized slip promotes the early stages of PSB formation.

Indicator of dominant parallel slip. Since parallel slip configurations are identified as precursors to crack nucleation, it is important to develop an indicator for dominant parallel slip. This indicator should identify coherent twin boundaries and subsequently search for grain-pairs, oriented such that the resolved shear stress favors slip along the shared boundary plane. The following procedure is adopted to account for these conditions and lead up to a definition for the parallel slip indicator.

The criteria used to determine if two grains are in a $\Sigma 3$ twinned configuration is based on two conditions. The first condition is that the misorientation angle must be $60^\circ \pm 2^\circ$, and the second requires that the absolute value of the inner product between the shared slip plane normals is >0.985 . For each grain pair, the shared slip plane normal is defined by the unit vector average of the two slip plane normals, whose inner product has the largest absolute value. Specifically, these vectors are determined by the following process.

For two adjacent grains g_1 and g_2 in the microstructure, the two respective slip planes that most closely share the same normal are selected according to the criterion:

$$(l^*, k^*) = \arg \max_{l,k} (|\mathbf{n}_l^{g_1} \cdot \mathbf{n}_k^{g_2}|) \quad (4)$$

where l and k enumerate the four possible slip planes in the $\langle 111 \rangle$ family. Correspondingly, the shared slip plane normal for the grains g_1 and g_2 are defined as:

$$\mathbf{n}_p^{g_1, g_2} = \frac{\mathbf{n}_{l^*}^{g_1} + \mathbf{n}_{k^*}^{g_2}}{|\mathbf{n}_{l^*}^{g_1} + \mathbf{n}_{k^*}^{g_2}|} \quad (5)$$

where $\mathbf{n}_{l^*}^{g_1}$ is the normal to the l^* slip plane of grain g_1 and $\mathbf{n}_{k^*}^{g_2}$ is the normal to the k^* slip plane of grain g_2 .

The existence of a shared plane allows for a partition of the standard octahedral slip systems of FCC materials. Let \mathcal{A} denote the set of all its slip systems for a grain that shares a slip system with an adjacent grain. Also, let \mathcal{A}_p be the set of slip systems that contain the shared slip plane characterized by $\mathbf{n}_p^{g_1, g_2}$ in Eq. 5. Finally, let \mathcal{A}_0 be the complement of \mathcal{A}_p in \mathcal{A} , i.e. the complementary set of slip systems that do not contain the shared slip plane.

Two special resolved shear stresses (RSS) are defined to determine when parallel slip is dominant for a given twin boundary. The maximum RSS is given as $\tau_* = \max_{\alpha \in \mathcal{A}} (|\mathbf{M} : \mathbf{s}_\alpha^a|)$, where $\mathbf{s}_\alpha^a = \mathbf{m}_\alpha^a \otimes \mathbf{n}_\alpha^a$ is the Schmid tensor for the α slip system, \mathbf{m}_α^a is the slip direction, and \mathbf{n}_α^a is the slip plane normal in the reference configuration. Correspondingly, the parallel slip RSS and the out-of-plane RSS are respectively defined as:

$$\tau_p = \max_{\alpha \in \mathcal{A}_p} (|\mathbf{M} : \mathbf{s}_\alpha^a|) \quad \text{and} \quad \tau_o = \max_{\alpha \in \mathcal{A}_0} (|\mathbf{M} : \mathbf{s}_\alpha^a|) \quad (6)$$

The parallel slip RSS is the maximum RSS on the shared slip plane, while the out-of-plane RSS is the maximum over the remaining three slip planes. When $\tau_p > \tau_o$, primary slip occurs along the twin boundary, and out-of-

plane dislocations do not disrupt this plastic flow. Therefore, parallel slip dominates the deformation process. This condition is combined with the coherent twin and a plastic flow requirement to define a binary parallel slip indicator $\Xi \equiv \xi$ or $\bar{\xi}$ for a given material volume. The parallel slip indicator is activated with $\Xi = \xi$, if all of the following three requirements are met simultaneously:

1. The material volume is located on a coherent twin boundary;
2. The parallel slip plane is dominant, i.e. $\tau_p > \tau_o$;
3. The maximum plastic slip-rate exceeds a minimum threshold, i.e. $\dot{\gamma} > \dot{\gamma}_m$, where $\dot{\gamma}_m$ is a minimum threshold slip-rate.

The threshold slip-rate is selected as $\dot{\gamma}_m = 10^{-5} \text{s}^{-1}$ to indicate that the material is undergoing plasticity. If any one of the three conditions are not met, then $\Xi = \bar{\xi}$. The parallel slip indicator is designed to signal time-evolving, point-wise occurrence of parallel slip throughout the microstructure. It is analogous to the time-independent, grain-wise-homogeneous parallel slip configuration definition in^{1,8,9,15}. The importance of Ξ is that it acts as a sufficiency condition for crack nucleation within the experimental data set of microstructures. This empirical data is exploited in the probabilistic model framework.

Formulation of the Bayesian model for classifying crack nucleation

Let a discrete random variable Θ represent the possible observation of a crack event in a fixed, reference material volume V_* , and let a continuous k -dimensional random vector \mathbf{X} denote the set of core state variables pertaining to that particular region. The entire microstructural field is classified by one of two distinct states, viz. (i) with a crack event $\Theta = C$, and (ii) without a crack event $\Theta = \bar{C}$. Additionally, let the previously defined parallel slip indicator Ξ in the microstructure be a discrete random variable with two possible states, viz. (i) $\Xi = \xi$ indicating a parallel slip configuration, and (ii) $\Xi = \bar{\xi}$ otherwise. In accordance with experimental observations, it is assumed that all cracked states C require that a material volume is in a parallel slip configuration $\Xi = \xi$. Therefore, the probability of observing a crack nucleation event C in the microstructure with given state variables \mathbf{x} , homogeneous over some small volume V_* , is given by Bayes' theorem. For the case of $\Xi = \xi$, this is stated as:

$$P(\Theta = C | \mathbf{X} = \mathbf{x}) = p_{\Theta|\mathbf{X}}(C | \mathbf{x}) = p_{\Theta|\Xi, \mathbf{X}}(C | \xi, \mathbf{x}) = \frac{p_{\mathbf{X}|\Theta, \Xi}(\mathbf{x} | C, \xi) p_{\Theta|\Xi}(C | \xi)}{p_{\mathbf{X}|\Xi}(\mathbf{x} | \xi)} \quad (7)$$

where $p_{\Theta|\mathbf{X}}(C | \mathbf{x})$ and $p_{\Theta|\Xi, \mathbf{X}}(C | \xi, \mathbf{x})$ are both posterior probability density functions (PDF's), $p_{\mathbf{X}|\Theta, \Xi}(\mathbf{x} | C, \xi)$ is a likelihood function, $p_{\Theta|\Xi}(C | \xi)$ is the prior belief of observing a crack nucleation event, and $p_{\mathbf{X}|\Xi}(\mathbf{x} | \xi)$ is the marginal state variable PDF, which normalizes the expression. For the complementary event $\Xi = \bar{\xi}$, the posterior PDF $p_{\Theta|\mathbf{X}}(C | \mathbf{x}) = 0$. Hereafter, the subscripts on the PDF's will be omitted to eliminate notational clutter, but may be used occasionally for providing beneficial clarity. Applying the law of total probability to Eq. 7 yields:

$$p(C | \mathbf{x}) = \frac{p(\mathbf{x} | C, \xi) p(C | \xi)}{p(\mathbf{x} | C, \xi) p(C | \xi) + p(\mathbf{x} | \bar{C}, \bar{\xi}) p(\bar{C} | \bar{\xi})} \quad (8)$$

The posterior PDF $p(C | \mathbf{x})$ is the primary quantity of interest to be determined for the crack nucleation model. From Eq. 8, it is seen that four PDF's must be calculated to fully characterize this posterior, viz. $p(\mathbf{x} | C, \xi)$, $p(\mathbf{x} | \bar{C}, \bar{\xi})$, $p(C | \xi)$, and $p(\bar{C} | \bar{\xi})$. Each of these distributions is systematically constructed from the experimental data set in the following sections.

Determination of the prior distributions $p(C | \xi)$ and $p(\bar{C} | \bar{\xi})$. The prior belief that a crack event occurs in a material volume is constructed by calculating the ratio of cracked volume to total volume over all experimental microstructures. For the available experiments, let $M = 12$ denote the number of independent microstructures in the data set, and let $V_C^{(i)}$ and $V_{\text{tot}}^{(i)}$ be respectively the total cracked volume and total overall volume of the i th microstructure, where $\Xi = \xi$. The crack nucleation prior is correspondingly defined as:

$$p(C | \xi) = \frac{\sum_i^M V_C^{(i)}}{\sum_i^M V_{\text{tot}}^{(i)}} \quad (9)$$

The uncracked prior is constrained by the complementary event, and is therefore determined as $p(\bar{C} | \bar{\xi}) = 1 - p(C | \xi)$. Additionally, let the

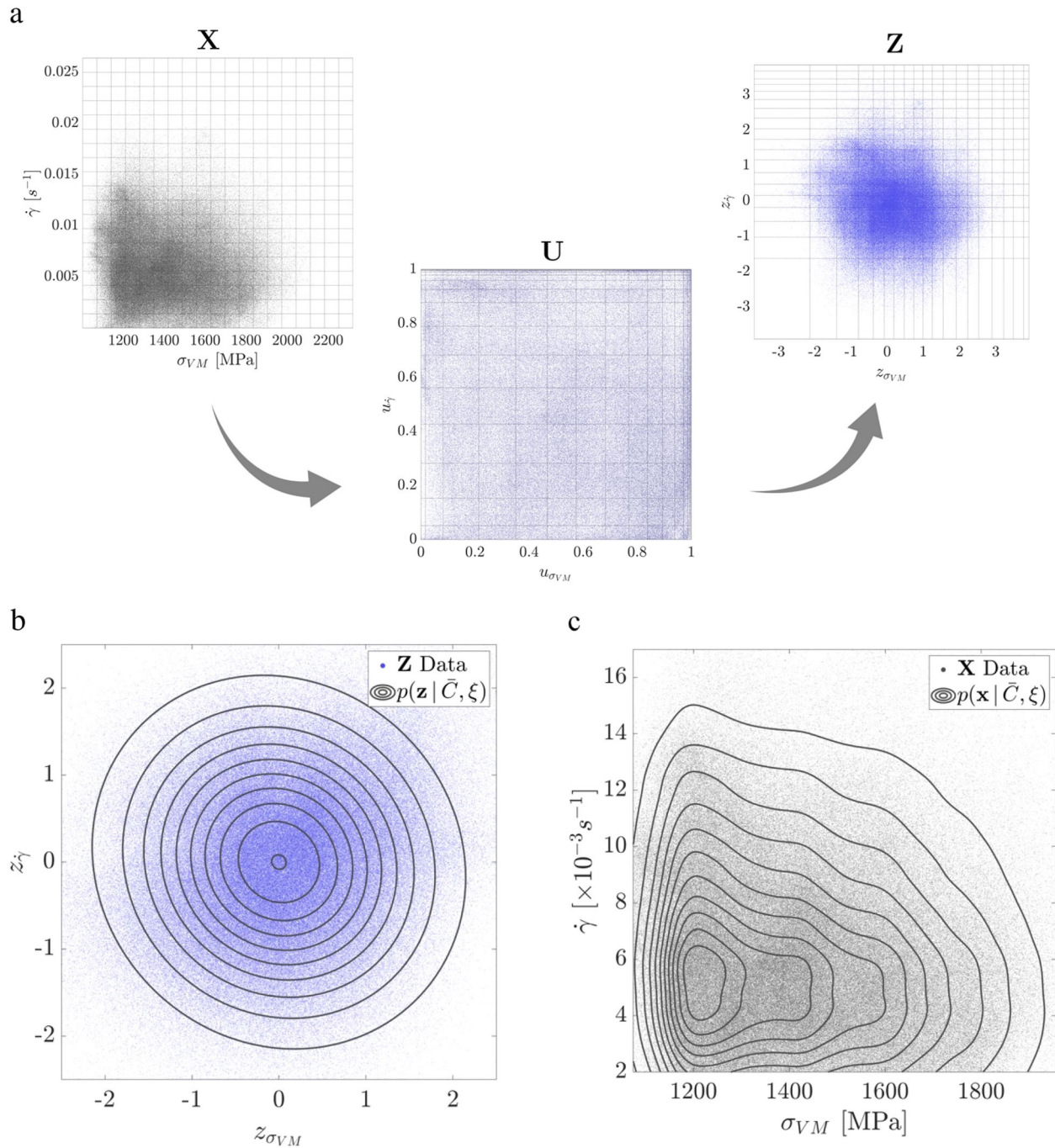


Fig. 8 Illustration of the Nataf transformation for two variables. **a** Example action of the Nataf transformation for the bivariate $\mathbf{x} = [\sigma_{VM}, \dot{\gamma}]$. Approximately 500,000 data points from all 12 microstructures are transformed from the observed \mathbf{X} space to the marginally uniform \mathbf{U} space to the standardized \mathbf{Z} space. **b** Iso-probabilistic contours of the bivariate normal model fit for the uncracked likelihood $p(\mathbf{z}|\bar{C})$ in the bivariate example case of $\mathbf{x} = [\sigma_{VM}, \dot{\gamma}]$ and **c** Iso-probabilistic contours of the model fit for $p(\mathbf{z}|\bar{C})$ by transforming the standardized model back to the \mathbf{X} space.

reference volume for the model be defined as:

$$V_* = \frac{1}{M} \sum_i^M V_C^{(i)} \quad (10)$$

which corresponds to the average crack volume over all microstructures.

Determination of the likelihood distributions $p(\mathbf{x}|C, \xi)$ and $p(\mathbf{x}|\bar{C}, \xi)$. In general, the likelihood PDF's, $p(\mathbf{x}|C, \xi)$ and $p(\mathbf{x}|\bar{C}, \xi)$, are high-dimensional, highly covariant joint distributions. The Nataf transformation method^{62–64} is used in this work to construct explicit functional

representations of both $p(\mathbf{x}|C, \xi)$ and $p(\mathbf{x}|\bar{C}, \xi)$. It has been shown through Sklar's theorem that a multivariate joint distribution can be decomposed into a composition of component-wise univariate marginal distributions and a multivariate copula, which retains the covariance structure of the random vector. For the Nataf transformation, the joint distribution is approximated using a Gaussian copula to permit a straightforward analytical computation. The following construction is formulated in the context of $p(\mathbf{x}|C, \xi)$. However for $p(\mathbf{x}|\bar{C}, \xi)$, analogous steps are performed by substituting \bar{C} for C .

The Nataf transformation is comprised of three sequential transformations. First, an observed data point $\mathbf{x} = [x_1, x_2, \dots, x_k]$ of the random set \mathbf{X} of

state variables is transformed component-wise into a vector of uniformly distributed random variables $\mathbf{u} = [u_1, u_2, \dots, u_k]$, where

$$u_i = F_{X_i|\Theta, \Xi}(x_i | C, \xi) \quad (11)$$

$F_{X_i|\Theta, \Xi}$ is the marginal cumulative distribution function (CDF) for the i th component of \mathbf{X} . Secondly, each component of \mathbf{u} is standardized by applying the inverse standard normal CDF, specified by a mean of zero and a variance of one. Let the standardized vector $\mathbf{z} = [z_1, z_2, \dots, z_k]$ be characterized by this transformation as:

$$z_i = \Phi^{-1}(u_i) \quad (12)$$

Finally, a k -dimensional normal distribution is fit to the observations in the standardized space. This sequential application of the ingredients of the Nataf transformation is summarized as:

$$F(\mathbf{x} | C, \xi) = \Phi_k(\mathbf{z}) = \Phi_k(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \dots, \Phi^{-1}(u_k)) \quad (13)$$

where the joint CDF $F(\mathbf{x} | C, \xi)$ corresponds to the likelihood $p(\mathbf{x} | C, \xi)$ and Φ_k is the k -dimensional multivariate normal CDF. Differentiating Eq. 13 yields:

$$p(\mathbf{x} | C, \xi) = p(\mathbf{z} | C, \xi) \prod_i \frac{p_{X_i|\Theta, \Xi}(x_i | C, \xi)}{\phi(z_i)} \quad (14)$$

where $p_{X_i|\Theta, \Xi}(x_i | C)$ is the marginal PDF for the i th component, ϕ is the standard normal distribution, and $p(\mathbf{z} | C)$ is a multivariate normal distribution ϕ_k given as:

$$p(\mathbf{z} | C, \xi) = \phi_k(\mathbf{z}) = \frac{1}{\sqrt{(2\pi)^k \det \Sigma}} \exp\left(-\frac{1}{2} \mathbf{z} \cdot (\Sigma^{-1} \mathbf{z})\right) \quad (15)$$

with the covariance matrix Σ . The computation of $p(\mathbf{x} | C, \xi)$ for the state variable data set only requires the determination of $F_{X_i|\Theta, \Xi}$, $p_{X_i|\Theta, \Xi}$, and Σ . $F_{X_i|\Theta, \Xi}$ and $p_{X_i|\Theta, \Xi}$ are represented by a regularized, weighted empirical CDF and a weighted kernel density estimation from the observed \mathbf{x} data, respectively. The covariance matrix Σ , defined as: $\Sigma = \mathbb{E}[\mathbf{z} \mathbf{z}^T]$, is approximated by a weighted unbiased estimator of the observed data as:

$$\Sigma = \frac{1}{1 - \omega} \sum_i^{N_e} w^{(i)} \mathbf{z}^{(i)} \otimes \mathbf{z}^{(i)} \quad (16)$$

where N_e is the total number of discretized volumes (finite elements) over all M microstructures in a state of $\Xi = \xi$, $w^{(i)}$ are observation weights corresponding to the normalized material volume of each point such that $\sum_i^{N_e} w^{(i)} = 1$, and $\omega = \sum_i^{N_e} w^{(i)2}$ is the sum of squares of the observation weights.

An example of the action of the Nataf transformation for the uncracked likelihood PDF $p(\mathbf{x} | \bar{C}, \xi)$ is shown in Fig. 8a for two representative state variables. The von Mises stress and maximum plastic slip-rate data points $\mathbf{x} = [\sigma_{VM}, \dot{\gamma}]$, are depicted in the original \mathbf{X} space, transformed into the uniform \mathbf{U} space, and then transformed again into the standardized \mathbf{Z} space. Additionally, grid lines are added and transformed accordingly to illustrate the component-wise warping of each observation space.

Figure 8b demonstrates the application of the model described in Eq. 15 to determine a bivariate normal distribution as an approximation for the standardized observations \mathbf{z} . This approximation is achieved through the computation of Σ , using Eq. 16. It is seen that the preservation of the isoprobabilistic property of the transformation and of the covariance structure are determined by how well the standard bivariate normal fits the data in the standardized \mathbf{Z} space, shown in Fig. 8b. The original joint PDF, $p(\mathbf{x} | \bar{C}, \xi)$, is recovered in the \mathbf{X} space by applying Eq. 14. The outcome, shown in Fig. 8c, is an excellent model for a joint distribution, which would typically be complicated to analytically represent and infer from the data.

DATA AVAILABILITY

The input data sets for the training and testing of the Bayesian model will be available from the corresponding author upon reasonable request.

CODE AVAILABILITY

The codes to train and test the Bayesian model will be available from the corresponding author in a suitable format upon reasonable request.

Received: 13 October 2021; Accepted: 13 February 2022;
Published online: 14 March 2022

REFERENCES

- Miao, J., Pollock, T. & Jones, W. Crystallographic fatigue crack initiation in nickel-based superalloy René 88DT at elevated temperature. *Acta Mater.* **57**, 5964–5974 (2009).
- McDowell, D. Viscoplasticity of heterogeneous metallic materials. *Mat. Sci. Eng. R. Rep.* **62**, 67–123 (2008).
- Stein, C., Lee, S. & Rollett, A. An analysis of fatigue crack initiation using 2D orientation mapping and full-field simulation of elastic stress response. *Superalloys* **2012**, 439–444 (2012).
- Mineur, M., Villechaise, P. & Mendez, J. Influence of the crystalline texture on the fatigue behavior of a 316L austenitic stainless steel. *Mat. Sci. Eng. A* **286**, 257–268 (2000).
- Chen, Q. et al. Small crack behavior and fracture of Nickel-based superalloy under ultrasonic fatigue. *Int. J. Fatigue* **27**, 1227–1232 (2005).
- Thompson, N., Wadsworth, N. & Louat, N. The origin of fatigue fracture in copper. *Philos. Mag.* **1**, 113–126 (1956).
- Boettner, R., McEvily Jr, A. & Liu, Y. On the formation of fatigue cracks at twin boundaries. *Philos. Mag.* **10**, 95–106 (1964).
- Miao, J., Pollock, T. & Jones, W. Microstructural extremes and the transition from fatigue crack initiation to small crack growth in a polycrystalline nickel-base superalloy. *Acta Mater.* **60**, 2840–2854 (2012).
- Stinville, J., Lenthe, W., Miao, J. & Pollock, T. A combined grain scale elastic–plastic criterion for identification of fatigue crack initiation sites in a twin containing polycrystalline nickel-base superalloy. *Acta Mater.* **103**, 461–473 (2016).
- Alam, Z., Eastman, D., Weber, G., Ghosh, S. & Hemker, K. Microstructural aspects of fatigue crack initiation and short crack growth in René 88DT. *Superalloys* **2016**, 561–568 (2016).
- Heinz, A. & Neumann, P. Crack initiation during high cycle fatigue of an austenitic steel. *Acta Met. Mater.* **38**, 1933–1940 (1990).
- Wong, S. & Dawson, P. Influence of directional strength-to-stiffness on the elastic–plastic transition of fcc polycrystals under uniaxial tensile loading. *Acta Mater.* **58**, 1658–1678 (2010).
- Stinville, J. et al. Sub-grain scale digital image correlation by electron microscopy for polycrystalline materials during elastic and plastic deformation. *Exp. Mech.* **56**, 197–216 (2016).
- Stein, C. et al. Fatigue crack initiation, slip localization and twin boundaries in a nickel-based superalloy. *Curr. Opin. Solid State Mater. Sci.* **18**, 244–252 (2014).
- Stinville, J., Vanderesse, N., Bridier, F., Bocher, P. & Pollock, T. High resolution mapping of strain localization near twin boundaries in a nickel-based superalloy. *Acta Mater.* **98**, 29–42 (2015).
- Shyam, A. et al. Development of ultrasonic fatigue for rapid, high temperature fatigue studies in turbine engine materials. *Superalloys* **2004**, 259–268 (2004).
- Healy, J., Grabowski, L. & Beevers, C. Short-fatigue-crack growth in a nickel-base superalloy at room and elevated temperature. *Int. J. Fatigue* **13**, 133–138 (1991).
- Li, K., Ashbaugh, N. & Rosenberger, A. Crystallographic initiation of nickel-base superalloy IN100 at RT and 538 °C under low cycle fatigue conditions. *Superalloys* **2004**, 251 (2004).
- Davidson, D., Tryon, R., Oja, M., Matthews, R. & Chandran, K. Fatigue crack initiation in Waspaloy at 20°C. *Met. Mater. Trans. A* **38**, 2214–2225 (2007).
- Lukáš, P. & Kunz, L. Role of persistent slip bands in fatigue. *Philos. Mag.* **84**, 317–330 (2004).
- Rovinelli, A. et al. Predicting the 3D fatigue crack growth rate of small cracks using multimodal data via Bayesian networks: in-situ experiments and crystal plasticity simulations. *J. Mech. Phys. Solids* **115**, 208–229 (2018).
- Hochhalter, J. et al. A geometric approach to modeling microstructurally small fatigue crack formation: II. Physically based modeling of microstructure-dependent slip localization and actuation of the crack nucleation mechanism in AA 7075-T651. *Model Simul. Mater. Sci.* **18**, 045004 (2010).
- Chen, B., Jiang, J. & Dunne, F. Is stored energy density the primary meso-scale mechanistic driver for fatigue crack nucleation? *Int. J. Plast.* **101**, 213–229 (2018).
- Shenoy, M., Zhang, J. & McDowell, D. Estimating fatigue sensitivity to polycrystalline Ni-base superalloy microstructures using a computational approach. *Fatigue Fract. Eng. Mater. Struct.* **30**, 889–904 (2007).
- Ozturk, D., Shahba, A. & Ghosh, S. Crystal plasticity FE study of the effect of thermo-mechanical loading on fatigue crack nucleation in titanium alloys. *Fatigue Fract. Eng. Mater. Struct.* **39**, 752–769 (2016).
- Yeratapally, S., Glavicic, M., Hardy, M. & Sangid, M. Microstructure based fatigue life prediction framework for polycrystalline nickel-base superalloys with

- emphasis on the role played by twin boundaries in crack initiation. *Acta Mater.* **107**, 152–167 (2016).
27. Korsunsky, A., Dini, D., Dunne, F. & Walsh, M. Comparative assessment of dissipated energy and other fatigue criteria. *Int. J. Fatigue* **29**, 1990–1995 (2007).
 28. Bandyopadhyay, R., Prithvirajan, V., Peralta, A. & Sangid, M. Microstructure-sensitive critical plastic strain energy density criterion for fatigue life prediction across various loading regimes. *Proc. Math. Phys.* **476**, 20190766 (2020).
 29. Pierson, K., Hochhalter, J. & Spear, A. Data-driven correlation analysis between observed 3d fatigue-crack path and computed fields from high-fidelity, crystal-plasticity, finite-element simulations. *JOM* **70**, 1159–1167 (2018).
 30. Sankararaman, A., Ling, Y. & Mahadevan, S. Uncertainty quantification and model validation of fatigue crack growth prediction. *Eng. Fract. Mech.* **78**, 1487–1504 (2011).
 31. Yeratapally, S., Glavicic, M., Argyrakis, C. & Sangid, M. Bayesian uncertainty quantification and propagation for validation of a microstructure sensitive model for prediction of fatigue crack initiation. *Reliab. Eng. Syst. Saf.* **164**, 110–123 (2017).
 32. Ghosh, S. & Kubair, D. Exterior statistics based boundary conditions for representative volume elements of elastic composites. *J. Mech. Phys. Solids* **95**, 1–24 (2016).
 33. Kubair, D., Pinz, M., Kollins, K., Przybyla, C. & Ghosh, S. Role of exterior statistics-based boundary conditions for property-based statistically equivalent RVEs of polydispersed elastic composites. *J. Compos. Mater.* **52**, 2919–2928 (2018).
 34. Bagri, A. et al. Microstructure and property-based statistically equivalent representative volume elements for polycrystalline Ni-based superalloys containing annealing twins. *Met. Mater. Trans. A* **49**, 5727–5744 (2018).
 35. Weber, G. & Ghosh, S. Thermo-mechanical deformation evolution in polycrystalline Ni-based superalloys by a hierarchical crystal plasticity model. *Mater. High. Temp.* **33**, 401–411 (2016).
 36. Keshavarz, S., Ghosh, S., Reid, A. & Langer, S. A non-Schmid crystal plasticity finite element approach to multi-scale modeling of nickel-based superalloys. *Acta Mater.* **114**, 106–115 (2016).
 37. Keshavarz, S. & Ghosh, S. Multi-scale crystal plasticity finite element model approach to modeling nickel-based superalloys. *Acta Mater.* **61**, 6549–6561 (2013).
 38. Kotha, S., Ozturk, D. & Ghosh, S. Parametrically homogenized constitutive models (PHCMs) from micromechanical crystal plasticity FE simulations. Part I: sensitivity analysis and parameter identification for Titanium alloys. *Int. J. Plasticity* **120**, 296–319 (2019).
 39. Shen, J., Kotha, S., Noraas, R., Venkatesh, V. & Ghosh, S. Developing parametrically upscaled constitutive and crack nucleation models for the alpha/beta Ti64 alloy. *Int. J. Plasticity* **151**, 103182 (2022).
 40. Ozturk, D., Kotha, S. & Ghosh, S. An uncertainty quantification framework for multiscale parametrically homogenized constitutive models (PHCMs) of polycrystalline Ti alloys. *J. Mech. Phys. Solids* **148**, 104294 (2021).
 41. Kotha, S., Ozturk, D. & Ghosh, S. Uncertainty-quantified parametrically homogenized constitutive models (UQ-PHCMs) for dual-phase α/β titanium alloys. *Npj Comp. Mater.* **6**, 117 (2020).
 42. Ozturk, D., Kotha, S., Pilchak, A. & Ghosh, S. Two-way multi-scaling for predicting fatigue crack nucleation in Titanium alloys using parametrically homogenized constitutive models. *J. Mech. Phys. Solids* **128**, 181–207 (2019).
 43. Weber, G., Pinz, M. & Ghosh, S. Machine learning-aided parametrically homogenized crystal plasticity model (PHCPM) for single crystal Ni-based superalloys. *JOM* **72**, 4404–4419 (2020).
 44. Keshavarz, S. & Ghosh, S. Hierarchical crystal plasticity FE model for nickel-based superalloys: sub-grain microstructures to polycrystalline aggregates. *Int. J. Solids Struct.* **55**, 17–31 (2015).
 45. Krueger, D., Kissinger, R. & Menzies, R. Development and introduction of a damage tolerant high temperature nickel-base disk alloy, Rene'88DT. *Superalloys* **1992**, 277–286 (1992).
 46. Chang, D., Krueger, D. & Sprague, R. Superalloy powder processing, properties and turbine disk applications. *Superalloys* **1984**, 245–273 (1984).
 47. Wlodek, S., Kelly, M. & Alden, D. The structure of Rene 88 DT. *Superalloys* **1996**, 129–136 (1996).
 48. Bataille, A. & Magnin, T. Surface damage accumulation in low-cycle fatigue: physical analysis and numerical modelling. *Acta Met. Mater.* **42**, 3817–3825 (1994).
 49. Groeber, M. & Jackson, M. DREAM.3D: a digital representation environment for the analysis of microstructure in 3D. *Integr. Mater. Manuf. Innov.* **3**, 56–72 (2014).
 50. Stinville, J. et al. Fatigue deformation in a polycrystalline nickel base superalloy at intermediate and high temperature: Competing failure modes. *Acta Mater.* **152**, 16–33 (2018).
 51. Stinville, J. et al. Microstructural statistics for fatigue crack initiation in polycrystalline nickel-base superalloys. *Int. J. Fract.* **208**, 221–240 (2017).
 52. Jaynes, E. *Probability Theory: The Logic of Science* (Cambridge University Press, 2003).
 53. Clayton, J. *Nonlinear Mechanics of Crystals*, Vol. 177 (Springer Science & Business Media, 2010).
 54. Pinz, M. et al. Microstructure and property based statistically equivalent RVEs for intragranular γ - γ' microstructures of Ni-based superalloys. *Acta Mater.* **157**, 245–258 (2018).
 55. Paquet, D., Dondeti, P. & Ghosh, S. Dual-stage nested homogenization for rate-dependent anisotropic elasto-plasticity model of dendritic cast aluminum alloys. *Int. J. Plasticity* **27**, 1677–1701 (2011).
 56. Ghosh, S., Bai, J. & Paquet, D. Homogenization-based continuum plasticity-damage model for ductile failure of materials containing heterogeneities. *J. Mech. Phys. Solids* **57**, 1017–1044 (2009).
 57. Spear, A. et al. A method to generate conformal finite-element meshes from 3D measurements of microstructurally small fatigue-crack propagation. *Fatigue Fract. Eng. Mater. Struct.* **39**, 737–751 (2016).
 58. Prithvirajan, V. & Sangid, M. Examining metrics for fatigue life predictions of additively manufactured IN718 via crystal plasticity modeling including the role of simulation volume and microstructural constraints. *Mat. Sci. Eng. A* **783**, 139312 (2020).
 59. Turner, T. J. & Semiatin, S. L. Modeling large-strain deformation behavior and neighborhood effects during hot working of a coarse-grain nickel-base superalloy. *Model. Simul. Mater. Sci.* **19**, 065010 (2011).
 60. Turner, T. J., Shade, P. A., Schuren, J. C. & Groeber, M. A. The influence of microstructure on surface strain distributions in a nickel micro-tension specimen. *Model. Simul. Mater. Sci.* **21**, 015002 (2013).
 61. Simmetrix Inc. *Simulation Modeling Suite*. www.simmetrix.com (2021).
 62. Lebrun, R. & Dutfoy, A. An innovating analysis of the Nataf transformation from the copula viewpoint. *Probabilist. Eng. Mech.* **24**, 312–320 (2009).
 63. Lebrun, R. & Dutfoy, A. A generalization of the Nataf transformation to distributions with elliptical copula. *Probabilist. Eng. Mech.* **24**, 172–178 (2009).
 64. Kim, H. & Shields, M. Modeling strongly non-Gaussian non-stationary stochastic processes using the iterative translation approximation method and Karhunen–Loève expansion. *Comput. Struct.* **161**, 31–42 (2015).

ACKNOWLEDGEMENTS

This work has been supported through Grant No. CMMI-1825115 from the National Science Foundation awarded by the Mechanics of Materials and Structures (MOMS) Program (Program Manager: Dr. Siddiq Qidwai). Early stages of this work was supported through a grant No. FA9550-12-1-0445 to the Center of Excellence on Integrated Materials Modeling (CEIMM) at Johns Hopkins University awarded by the AFOSR/RL Computational Mathematics Program (Managers Dr. Fariba Fahroo and Dr. A. Sayir). Computing support by the Maryland Advanced Research Computing Center (MARCC) is gratefully acknowledged.

AUTHOR CONTRIBUTIONS

M.P. and G.W. have developed the Bayesian framework, homogenized model, and have completed all of the simulation results of the paper, as well as writing a substantial portion. M.P. and G.W. wish to be represented as Co-First authors. J.C.S. and T.P. have contributed the experimental results, and have written the corresponding sections. S.G. has directed the study and has written a majority of the paper.

COMPETING INTERESTS

The authors declare no competing interests

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41524-022-00727-5>.

Correspondence and requests for materials should be addressed to Somnath Ghosh.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022