INVESTIGATION OF DIMENSION REDUCTION FOR BAYESIAN CALIBRATION OF A TI-6AL-4V CRYSTAL PLASTICITY MODEL

Promit Chakroborty
Department of Civil and Environmental Engineering
Vanderbilt University
Nashville, TN 37212
promit.chakroborty@vanderbilt.edu

Sankaran Mahadevan
Department of Civil and Environmental Engineering
Vanderbilt University
Nashville, TN 37212
sankaran.mahadevan@vanderbilt.edu

October 22, 2026

ABSTRACT

Numerical simulation models that accurately predict the behavior of complex materials such as the aerospace-grade Ti-6Al-4V are often highly computationally expensive and contain a large number of model parameters that must be inferred by comparing the model predictions with experimental observations. Often, such calibration is carried out using deterministic optimization procedures, but these methods do not account for physical, data, or modeling uncertainties and may produce results that cannot be meaningfully extrapolated to new operating conditions not covered by the calibration experiments. On the other hand, uncertainty-aware calibration using the Bayesian framework is more robust, but it is difficult to carry out for a large number of parameters because of the *curse of dimensionality* and the computational expense of evaluating the simulation model many times. With this context, the contributions of this work are twofold. First, we investigate the application of Bayesian calibration, enhanced with active subspace-based dimension reduction and surrogate modeling, to a crystal plasticity finite element model that predicts the mechanical properties of a forged Ti-6Al-4V alloy. In particular, we demonstrate how dimension reduction can enable calibration in high dimensions without the use of sophisticated sampling techniques or heuristic down-selecting of parameters. Next, we study the impact of the latent space dimension (constructed through dimension reduction) on the calibrated posteriors. Most importantly, we examine how optimal single-value estimates of the model parameters differ between deterministic calibration algorithms and Bayesian point estimates, underscoring the necessity of appropriately incorporating all physical, modeling, and approximation uncertainties to ensure robust calibration outcomes.

1 Introduction

Predicting the mechanical behavior of materials is an essential first step in designing and fabricating parts across engineering disciplines. Simulation models routinely study material time-, scale-, and microstructure-dependent response by numerically solving the equations governing their behavior. However, materials of interest often have complex microstructures and exhibit intricate physical mechanisms at a variety of scales. Thus, faithfully capturing their mechanical response under a range of operating conditions requires detailed modeling of the microstructure interactions. For example, the dual-phase α / β Ti-6Al-4V (Ti64), one of the most commonly used Titanium alloys and especially widespread in aerospace applications [1], has a basket-weave microstructure [2] characterized by a primary α phase of hexagonal close-packed (HCP) crystals interspersed with the transformed β phase of body-centered cubic (BCC) crystals [3], and may even include other minor phases [4]. The relative volume fractions of these phases depend on feedstock composition, manufacturing process, and subsequent heat treatment [5], making it harder to analyze parts made from the Ti64 alloy. Predicting its material behavior requires comprehensive micromechanical simulation models to capture the interplay among the different phases.

Such sophisticated numerical models, capable of undertaking high-fidelity simulations of complex materials like Ti64, involve several parameters arising from the multiple governing equations. Before the model can make accurate predictions, these parameters must be tuned to represent the specific behavior of the material of interest, characterized using experiments that can range from tensile tests [6] to in situ high-energy X-ray diffraction experiments [7, 8]. Typically, most model calibration procedures in the literature have deterministically minimized some measure of error (e.g., sum of squared errors) between model predictions and experimental observations using a range of techniques, from gradient-based optimization or genetic algorithms [6] to physics-informed iterative fitting procedures [9], and even heuristic tuning to reproduce selected experimental trends [10, 11]. These procedures yield single-valued parameter estimates without information about calibration uncertainty, do not account for measurement errors in the experiments, and do not incorporate any prior information about the model parameters. Furthermore, experimental observations are usually sparce, and deterministically calibrated parameter values from limited data often cannot be extended to new conditions not included in the experimental dataset.

In contrast, Bayesian inference methods incorporate prior knowledge and account for data sparsity and experimental uncertainty and compute a calibrated posterior distribution of potential parameter values based on their likelihood of replicating the observed behavior, thus increasing the robustness of the calibration process and providing a measure of uncertainty in the resultant parameter estimates. Except in a few special situations, Bayesian inference employs sampling techniques such as Markov chain Monte Carlo (MCMC) to construct the posterior distribution, which can easily become computationally unaffordable depending on the cost of evaluating the simulation model and the number of samples - and hence model evaluations - needed to resolve the shape of the posterior. Thus, surrogate models are often used for inexpensive approximations of the physics-based numerical model in the calibration framework [12, 13]. A growing body of work in materials models (e.g., crystal plasticity models) has begun to incorporate surrogate-based Bayesian calibration techniques to improve the parameter inference for their numerical models [14, 15, 16, 17].

Despite the use of surrogate models, calibrating large parameter sets using sample-based strategies remains challenging due to the *curse of dimensionality*, which causes nearly all sampling algorithms to struggle with degeneracy in high-dimensional spaces [18, 19, 20]. Thus, *dimension reduction* becomes necessary, where a low-dimensional latent space is constructed such that a small number of latent parameters capture the majority of the variability of the underlying physics. The simplest such strategies involve using global sensitivity analysis in the original, physical space to focus on only the most sensitive parameters [21, 22], while more sophisticated strategies such as the method of *active subspaces* [23, 24, 25] aim to identify a small set of new directions in the parameter space (resulting in a rotation of the coordinate frame) that capture some pre-decided fraction of the numerical model's output variability. Past studies have shown the benefits of dimension reduction in surrogate modeling [26, 27] and Bayesian calibration [28, 25], motivating their use in calibration problems where several model parameters must be calibrated simultaneously.

However, tools such as dimension reduction and surrogate modeling cannot be incorporated ad hoc. Care must be taken to ensure mathematical compatibility, and because they are approximation techniques, they introduce errors and uncertainties that must be accounted for in the calibration procedure to maintain robustness in the results. Further, even for the same physics model, differences in how uncertainties are treated and represented change the "best" model parameter values. Thus, it is imperative to understand how different decisions for representing the physics and the uncertainties affect the calibration problem.

1.1 Key contributions

In [29], we proposed a robust, efficient high-dimensional Bayesian model calibration framework. Our method (henceforth referred to as the CMF) couples active subspace discovery for dimension reduction with surrogate modeling to reduce simulation costs, and comprehensively includes all the approximation uncertainties in the calibrated posteriors. The CMF has been illustrated with simple, academic examples in [29], and the current manuscript explores its application to a highly nonlinear, high-dimensional, multi-scale material model. In the CMF, constructing the active subspace requires preselecting how much of the model's output variation must be captured in the latent space. We investigate in this work how the aforementioned choice, i.e., the degree of dimension reduction, affects the calibrated results, using the crystal plasticity finite element (CPFE) model for Ti-6Al-4V developed by Shen et al. [3] as a case study. Additionally, we study the difference in single-value model parameter estimates from deterministic calibration approaches versus point estimators that incorporate modeling uncertainties through the Bayesian paradigm. The key contributions of this manuscript are as follows:

1. We implement our CMF framework on a real-world materials model to demonstrate how dimension reduction and surrogate modeling can be coupled in a physically and mathematically compatible manner for Bayesian calibration of materials models with many parameters. In our framework, the calibrated posterior distributions comprehensively include all the uncertainties introduced by the various approximations.
2. We investigate the effect of latent space dimensionality on the calibrated posteriors.

3. We discuss physics-derived strategies for extracting meaningful output features that allow us to reconstruct the full stress-strain time history while also facilitating likelihood construction.
4. We highlight differences in optimal single-valued parameter estimates between deterministic and uncertainty-aware calibration, thus justifying the need to incorporate various uncertainty sources to ensure robustness.

In Section 2, we summarize the CPFE model used in this work [3], and discuss how to formulate the Bayesian calibration of this CPFE model's parameters. Section 3 details the steps involved in implementing the CMF to solve this problem. The numerical results from the calibration procedure are presented in Section 4, and the insights gained are discussed in Section 5.

2 Methodology: Calibration problem formulation

Shen et al. developed a comprehensive CPFE model for conventional cold-forged Ti64 in [3] and used a genetic algorithm to deterministically calibrate eleven parameter types by minimizing the difference between the stress-strain response predicted by volume-averaged CPFE model simulations and experimental observations, distinguishing between slip systems where relevant. These parameters, associated with the primary α phase of Ti64 and corresponding to tensile behavior, were chosen based on feasibility considerations, given that only a single tension test experiment was available. The rest of the parameters were held constant at values calibrated from experiments on the Titanium alloy Ti6242 [9, 11], due to the similarity between the two materials, and the fact that the β phase accounts for only $\approx 1.92\%$ of the volume. A more detailed justification of the assumptions guiding the parameter selection can be found in [3].

Ramesh Babu et al. [30] build on this work, conducting Bayesian calibration for the same parameters. Here, deterministic relationships are assumed between parameters describing the same quantity for different slip systems, based on observations made by Shen et al. [3], leaving a total of 11 independent parameters. After correlation and sensitivity analyses, they use a Gaussian Process surrogate to reduce the computational cost of the calibration procedure, and build the likelihood function in accordance with the seminal work of Kennedy & O'Hagan [12] (KOH). To address the challenge of high-dimensional sampling (even sampling in an 11-dimensional space is difficult), they use Metropolis-Hastings (MH) with a mixture proposal kernel consisting of both a local and an independent (i.e., global) proposal component. The parameters for the MH algorithm are tuned manually based on engineering judgment. Further, separate calibration analyses are conducted for subsets of the 11 parameters based on their Sobol' index sensitivity contributions; undertaking sampling in these lower-dimensional spaces allowed more robust posteriors to be estimated for the most sensitive model parameters.

In this work, we calibrate the same set of 11 parameters using the CMF calibration framework [29], which integrates derivative-based sensitivity analysis with active subspace dimension reduction and stochastic field surrogate modeling to enhance the efficiency and robustness of Bayesian calibration. By removing the need for heuristic tuning of solution algorithm parameters and explicitly accounting for the uncertainties stemming from the dimension reduction and the surrogate approximation, this framework provides a comprehensive methodology to incorporate the uncertainty arising from the dimension reduction and surrogate approximation within the calibration results for all 11 parameters, while reducing the computational complexity and cost through the use of a low-dimensional surrogate that supports generating samples from the posterior distribution in the latent space. Note that Ramesh Babu et al.'s [30] calibration is performed in the original physical space, whereas CMF calibration is done in the latent space.

Section 2.1 provides a short summary of the CPFE model, while Section 2.3 describes the experimental data used for calibration. Sections 2.2 and 2.3 present details on the chosen calibration parameters and the CPFE outputs, respectively. Finally, Section 2.4 discusses the simulations conducted using the CPFE model, as well as specifics on how this simulation data is pre-processed for use in the Bayesian calibration framework.

2.1 Brief overview of the Ti-6Al-4V CPFE model

The CPFE model developed by Shen et al. [3] to predict the micro-mechanical behavior of Ti64 is a rate- and size-dependent crystal plasticity constitutive model based on earlier models of α / β dual-phase Titanium alloys [9, 11, 31]. Microstructural image-based statistically equivalent representative volume elements (SERVEs) - optimally sized virtual microstructure domains whose morphological and crystallographic statistics match scans from electron backscatter diffraction (EBSD) experiments - are used to model the microstructure and simulate the intra- and inter-granular interactions. The transformed β phase is treated using an effective homogenized constitutive model randomly embedded in the domain with the appropriate volume fraction. Each SERVE is generated using the DREAM.3D software [32], then meshed into 370396 four-node linear tetrahedral elements (TET4) with periodic boundary conditions to approximate the effect of large microstructural volumes in the finite-element homogenization process.

To accurately predict stress localizations, crack nucleation, dwell fatigue, intragranular plasticity, intergranular interactions, and other complex phenomena strongly affected by microstructure properties such as grain orientation, morphology, and phase distribution, the CPFE model must explicitly capture a variety of mechanisms. Several key features are thus deemed to be necessary to accurately predict the material response of Ti64 under a variety of loading and microstructure conditions. The anisotropy in the stress-strain relationships is characterized using the fourth-order elasticity tensor, which has transversely isotropic symmetry in the α (HCP) phase and cubic symmetry in the β (BCC) phase. Plastic deformation evolves according to a rate-sensitive power-law slip-rate formulation adopted from prior work [31, 33], and includes the Hall-Petch effect [34, 35] for grain-size-dependent strengthening and the Armstrong-Frederick law [36] for back stress evolution. The α phase is modeled with 30 HCP slip systems, while the effective homogenized representation for the transformed β phase has a total of 78 slip systems - 30 HCP and 48 BCC - corresponding to the alternating α and β laths. Geometrically necessary dislocations arising from the crystal lattice curvature caused by plastic strain gradients near the grain boundaries of polycrystalline aggregates are also accounted for, as is the transient yield point phenomenon associated with rapid dislocation multiplication in the microstructure. Overall, the CPFE model has over 150 parameters; a detailed account of these and a deeper discussion of the model formulation and development can be found in Shen et al. [3].

2.2 Chosen calibration parameters

Of the total 150 independent parameters involved in the CPFE model, Shen et al. [3] and Ramesh Babu et al. [30] chose the following:

- Elasticity matrix components C_{11}^{α}, C_{33}^{α}, C_{55}^{α}, C_{12}^{α}, and C_{13}^{α}, where the variables correspond to the matrix components in Voigt notation when written in material symmetry coordinates. C_{22}^{α}, C_{66}^{α}, and C_{23}^{α} are also calibrated automatically due to the below relationships:

$$C_{22}^{\alpha} = C_{11}^{\alpha}, \qquad C_{66}^{\alpha} = C_{55}^{\alpha}, \qquad C_{23}^{\alpha} = C_{13}^{\alpha} \tag{1}$$

  Additional constraints are enforced on these parameters to ensure physical realism and consistency with observed material behavior. Eq. (2) ensures positive stiffness in the longitudinal and transverse directions, while Eq. (3) enforces that the former exceeds the latter, making the model consistent with most engineering materials, where the primary loading direction has higher stiffness than the transverse directions. Finally, Eq. (4) imposes positive definiteness on the stiffness matrix for both mechanical and numerical stability.

$$C_{11}^{\alpha} > 0 \qquad \& \qquad C_{55}^{\alpha} > 0 \tag{2}$$

$$C_{11}^{\alpha} > C_{12}^{\alpha} \tag{3}$$

$$\left(C_{11}^{\alpha} + C_{12}^{\alpha}\right) C_{33}^{\alpha} - 2\left(C_{13}^{\alpha}\right)^2 > 0 \tag{4}$$

- α-phase Hardening parameters g_0^{α}, K^{α}, h_0^{α}, and $\tilde{g}_s^{\alpha}$. K^{α} and h_0^{α} are assumed to be the same for all 30 slip systems, while g_0^{α} and $\tilde{g}_s^{\alpha}$ take different values based on one of four different slip families: Basal <a>, Prismatic <a>, Pyramidal <a>, and Pyramidal <c+a>. Shen et al. [3] calibrate the four g_0^{α} independently, but enforce fixed ratios between the $\tilde{g}_s^{\alpha}$ parameters, as shown in Eq. (5). However, Ramesh Babu et al. [30] also enforce physical constraints between the g_0^{α} cases, as per Eq. (5).

$$\begin{aligned}
\left(g_0^{\alpha}\right)_{\text{prismatic <a>}} &= 0.9673\left(g_0^{\alpha}\right)_{\text{basal <a>}}, \qquad \left(\tilde{g}_s^{\alpha}\right)_{\text{prismatic <a>}} = 1.2127\left(\tilde{g}_s^{\alpha}\right)_{\text{basal <a>}} \\
\left(g_0^{\alpha}\right)_{\text{pyramidal <a>}} &= 1.4727\left(g_0^{\alpha}\right)_{\text{basal <a>}}, \qquad \left(\tilde{g}_s^{\alpha}\right)_{\text{pyramidal <a>}} = 1.2127\left(\tilde{g}_s^{\alpha}\right)_{\text{basal <a>}} \\
\left(g_0^{\alpha}\right)_{\text{pyramidal <c+a>}} &= 1.9273\left(g_0^{\alpha}\right)_{\text{basal <a>}}, \qquad \left(\tilde{g}_s^{\alpha}\right)_{\text{pyramidal <c+a>}} = 3.2979\left(\tilde{g}_s^{\alpha}\right)_{\text{basal <a>}}
\end{aligned} \tag{5}$$

- Yield point phenomenon parameter $\tilde{\tau}_p^{\alpha}$, which is assumed to be the same for all 30 slip systems.
- Rate exponent m.

After these simplifications, 11 independent parameters are considered for calibration. Thus, we write the calibration parameter vector $\mathbf{v}$ for our work in this manuscript as

$$\mathbf{v} = \begin{bmatrix} v_1 & \dots & v_{11} \end{bmatrix}^T = \begin{bmatrix} C_{11}^{\alpha} & C_{33}^{\alpha} & C_{55}^{\alpha} & C_{12}^{\alpha} & C_{13}^{\alpha} & g_0^{\alpha} & K^{\alpha} & h_0^{\alpha} & \tilde{g}_s^{\alpha} & \tilde{\tau}_p^{\alpha} & m \end{bmatrix}^T \tag{6}$$

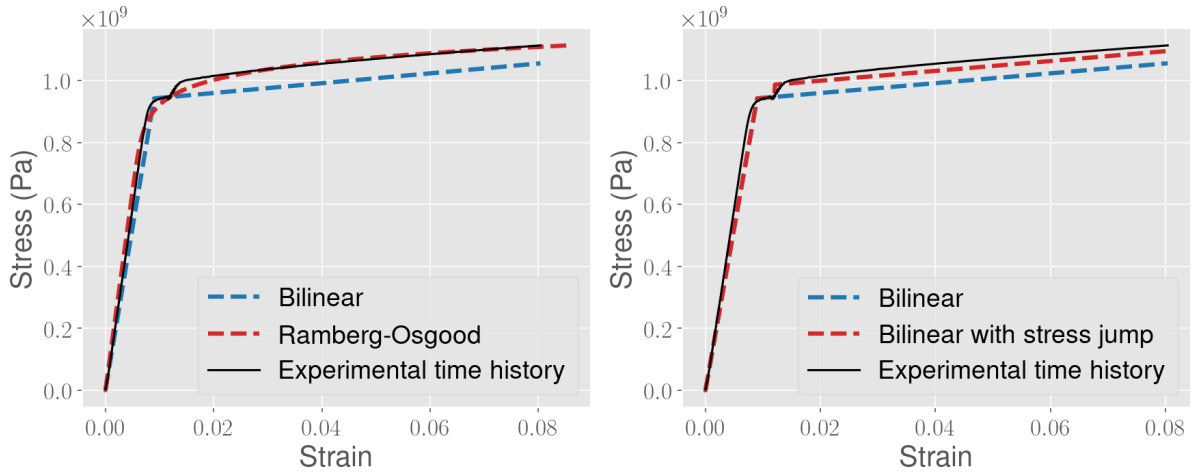
where g_0^{α} and $\tilde{g}_s^{\alpha}$ denote $\left(g_0^{\alpha}\right)_{\text{basal <a>}}$ and $\left(\tilde{g}_s^{\alpha}\right)_{\text{basal <a>}}$, respectively. Following [30], the priors $\pi\left(v_i\right)$ are taken to be independent Uniform, centered at the deterministically calibrated (nominal) values of each parameter from [3] and ranging between $\pm 30\%$ of the nominal value, as shown in Table 1. These ranges reflect, to some extent, the physically admissible values for these parameters.

Parameter	Distribution	Lower Bound	Upper Bound
C_{11}^{α}	Uniform	129.9×10^{9}	194.9×10^{9}
C_{33}^{α}	Uniform	144.6×10^{9}	216.8×10^{9}
C_{55}^{α}	Uniform	37.4×10^{9}	56.0×10^{9}
C_{12}^{α}	Uniform	73.6×10^{9}	110.4×10^{9}
C_{13}^{α}	Uniform	55.2×10^{9}	82.8×10^{9}
$\tilde{g}_{0}^{\alpha}$	Uniform	220.0×10^{6}	330.0×10^{6}
$\tilde{g}_{s}^{\alpha}$	Uniform	376.0×10^{6}	564.0×10^{6}
h_{0}^{α}	Uniform	120.0×10^{6}	180.0×10^{6}
K^{α}	Uniform	131.6×10^{3}	197.4×10^{3}
$\tilde{\tau}_{p}^{\alpha}$	Uniform	48.0×10^{6}	72.0×10^{6}
m	Uniform	8.4×10^{-3}	12.6×10^{-3}

Table 1: Prior distributions for the chosen calibration parameters.

2.3 CPFE model output and experimental data

The experimental data used for calibration comes from a single strain-controlled uniaxial tensile test on a coupon at room temperature until specimen failure, in accordance with ASTM standards [3]. A controlled strain rate was applied in the longitudinal direction, with $\dot{\varepsilon} = 1.0 \times 10^{-4}\text{s}^{-1}$ until the total strain reached $\varepsilon \approx 1.2\%$, after which the strain rate was made to undergo a discrete jump to $\dot{\varepsilon} = 5.0 \times 10^{-3}\text{s}^{-1}$. Figure 1 plots the experimentally observed stress-strain curve. This loading condition is matched and provided as input to the CPFE model. Since the test measures the material's bulk response, we use the SERVE-averaged stress-strain response predicted by the CPFE model for calibration.



(a) (b)

Figure 1: Plot of the stress-strain history measured from the uniaxial tensile test [3], along with idealized representations. (a) Compares the original experimental time history (solid black line) with the bilinear (dashed blue line) and Ramberg-Osgood (dashed red line) idealizations. (b) Compares the original experimental time history (solid black line) with the standard bilinear idealization (dashed blue line) and a bilinear trend with a stress jump accounting for the effect of the discrete strain-rate change. A clear kink can be observed in the experimental time history around 1.2% strain, corresponding to the discrete jump in the applied strain rate.

However, this stress-strain curve, measured as a time-discretized array, is high-dimensional, making it difficult to use for Bayesian calibration. Extracting a small set of important features from high-dimensional model outputs can be achieved through functional data analysis [37], data-driven dimension reduction techniques like principal component analysis (PCA) [27], or through the selection of carefully chosen physically interpretable metrics (e.g., characteristic material response quantities [38]). Because the stress-strain behavior of metals is well studied and consistent across a

wide range of materials, the scientific community accepts some common idealizations for this curve. Using such an idealization allows us to effectively represent the entire time history using a few explainable features while retaining the ability to approximately reconstruct the stress-strain behavior. Here, we evaluate the bilinear and the Ramberg-Osgood power-law trends - two of the most common idealizations for the mechanical behavior of metals.

The bilinear trend is characterized by three parameters: the Young's modulus or modulus of elasticity E, which is the slope of the linear-elastic region; the yield strength σ_y, which marks the onset of plastic deformation; and the hardening slope H, which models plastic strain-hardening behavior. All three parameters can be estimated using practical and physically interpretable techniques. E and H are simply the slopes of the elastic and strain-hardening regions and can be determined by linear regression of the respective regions, while the yield strength σ_y is commonly found by locating the intersection point of the stress-strain curve with the offset line (a line parallel to the linear-elastic region starting at 0.2% strain) - a procedure known as the 0.2% offset method. On the other hand, the Ramberg-Osgood power-law trend, which is given as

$$\varepsilon = \frac{\sigma}{E} + K\left(\frac{\sigma}{E}\right)^n \tag{7}$$

is easiest to fit through least-squares regression, and this is the procedure adopted here. It is also a three-parameter model, where E is again the Young's modulus, K is known as the amplitude stress, and n is called the shape factor parameter.

The kink in the stress-strain curve, caused by the discrete change in applied strain rate partway through the test, adversely affects the applicability of both the bilinear and Ramberg-Osgood trends, as shown in Figure 1a. Because both idealizations assume a constant strain rate across the time history, they cannot account for the sharp jump in stress at the point where the strain rate changes. The bilinear model, whose features are extracted through engineering procedures, underestimates the stress throughout the plastic regime after $\varepsilon \approx 1.2\%$ by an amount equal to the strain rate change induced stress jump. In contrast, the Ramberg-Osgood model, fit using least-squares regression, matches more closely at first glance but adopts a far more rounded transition from elasticity to plasticity than is expected for the material. Additionally, although not shown here, we found the Ramberg-Osgood fitting procedure to be highly variable because of the kink in the time history, causing the power-law fit to poorly represent the elastic regime for several of the CPFE simulations.

To better capture the observed (and CPFE-predicted) stress-strain behavior, we modify the bilinear model to account for the stress jump, as shown in Figure 1b. Because we extract the bilinear trend features using simple numerical procedures localized to specific regions of the time history, we can add a simple stress correction to the bilinear trend and reproduce the observed kink. This is done using the strain-rate sensitivity parameter n, which is defined as

$$n = \frac{\partial \sigma}{\partial \dot{\varepsilon}} \tag{8}$$

Since the strain rate undergoes a discrete jump, we can use Eq. (8) to estimate the near-instantaneous increase in stress as

$$\Delta\sigma = \sigma_{1.2\%}\left[\left(\frac{\dot{\varepsilon}_{\text{hi}}}{\dot{\varepsilon}_{\text{lo}}}\right)^n - 1\right] \tag{9}$$

where $\sigma_{1.2\%}$ is the stress at $\varepsilon = 1.2\%$ (the point at which the strain rate changes), $\dot{\varepsilon}_{\text{hi}} = 5 \times 10^{-3}\text{s}^{-1}$ is the new strain rate after the change, and $\dot{\varepsilon}_{\text{lo}} = 1 \times 10^{-4}\text{s}^{-1}$ is the original strain rate before the change. To apply this correction to the bilinear model, we take $\sigma_{1.2\%}$ as the stress predicted by the bilinear trend at $\varepsilon = 1.2\%$. While n can be treated as an additional parameter to be extracted from the stress-strain curve, resulting in a four-parameter "corrected bilinear" trend, in our case we expect the rate exponent m to be a good approximation for n. (The corrected bilinear graph in Figure 1b uses the deterministically calibrated m value from [3] as a proxy for n, and we see that the stress jump is acceptably captured.) Thus, even the three-parameter bilinear curve can be used to represent the stress-strain curve produced by our modeling conditions after the correction, and the vector $\begin{bmatrix} E & \sigma_y & H \end{bmatrix}^T$ contains our output features of interest from the CPFE simulations, to be compared against the experiment in the calibration procedure.

Here, we note that both σ_y and H are strain rate dependent quantities, and the specimen technically begins plastic deformation before the strain rate changes. Thus, strictly speaking, σ_y and H are not constant across the kink. However, we argue that in this case the change in σ_y and H can be ignored without introducing significant error, for two main reasons. First, both strain rates ($1 \times 10^{-4}\text{s}^{-1}$ and $5 \times 10^{-3}\text{s}^{-1}$) fall within the quasi-static regime, implying that the expected change in both σ_y and H should be small. Further, the specimen spends very little time undergoing plastic deformation under the initial strain rate. Thus, even if H changed significantly with the increase in strain rate, using the value of H for $\varepsilon > 1.2\%$ to represent the short region of plastic deformation where $\varepsilon < 1.2\%$ will not lead to a meaningful deviation in the predicted stress values, since there is not much room for deviation between the "wrong" and "right" curves. Additionally, because the elastic regime ends before the strain rate changes, variation in σ_y does

not affect the constructed bilinear trend. Thus, we use the 0.2% offset method to estimate σ_y, and extract H from the strain-hardening region post $\varepsilon > 1.2\%$, and ignore the change in these values due to the change in strain rate.

2.4 Data collection and pre-processing

Due to the complex micro-mechanical simulations involved, each CPFE run takes several hours; Ramesh Babu et al. [30] show that on average, a single model evaluation takes 2 to 15 hours of wall-clock time. This limits the total number of simulation runs affordable given the computational resource limitations. Enough information about the variability of the CPFE predictions with respect to the 11 calibration parameters must be collected within this computational budget to support a thorough and comprehensive calibration study. The surrogate model embedded in the CMF [29], which we use here, eliminates the need to run the CPFE model at each candidate MCMC sample in generating the posterior distribution during the Bayesian calibration. Instead, the full-physics simulations need only support the dimension reduction and surrogate model construction steps for the CMF. A Latin hypercube design-of-experiments (LHS-DOE) with $D = 250$ samples was generated from the prior distribution $\pi(\mathbf{v})$ in [30], and this is sufficient for the CMF (see Section 3.1 for further details). Thus, in this work, we directly borrow the 250 CPFE runs and the corresponding LHS-DOE from [30].

The simulation data is obtained as full stress-strain curves corresponding to samples of the parameter vector $\mathbf{v}$ as defined in Section 2.2. Pre-processing steps are necessary for both the input parameters and the outputs before they can be used in the CMF. The CMF's dimension reduction step strongly prefers parameters to follow standard Gaussian distributions, whereas $\pi(\mathbf{v})$ is jointly uniform. Hence, we construct a "*standardized calibration vector*" $\boldsymbol{\theta} = \begin{bmatrix}\theta_1 & \dots & \theta_{11}\end{bmatrix}^T$, such that $\pi(\boldsymbol{\theta})$ follows the probability density function (pdf) of the 11-dimensional standard normal distribution $\mathcal{N}(\mathbf{0}_{11}, \mathbf{1}_{11})$, and

$$v_i = l_i + (r_i - l_i)\,\Phi(\theta_i) \tag{10}$$

where $\Phi(\cdot)$ is the cumulative distribution function of a univariate standard normal distribution, and l_i and r_i are the left and right limits of the interval support of v_i, respectively, as listed in Table 1. In essence, θ_i is obtained by applying the inverse CDF method to v_i to transform the specified uniform prior into a standard Gaussian prior. We also denote its dimensionality by $d_{\boldsymbol{\theta}} = 11$. (Henceforth, we use $d_{\boldsymbol{a}}$ to denote the dimension of any vector $\boldsymbol{a}$.)

For each CPFE simulation, the features E, σ_y, and H are extracted from the predicted stress-strain curves, following the procedure described in Section 2.3. However, these features are correlated, whereas we prefer uncorrelated output features for mathematical simplicity in the subsequent analysis. Thus, we compute the principal components of the vector $\begin{bmatrix}E & \sigma_y & H\end{bmatrix}^T$ (after the usual mean-centering and standard-deviation-normalization) using standard principal component analysis (PCA) and retain all three axes after rotation, obtaining uncorrelated features without information loss. Consequently, we write our $d_{\boldsymbol{f}} = 3$ dimensional output vector as

$$\boldsymbol{f}(\boldsymbol{\theta}) = \begin{bmatrix}f_1(\boldsymbol{\theta}) & f_2(\boldsymbol{\theta}) & f_3(\boldsymbol{\theta})\end{bmatrix}^T = \mathbf{T}^T \cdot \begin{bmatrix}E & \sigma_y & H\end{bmatrix}^T \tag{11}$$

The matrix $\mathbf{T}$ contains as its columns the eigenvectors of the covariance matrix of $\begin{bmatrix}E & \sigma_y & H\end{bmatrix}^T$, estimated using the 250 LHS-DOE samples. Since we use standard PCA to find $\mathbf{T}$, we also obtain the fractional variance contribution of each of the three features, which indicates their relative importance. These values, collected in the vector $\boldsymbol{\nu} = \begin{bmatrix}\nu_1 & \nu_2 & \nu_3\end{bmatrix}^T$, are reported in Table 2.

As explained in Section 3.1, the dimension reduction component of the CMF requires evaluations of the gradient of the negative log-likelihood function with respect to the calibration parameters, which involves gradients of the CPFE model outputs, i.e., in this case $\nabla_{\boldsymbol{\theta}} \boldsymbol{f}(\boldsymbol{\theta})$. Computing these gradients numerically would require, at minimum, an additional 11 CPFE evaluations at each LHS-DOE location, which is clearly infeasible. Instead, we build an initial surrogate to approximate the gradients, using an analogous procedure to that used in [30]. Gaussian Process (GP) regression is used to build highly flexible surrogates by treating the model prediction as an instance of a Gaussian process with a specified autocorrelation structure. Details on constructing and training GPs can be found in [39]. Here, we build an individual GP $g_i(\boldsymbol{\theta})$ for each output feature $f_i(\boldsymbol{\theta})$ $(i = 1, 2, 3)$, using a white noise kernel added to a Matérn 5/2 kernel multiplied by a magnification constant for the covariance matrix. The vector $\boldsymbol{g}(\boldsymbol{\theta}) = \begin{bmatrix}g_1(\boldsymbol{\theta}) & g_2(\boldsymbol{\theta}) & g_3(\boldsymbol{\theta})\end{bmatrix}^T$ thus serves as an approximation for $\boldsymbol{f}(\boldsymbol{\theta})$. The 250 simulation samples are split into a set of 210 training data and 40 test data points; we use the test points to validate the GP constructed using the training data set, with Table 2 presenting some standard error metrics to quantify this goodness-of-fit. We find that while the error in our initial GPs seems high when predicting the decorrelated parameters, they perform similarly to those in [30] when predicting the features of

interest, making them sufficient for our case. (The higher errors when predicting the decorrelated output features stem from their decreasing importance.) We use these GPs primarily to predict the gradients needed for the CMF. Section 3.1 discusses how these gradients are used.

Output features		Variance contribution	R^2	MAPE	RMSE
Decorrelated features	$f_1(\boldsymbol{\theta})$	61.02%	0.97	43.47%	0.25
	$f_2(\boldsymbol{\theta})$	28.12%	0.98	22.53%	0.12
	$f_3(\boldsymbol{\theta})$	10.85%	0.69	217.88%	0.27
Original features	E	-	0.99	0.82%	1.25×10^9
	σ_y	-	0.98	0.66%	7.17×10^6
	H	-	0.90	1.92%	3.73×10^7

Table 2: Goodness-of-fit results for Gaussian process surrogates used to estimate gradients at the LHS-DOE points. Values for coefficient of determination (R^2), mean absolute percentage error (MAPE), and root-mean-square error (RMSE) are reported for both the principal component, i.e., decorrelated output vector $\boldsymbol{f}(\boldsymbol{\theta})$ and the physical features of interest. Additionally, the variance contribution of the decorrelated output components $f_i(\boldsymbol{\theta})$ $(i = 1, 2, 3)$ are provided as an indicator of their relative importance.

2.5 Bayesian Calibration Formulation

Section 2.4 defined $\boldsymbol{\theta}$ and $\boldsymbol{f}(\boldsymbol{\theta})$ to facilitate the implementation of the CMF; we construct the Bayesian inference problem in terms of these quantities. The posterior density of the parameters $\boldsymbol{\theta}$, denoted $p(\boldsymbol{\theta})$, is given by

$$p(\boldsymbol{\theta}) \propto \mathcal{L}(\boldsymbol{\theta})\, \pi(\boldsymbol{\theta}) \tag{12}$$

where $\pi(\boldsymbol{\theta})$ is the prior distribution of $\boldsymbol{\theta}$ as defined before, and $\mathcal{L}(\boldsymbol{\theta})$ is the likelihood function.

To construct the likelihood function, let us first define $\mathbf{y}$ as the vector of decorrelated output features of interest from the experimental time history, i.e., $\mathbf{y} = \mathbf{T}^T \cdot \begin{bmatrix} E^{(\text{obs})} & \sigma_y^{(\text{obs})} & H^{(\text{obs})} \end{bmatrix}^T$ (analogous to Eq. (11)), where $E^{(\text{obs})}$, $\sigma_y^{(\text{obs})}$, and $H^{(\text{obs})}$ are the Young's modulus, yield strength, and hardening slope extracted from the experimental stress-strain curve, respectively. Due to modeling inadequacies in the CPFE model and observation errors in the experimental setup, $\mathbf{y}$ will deviate from the CPFE prediction $\boldsymbol{f}(\boldsymbol{\theta}_T)$ evaluated using the "true" (perfectly calibrated) model parameters $\boldsymbol{\theta}_T$. Hence, following the work by Kennedy and O'Hagan [12],

$$\mathbf{y} = \boldsymbol{f}(\boldsymbol{\theta}_T) + \boldsymbol{\delta} + \boldsymbol{\epsilon} \tag{13}$$

where the unknown constant vector $\boldsymbol{\delta} = \begin{bmatrix} \delta_1 & \delta_2 & \delta_3 \end{bmatrix}^T$ measures the model discrepancy / inadequacy and the random vector $\boldsymbol{\epsilon}$ accounts for the observation error. We do not have measurements of the observation noise from the experimental setup, so here we assume $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon)$. Since we know the variance fractions $\boldsymbol{\nu}$ of the decorrelated features (see Section 2.3), we specify $\boldsymbol{\Sigma}_\epsilon = \text{Diag}(\eta \cdot \boldsymbol{\nu})$, where $\text{Diag}(\cdot)$ indicates a diagonal matrix, and η is an unknown noise magnitude term. This strategy ensures that adding observation noise does not break the known relative importance of the decorrelated output features, unlike if an independent noise variance was added to each feature.

In [12], the unknown model discrepancy and noise magnitude are inferred along with the parameters $\boldsymbol{\theta}$. To do so, we first select prior densities for $\boldsymbol{\delta}$ and η based on the problem and prior knowledge. Since these terms may not necessarily have Gaussian priors, we define $\tilde{\boldsymbol{\delta}}$ and $\tilde{\eta}$, which are obtained through the necessary transformations on $\boldsymbol{\delta}$ and η such that $\tilde{\boldsymbol{\delta}}$ and $\tilde{\eta}$ can always be assumed to have standard normal priors regardless of the prior distribution on $\boldsymbol{\delta}$ and η. (Of course, $\tilde{\boldsymbol{\delta}}$ has the standard normal prior in $d_{\boldsymbol{\delta}} = d_{\boldsymbol{f}} = 3$ dimensions, while $\tilde{\eta}$ has the one-dimensional standard normal prior.) Then, the augmented calibration vector is constructed as $\boldsymbol{\Theta} = \begin{bmatrix} \boldsymbol{\theta} & \tilde{\boldsymbol{\delta}} & \tilde{\eta} \end{bmatrix}^T$. Since each component of $\boldsymbol{\Theta}$ has a standard normal prior by construction, $\pi(\boldsymbol{\Theta})$ is the standard normal density in $d_{\boldsymbol{\Theta}} = 15$ dimensions. The posterior $p(\boldsymbol{\Theta})$ follows Eq. (12) with the likelihood $\mathcal{L}(\boldsymbol{\Theta})$.

The relationship from Eq. (13) can be extended to measure the deviation between the experimental observation and the CPFE prediction for any set of parameter realizations $\boldsymbol{\theta}$, not just the "true" $\boldsymbol{\theta}_T$. Further, since we numerically compute the gradients of the negative log-likelihood function using predictions from the GPs described in Section 2.4, we replace $\boldsymbol{f}(\boldsymbol{\theta})$ with $\boldsymbol{g}(\boldsymbol{\theta})$ to capture the prediction uncertainty in the gradients. Since both $\boldsymbol{g}(\boldsymbol{\theta})$ and $\boldsymbol{\epsilon}$ are Gaussian,

$$\mathcal{L}(\boldsymbol{\Theta}) \propto \exp\left[-\frac{1}{2} \left(\mathbf{y} - \boldsymbol{g}(\boldsymbol{\theta}) - \boldsymbol{\delta} \right)^T \cdot \left(\boldsymbol{\Sigma}(\boldsymbol{\theta}) \right)^{-1} \cdot \left(\mathbf{y} - \boldsymbol{g}(\boldsymbol{\theta}) - \boldsymbol{\delta} \right) \right] \tag{14}$$

where $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_\epsilon + \boldsymbol{\Sigma}_g(\boldsymbol{\theta})$, with $\boldsymbol{\Sigma}_g(\boldsymbol{\theta}) = \text{Diag}\left(\begin{bmatrix}\sigma^2_{g_1}(\boldsymbol{\theta}) & \sigma^2_{g_2}(\boldsymbol{\theta}) & \sigma^2_{g_3}(\boldsymbol{\theta})\end{bmatrix}\right)$ denoting the matrix of predicted variances from the GPs. We emphasize here that the only reason for building the GPs for the output features and using $\boldsymbol{g}(\boldsymbol{\theta})$ in the likelihood definition is because gradients are not available directly from the simulation model. (The need for gradients will be explained in Section 3.1.) Therefore, surrogates are necessary to numerically evaluate the gradients, and thus the likelihood should account for the uncertainty of the GPs to maintain robustness in the uncertainty quantification and Bayesian calibration results.

In Section 3, we will also extensively use the *misfit function* $\mathcal{M}(\boldsymbol{\Theta})$, which is defined as the negative log-likelihood from Eq. (14), i.e.,

$$\mathcal{L}(\boldsymbol{\Theta}) \propto \exp\left(-\mathcal{M}(\boldsymbol{\Theta})\right) \tag{15}$$

$$\text{where } \mathcal{M}(\boldsymbol{\Theta}) = \frac{1}{2}\left(\mathbf{y} - \boldsymbol{g}(\boldsymbol{\theta}) - \boldsymbol{\delta}\right)^T \cdot \left(\boldsymbol{\Sigma}(\boldsymbol{\theta})\right)^{-1} \cdot \left(\mathbf{y} - \boldsymbol{g}(\boldsymbol{\theta}) - \boldsymbol{\delta}\right) \tag{16}$$

By virtue of being in the log-space, the misfit function has better numerical scaling than the likelihood function. The squared-exponential nature of the likelihood also has a strong smoothing effect on the difference between the observed and predicted outputs, while the misfit function is more sensitive to these differences.

3 Methodology: Calibration problem solution

In this section, we discuss the implementation details of the efficient high-dimensional calibration framework CMF [29] to calibrate our Ti64 CPFE model of interest. Brief overviews of the mathematical formulation are provided to support the discussion of the implementation; for a more thorough treatment of the CMF formulation, the reader is referred to [29].

3.1 Dimension reduction with likelihood-informed subspaces (LIS)

The CMF follows [25] in achieving its dimension reduction by constructing an active subspace for the misfit function $\mathcal{M}(\boldsymbol{\Theta})$. The method of active subspaces (AS) [23] is a technique for reducing the dimensionality of the input space of a function of interest, such that the variability of the function captured within the latent space is nearly equal to that of the function in the original space. This is done by identifying the principal components of the covariance matrix of the function's gradient, which can be shown to be the directions that maximize the derivative-based global sensitivity measure (DGSM) of the function, i.e., the expected value of the function's squared directional derivative. When an active subspace is constructed for the misfit function to be used to reduce the dimensionality of a Bayesian calibration framework, it is termed the Likelihood-informed subspace (LIS) [25, 29].

To identify the active subspace, we first form the gradient data matrix $\hat{\mathbf{B}}_{\mathcal{M}}$

$$\hat{\mathbf{B}}_{\mathcal{M}} = \frac{1}{\sqrt{M}}\left[\nabla\mathcal{M}(\boldsymbol{\Theta}_1) \quad \dots \quad \nabla\mathcal{M}(\boldsymbol{\Theta}_M)\right] \tag{17}$$

Here, M is the number of samples of $\boldsymbol{\Theta}$ at which the gradient is evaluated, and $\boldsymbol{\Theta}_k$ are samples from $\pi(\boldsymbol{\Theta})$. Since the initial GPs must necessarily be used to estimate the gradients, there is no benefit to evaluating the gradients specifically at the LHS-DOE points (after suitable transformations), although this can be done and augmented with samples of $\boldsymbol{\delta}$. Here, we generate a set of M samples directly from the prior. Again, we emphasize that if gradient evaluations were directly available from the simulation model, we would not build the initial GPs for the output features, and would use the simulation outcomes to form the gradient data matrix $\hat{\mathbf{B}}_{\mathcal{M}}$.

Then, the singular value decomposition (SVD) of $\hat{\mathbf{B}}_{\mathcal{M}}$ produces the estimated eigenpairs $\left(\hat{\lambda}_j, \hat{\mathbf{w}}_j\right)$, $j = 1, \dots, d_{\boldsymbol{\Theta}}$, as the diagonal elements of the matrix of estimated eigenvalues $\hat{\boldsymbol{\Lambda}}$ and the columns of the matrix of estimated eigenvectors $\hat{\mathbf{W}}$, respectively, where

$$\hat{\mathbf{B}}_{\mathcal{M}} = \hat{\mathbf{W}}\sqrt{\hat{\boldsymbol{\Lambda}}}\hat{\boldsymbol{\mathcal{W}}}^T \tag{18}$$

and $\hat{\boldsymbol{\mathcal{W}}}$ is a unitary matrix, which is not used further. (Note here that the eigenvalues in $\hat{\boldsymbol{\Lambda}}$ are arranged in descending order.) The eigenvectors correspond to the active directions in the original parameter space, while the eigenvalues measure the function gradient variance fraction explained by the corresponding active vector.

To estimate the number of gradient evaluations M necessary to ensure that all the eigenvectors are estimated with high confidence, we follow an empirical guidance from [23], viz.,

$$M = \gamma d_{\boldsymbol{\Theta}} \ln d_{\boldsymbol{\Theta}} \tag{19}$$

$\gamma \in [2, 10]$ is an oversampling factor, and the interval is suggested in [23] based on empirical observations. For this work, we choose $\gamma = 3$.

Using the matrix of eigenvectors, we can define the "*rotated parameter vector*" $\boldsymbol{Z} = \begin{bmatrix} Z_1 & \dots & Z_{d_{\boldsymbol{\Theta}}} \end{bmatrix}^T$ as

$$\boldsymbol{Z} = \hat{\mathbf{W}}^T \boldsymbol{\Theta}, \tag{20}$$

which is then partitioned into the latent vector (or the low-dimensional parameter vector) $\boldsymbol{Z}_{d_l} = \begin{bmatrix} Z_1 & \dots & Z_{d_l} \end{bmatrix}^T$ and the vector of inactive parameters $\boldsymbol{Z}_\perp = \begin{bmatrix} Z_{d_l+1} & \dots & Z_{d_{\boldsymbol{\Theta}}} \end{bmatrix}^T$ (with the integer $1 \leq d_l \leq d_{\boldsymbol{\Theta}}$). These can equivalently be defined

$$\boldsymbol{Z}_{d_l} = \hat{\mathbf{W}}_{d_l}^T \boldsymbol{\Theta} \quad \text{where} \quad \hat{\mathbf{W}}_{d_l} = \begin{bmatrix} \hat{\mathbf{w}}_1 & \dots & \hat{\mathbf{w}}_{d_l} \end{bmatrix} \tag{21}$$

$$\boldsymbol{Z}_\perp = \hat{\mathbf{W}}_\perp^T \boldsymbol{\Theta} \quad \text{where} \quad \hat{\mathbf{W}}_\perp = \begin{bmatrix} \hat{\mathbf{w}}_{d_l+1} & \dots & \hat{\mathbf{w}}_{d_{\boldsymbol{\Theta}}} \end{bmatrix} \tag{22}$$

$$\text{and } \hat{\mathbf{W}} = \begin{bmatrix} \hat{\mathbf{W}}_{d_l} & \hat{\mathbf{W}}_\perp \end{bmatrix} \tag{23}$$

The dimensionality of the latent space d_l, i.e., the number of active directions retained after dimension reduction, is chosen based on the spectral gap $\hat{\lambda}_{d_l} - \hat{\lambda}_{d_l+1}$, which bounds the error in the estimated active subspace, and the sum of the retained eigenvalues $\sum_{j=1}^{d_l} \hat{\lambda}_j$, which captures the percentage of the function's gradient variance explained by the active variables. More details can be found in [29, 25]. In this work, we select d_l such that $\sum_{j=d_l+1}^{d_{\boldsymbol{\Theta}}} \hat{\lambda}_j < \kappa$, where κ is a pre-selected threshold.

Finally, to account for the uncertainty in estimating the eigenvectors, we use a bootstrapping procedure to resample the M gradient evaluations, build N bootstrapped replicates of the gradient data matrix $\hat{\mathbf{B}}_{\mathcal{M}}$ as $\hat{\mathbf{B}}_{\mathcal{M}}^{(i)}$, and use SVD to generate replicates of the eigenvector matrix as $\hat{\mathbf{W}}^{(i)}$. However, we do not generate replicates of d_l to maintain stability in the subsequent procedure. Instead, we use the value of d_l obtained from the nominal eigenvalue matrix $\hat{\boldsymbol{\Lambda}}$ when using the replicates $\hat{\mathbf{W}}_{d_l}^{(i)}$ to project samples into the latent space or reconstruct latent samples back to the original parameter space. A more thorough discussion of this procedure is in [29].

3.2 Evaluating the generalized likelihood

Next, we build a surrogate for the misfit function in the latent space. Chakroborty and Mahadevan [29] showed that this surrogate, which we denote as $\mathcal{G}_{\mathcal{M}}(\boldsymbol{Z}_{d_l})$, predicts a Gamma random field with local distribution $\Gamma\left(\alpha(\boldsymbol{Z}_{d_l}) + \frac{d_{\boldsymbol{f}}}{2}, c(\boldsymbol{Z}_{d_l})\right)$, where the parameter functions $\alpha(\boldsymbol{Z}_{d_l})$ and $c(\boldsymbol{Z}_{d_l})$ can be written in terms of the mean and variance functions of $\mathcal{G}_{\mathcal{M}}(\boldsymbol{Z}_{d_l})$. Let the mean of the Gamma distribution underlying $\mathcal{G}_{\mathcal{M}}(\boldsymbol{Z}_{d_l})$ at a given point $\boldsymbol{Z}_{d_l}$ be denoted $\mu_{\mathcal{M}}(\boldsymbol{Z}_{d_l}) = \mathbb{E}[\mathcal{G}_{\mathcal{M}}(\boldsymbol{Z}_{d_l}) | \boldsymbol{Z}_{d_l}]$, and the variance at $\boldsymbol{Z}_{d_l}$ be denoted $\sigma^2_{\mathcal{M}}(\boldsymbol{Z}_{d_l}) = \mathbb{V}\text{ar}[\mathcal{G}_{\mathcal{M}}(\boldsymbol{Z}_{d_l}) | \boldsymbol{Z}_{d_l}]$. Then,

$$c(\boldsymbol{Z}_{d_l}) = \frac{\sigma^2_{\mathcal{M}}(\boldsymbol{Z}_{d_l})}{\mu_{\mathcal{M}}(\boldsymbol{Z}_{d_l})} \tag{24}$$

$$\alpha(\boldsymbol{Z}_{d_l}) = \frac{(\mu_{\mathcal{M}}(\boldsymbol{Z}_{d_l}))^2}{\sigma^2_{\mathcal{M}}(\boldsymbol{Z}_{d_l})} - \frac{d_{\boldsymbol{f}}}{2} \tag{25}$$

The mean function $\mu_{\mathcal{M}}(\boldsymbol{Z}_{d_l})$ is approximated by fitting a Gaussian process (GP) surrogate to the misfit function in the latent space. D samples of the parameter vector are drawn from the prior $\pi(\boldsymbol{\Theta})$ as $\boldsymbol{\Theta}_k$ $(k = 1, \dots, D)$, and both the misfit value $\mathcal{M}(\boldsymbol{\Theta}_k)$ and the low-dimensional representation $\boldsymbol{Z}_{d_l,k}$ is computed for each $\boldsymbol{\Theta}_k$ (the latter using Eq. (21)). Then, the GP is trained with the input-output pairs $(\boldsymbol{Z}_{d_l,k}, \mathcal{M}(\boldsymbol{\Theta}_k))$, $k = 1, \dots, D$. Since multiple points in the original space with slightly varying misfit values can map to the same point in the latent space, the misfit value at each latent-space location has some uncertainty. The GP captures this by adding a white noise kernel to the covariance matrix formed using a Matérn 5/2 kernel scaled by a magnification constant; the added white noise component improves the mean fit and yields more robust variance prediction when the training data is noisy. Thus, the distribution predicted by the GP at each point is $\mathcal{N}(\mu_1(\boldsymbol{Z}_{d_l}), \sigma_1(\boldsymbol{Z}_{d_l}))$, and

$$\mu_{\mathcal{M}}(\boldsymbol{Z}_{d_l}) = \mu_1(\boldsymbol{Z}_{d_l}) \tag{26}$$

However, the above GP does not account for the uncertainty in estimating the active subspace itself, which increases the variance in the misfit values at any low-dimensional sample point $\boldsymbol{Z}_{d_l}$ beyond $(\sigma_1(\boldsymbol{Z}_{d_l}))^2$. This additional variance is captured using a second GP which also uses a Matérn 5/2 kernel scaled by a magnification constant and an additive white noise kernel. At each of the latent space points $\boldsymbol{Z}_{d_l,k}$, the additional variance stemming from the uncertainty in estimating the active subspace can be estimated using the bootstrapped replicates of the latent space transformation matrix $\hat{\mathbf{W}}_{d_l}^{(i)}$ and a simple first-order Taylor expansion, as follows.

$$V_k = \frac{1}{N-1}\sum_{i=1}^{N}\left[\mathsf{M}_k^{(i)} - \left(\frac{1}{N}\sum_{i=1}^{N}\mathsf{M}_k^{(i)}\right)\right]^2 \tag{27}$$

$$\text{where} \quad \mathsf{M}_k^{(i)} = \mathcal{M}(\boldsymbol{\Theta}_k) + \left[\nabla\mathcal{M}(\boldsymbol{\Theta}_k)\cdot\left(\hat{\mathbf{W}}_{d_l}^T\boldsymbol{\Theta}_k - \left(\hat{\mathbf{W}}_{d_l}^{(i)}\right)^T\boldsymbol{\Theta}_k\right)\right] \tag{28}$$

The second GP is trained directly on these variance estimates, i.e., on the input-output pairs $(\boldsymbol{Z}_{d_l,k}, V_k)$, $k = 1,\ldots,D$, and predicts the distribution $\mathcal{N}(\mu_2(\boldsymbol{Z}_{d_l}), \sigma_2(\boldsymbol{Z}_{d_l}))$ at each $\boldsymbol{Z}_{d_l}$ for the additional variance caused by the active subspace estimation uncertainty. Therefore,

$$\sigma_{\mathcal{M}}^2(\boldsymbol{Z}_{d_l}) = \mu_2(\boldsymbol{Z}_{d_l}) + (\sigma_1(\boldsymbol{Z}_{d_l}))^2 \tag{29}$$

From Eqs. (26) and 29, coupled with Eqs. (24) and (25), we can fully construct the Gamma random field $\Gamma\left(\alpha(\boldsymbol{Z}_{d_l}) + \frac{d_{\boldsymbol{f}}}{2}, c(\boldsymbol{Z}_{d_l})\right)$ that defines the misfit surrogate $\mathcal{G}_{\mathcal{M}}(\boldsymbol{Z}_{d_l})$. A more comprehensive discussion of the intuition underpinning the above constructions and their derivations is available in [29].

The misfit predictions made by the surrogate cannot be directly used in Eq. (15), since neither the prediction uncertainty of $\mathcal{G}_{\mathcal{M}}(\boldsymbol{Z}_{d_l})$ nor any fitting error in the surrogate is accounted for in Eq. (15). We capture the latter using an additive surrogate error term $G_\zeta \sim \Gamma(\zeta, c(\boldsymbol{Z}_{d_l}))$, such that

$$\mathcal{M}(\boldsymbol{\Theta}) \approx \mathcal{G}_{\mathcal{M}}(\boldsymbol{Z}_{d_l}) + G_\zeta = \mathcal{G}_{\mathcal{M},\zeta}(\boldsymbol{Z}_{d_l},\zeta) \sim \Gamma\left(\zeta + \alpha(\boldsymbol{Z}_{d_l}) + \frac{d_{\boldsymbol{f}}}{2}, c(\boldsymbol{Z}_{d_l})\right) \tag{30}$$

Then, the generalized likelihood formulation from [29] is used with $\mathcal{G}_{\mathcal{M},\zeta}(\boldsymbol{Z}_{d_l},\zeta)$ to obtain

$$\mathcal{L}(\boldsymbol{\Theta},\zeta) \approx \mathcal{L}(\boldsymbol{Z}_{d_l},\zeta) = \mathbb{E}_{\mathcal{G}_{\mathcal{M},\zeta}}\left[\exp\left(-\mathcal{G}_{\mathcal{M},\zeta}(\boldsymbol{Z}_{d_l},\zeta)\right)\right] \tag{31}$$

Here, ζ is a final unknown constant inferred alongside the calibration parameters during the Bayesian calibration procedure.

3.3 Calibration outcomes and the posterior

The benefit of the joint Gaussian prior $\pi(\boldsymbol{\Theta})$ becomes clear when one considers that the standard multivariate Gaussian is rotationally invariant. Thus, we can write

$$\pi(\boldsymbol{\Theta}) = \pi(\boldsymbol{Z}_{d_l})\cdot\pi(\boldsymbol{Z}_\perp|\boldsymbol{Z}_{d_l}) \tag{32}$$

where $\pi(\boldsymbol{Z}_{d_l})$ follows a d_l dimensional standard Gaussian density function, and $\pi(\boldsymbol{Z}_\perp|\boldsymbol{Z}_{d_l}) \equiv \pi(\boldsymbol{Z}_\perp)$ follows a $d_{\boldsymbol{\Theta}} - d_l$ dimensional standard Gaussian density function. This decomposition can be used to formulate the Bayesian calibration problem solely in the latent space as

$$p(\boldsymbol{Z}_{d_l}) \propto \mathcal{L}(\boldsymbol{Z}_{d_l},\zeta)\,\pi(\boldsymbol{Z}_{d_l}) \tag{33}$$

and since $\mathcal{L}(\boldsymbol{\Theta},\zeta) \approx \mathcal{L}(\boldsymbol{Z}_{d_l},\zeta)$, the full posterior, accounting for the uncertainty in $\hat{\mathbf{W}}$, is given as

$$p(\boldsymbol{\Theta}) = \frac{1}{N}\sum_{i=1}^{N}\mathcal{L}\left(\left(\hat{\mathbf{W}}_{d_l}^{(i)}\right)^T\boldsymbol{\Theta},\zeta\right)\phi_{d_l}\left(\left(\hat{\mathbf{W}}_{d_l}^{(i)}\right)^T\boldsymbol{\Theta}\right)\phi_{(d_{\boldsymbol{\Theta}}-d_l)}\left(\left(\hat{\mathbf{W}}_{\perp}^{(i)}\right)^T\boldsymbol{\Theta}\right) \tag{34}$$

where $\phi_d(\cdot)$ denotes the d-dimensional standard normal density function. Note here that the likelihood only involves the d_l active variables, implying that only they have their density updated from the prior to the posterior. The remaining $d_{\boldsymbol{\Theta}} - d_l$ variables are not updated, and their prior density appears unchanged in the expression of the posterior density $p(\boldsymbol{\Theta})$.

In practice, we generate samples from the latent space posterior $p\left(\boldsymbol{Z}_{d_l}\right)$, which are difficult to convert into samples from the original space posterior $p\left(\boldsymbol{\Theta}\right)$. Instead, we can draw samples from densities corresponding to each term in the sum in Eq. (34), which we denote $p^{(i)}\left(\boldsymbol{\Theta}\right)$

$$p^{(i)}\left(\boldsymbol{\Theta}\right) = \mathcal{L}\left(\left(\hat{\mathbf{W}}_{d_l}^{(i)}\right)^T \boldsymbol{\Theta}, \delta\right) \phi_{d_l}\left(\left(\hat{\mathbf{W}}_{d_l}^{(i)}\right)^T \boldsymbol{\Theta}\right) \phi_{(d_{\boldsymbol{\Theta}}-d_l)}\left(\left(\hat{\mathbf{W}}_{\perp}^{(i)}\right)^T \boldsymbol{\Theta}\right) \tag{35}$$

Since each $p^{(i)}\left(\boldsymbol{\Theta}\right)$ only varies based on which bootstrapped replicate $\hat{\mathbf{W}}^{(i)}$ of the active subspace eigenvector matrix is used, we term $p^{(i)}\left(\boldsymbol{\Theta}\right)$ as the *replicate posterior densities*.

Samples from each replicate posterior density can be generated as follows:

1. Use MCMC to draw $\mathfrak{n}$ samples $\boldsymbol{Z}_{d_l,k} \sim p\left(\boldsymbol{Z}_{d_l}\right)$, $k = 1, \ldots, \mathfrak{n}$.
2. Use basic Monte Carlo sampling to draw a corresponding $\mathfrak{n}$ samples $\boldsymbol{Z}_{\perp,k} \sim \pi\left(\boldsymbol{Z}_{\perp}\right)$, $k = 1, \ldots, \mathfrak{n}$.
3. Put these samples together as $\boldsymbol{Z}_k = \begin{bmatrix}\boldsymbol{Z}_{d_l,k} & \boldsymbol{Z}_{\perp,k}\end{bmatrix}$, $k = 1, \ldots, \mathfrak{n}$.
4. Then, $\boldsymbol{\Theta}_k^{(i)} = \hat{\mathbf{W}}^{(i)}\boldsymbol{Z}_k$, $\forall k = 1, \ldots, \mathfrak{n}$, are $\mathfrak{n}$ samples from $p^{(i)}\left(\boldsymbol{\Theta}\right)$.

The overall posterior ($p\left(\boldsymbol{\Theta}\right)$ from Eq. (34)) can be constructed by averaging over the densities for each replicate posterior density obtained from the generated samples, although this is computationally difficult. However, we can evaluate any statistic of the true posterior straightforwardly. For example, the mean value of an arbitrary function $h\left(\boldsymbol{\Theta}\right)$, where $\boldsymbol{\Theta}$ follows $p\left(\boldsymbol{\Theta}\right)$, can be estimated as

$$\hat{\mathbb{E}}_p\left[h\left(\boldsymbol{\Theta}\right)\right] = \sum_{i=1}^{N} \frac{1}{N}\left[\sum_{k=1}^{\mathfrak{n}} \frac{h\left(\boldsymbol{\Theta}_k^{(i)}\right)}{\mathfrak{n}}\right] \quad \text{, where } \boldsymbol{\Theta}_k^{(i)} \in \mathfrak{C}_{\boldsymbol{\Theta}}^{(i)} \tag{36}$$

We can see from Eq. (34) that the posterior $p\left(\boldsymbol{\Theta}\right)$ retains the complete prior uncertainty of the $d_{\boldsymbol{\Theta}} - d_l$ inactive variables. By virtue of being inactive, these variables do not affect the likelihood function and therefore are not updated by the experimental observation. Consequently, calibration after larger dimension reduction retains more prior uncertainty. This is of course not desirable from an engineering perspective; we want maximally confident inference of the model parameters, which should be supported, not hindered, by knowledge of inactive directions in the parameter space. For this reason, Chakroborty and Mahadeven [29] introduce the *conditional active posterior* (CAP), which balances inference uncertainty and parameter informativeness by fixing the inactive components at their mean value. Using the CAP lets us quantify only that uncertainty in the calibrated posteriors that contributes to variability in the CPFE model output.

Again, sampling directly from the CAP is difficult for the same reasons as for the true posterior. Instead, we can generate samples from each replicate CAP density using the same steps as above, and these samples can be used to estimate the empirical CAP density or its statistics in an analogous way to the true posterior. The steps to sample from the CAP are as follows:

1. Use MCMC to draw $\mathfrak{n}$ samples $\boldsymbol{Z}_{d_l,k} \sim p\left(\boldsymbol{Z}_{d_l}\right)$, $k = 1, \ldots, \mathfrak{n}$.
2. Fix the inactive variables at their prior mean, i.e., 0.
3. Put these samples together as $\boldsymbol{Z}_k = \begin{bmatrix}\boldsymbol{Z}_{d_l,k} & \mathbf{0}\end{bmatrix}$, $k = 1, \ldots, \mathfrak{n}$.
4. Then, $\boldsymbol{\Theta}_k^{(i)} = \hat{\mathbf{W}}^{(i)}\boldsymbol{Z}_k$, $\forall k = 1, \ldots, \mathfrak{n}$, are $\mathfrak{n}$ samples from the replicate CAP densities.

Single-point estimates of the parameters can be extracted from the calibrated posteriors to compare with classical, deterministic calibration procedures. The posterior mean, which can be estimated using Eq. (36), has been shown to minimize the mean squared error between a parameter estimate and its underlying true value [40], and is thus called the minimum mean squared error (MMSE) Bayes' point estimate. Other point estimates may be obtained as important statistics of the posterior, like the median or specific quantiles that correspond to minimizing alternate error measures between the estimated and the true parameter value, or be chosen based on locating optima of important functions, such as the maximum likelihood estimate or the maximum a posteriori estimate (i.e., the mode of the posterior). However, due to the multi-dimensional nature of the joint posterior and the inability to directly generate samples from it, these alternate points are difficult to estimate numerically. Similarly, confidence/probability intervals are difficult to compute because the posterior is high-dimensional, and samples are available only from replicate posterior densities, not the true posterior density. Estimating quantiles and probability intervals from samples of the replicate posterior densities remains an important future direction of methodological development.

4 Numerical results

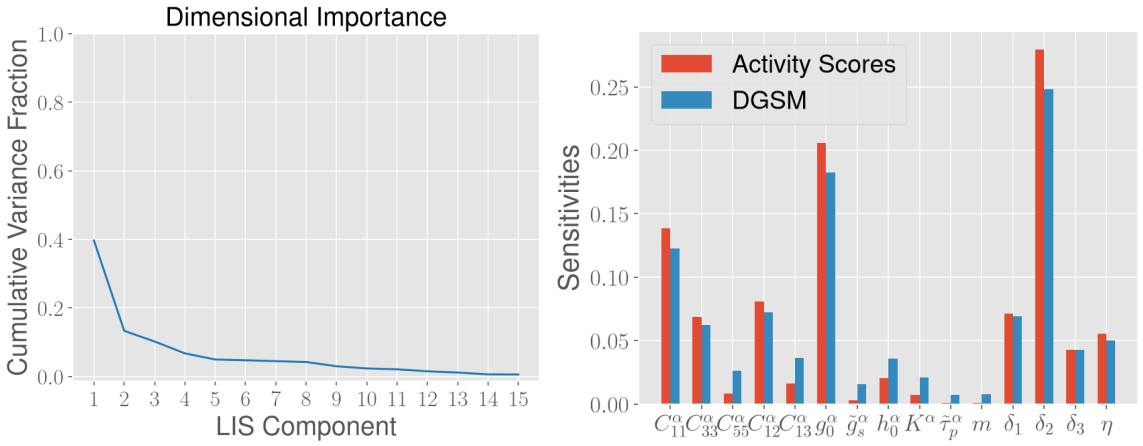
In this section we provide a detailed discussion of the results from applying the CMF to the Ti64 CPFE model. Two different cases are analyzed. The difference between the two cases is the misfit gradient variance fraction threshold κ associated with the dimension reduction procedure, with $\kappa = 80\%$ and $= 90\%$ for cases 1 and 2, respectively. For both cases, $\boldsymbol{\delta} = \tilde{\boldsymbol{\delta}}$ and $\eta = 0.01 \exp(\tilde{\eta})$ (see Section 2.5 for notation), such that the discrepancy $\boldsymbol{\delta}$ follows a 3-dimensional standard normal prior and the noise magnitude η follows a lognormal prior.

We begin our discussion in Section 4.1 with a look at the sensitivities of the individual model parameters. The active directions for cases 1 and 2 are also visualized here. Section 4.2 then compares the CMF in the latent space against direct calibration results in the original space, using simple Metropolis-Hastings with a standard multivariate normal proposal density to generate samples in both cases. The comparison clearly shows the necessity of dimension reduction when not using sophisticated, difficult-to-tune samplers. Then, the conditional active posteriors are compared against the true posteriors to justify the former's benefits over the latter. Finally, the calibration results for all cases are collected in Sections 4.3 and 4.4. The findings specific to Ti64 are summarized in Section 4.5.

4.1 Sensitivities and the likelihood-informed subspace (LIS)

Case	d_l	R^2	MAPE
Case 1 ($\kappa = 80\%$)	6	0.88	75.70%
Case 2 ($\kappa = 90\%$)	8	0.87	118.94%

Table 3: Goodness-of-fit results for mean misfit surrogate prediction. Values for coefficient of determination (R^2) and mean absolute percentage error (MAPE) are reported. Additionally, the dimension of the latent space discovered, i.e., d_l, is also listed.



(a) (b)

Figure 2: Plots characterizing the sensitivity behavior of the parameter space. Figure (a) shows the relative importance of each eigenvector of the parameter vector $\boldsymbol{\Theta}$ as quantified by the eigenvalues (which capture the misfit function gradient variance fraction along the corresponding direction). Figure (b) shows the sensitivities of each model parameter, with the activity scores in red and DGSM in blue, each normalized by their respective total sums.

The first step of the CMF is to construct the likelihood-informed subspace and the low-dimensional misfit surrogate. Table 3 lists the latent space dimension d_l for each case, along with goodness-of-fit results for the misfit surrogate. Importantly, the results only compare the mean of the Gamma random field $\mathcal{G}_{\mathcal{M}}(\boldsymbol{Z}_{d_l})$ against the misfit as defined in Eq. (16). The variance in the predictions is not accounted for, leading to a positive bias in the mean absolute percentage error (MAPE) values. However, even without correcting for the stochastic nature of the prediction, the R^2 values for the misfit surrogate are sufficient in both cases. We further observe that increasing the active subspace variance fraction threshold κ increases the error. Although increasing κ captures more of the misfit function's variation, the higher latent space dimensionality makes surrogate training harder, reducing overall surrogate accuracy. This suggests that it is important to balance active subspace accuracy against the misfit approximation accuracy.

One benefit of the CMF is that the framework directly provides sensitivities for the model parameters. Since the latent space is constructed from the misfit function, the parameter sensitivities are computed with respect to the misfit function as well, which combines the effects of each output feature into a single scalar measure. The effects of the parameters on each individual output feature must be computed separately. A parameter's importance in the LIS is measured by its *activity score* [41], defined as

$$\rho_i = \sum_{j=1}^{d_l} \hat{\lambda}_j \hat{w}_{i,j}^2 \ , \quad i = 1, \ldots, d_{\boldsymbol{\Theta}} \tag{37}$$

where $\hat{w}_{i,j}$ is the i-th component of $\hat{\mathbf{w}}_j$, i.e., the weight associated with θ_i along the direction $\hat{\mathbf{w}}_j$. The derivative-based global sensitivity measure (DGSM) of the parameters are also available for free, as

$$\psi_i = \sum_{j=1}^{d_{\boldsymbol{\Theta}}} \hat{\lambda}_j \hat{w}_{i,j}^2 \ , \quad i = 1, \ldots, d_{\boldsymbol{\Theta}} \tag{38}$$

The activity scores and DGSM for both cases are shown in Figure 2b, with both measures normalized by their respective total sums (i.e., $\rho_i / \sum_{i=1}^{d_{\boldsymbol{\Theta}}} \rho_i$ and $\psi_i / \sum_{i=1}^{d_{\boldsymbol{\Theta}}} \psi_i$ are plotted). Cases 1 & 2 have the same sensitivities, as they have the same parameter vector. The metrics are transformed to directly represent the sensitivities of the model parameters themselves, instead of the standardized parameter vector $\boldsymbol{\theta}$ components. Among the model parameters, g_0^α has the highest sensitivity, followed by C_{11}^α. C_{33}^α and C_{12}^α contribute moderately, while the remaining parameters are largely insensitive. The additional uncertainty modeling terms, i.e., the discrepancy terms and noise magnitude, all have low-to-moderate sensitivities except δ_2, which has the highest sensitivity across all the calibrated parameters.

4.2 Benefits of the CMF and the conditional active posterior (CAP)

Before discussing the conditional active posteriors, which we present as our primary calibration outcomes, we first showcase the benefits of the CMF (in particular, the dimension reduction) and the CAP density. To do so, we run the CMF, construct both the full posteriors and the CAP density, and also conduct calibration in the original parameter space without dimension reduction. The basic Metropolis-Hastings (MH) algorithm is used in both cases, with a standard normal proposal density, to highlight the robustness of the CMF without relying on sophisticated samplers. Figure 3 plots histograms of the features of interest E, σ_y, and H for case 1. As shown in Figures 3a-3c, the output feature distributions for the posteriors calibrated in the original space are of very poor quality and do not converge near the experimentally observed values. This is because the parameter posterior is expected to have a complex shape (as the CMF results reveal in Section 4.3), and simple samplers like MH struggle to remain robust even in 15 dimensions. By contrast, the output feature distributions for both the CMF posteriors (the full posterior density and the CAP density) are much higher quality and centered around the experimental values. This significant improvement in sample quality, despite using the same MH sampler, is attributable almost entirely to sampling in a much lower-dimensional space ($6+1$ dimensions for the LIS with surrogate model error as opposed to 15 for the original space). This clearly underscores the importance of dimension reduction when calibrating the Ti64 CPFE model, as well as other models with a large number of parameters.

Also, comparing the full misfit posterior density (Figure 3j) with the CAP misfit density (Figure 3k), we see that the CAP samples and the full posterior samples capture the output behavior equally well. This is further supported by the histograms of the full output feature posterior densities (Figures 3d-3f) and the CAP output feature densities (Figures 3g-3i), which show no perceptible difference in the histogram of σ_y between the two cases, and only a minor difference for E and H. On the other hand, we see clear differences between the full posterior density and the CAP density for the model parameters themselves. In Figure 4, we examine the calibration of a highly sensitive parameter (g_0^α), a moderately sensitive parameter C_{33}^α, and an insensitive parameter (m). Figures 4d-4f show the marginal CAP densities of these three parameters, while Figures 4a-4c show their marginal full posterior densities. For the highly sensitive g_0^α, both histograms look the same. However, as parameter sensitivity decreases, first with C_{33}^α, then further with m (meaning they are less informative about the CPFE model output), the full posteriors shift less from the prior, while the CAP densities become much sharper. This occurs because the CAP density removes the contribution of the inactive directions, whereas the prior uncertainty from the inactive directions remains unchanged in the full posteriors. Figure 3 shows that the output behavior does not change between the full posterior and the CAP, and Figure 4 completes the picture by showing that using the CAP leads to much more confident parameter calibration. Thus, in the next section, we only focus on the CAP densities.

4.3 Calibration outcomes

In this section, we present the calibration results in terms of the marginal CAP densities for each model parameter, including the model discrepancy and noise magnitude terms, as well as the marginal posterior densities of the output

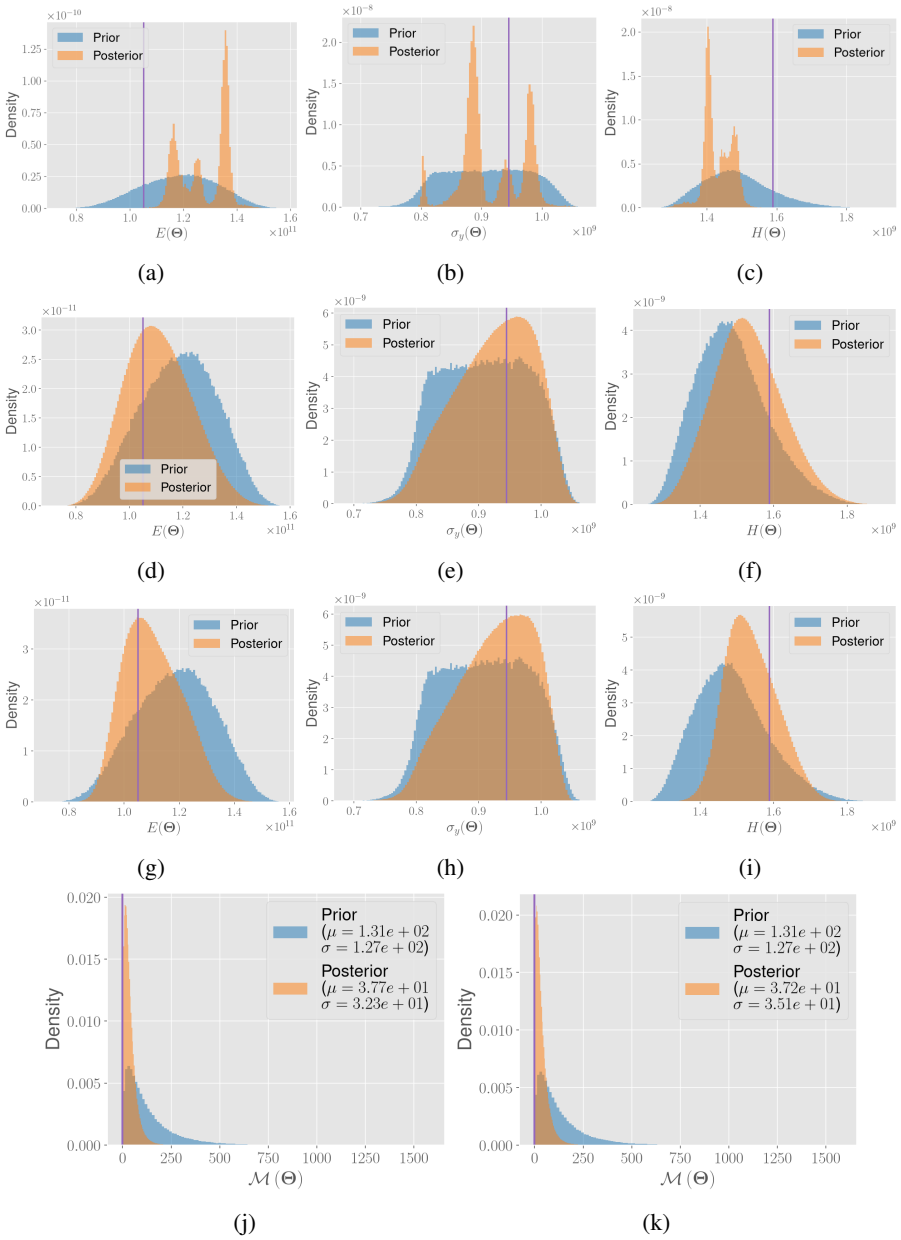


Figure 3: Plots comparing the posteriors generated from direct calibration in the original space and the CMF for case 1. Figures (a)-(c) show histograms for the densities induced by the posteriors from the direct calibration for E, σ_y, and H, respectively. Figures (d)-(f) show the corresponding histograms induced by the full posterior from the CMF. Figures (g)-(i) plot the output feature histograms resulting from the CAP densities. Figures (j) and (k) show the histograms of the misfit function induced by the full posterior and the CAP density, respectively. In all figures, the prior density is shown in blue, the posterior in orange, and the purple line represents the experimental observation.

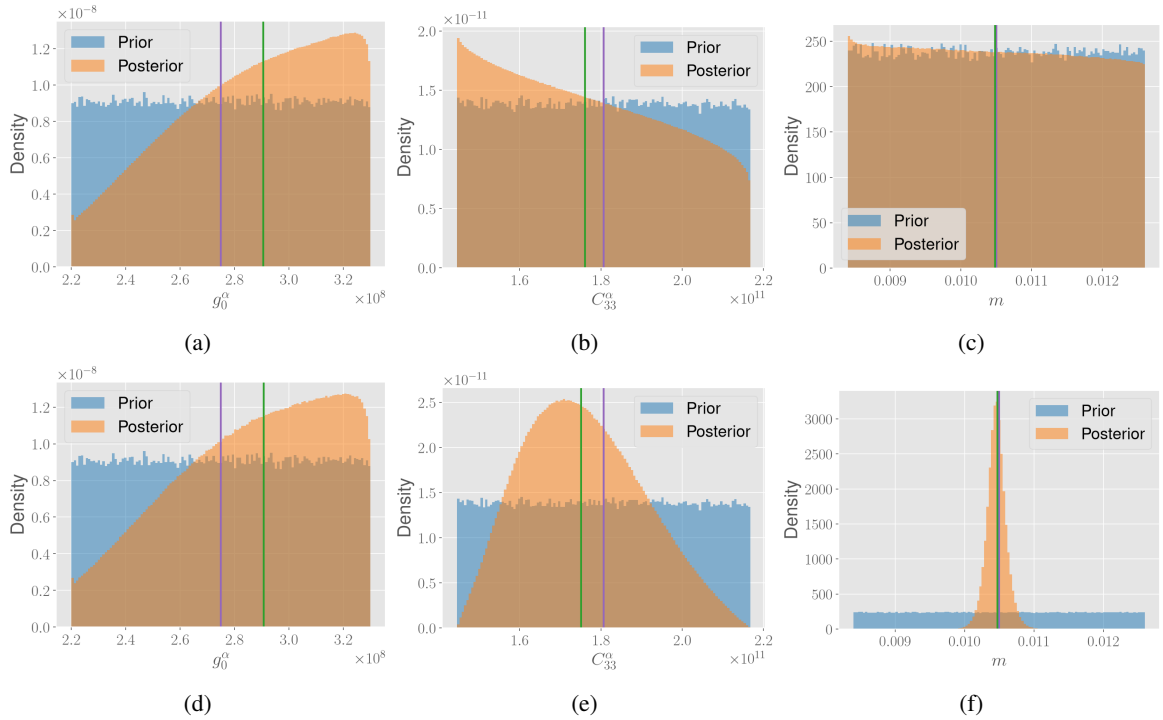


Figure 4: Plots comparing the marginal full posterior densities and marginal CAP densities of g_0^α, C_{33}^α, and m, generated from the CMF for case 1. Figures (a)-(c) show histograms for the marginal densities for g_0^α, C_{33}^α, and m from the full posterior, respectively. Figures (d)-(f) show the corresponding histograms for the CAP densities from the CMF. In all figures, the prior density is shown in blue, and the posterior in orange. The green and purple lines, indicating the posterior and prior means, respectively, are discussed further in Section 4.3.

features and the misfit function induced by the CAP density. Figure 5 collects plots of the marginals of C_{11}^α to m, respectively, following the order in the definition of the calibration parameter vector $\mathbf{v}$ (Eq. (6)), as well as the discrepancy $\boldsymbol{\delta}$ and noise magnitude η, for case 1. Figure 6 does the same for case 2. Finally, the marginals for all three output features (E, σ_y, H), along with the misfit function, are collected in Figure 7.

The key observation is that, as the variance fraction threshold κ for LIS construction increases (case 1 vs. case 2), the calibration gets less effective. The mean and standard deviation of the misfit CAP density in Figures 7g and 7h quantify this deterioration; the average misfit in case 2 is nearly twice that in case 1, and the standard deviation increases by a factor of ≈ 2 as well. The posteriors for all model parameters and output features also either widen or remain closer to the prior shape. However, the degree varies, with the effect being most pronounced for moderately sensitive parameters such as C_{33}^α (Figure 5b & 6b) and even some of the less sensitive parameters like C_{13}^α (Figure 5e & 6e) and K^α Figure 5i & 6i). Although the latent space representation improves for higher κ (Table 3), both the misfit approximation and the posterior sampling become less effective as the LIS dimensionality increases (Table 3). This is because sampling becomes harder in higher dimensions, leading to poorer exploration of the domain and, consequently, wider posteriors. Thus, it is desirable to reduce dimensionality as much as possible in a manner that balances the accuracy of the misfit approximation against sampling efficiency.

4.4 Point and interval estimates

As described in Section 3.3, the posterior mean has been shown to be an optimal single-value estimate of the model parameters, since it minimizes the mean squared error of the parameter estimate with respect to the underlying *true values* that describe the experimental data [40]. This is in contrast with deterministic calibration approaches, whose single-value estimates are not tied to the underlying true values but simply aim to minimize some error measure between the model predictions and the experimental observations. Comparing the posterior mean (i.e., the MMSE Bayes' point estimate, shown in green in Figures 5 & 6) to the genetic algorithm-based deterministically calibrated values from Shen et al. [3] (shown in purple in Figures 5 & 6), we see that several parameters, particularly the more sensitive ones, show significant differences between the two values. This highlights the importance of comprehensively accounting for

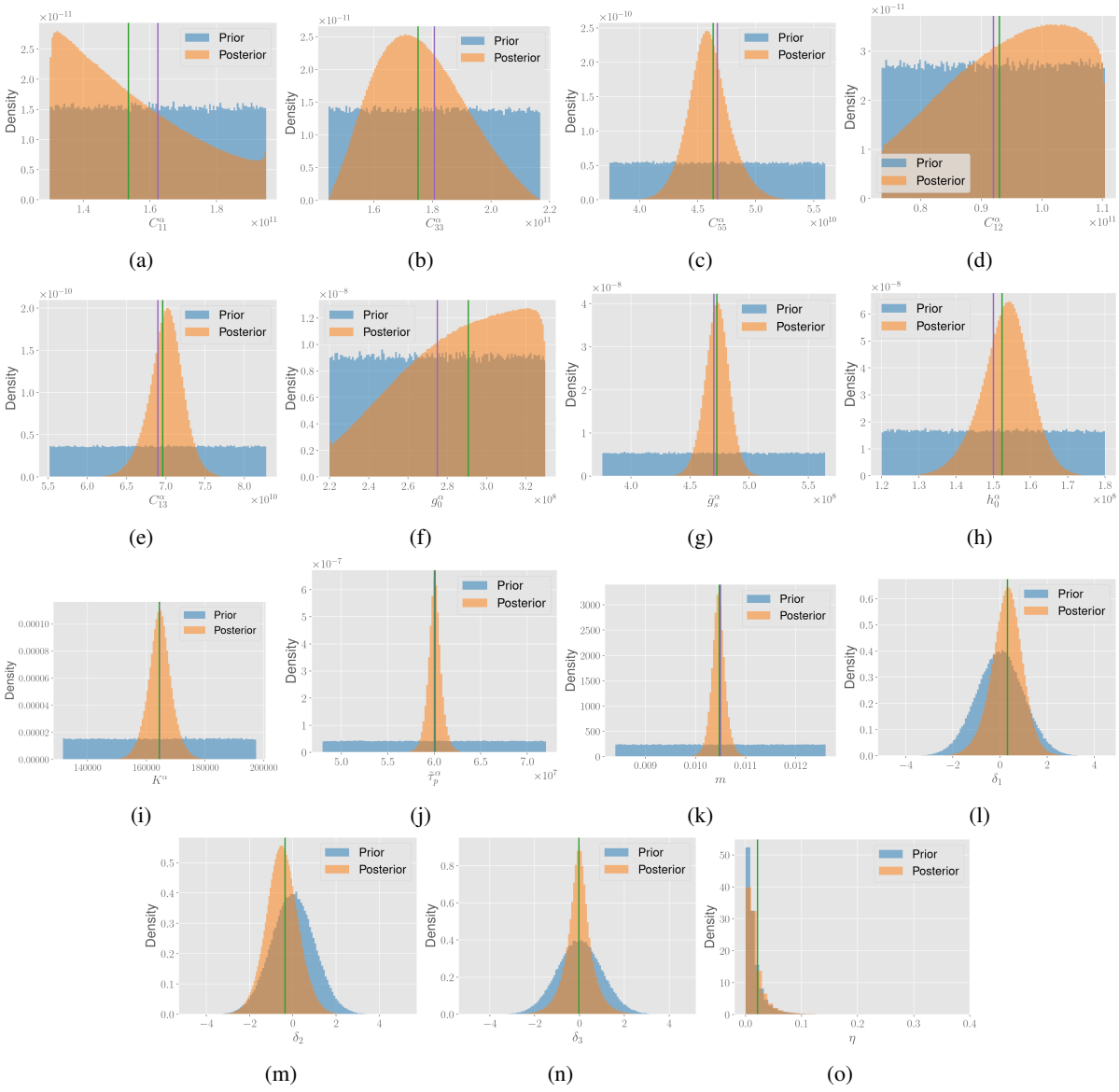


Figure 5: Marginal conditional active posterior (CAP) densities for case 1. Figures (a)-(k) show histograms for C_{11}^{α} to m, following the order in the definition of the calibration parameter vector $\mathbf{v}$ (Eq. (6)). Figures (l)-(n) show the histograms for δ_1 - δ_3, respectively, and figure (o) shows the histogram for η. In all figures, the prior density is shown in blue, and the posterior in orange. The green line indicates the posterior mean, while the purple line marks the deterministically calibrated value from Shen et al. [3].

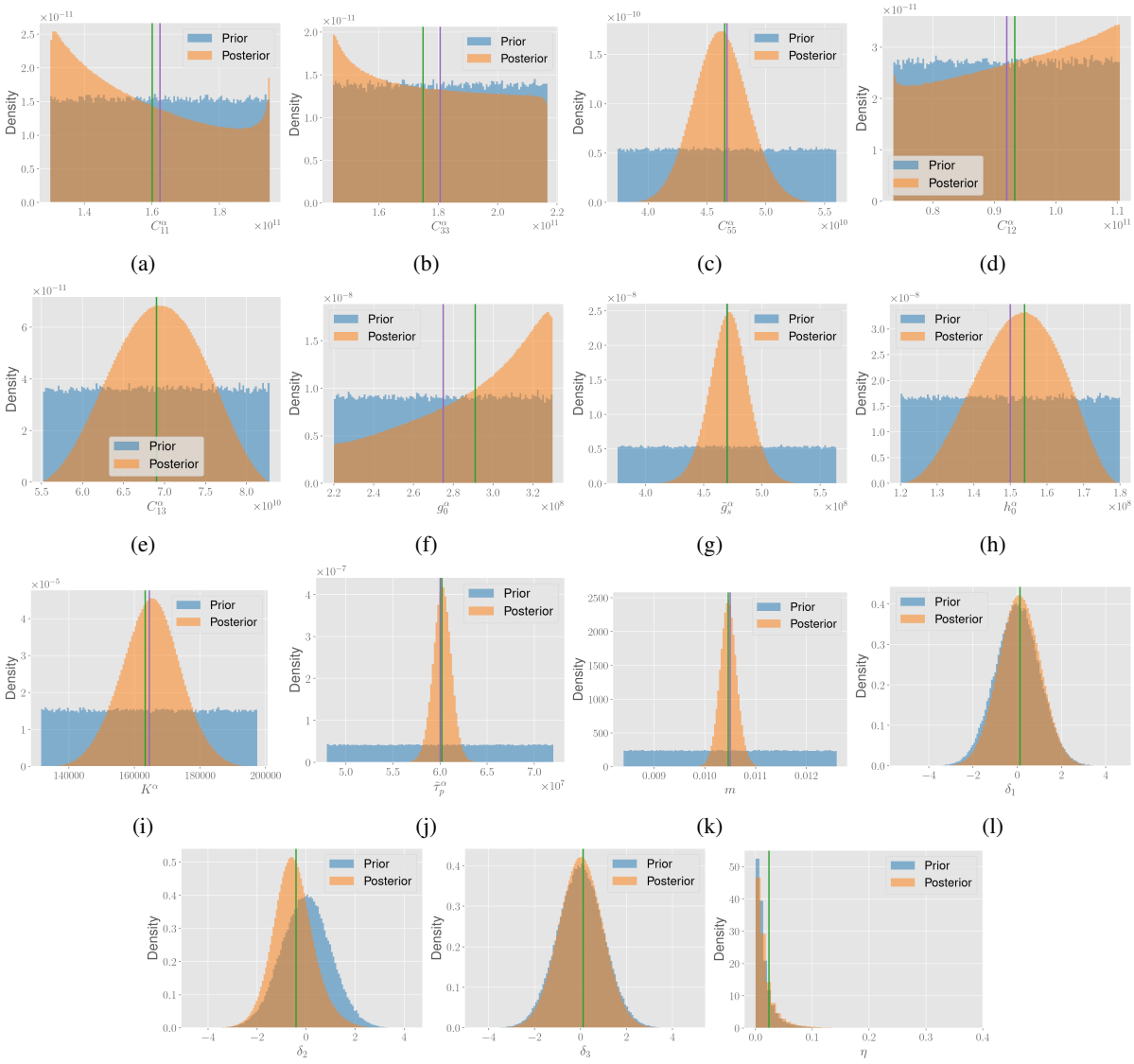


Figure 6: Marginal conditional active posterior (CAP) densities for case 2. Figures (a)-(k) show histograms for C_{11}^{α} to m, following the order in the definition of the calibration parameter vector $\mathbf{v}$ (Eq. (6)). Figures (l)-(n) show the histograms for δ_1 - δ_3, respectively, and figure (o) shows the histogram for η. In all figures, the prior density is shown in blue, and the posterior in orange. The green line indicates the posterior mean, while the purple line marks the deterministically calibrated value from Shen et al. [3].

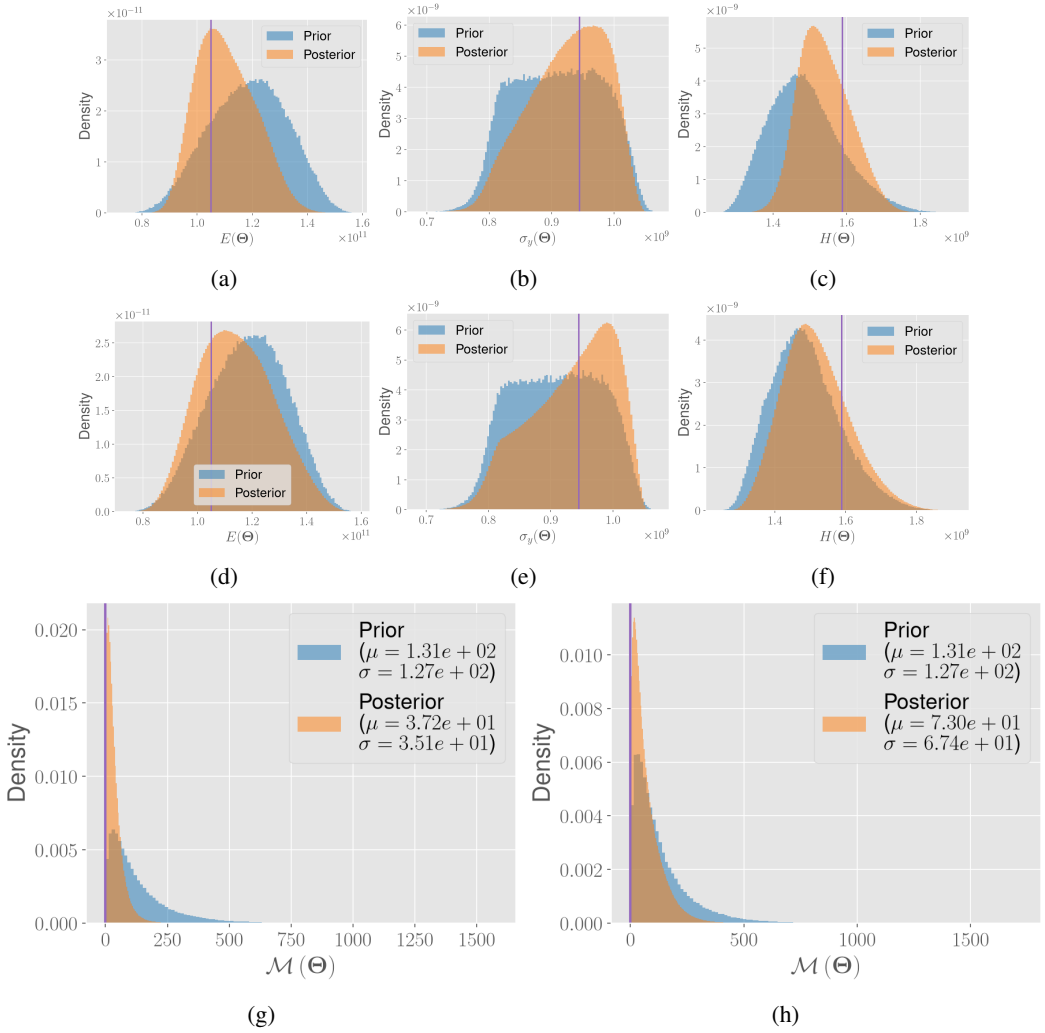


Figure 7: Output feature marginal densities induced by the CAP density. Figures (a)-(c) show histograms for E, σ_y, and H, respectively, for case 1. Similarly, figures (d)-(f) show histograms for E, σ_y, and H, respectively, for case 2. Finally, figures (g) and (h) show histograms for the true misfit function $\mathcal{M}(\boldsymbol{\Theta})$ (Eq. (16)) for cases 1 and 2, respectively. In all figures, the prior density is shown in blue, the posterior in orange, and the purple line represents the experimental observation. For figures (g) and (h), we also report the mean (μ) and standard deviation (σ) of the misfit function under the prior and CAP densities.

the uncertainties in the calibration process, as the deterministically calibrated parameter values are unable to match the uncertainty-aware MMSE Bayes' point estimate, which is theoretically optimal (provided all uncertainties are accounted for in the posterior).

Evaluating probability intervals around the Bayes' estimate is difficult in high dimensions, and the required techniques involve complex computations. The common engineering practice of extracting individual probability intervals from the marginal posteriors of each parameter relies on a critical assumption of independence between the model parameters, which is almost never satisfied in engineering problems. Consequently, such probability intervals are difficult to justify, as the true probability intervals depend strongly on the shape of the joint posterior distribution in the high-dimensional parameter space, which is ignored by the marginals. This is further complicated by the lack of direct samples from the posterior, since we can only generate samples from the replicate CAP densities. On the other hand, the uncertainty encoded by these replicate densities is significant and cannot be ignored. Figure 8 visualizes the histograms generated from these replicate samples for case 1. The nominal CAP density is drawn in blue, with the orange lines plotting the $N = 250$ replicate sample sets generated corresponding to each replicate rotation matrix $\hat{\mathbf{W}}^{(i)}$. The large orange band underscores the contribution of this source of uncertainty; probability intervals drawn purely from a single, nominal replicate will not appropriately capture the full uncertainty. Efficiently treating this large collection of densities spanning a wide range of shapes and behaviors to estimate probability intervals is an important future step to improve the engineering treatment of complex materials models, like the current CPFE model of Ti64, using robust Bayesian frameworks such as the CMF.

4.5 Summary of findings for Ti64

Aside from the above observations about the benefits of dimension reduction and the need to carefully select the latent dimensionality such that the misfit prediction accuracy and posterior sampling quality are balanced against the completeness of the latent space representation, we can also infer several findings regarding the CPFE model of Ti64 constructed by Shen et al. [3]. Of the chosen 11 independent parameters, g_0^α and C_{11}^α affect the material behavior the most, with C_{33}^α and C_{12}^α also making moderate contributions. The rest are largely uninformative, at least from the point of view of calibration, as they do not cause the CPFE model-predicted stress-strain curve to deviate meaningfully from the experimentally recorded one. These four parameters also retain the most uncertainty and have the highest difference between the MMSE Bayes' estimate and the deterministically calibrated values from Shen et al. [3]. Moreover, the shapes of the posteriors of C_{11}^α, C_{12}^α, and g_0^α indicate that they may have significant density outside the interval chosen by Ramesh Babu et al. [30] for the uniform priors. Thus, future work may conduct calibration over a wider domain to locate regions of high likelihood further away from the deterministically calibrated values reported by Shen et al. [3].

5 Conclusion

In this manuscript, we investigated dimension reduction for effective Bayesian calibration of materials simulation models with a large number of parameters and examined the impact of various uncertainty modeling choices in the calibration framework formulation. Bayesian calibration with active subspace-based surrogate modeling of the likelihood was applied to a CPFE model for Ti-6Al-4V [3] to calibrate a subset of 11 α-phase parameters and compare our uncertainty-aware results with the deterministic calibration results in [3]. Key physically explainable features were extracted from the experimental data and the CPFE model output using a bilinear idealization for the stress-strain curve, after adding a physical law-derived correction to account for the strain rate jump in the experiment. The likelihood was constructed in accordance with the Kennedy and O'Hagan framework [12], and the associated misfit function (negative log-likelihood) was approximated using active subspace dimension reduction followed by a Gamma random field [29]. Both the latent space estimation uncertainty and the surrogate predictive uncertainty were incorporated in the calibration procedure, and the *conditional active posteriors*, which balance parameter informativeness with the prior uncertainty, were studied as the calibration outcomes.

Comparing two levels of active subspace accuracy (based on the percentage of misfit gradient variance captured), we find that a more accurate latent space predictably requires a higher-dimensional representation, but it reduces the effectiveness of MCMC sampling and thus the robustness of the calibration outcomes. This indicates that a key consideration when implementing dimension reduction for Bayesian calibration is balancing the accuracy of the low-dimensional approximation with the resulting dimensionality of the latent space.

Finally, we investigate how the inclusion of modeling uncertainties through a Bayesian lens causes a clear shift in the optimal point estimates of the model parameters (in a mean squared error sense) compared to deterministically calibrated results. This clearly justifies the necessity of incorporating uncertainties in a calibration procedure to ensure robustness even when desiring single-valued estimates. Additionally, it highlights the importance of a comprehensive

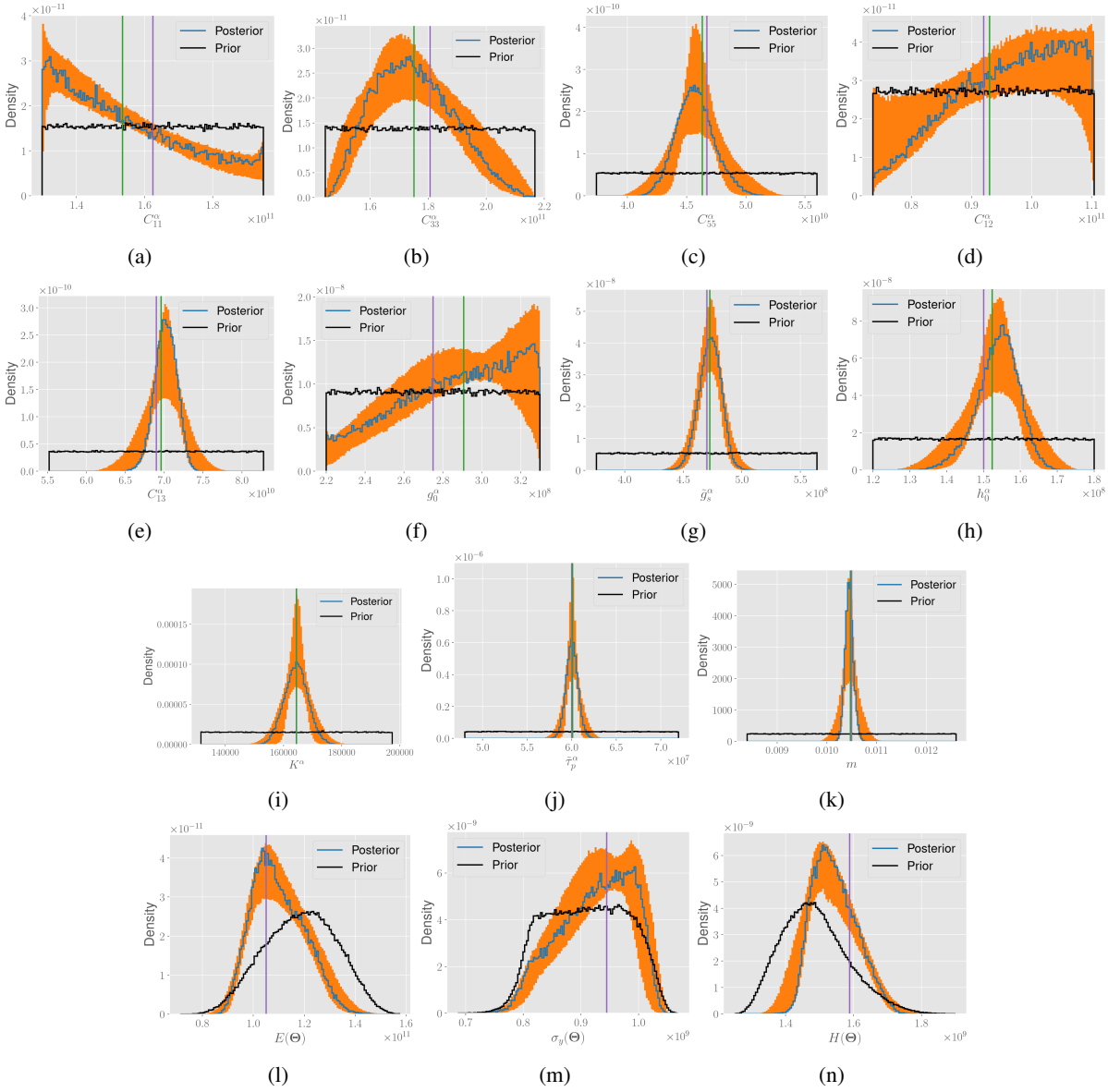


Figure 8: Plots showing the nominal and the replicate model parameter marginal CAP densities and output feature marginal CAP densities for case 1. Figures (a)-(k) show histograms for C_{11}^{α} to m, following the order in the definition of the calibration parameter vector $\mathbf{v}$ (Eq. (6)). Figures (l)-(m) show the histograms for E, σ_y, and H, respectively. In all figures, the prior density is shown in black, the nominal posterior in blue, and the orange lines represent the $N = 250$ replicate densities, providing a visual estimate of the uncertainty introduced by the LIS estimation procedure. The purple lines mark the deterministically calibrated parameter value from Shen et al. [3] in figures (a)-(k), and the experimentally observed output features in figures (l)-(m). The green line in figures (a)-(k) indicates the posterior mean.

inclusion of the uncertainties arising from every component of the modeling framework - not just the uncertainties present in the original physical system, but also those introduced by numerical approximations.

For the specific case of the Ti-6Al-4V CPFE model, our work shows that deterministically calibrated parameter values are not robust to the existence of uncertainties, and that accounting for modeling uncertainty results in significantly different optimum parameter values. In place of deterministic approaches, the Bayes' point estimates can be extracted from the posteriors of a Bayesian calibration algorithm to obtain uncertainty-aware optimal single-valued estimates of the model parameters. Further, while this work and past works [3, 30] only consider 11 parameters, the original model has ≈ 150 parameters that must all be calibrated before this CPFE model can be used in practical applications. This is too large a set to be handled by conventional Bayesian calibration procedures that rely on sampling in the original space. However, the CMF used in this work is directly extensible to the full 150-dimensional case. While the increase in dimension will require a higher computational effort in the dimension reduction (i.e., active subspace discovery) phase, the number of model evaluations scales as $d_{\mathbf{\Theta}} \log(d_{\mathbf{\Theta}})$ (see Eq. (19), Section 3.1), which is within the realm of feasibility. In contrast, most surrogate modeling techniques can be expected to fail in sufficiently approximating the CPFE model response without a significantly larger data set, and even if they were to succeed, the *curse of dimensionality* would remain a critical challenge for sampling from the posteriors. Thus, the CMF provides a key solution strategy for calibrating the full set of parameters for this CPFE model from [3], and other similarly complex models for materials of aerospace interest, such as other Titanium alloys (e.g., Ti6242) or Nickel alloy-718. The only key bottleneck to the application of the CMF is the necessity for evaluating the gradient of the model output with respect to the parameters. The current strategy of using a Gaussian process surrogate to numerically approximate the gradients prior to applying the CMF is not viable in 150 dimensions. Thus, future work in calibrating this and similar models would require a focus on obtaining accurate gradient evaluations within the computational budget. If gradients can be obtained, the CMF is enabled, and robust Bayesian calibration is achievable even for materials models with hundreds of parameters.

6 Acknowledgement

The authors acknowledge financial support from the National Aeronautics and Space Administration-Space Technology Research Institute (NASA STRI project 35, Award No. 80NSSC23K1342. Technical monitor: Timothy Poe).

Dr. J. Ramesh Babu (Vanderbilt University and University of Oklahoma) produced the simulation dataset used in this work, and Prof. Pranav Karve (Vanderbilt University) contributed valuable discussions. The simulations were conducted using the crystal plasticity code provided by Prof. Somnath Ghosh at Johns Hopkins University. All these contributions are gratefully acknowledged.

References

[1] Ikuhiro Inagaki, Tsutomu Takechi, Yoshihisa Shirai, and Nozomu Ariyasu. Application and features of titanium for the aerospace industry. *Nippon steel & sumitomo metal technical report*, 106(106):22–27, 2014.

[2] FJ Gil, MP Ginebra, JM Manero, and JA Planell. Formation of α-widmanstätten structure: effects of grain size and cooling rate on the widmanstätten morphologies and on the mechanical properties in ti6al4v alloy. *Journal of Alloys and Compounds*, 329(1-2):142–152, 2001.

[3] Jinlei Shen, Shravan Kotha, Ryan Noraas, Vasisht Venkatesh, and Somnath Ghosh. Developing parametrically upscaled constitutive and crack nucleation models for the α/β ti64 alloy. *International Journal of Plasticity*, 151:103182, 2022.

[4] Jingjing Yang, Hanchen Yu, Jie Yin, Ming Gao, Zemin Wang, and Xiaoyan Zeng. Formation and control of martensite in ti-6al-4v alloy produced by selective laser melting. *Materials & Design*, 108:308–318, 2016.

[5] M Pinz, S Storck, T Montalbano, B Croom, N Salahudin, M Trexler, and S Ghosh. Efficient computational framework for image-based micromechanical analysis of additively manufactured ti-6al-4v alloy. *Additive Manufacturing*, 60:103269, 2022.

[6] J Thomas, M Groeber, and S Ghosh. Image-based crystal plasticity fe framework for microstructure dependent properties of ti–6al–4v alloys. *Materials Science and Engineering: A*, 553:164–175, 2012.

[7] Kartik Kapoor, Ryan Noraas, Venkat Seetharaman, and Michael D Sangid. Modeling strain localization in microtextured regions in a titanium alloy: Ti–6al–4v. *Integrating Materials and Manufacturing Innovation*, 8(4):455–467, 2019.

[8] Kartik Kapoor, Priya Ravi, Ryan Noraas, Jun-Sang Park, Vasisht Venkatesh, and Michael D Sangid. Modeling ti–6al–4v using crystal plasticity, calibrated with multi-scale experiments, to understand the effect of the orientation

and morphology of the α and β phases on time dependent cyclic loading. *Journal of the Mechanics and Physics of Solids*, 146:104192, 2021.

[9] Dhyanjyoti Deka, Deepu S Joseph, Somnath Ghosh, and Michael J Mills. Crystal plasticity modeling of deformation and creep in polycrystalline ti-6242. *Metallurgical and materials transactions A*, 37(5):1371–1388, 2006.

[10] JR Mayeur and DL McDowell. A three-dimensional crystal plasticity model for duplex ti–6al–4v. *International journal of plasticity*, 23(9):1457–1485, 2007.

[11] G Venkatramani, S Ghosh, and M Mills. A size-dependent crystal plasticity finite-element model for creep and load shedding in polycrystalline titanium alloys. *Acta materialia*, 55(11):3971–3986, 2007.

[12] Marc C Kennedy and Anthony O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001.

[13] John McFarland, Sankaran Mahadevan, Vicente Romero, and Laura Swiler. Calibration and uncertainty analysis for computer simulations with multivariate output. *AIAA journal*, 46(5):1253–1265, 2008.

[14] Aditya Venkatraman, David L McDowell, and Surya R Kalidindi. Bayesian analysis of parametric uncertainties and model form probabilities for two different crystal plasticity models of lamellar grains in α+ β titanium alloys. *International Journal of Plasticity*, 154:103289, 2022.

[15] Aaron E Tallman, Krzysztof S Stopka, Laura P Swiler, Yan Wang, Surya R Kalidindi, and David L McDowell. Gaussian-process-driven adaptive sampling for reduced-order modeling of texture effects in polycrystalline alpha-ti: Tallman, stopka, swiler, wang, kalidindi, and mcdowell. *JOM*, 71(8):2646–2656, 2019.

[16] Joshua D Pribe, Brodan Richter, Patrick E Leser, Saikumar R Yeratapally, George R Weber, Andrew R Kitahara, and Edward H Glaessgen. A process-structure-property simulation framework for quantifying uncertainty in additive manufacturing: application to fatigue in ti-6al-4v. *Integrating Materials and Manufacturing Innovation*, 12(3):231–250, 2023.

[17] Joshua D Pribe, Patrick E Leser, Saikumar R Yeratapally, and George Weber. Probabilistic calibration of crystal plasticity material models with synthetic global and local data. *arXiv preprint arXiv:2603.20241*, 2026.

[18] Thomas Bengtsson, Peter Bickel, and Bo Li. *Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems*, page 316–334. Institute of Mathematical Statistics, 2008.

[19] Chris Snyder, Thomas Bengtsson, Peter Bickel, and Jeff Anderson. Obstacles to high-dimensional particle filtering. *Monthly Weather Review*, 136(12):4629 – 4640, 2008.

[20] S. Agapiou, O. Papaspiliopoulos, D. Sanz-Alonso, and A. M. Stuart. Importance sampling: Intrinsic dimension and computational cost. *Statistical Science*, 32(3):405–431, 2017.

[21] Ilya M Sobol. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and computers in simulation*, 55(1-3):271–280, 2001.

[22] Stefano Tarantola, D Gatelli, SS Kucherenko, Wolfgang Mauntz, et al. Estimating the approximation error when fixing unessential factors in global sensitivity analysis. *Reliability engineering & system safety*, 92(7):957–960, 2007.

[23] Paul G. Constantine. *Active Subspaces*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2015.

[24] Paul G Constantine, Eric Dow, and Qiqi Wang. Active subspace methods in theory and practice: applications to kriging surfaces. *SIAM Journal on Scientific Computing*, 36(4):A1500–A1524, 2014.

[25] Paul G Constantine, Carson Kent, and Tan Bui-Thanh. Accelerating markov chain monte carlo with active subspaces. *SIAM Journal on Scientific Computing*, 38(5):A2779–A2805, 2016.

[26] Yulin Guo, Sankaran Mahadevan, Shunsaku Matsumoto, Shunsuke Taba, and Daigo Watanabe. Investigation of surrogate modeling options with high-dimensional input and output. *AIAA journal*, 61(3):1334–1348, 2023.

[27] Yulin Guo, Paromita Nath, Sankaran Mahadevan, and Paul Witherell. Active learning for adaptive surrogate model improvement in high-dimensional problems. *Structural and Multidisciplinary Optimization*, 67(7):122, 2024.

[28] Tiangang Cui, James Martin, Youssef M Marzouk, Antti Solonen, and Alessio Spantini. Likelihood-informed dimension reduction for nonlinear inverse problems. *Inverse Problems*, 30(11):114015, 2014.

[29] Promit Chakroborty and Sankaran Mahadevan. Efficient bayesian calibration of many-parameter system models. *arXiv preprint arXiv:2608.08488*, 2026.

[30] J. Ramesh Babu, Pranav M. Karve, Sankaran Mahadevan, Kishore A. Nair, Tawqeer N. Tak, and Somnath Ghosh. Bayesian calibration of crystal plasticity finite element model parameters for ti-6al-4v alloy. *International Journal of Plasticity*, (Accepted for publication, 2026).

[31] Deniz Ozturk, Ahmad Shahba, and Somnath Ghosh. Crystal plasticity fe study of the effect of thermo-mechanical loading on fatigue crack nucleation in titanium alloys. *Fatigue & Fracture of Engineering Materials & Structures*, 39(6):752–769, 2016.

[32] Michael A Groeber and Michael A Jackson. Dream. 3d: a digital representation environment for the analysis of microstructure in 3d. *Integrating materials and manufacturing innovation*, 3(1):56–72, 2014.

[33] Ahmad Shahba and Somnath Ghosh. Crystal plasticity fe modeling of ti alloys for a range of strain-rates. part i: A unified constitutive model and flow rule. *International Journal of Plasticity*, 87:48–68, 2016.

[34] Edwin O Hall. The deformation and ageing of mild steel: Iii discussion of results. *Proceedings of the Physical Society. Section B*, 64(9):747–753, 1951.

[35] Norman J Petch. The cleavage strength of polycrystals. *J. Iron Steel Inst.*, 174:25–28, 1953.

[36] Peter J Armstrong, Charles O Frederick, et al. *A mathematical representation of the multiaxial Bauschinger effect*, volume 731. Berkeley Nuclear Laboratories Berkeley, CA, 1966.

[37] James O Ramsay and Bernard W Silverman. *Functional data analysis*. Springer, 1997.

[38] Manav Vohra, Paromita Nath, Sankaran Mahadevan, and Yung-Tsun Tina Lee. Fast surrogate modeling using dimensionality reduction in model inputs and field output: Application to additive manufacturing. *Reliability engineering & system safety*, 201:106986, 2020.

[39] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2005.

[40] Edwin Thompson Jaynes. *Probability theory: the logic of science*. Washington University, 1996.

[41] Paul G Constantine and Paul Diaz. Global sensitivity metrics from active subspaces. *Reliability Engineering & System Safety*, 162:1–13, 2017.