



Delft University of Technology

Mixing Data-Driven and Physics-Based Constitutive Models Using Uncertainty-Driven Phase Fields

Storm, Joep; Sun, WaiChing; Rocha, Iuri B.C.M.; van der Meer, Frans P.

DOI

[10.1002/nme.70162](https://doi.org/10.1002/nme.70162)

Publication date

2025

Document Version

Final published version

Published in

International Journal for Numerical Methods in Engineering

Citation (APA)

Storm, J., Sun, W., Rocha, I. B. C. M., & van der Meer, F. P. (2025). Mixing Data-Driven and Physics-Based Constitutive Models Using Uncertainty-Driven Phase Fields. *International Journal for Numerical Methods in Engineering*, 126(20), Article e70162. <https://doi.org/10.1002/nme.70162>

Important note

To cite this publication, please use the final published version (if applicable).

Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.

We will remove access to the work immediately and investigate your claim.

RESEARCH ARTICLE OPEN ACCESS

Mixing Data-Driven and Physics-Based Constitutive Models Using Uncertainty-Driven Phase Fields

Joep Storm¹  | WaiChing Sun²  | Iuri B. C. M. Rocha¹ | Frans P. van der Meer¹

¹Department of Civil Engineering and Geosciences, Delft University of Technology, Delft, the Netherlands | ²Department of Civil Engineering and Engineering Mechanics, Columbia University, New York, New York, USA

Correspondence: Joep Storm (j.storm@tudelft.nl)

Received: 23 April 2025 | **Revised:** 24 September 2025 | **Accepted:** 7 October 2025

Funding: This work was supported by TU Delft AI Labs.

Keywords: adaptive modeling | Gaussian Process | phase field | surrogate modeling

ABSTRACT

There is a high interest in accelerating multiscale models using data-driven surrogate modeling techniques. Creating a large training dataset encompassing all relevant load scenarios is essential for a good surrogate, yet the computational cost of producing this data quickly becomes a limiting factor. Commonly, a pre-trained surrogate is used throughout the computational domain. We introduce an alternative adaptive mixture approach that uses a fast probabilistic surrogate model as a constitutive model when possible, but resorts to the true high-fidelity model when necessary. The surrogate is thus not required to be accurate for every possible load condition, enabling a significant reduction in the data collection time. We achieve this by creating phases in the computational domain corresponding to the different models. These phases evolve using a phase-field model driven by the surrogate uncertainty. When the surrogate uncertainty becomes large, the phase-field model causes a local transition from the surrogate to the high-fidelity model, maintaining a highly accurate simulation. We discuss requirements for accuracy and numerical stability and compare the phase-field model to a local approach that does not enforce spatial smoothness in phase mixing. Using a Gaussian Process surrogate for an elasto-plastic material, we demonstrate the potential of this mixture of models to accelerate multiscale simulations.

1 | Introduction

Advanced manufacturing techniques allow for the creation of materials with properties tailored to their use. Multiscale modeling can accurately capture the behavior of these materials, yet its use is limited by the high computational cost of simulations. The need for more computationally efficient models, combined with the progress in machine learning, has resulted in numerous data-driven models [1–6]. Given sufficient data, these models often show an impressive ability to capture the behavior of complex history-dependent load paths. They can therefore be used as surrogates for the representative volume element (micromodel) in a multiscale analysis. However, obtaining the necessary training data from

This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *International Journal for Numerical Methods in Engineering* published by John Wiley & Sons Ltd.

microscale simulations is very costly. Because purely data-driven methods often extrapolate poorly, the data required to ensure fidelity tends to be large. In addition, it is often challenging to predict which loading scenarios will be relevant. The surrogate, therefore, needs to be trained across a broad range of inputs to generalize for unseen data, many of which may never be relevant during the target simulation.

We can broadly categorize two (non-exclusive) approaches to dealing with this problem. The first is to embed a prior in the surrogate design, allowing the model to extrapolate constitutive responses from small datasets. Examples include leveraging geometry or embedding physics into the model as an inductive bias [7–10]. These models often have intermediate quantities or additional outputs with a physical meaning that can be leveraged during the training and inference process.

The second approach is to use active learning rather than rely solely on a pre-trained surrogate. In active learning, the surrogate is continuously retrained whenever it makes a poor prediction, avoiding the need for an all-encompassing dataset. Authors in [11] propose using the uncertainty of a Gaussian Process (GP) surrogate to inform on-the-fly adaptivity. There, several fully-solved anchor models are used throughout the simulation to update the GP, triggered by the GP uncertainty. A similar approach, FE^{ANN} , stores all deformation gradients and, by comparing invariants of each new deformation to this dataset, it detects unseen deformations [12]. In that case, the method runs new simulations, enriches the dataset, and retrains the network. Adaptively updating the surrogate during the simulation avoids the necessity for all scenarios to be covered in an extensive pre-training dataset. However, active learning does require the surrogate to be suitable for retraining. Furthermore, updating the surrogate can cause instabilities or slower convergence of the macroscopic solver.

Instead of using the high-fidelity model just to update the surrogate, a mixture of several models can be applied. A common strategy is clustering, where elements are grouped by response after an initial solve of the system [13]. Alternatively, a mixture of a pre-trained neural network (NN) and a reduced order model (ROM) can be used throughout the domain [14]. After solving an FE^2 simulation with the NN surrogate at every point, a second run is performed where all points that reached values outside the NN training region during the first run are replaced with the ROM. To minimize unnecessary evaluations, the second run can also start with the NN at all points, and switch to the ROM only when strain values outside the NN training region occur. One of the challenges of these methods is avoiding the instability of the finite element method (FEM) caused by stress jumps at boundaries between two models. One option to avoid these jumps between models is to introduce a transition zone in front of a crack based on the thick level set method [15]. From a machine learning perspective, these hybrid approaches are analogous to a mixture-of-experts model [16], where we use domain-specific information to choose the expert. Alternatively, this also bears a resemblance to data assimilation methods, such as using a weighted least-squares finite element method to assimilate experimental data with numerical models [17].

In this work, we allow for adaptively using multiple constitutive models when and where necessary. Specifically, a probabilistic surrogate and the original model it was trained to replace are aggregated to make predictions. The main idea is to use the uncertainty in surrogate predictions to guide whether the surrogate or the original model should be used. For our motivation of accelerating multiscale simulations, the fast surrogate should be used when it has a low uncertainty, and the original model should be used otherwise. We use the idea of a weighted average transition zone and implement it using a phase field. The phase field promotes numerical stability, as it naturally introduces a spatial transition zone that is used to switch between the models. Phase-field models have been used to simulate the evolution of, among others, solidification [18], grain growth [19], and fracture [20] for several decades now. In a phase field, a scalar variable ϕ evolves to represent, e.g., liquid-solid transitions or crack formation. Here, we use the phases for the different constitutive models. The uncertainty coming from the probabilistic surrogate drives the phase field evolution. At the diffuse interface between phases, we take a weighted average between model predictions to smoothly transition between the models.

Using the phase field brings the cost of solving an additional partial differential equation (PDE) every time step. In FE^2 simulations, the costs associated with solving microstructures are generally so high that the cost of solving the phase field is negligible. When a surrogate is trained on a well-curated dataset that allows it to predict the full behavior of the simulation with little uncertainty, the phase field remains zero, and only the surrogate is used. In this case, this approach only requires the additional computation of the surrogate uncertainty. The mixture of constitutive models preserves accuracy when this condition is not met.

The width of the transition zone can play an important role in the stability of the mechanical and phase-field problems. A narrow transition zone requires fewer expensive model evaluations, making it computationally desirable. However, this might lead to undesirable stress jumps or make the solver require more Newton–Raphson iterations. To study this, we compare the phase-field approach to a purely local approach, which does not require solving another PDE. By letting the

uncertainty directly determine which constitutive model is used, we can essentially let the transition zone vanish and study the impact on the stability of the mechanical problem.

The outline of this paper is as follows. In Section 2, we discuss using FEM for solving the behavior of structures undergoing loading. We provide a detailed overview of our phase-field-based approach for mixing models in Section 3. There, we also discuss using a GP as a surrogate constitutive model. We present the results of the different approaches in Section 4. Finally, we conclude with an overview of our findings in Section 5.

2 | Finite Element Analysis

FEM is used to model the behavior of a structure or material under load. Specifically, the aim is to find the displacement field \mathbf{u} of a structure subject to boundary conditions. We obtain equilibrium in the domain by finding the displacement field \mathbf{u} satisfying

$$\nabla \cdot \boldsymbol{\sigma} = 0. \quad (1)$$

Here, $\boldsymbol{\sigma}$ is the stress, and $\nabla \cdot$ indicates the divergence operator. \mathbf{u} is related to the strains $\boldsymbol{\varepsilon}$ as

$$\boldsymbol{\varepsilon} = \frac{1}{2} (\nabla \mathbf{u} + (\nabla \mathbf{u})^T), \quad (2)$$

and $\boldsymbol{\varepsilon}$ can be related to $\boldsymbol{\sigma}$ through a constitutive relation:

$$\boldsymbol{\sigma} = \mathcal{C}(\boldsymbol{\varepsilon}, \boldsymbol{\alpha}). \quad (3)$$

The material-dependent constitutive operator \mathcal{C} is often nonlinear and can depend on the loading history. The internal variables $\boldsymbol{\alpha}$ can account for this loading history. Due to the nonlinear nature of the constitutive model, Equation (1) cannot be solved directly. Instead, an iterative Newton–Raphson scheme is employed. To obtain quadratic convergence, we require not just the outcome of the constitutive relation ($\boldsymbol{\sigma}$), but also its derivative (\mathbf{D}).

To find equilibrium, the FEM discretizes the continuous domain of the structure into a number of elements. The boundary conditions, in the form of forces or prescribed displacements, are applied to the nodes of these elements. Within each element, the displacement field \mathbf{u} is approximated using simple *basis functions*, often polynomials. The Gauss quadrature rule is used to evaluate the integrals arising from the discretization, with strategically placed integration points (IPs) within each element. The constitutive relation from Equation (3) is evaluated for each integration point.

We consider quasi-static loading, where the load is applied in consecutive time steps without causing dynamic effects. This allows for obtaining the full load path of the structure of interest. For a single simulation, we thus have to compute the constitutive model for all quadrature points in the elements, for each Newton–Raphson iteration of all loading steps.

For certain materials, such as composites, the constitutive relation can depend on the geometry at the microscale. Finding accurate constitutive relations for these materials is challenging, even if the constitutive relations of each constituent are known. In multiscale simulations using FE², a microscopic FEM problem is solved instead, where $\boldsymbol{\varepsilon}$ is imposed as boundary conditions on a microstructure. We solve this system as in FEM since the constituents' properties are well described, then pass the homogenized stress $\boldsymbol{\sigma}$ back to the macroscale. However, solving a boundary value problem at every macroscopic quadrature point and time step brings a considerable computational cost. This cost is so high that, even with parallelization, FE² is used only sparingly in practice, motivating acceleration strategies.

To make it computationally feasible to study the approach in detail, we limit ourselves to single-scale problems in this work. We use an elasto-plastic material with a von Mises yield criterion as our high-fidelity (HF) model, which serves as the ground truth. To quantify the acceleration of our approach, it thus makes little sense to compare the computational time—as evaluating this elasto-plasticity model is about as fast as the surrogate we use to replace it. Instead, we measure the reduction in the number of HF model evaluations. The simplification of using an analytical constitutive model, therefore, does not affect our upcoming conclusions about the approach.

For complex problems, the Newton–Raphson solver can fail to converge for a given load increment. Adaptive load stepping is then required to obtain a solution. In this work, we use an adaptive load step strategy that stays constant as long as the solver converges. If it fails to converge, the load step is reduced by a factor γ . This is repeated until the solver converges,

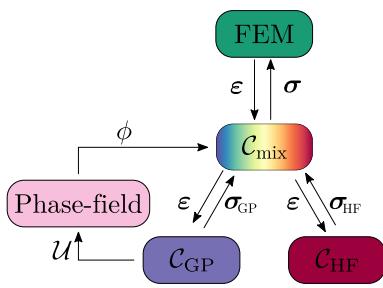


FIGURE 1 | Overview of how the mixture constitutive model interacts with the other models. The tangent stiffness \mathbf{D} is omitted for clarity.

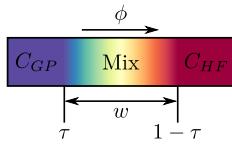


FIGURE 2 | As ϕ increases, the constitutive model C_{mix} gradually switches from C_{GP} to C_{HF} .

or until a minimum load step value is reached. After reaching the minimum value, increased load steps are attempted, up to a maximum value, after which the simulation is terminated and the problem is considered unsolved. If it converges, the load step remains constant for one step before gradually increasing back to the initial size, using a factor $\frac{1}{\gamma}$.

3 | Mixture of Constitutive Models

We solve the mechanical problem by combining constitutive models of distinct accuracies and computation speeds in the macroscopic domain. The analytical elasto-plastic material model serves as the high-fidelity model (C_{HF}), and we use a GP surrogate model (C_{GP}). Using a GP is a pragmatic choice, as it naturally provides an uncertainty. A limitation of a GP is its inability to handle path-dependency. Other types of constitutive models and surrogates could be used, as long as the surrogate provides an uncertainty in its prediction. The general schematic of the approach in Figure 1 shows what information the mixture model obtains from other models. Ideally, C_{GP} is used throughout most of the domain, and C_{HF} only for a small subset of points. Our hybrid constitutive model depends on a phase-field variable ϕ to determine which constitutive model is used at each quadrature point:

$$C_{\text{mix}} = \begin{cases} C_{\text{GP}} & \text{if } \phi < \tau, \\ (1 - \phi)C_{\text{GP}} + \phi C_{\text{HF}} & \text{if } \tau < \phi < 1 - \tau, \\ C_{\text{HF}} & \text{if } \phi > 1 - \tau. \end{cases} \quad (4)$$

At the interface between the two phases, the hybrid constitutive model is thus a weighted average of C_{HF} and C_{GP} , as visualized in Figure 2. In practice, we take weighted averages over both σ and \mathbf{D} . We use a small cutoff value, τ , to avoid simulating C_{HF} when the phase field is close to but not exactly zero. The upper boundary of $1 - \tau$ is of lesser importance since the cost of solving C_{GP} is negligible, and its uncertainty is still required when updating the mixture. We use $\tau = 0.01$ in all numerical experiments. Algorithm 1 shows this procedure in more detail. The high-fidelity model generally tracks internal variables α to, for example, account for plasticity. These internal variables are only committed when moving to the next time step. When switching from the surrogate model to the full model (when ϕ crosses the threshold value τ), the internal variables of the full model are unknown yet necessary. In this scenario, the history path is recomputed using the full model to obtain the internal variables. We call this process the retracing of history. Although this increases the number of evaluations, we only need to retrace along the converged solutions. Thus, this still requires fewer evaluations than if the full model were used for this point from the start. We could, in principle, use larger time step sizes when reconstructing, but here we pragmatically opt for retracing every increment missed by the full model. This also means we have to store converged strain values at every fully-reduced IP for all time steps in the event a retracing is needed.

ALGORITHM 1 | Mixture of constitutive models.

Input: ϵ, ϕ
Output: σ, D
Initialize: Internal variables α
if $\phi < \tau$ **then**
 Use only the GP model: $\sigma, D = C_{\text{GP}}(\epsilon)$
else
 Retrace α to the current timestep if it is outdated
 if $\phi > 1 - \tau$ **then**
 Use only the high-fidelity model: $\sigma, D = C_{\text{HF}}(\epsilon, \alpha)$
 else
 $\sigma_{\text{GP}}, D_{\text{GP}} = C_{\text{GP}}(\epsilon)$
 $\sigma_{\text{HF}}, D_{\text{HF}} = C_{\text{HF}}(\epsilon, \alpha)$
 $\sigma = \phi\sigma_{\text{HF}} + (1 - \phi)\sigma_{\text{GP}}$
 $D = \phi D_{\text{HF}} + (1 - \phi)D_{\text{GP}}$
 end if
end if

3.1 | Phase-Field Approach

The main purpose of the phase field is to control how the surrogate uncertainty influences the mixture. Specifically, it allows us to create a controllable interface width and determine the minimum uncertainty required to initiate the switch. While monolithic schemes exist [21], staggered schemes are more common [22] and are adopted here. The variational form used in our numerical implementation is as follows:

$$F(\phi, v) = - \underbrace{\int_{\Omega} [\mathcal{U}v] \, d\Omega}_{\text{driving force}} + b \underbrace{\int_{\Omega} v \, d\Omega}_{\text{opposing force}} + \underbrace{\int_{\Omega} [\epsilon^2 \nabla \phi \cdot \nabla v] \, d\Omega}_{\text{interface energy}} + \underbrace{\int_{\Omega} [\omega \phi(1 - \phi)(1 - 2\phi)v] \, d\Omega}_{\text{double-well}} \quad (5)$$

Here, v represents the test function. The first term is the driving force \mathcal{U} , causing ϕ to increase as the uncertainty increases. Because we use a staggered approach, the driving force does not necessarily have to be continuous with respect to the strains to maintain stability. The second term is a constant opposing force, acting as a threshold that the driving force needs to overcome. The third term represents the interface energy, penalizing sharp gradients in the phase-field variable. By influencing the interface width, this term may affect the stability of the problem. The fourth term is the derivative of the double-well potential, ensuring the phase-field variable favors values near 0 or 1. The phase-field formulation does not depend on its previous state, as we do not require our phase field to be smooth in time. This avoids issues related to path dependency, where a different loading step size would lead to different behavior.

3.2 | Staggered Updating Scheme

We solve the phase-field and mechanical problem in a decoupled, staggered manner. In Figure 3, a schematic overview of the staggered updating scheme is shown. Although either field can be updated first [22], we choose to first update the phase field based on the last known surrogate model uncertainty. The phase field is then fixed while solving the mechanical problem.

Solving each problem once per time step is desirable, but it has two drawbacks. First, the uncertainty from the previous converged step might not be a good indicator of the performance at the current step. Second, this can introduce a load step size dependency, where taking different step sizes leads to different solutions. To overcome these issues, we iterate several times over this staggered scheme before moving on to the next step. We consider this approach to be converged when the norm of the change in \mathbf{u} is below a threshold value ϵ_{it}^u . In certain scenarios, such as for the first few time steps when $\phi = 0$ for all points, the phase field will not change, even when \mathbf{u} changes. We then know that \mathbf{u} will not change in the next iteration and can consider the current solution converged based on ϕ , avoiding redundant iterations. By employing these iterations over the staggered approach, we are adding additional solves of the mechanical problem in which we generally need to perform costly C_{HF} evaluations. However, by using the result of each iteration as the initial prediction for the

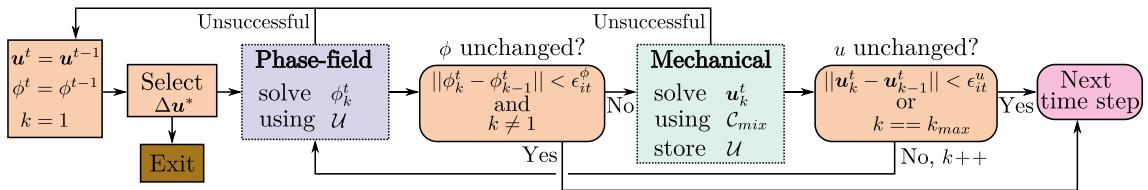


FIGURE 3 | The staggered updating approach used in this work. A prescribed displacement Δu^* is selected using an adaptive time-step function, before updating the phase field based on \mathcal{U} . Then the mechanical problem is solved. The phase-field and mechanical problem are updated iteratively until convergence, or k_{\max} is reached.

subsequent iterations, these should converge using fewer Newton–Raphson iterations than they would if starting from \mathbf{u}_{t-1} .

Instead of relying on the uncertainty coming from the mixed model (which entails expensive high-fidelity simulations), one might instead be inclined to obtain the uncertainty from a preliminary surrogate-only solution. One would first solve the mechanical problem using only the surrogate model, update the phase field based on the resulting uncertainty, and then solve the mechanical problem again with the mixture model. However, we find empirically that this approach is not viable: Firstly, it heavily depends on reasonable surrogate performance and can fail to converge for inaccurate predictions; secondly, it reintroduces load step size dependency. Due to these reliability concerns, we do not present results using this approach.

3.3 | Gaussian Processes as Surrogate Constitutive Models

Bayesian surrogate models provide powerful tools for approximating complex functions. While several surrogate models provide uncertainty estimates (neural network ensembles, dropout networks [23], bootstrapping approaches) and are therefore suitable for our approach, we focus on GP regression. This is a pragmatic choice to demonstrate our framework for mixing constitutive models. In more realistic applications, GPs might be insufficient as they are unable to handle path-dependency, and scale poorly with the number of training points. Other surrogates that provide a measure of uncertainty could be used instead. In Section A, we provide a general description of GPs for regression that serves as background to this section. Here, we describe how we use GPs specifically as constitutive models.

We create a separate GP for each stress component. Since our experiments are 2-dimensional, we thus have separate GPs for σ_x , σ_y , and σ_{xy} , each taking the full strain tensor as input. The GPs use a radial basis function kernel, and their hyperparameters are individually tuned. Rather than predicting σ directly, we predict a correction term to a linear elastic model:

$$\sigma_i = [\mathbf{D}_e \boldsymbol{\epsilon}]_i + \text{GP}_i(\boldsymbol{\epsilon}), \quad (6)$$

where \mathbf{D}_e is the elasticity tensor. For predictions away from the data, the output thus reverts back to a linear elastic model, rather than to the zero prior. The scalar driving force for the phase field is computed as the maximum variance of the components:

$$\mathcal{U} = \max_i(\sqrt{\text{var}[\text{GP}_i(\boldsymbol{\epsilon})]}). \quad (7)$$

Since this GP formulation only depends on $\boldsymbol{\epsilon}$, it cannot capture unloading after plasticity has occurred. If unloading is to occur in the simulation, it is likely that the GP could make inaccurate predictions. For this reason, we focus on scenarios where the global load prescribed by the boundary conditions increases monotonically throughout the simulation.

As in related work, we assume that the surrogate variance correlates with the true error. If a surrogate is confidently wrong, the result can be inaccurate, as it would be if only the surrogate were used. Anomaly detection could be used to identify such scenarios, but this would come at a large computational cost.

4 | Results

We start by showing the behavior of our mixture of models on a simple dogbone structure loaded with a prescribed displacement in tension. In this dogbone experiment, we focus on the influence of performing several staggered iterations.

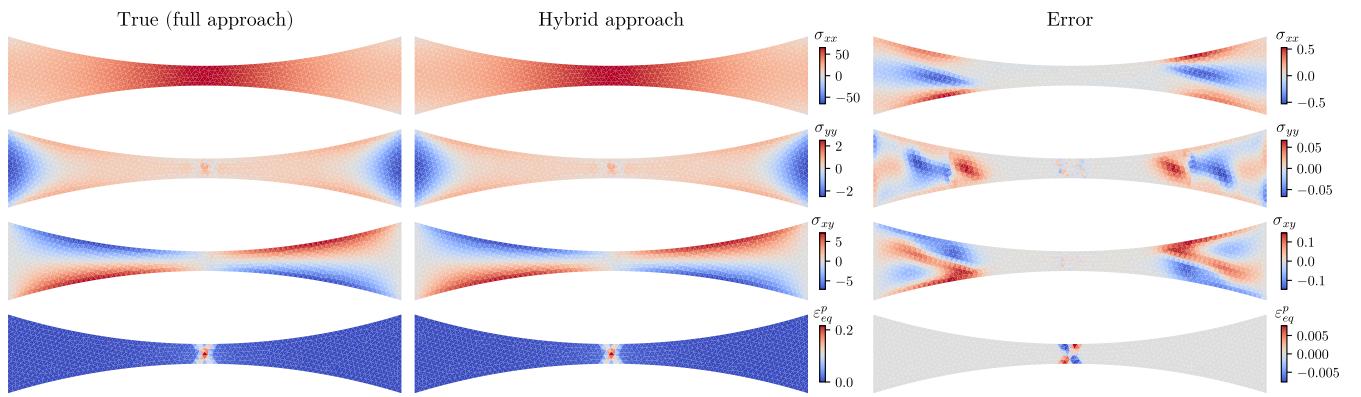


FIGURE 4 | Full-field stress and equivalent plastic strain ε_{eq}^p plots of the dogbone experiment at the maximum prescribed displacement. The full and hybrid model with $k_{max} = 3$ are shown, as well as the difference between them.

Then, moving to a more complex notched plate, we study the influence of the different phase-field parameters. We additionally compare the phase-field method to an alternative local approach. Finally, we investigate the potential of our approach to reduce the number of high-fidelity simulations on a plate with holes where a complex phase field is required.

4.1 | Numerical Setup

As discussed earlier, for computational feasibility, we do not perform FE² in this study, but instead use an analytical constitutive model as C_{HF} . For this reason, we do not compare the computational time of our approach to the full FE² model, but instead focus on the reduction in the number of high-fidelity model evaluations. We use an elasto-plastic material with von Mises plasticity in a plane-stress condition, a Young's modulus of 3130 [MPa], a Poisson's ratio of 0.37[–], and yield criterion $\sigma_y = 64.80 - 33.60 \cdot e^{\frac{\varepsilon_{eq}^p}{-0.003407}}$, where ε_{eq}^p is the equivalent plastic strain (corresponding to the internal variable α used earlier). These material properties and yield criterion are adopted from [24]. In all experiments, we use quadratic triangular elements (T6) for the mechanical problem and linear triangular elements (T3) for the phase field. For the dogbone structure, a constant displacement increment of 0.001 is sufficient to obtain convergence. In the remaining problems, the adaptive time step handling as described in Section 2 is used, with $\gamma = 0.5$.

We study the influence of the phase field with three different GP surrogates, *GP10*, *GP30*, and *GP100*, trained on different datasets with 10, 30, and 100 load curves, respectively. Each load curve consists of 20 monotonically increasing steps in a random strain direction, up to a strain norm of 10%. We choose 10% for pragmatic reasons, assuming we have no prior information about the strains that will occur during the simulation. Just because certain GPs have more data does not guarantee that their predictions during the simulation are always more accurate, since the loading directions of their data are random. Still, it can be reasonably expected that the GPs with more data improve accuracy and reduce reliance on the full model.

The framework is implemented using the *FEniCSx* finite element library [25] with the *dolfinx_materials* [26] package to define arbitrary constitutive models. *PETSc* [27] solvers are used to solve the mechanical and phase-field problem. The phase field is solved using a constrained solver with bounds [0, 1]. All code is available on GitHub at <https://github.com/SLIMM-Lab/phase-field-mixture>.

4.2 | Time Step Consistency

We start with a simple dogbone structure to study the influence of the staggered approach of Figure 3 on the time-step dependency. The dogbone structure is fixed on its left edge, and loaded by incrementally applying a prescribed displacement to the right edge, up to an elongation of 2%. Figure 4 shows the full-field responses for the stresses and equivalent plastic strain. The stresses mainly occur in the x-direction and concentrate at the center of the dogbone, where the structure is narrower than at its ends, with ensuing strain localization.

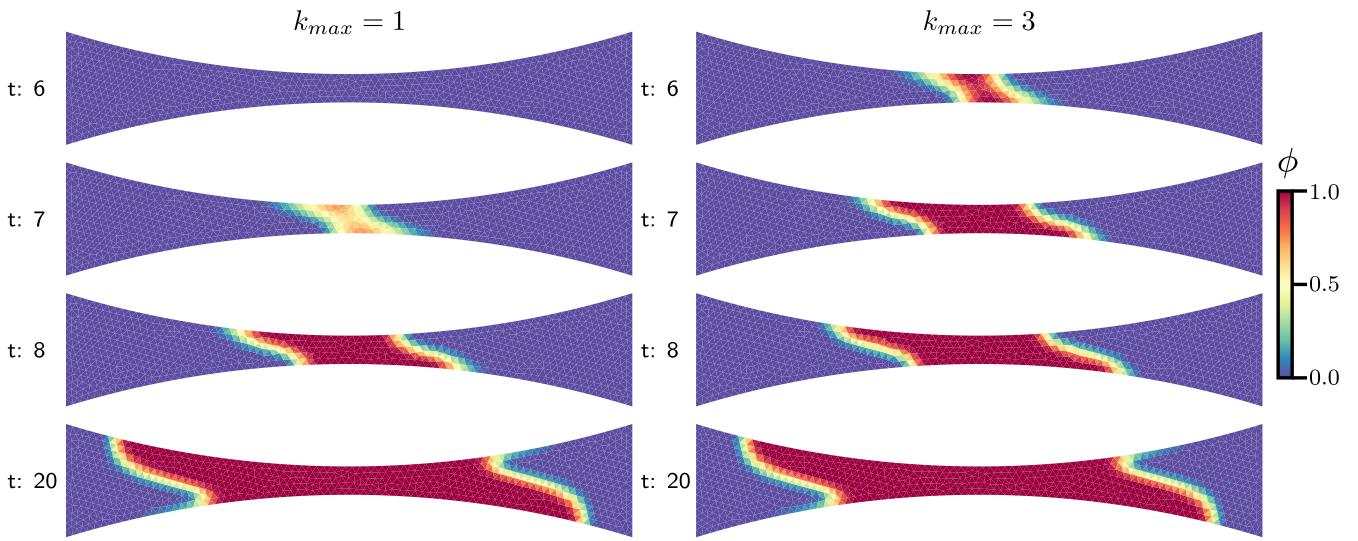


FIGURE 5 | Evolution of the phase field for the dogbone experiment. The phase field gradually grows until time step 17, after which it stays constant due to strain localization. The asymmetrical pattern occurs as a result of the asymmetrical shear stresses.

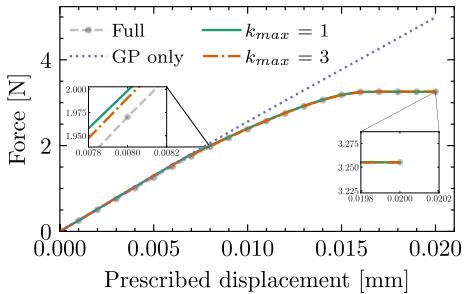


FIGURE 6 | The load-displacement curves for the dogbone experiment. While using only the surrogate model results in poor performance, the hybrid approach shows a large agreement with the baseline full model, with a slightly lower error for $k_{\max} = 3$ compared to $k_{\max} = 1$.

For the phase-field mixture we use the *GP10* surrogate with hyperparameters $\epsilon = 10^{-2}$, $\omega = 10^{-3}$, and $b = 1$. Figure 4 shows that the hybrid approach closely follows the full model. On the left side of Figure 5, we plot the phase-field variable for various time steps using $k_{\max} = 1$ (i.e., a single solve of each PDE per load increment, see Figure 3). At the start, when ϕ is zero everywhere, only the GP surrogate is used. Starting from time step seven, the uncertainty in the high-strain region at the center of the dogbone drives the phase field locally up to one. The phase field can switch from $\phi = 0$ to $\phi = 1$ in a single load step when the uncertainty increases quickly. This is necessary to prevent large errors when the GP suddenly becomes uncertain.

In Figure 6, we present the load-displacement curves for this problem. Using only the GP surrogate results in a poor prediction that does not capture the nonlinear behavior of the material, showing that the surrogate is inadequate for predicting the structural response by itself. The inability to capture non-linearity results from the very limited dataset used for *GP10*. Still, by switching away from the surrogate at the right time, the hybrid approach closely follows the full model path.

Figure 7 shows the mean absolute error of this hybrid model over the domain compared to using the high-fidelity model everywhere. The errors are computed for two fixed increment sizes Δu . When $k_{\max} = 1$, there is a clear difference in errors when using different increments. This difference can be explained by the phase field using uncertainty from the previous time step to determine the mixture of constitutive models. When using a different increment size, the phase field thus updates on different information, creating a time-step dependency. In contrast, we observe that using $k_{\max} = 3$ results in the same response for both loading increment sizes at the points where both are evaluated. Performing multiple iterations of the staggered approach per loading increment thus avoids the time-step dependency.

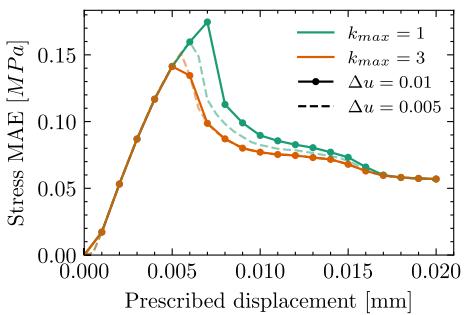


FIGURE 7 | The full-field stress error when using $k_{\max} = 1$ or $k_{\max} = 3$ for different load step increments.

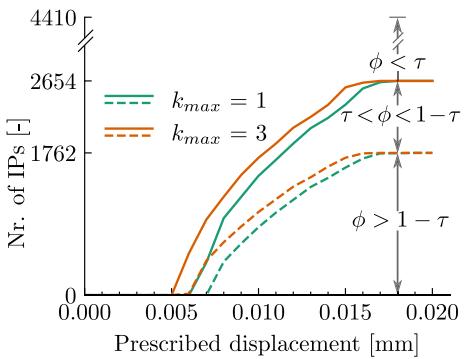


FIGURE 8 | Number of integration points (IPs) that use the GP surrogate ($\phi < \tau$), a mixture of both models ($\tau < \phi < 1 - \tau$), or the high-fidelity model only ($\phi > 1 - \tau$). The mesh contains 4410 IPs in total.

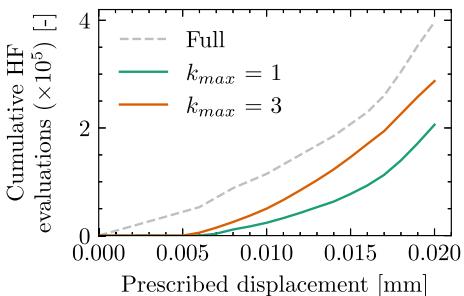


FIGURE 9 | Comparison of the cumulative number of HF evaluations. Retraced history computations are counted in the time step in which they are retraced. The significant difference between $k_{\max} = 1$ and $k_{\max} = 3$ comes mainly from the additional staggered iterations, requiring more Newton–Raphson steps.

With multiple iterations, the phase field is updated using information from the current load increment rather than from the previous iteration. In Figure 8, the evolution of ϕ throughout the simulation is plotted. It is clear from this plot that $k_{\max} = 1$ lags roughly one step behind $k_{\max} = 3$. This can also be observed by comparing the left and right sides of Figure 5.

The downside of performing several iterations of the staggered approach is that it requires more Newton-Raphson iterations of the mechanical problem, causing more high-fidelity constitutive model evaluations. Figure 9 shows the cumulative number of high-fidelity evaluations, where we observe that $k_{\max} = 3$ uses more evaluations than $k_{\max} = 1$. The reduction in high-fidelity evaluations compared to running the full model is relatively small in this example, as we intentionally use a GP trained on a very limited dataset to show the behavior of the phase field. Since it avoids the time-step-dependent behavior and improves the model's accuracy, we argue that performing several iterations is worthwhile for guaranteeing model robustness.

We further investigate the influence of the number of staggered iterations by running the model for more values of k_{\max} . Figure 10 shows the results by plotting the error and number of staggered iterations used throughout the simulation.

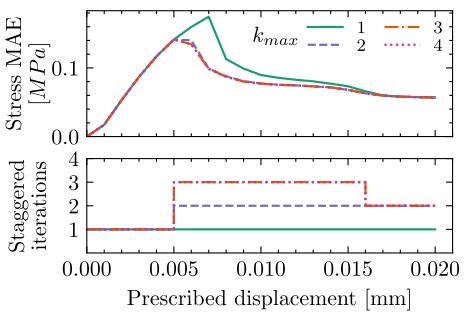


FIGURE 10 | The influence of k_{max} on the simulation. In this example, the model always converges within three iterations.

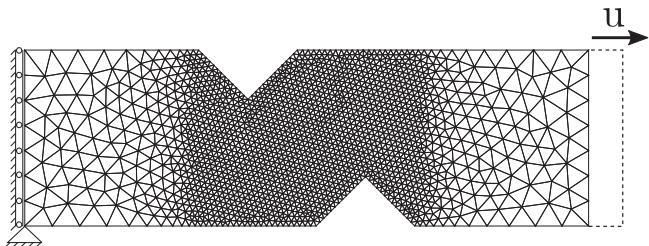


FIGURE 11 | Overview of the notched plate example.

Using two iterations already shows a significant difference in the error compared to using only one. Even with $k_{max} = 4$, the model always converges within three iterations for this experiment. However, it is possible for the problem to oscillate between iterations, never meeting the tolerances, so k_{max} should be set. We set $k_{max} = 3$ for the remainder of this work.

4.3 | Phase-Field Parameters

In this section, we explore the influence of the phase-field parameters in more detail. To highlight their influence, we create an experimental setup with a more complex phase-field evolution pattern.

4.3.1 | Experimental Setup

A plate with two notches diagonally opposite each other is loaded by incrementally displacing the right boundary by a prescribed amount. This setup is visualized in Figure 11, and we plot full-order simulation results in the left plots of Figure 12. It can be observed that stress concentrations occur at the tips of the notches, and a plastic strain band develops between these tips. The full-field errors of the hybrid approach with base settings show close agreement with the full-field model. Note how in this case the error with *GP10* is lower than with *GP100*, but this comes at the cost of running more high-fidelity simulations. We vary the element size over the domain to capture the strain localization accurately. Even though the global loading increases monotonically, we observe local unloading in some integration points close to the notch tips. As our GP surrogates cannot capture unloading, switching to the high-fidelity model at these points becomes essential. This is, however, not guaranteed by our method, particularly when using a high opposing force (through parameter b) that allows the GP to be used despite high uncertainty. Still, we empirically find that the limited local unloading has minimal impact on the results and does not affect our conclusions.

Several quantities are used to compare the influence of the hyperparameters. To evaluate the accuracy of the model, we compare force-displacement curves. Since we now use an adaptive stepper, the points of the force-displacement curve do not coincide if a certain step fails to converge, so we cannot directly compute their differences. Instead, this error is computed by first linearly interpolating each F-u curve to the fixed displacement points that the solution follows if it always converges. To compute the error, we sum the absolute force differences between the approximate model and the true model. In addition, we plot the average number of integration points (IPs) where the high-fidelity model is used (i.e., where $\phi > \tau$). This indicates how much the high-fidelity phase has grown during the simulation. As an indication of the stability, we compare the required number of Newton–Raphson iterations of the mechanical problem, obtained by

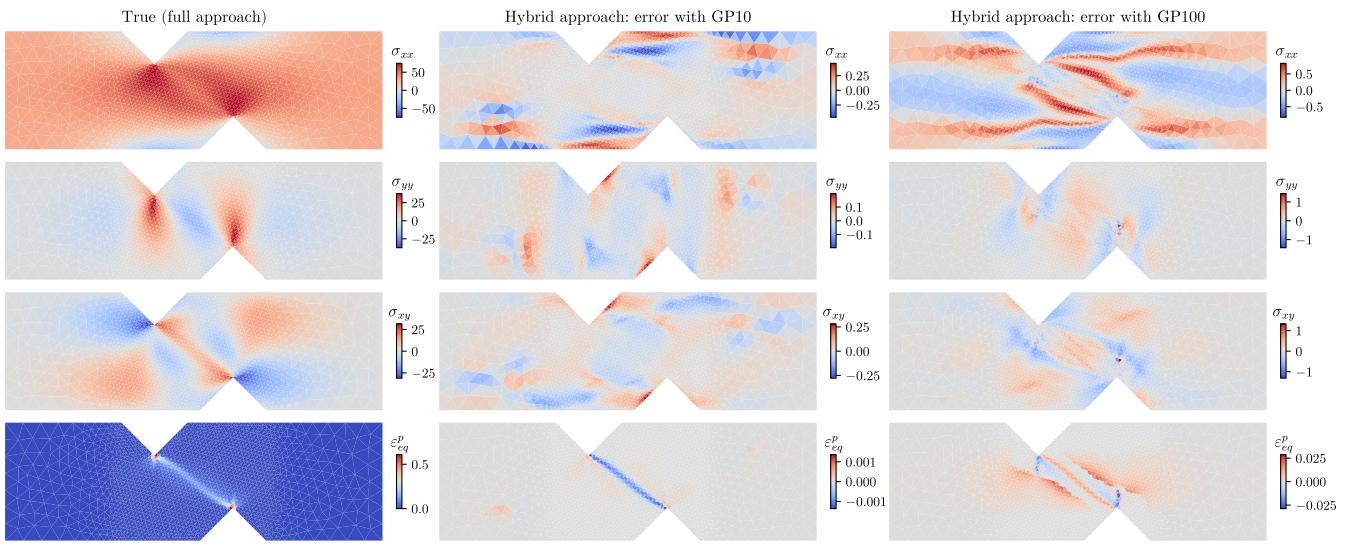


FIGURE 12 | Full-field stress and equivalent plastic strain ε_{eq}^p plots of the notched plate experiment at the maximum prescribed displacement. The full model results are plotted on the left. For the hybrid approaches, we plot the results for *GP10* and *GP100*. The differences are so small that we only plot the errors. Hyperparameters $b = 1$, $\epsilon = 10^{-2}$, and $\omega = 10^{-3}$ are used.

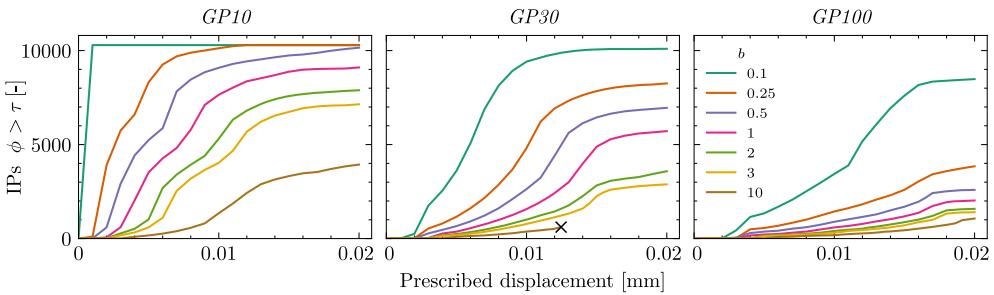


FIGURE 13 | Evolution of the phase field during the simulation for various b values. Default values $\epsilon = 10^{-2}$ and $\omega = 10^{-3}$ are used. The simulation with *GP30*, $b = 10$, failed to converge.

summing over all adaptive time steps (converged or unconverged), for all staggered approach loops. For the number of high-fidelity evaluations, we track each time the high-fidelity model is evaluated, including the retracing of load paths to obtain the internal variables (as described in Section 3.2).

4.3.2 | Phase-Field Opposing Force

We study the influence of the phase-field hyperparameter b on the mechanical problem. The parameter b determines the magnitude of the constant opposing force, as shown in Equation (5). The larger b is, the larger the uncertainty \mathcal{U} needs to be to drive the phase field up from zero towards one. Since the uncertainty of different surrogate models can behave very differently, the choice of b should be calibrated based on the behavior of the variance of the surrogate. Furthermore, it is challenging to directly compare different surrogates for the same b value. For example, the GPs we use in this work all have different hyperparameters that influence their maximum variance and how quickly the uncertainty increases away from training points. While this maximum variance could in principle be used to scale \mathcal{U} , making b an adimensional threshold, other surrogate methods often do not have an upper bound to their uncertainty. Therefore, we let b simply be the unscaled uncertainty.

We vary b while fixing the other phase field parameters from Equation (5) to $\epsilon = 10^{-2}$ and $\omega = 10^{-3}$. In Figure 13, we show how the phase field evolves throughout the simulation for the different GPs for different values of b . We observe that the uncertainty of GPs with more data grows more slowly and that the larger the value of b , the fewer IPs use the high-fidelity model. We visualize the influence of b on the error, the phase-field evolution, the required number of Newton–Raphson iterations, and the number of high-fidelity evaluations in Figure 14. The gray line in the plots refers to the values obtained

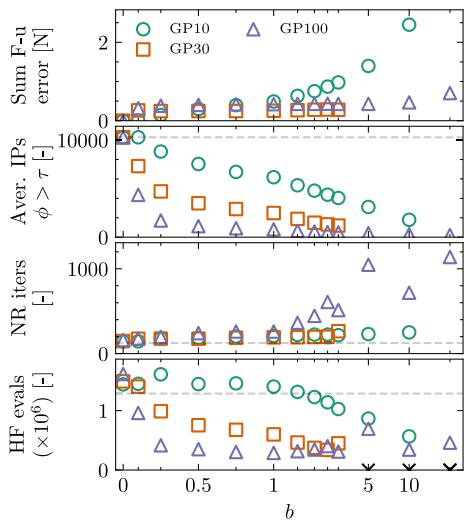


FIGURE 14 | Influence of the opposing force b on the mechanical problem. Note: that the x-axis is linear from zero to one, and then scales logarithmically.

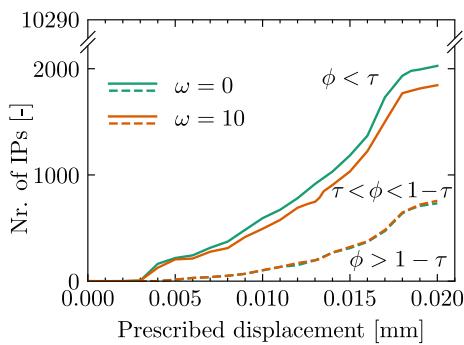


FIGURE 15 | Influence of the double-well term ω on the phase-field values. A higher ω leads to more elements having a ϕ value of either below τ or above $1 - \tau$. We use $GP100$, $\epsilon = 10^{-2}$, and $b = 1$.

when running the full model, with the high-fidelity constitutive model used in all IPs. Missing data points and the crosses in the bottom plot indicate that the setting failed to converge (black crosses indicate a failure of the mechanical problem, red crosses indicate a failure of the phase field). The error of the $GP10$ model increases consistently as we increase b . This indicates that we are using the insufficiently trained GP model even when it is highly uncertain. As expected, the GP models with more training data result in fewer $\phi > \tau$ elements. As we increase b , the required number of Newton–Raphson iterations of the mechanical problem generally increases, with several instances of $GP10$ and $GP30$ failing to reach the end of the simulation. This creates a trade-off where a higher b requires fewer high-fidelity evaluations but can increase the error and lead to instabilities.

4.3.3 | Phase-Field Double-Well

In Figure 15, we show the evolution of the phase-field values throughout a simulation for $\omega = 0$ and $\omega = 10$. We observe that this double-well term of the phase field pushes values away from $\phi = 0.5$ and towards $\phi = 0$ and $\phi = 1$. Increasing ω thus has the potential to increase the number of IPs with $\phi < \tau$, avoiding the high-fidelity model evaluation. Having fewer elements at the interface reduces the number of high-fidelity evaluations, leading to a larger acceleration. However, by setting the parameter ω too high, solving the phase field itself can become challenging. The difference for the $\phi = 1 - \tau$ boundary is smaller, indicating that there are fewer points with ϕ values close to that boundary.

The influence of ω on the result of the simulation for different GPs is shown in Figure 16. For $GP10$, the phase field reaches the part of the domain with larger mesh elements. The combination of $GP10$ being inaccurate and ω being high makes convergence more difficult. Values below $\omega < 10^{-1}$ appear to have little effect on any of the quantities. For larger values

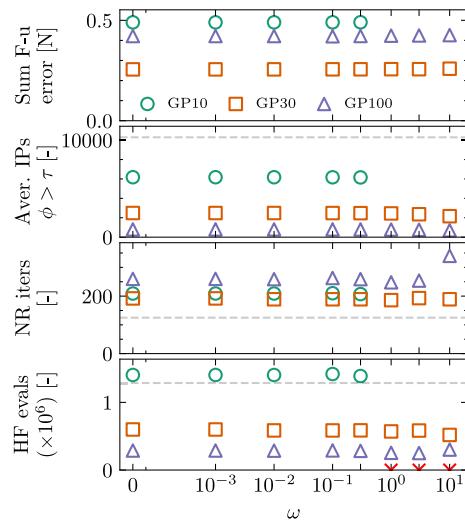


FIGURE 16 | Influence of the double-well term ω on the mechanical problem.

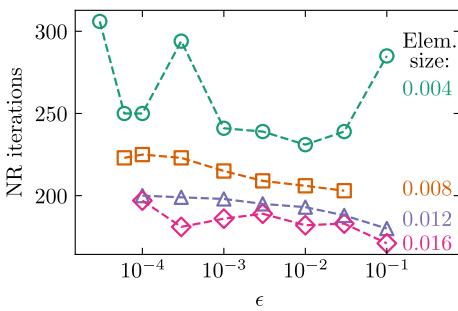


FIGURE 17 | Stability of the phase-field and mechanical problem concerning the element size and phase-field variable ϵ on the notched plate problem. All problems are evaluated on an $\epsilon \in [10^{-5}, 10^{-1}]$ interval, but only converged points are shown. We have used $\omega = 10^{-3}$ and the *GP10* model for all simulations.

of ω , slightly fewer C_{HF} evaluations are required, yet the phase field also becomes unstable, failing to converge for some cases, or requiring significantly more Newton–Raphson iterations. As the impact of this double-well term appears small, we conservatively set $\omega = 10^{-3}$ for the remainder of this work.

4.3.4 | Phase-Field Interface Width

The phase-field parameter ϵ influences the width of the interface between the phases and, therefore, the width of the transition region between the constitutive models (the interface width is equal to approximately 4ϵ). To study the influence of the interface more carefully, we perform tests with a constant element size throughout the domain. In Figure 17, the influence of ϵ on the required number of Newton–Raphson iterations is shown for different element sizes. A few corresponding full-field plots of the phase-field value for different mesh sizes are shown in Figure 18. We can observe two main trends in these figures. Firstly, finding solutions for small ϵ values requires smaller mesh sizes. A minimal interface width is thus required to solve the phase field (but the wider the interface, the more high-fidelity models are evaluated in each iteration). As we decrease ϵ such that the mesh element size is smaller than 4ϵ , the interface width stays one element until the phase field fails to converge. Secondly, the number of Newton–Raphson iterations of the mechanical problem decreases slightly as we increase ϵ . When the phase field fails for some steps, but the overall simulation still converges with adaptive steps, it causes the spikes observed for element size 0.004. The general trend indicates that having a transition region spanning multiple elements improves the numerical stability of the mechanical problem.

The mesh used for the other notched plate experiments in this work has a variable mesh element size to capture strain localization with fine elements (0.005) while having coarse elements (up to 0.04) in the rest of the domain, as shown in Figures 11 and 12. When using such a variable mesh, a separate mesh with a different characteristic element size could be implemented for the phase field as in [28]. This brings additional challenges and is not considered here.

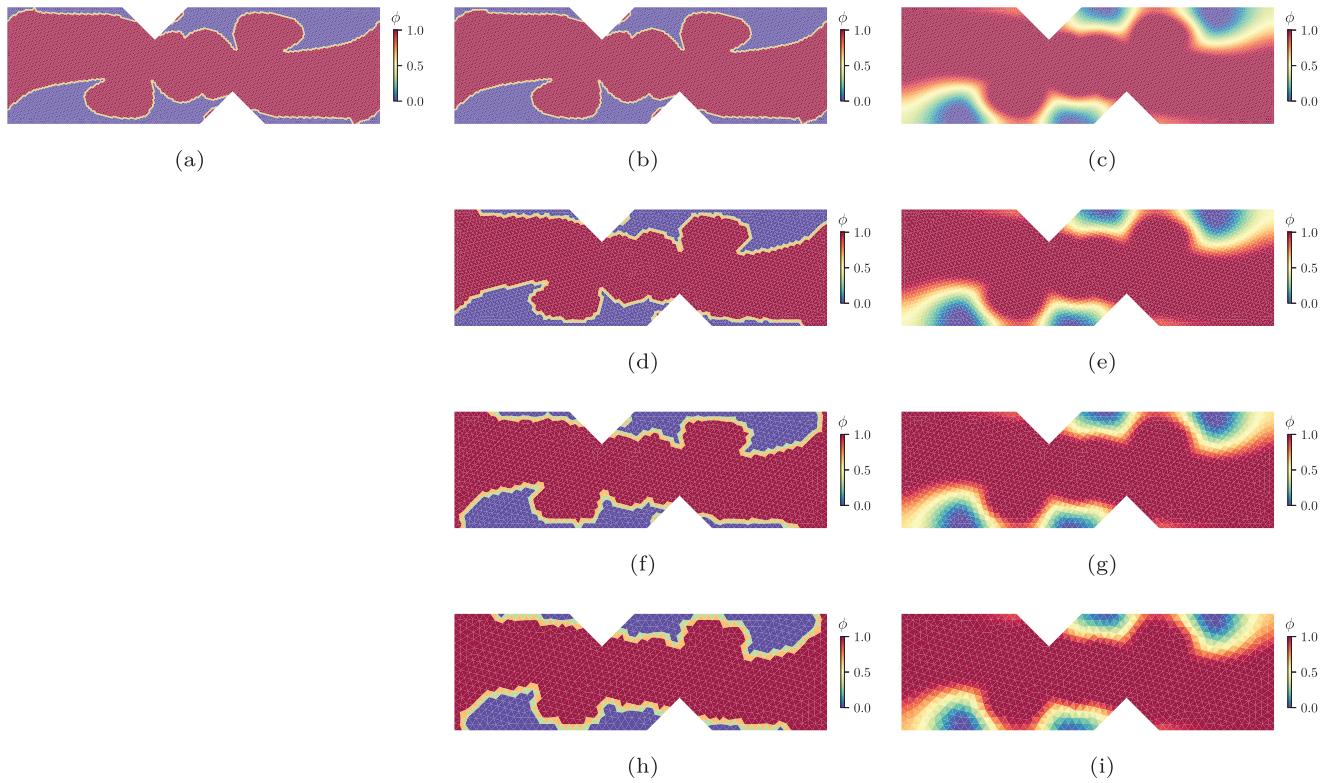


FIGURE 18 | The phase-field variable at the final time step of the notched plate problem for various element sizes and values of the phase-field interface parameter ϵ . (a) Element size: 0.004. $\epsilon = 0.00003$, (b) Element size: 0.004. $\epsilon = 0.0001$, (c) Element size: 0.004. $\epsilon = 0.03$, (d) Element size: 0.008. $\epsilon = 0.0001$, (e) Element size: 0.008. $\epsilon = 0.03$, (f) Element size: 0.012. $\epsilon = 0.0001$, (g) Element size: 0.012. $\epsilon = 0.03$, (h) Element size: 0.016. $\epsilon = 0.0001$, and (i) Element size: 0.016. $\epsilon = 0.03$.

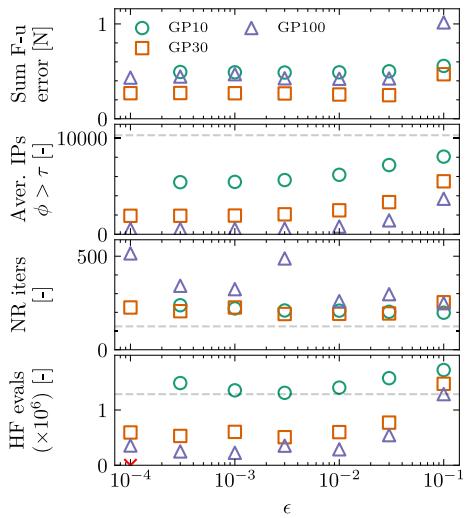


FIGURE 19 | The influence of the phase-field interface parameter ϵ on the mechanical and phase-field problem.

Figure 19 shows the influence of ϵ for the various GP surrogates on the domain with a variable mesh size. As ϵ decreases, the interface becomes narrower, leading to fewer elements where $\phi > \tau$. For $\epsilon = 10^{-1}$ the error increases, especially for $GP100$. This is because when there is a region with only a few points with high uncertainty, a large ϵ can cause the phase field to not fully switch, hindering accuracy. The difference in accuracy is generally small, and the number of high-fidelity evaluations reduces for smaller ϵ values. Still, as observed earlier, when ϵ is small, the mechanical and phase-field problems become less stable and require more NR iterations or do not converge. Therefore, the default value we have used so far of $\epsilon = 10^{-2}$ appears to be a sensible choice.

4.4 | Local Mixture

To avoid being limited by the convergence of the phase field for small ϵ values, we compare the phase field to a simple approach where no phase-field model is solved. In previous demonstrations, phase-field spatial smoothing, a non-local effect, was necessary for convergence. To better study the influence of the transition zone without being limited by having to solve the phase-field PDE, we consider an alternative case with no spatial smoothing. Here, we directly use the driving force to evolve the mixture. We refer to this as the local approach since the mixture of models in each IP is independent of the uncertainty in other IPs. In addition to removing the transition zone, another benefit is that we no longer need to solve the phase field, reducing the complexity of the method. We still use a similar formulation to the phase field, with b opposing the uncertainty. We consider two variations of the local approach. The first version linearly shifts between the models:

$$\phi = \mathcal{U} - b, \quad \phi \in [0, 1]. \quad (8)$$

In contrast to the phase field, this local shift does not necessarily lead to a spatial transition zone. When two points close in space have very different \mathcal{U} values, the spatial transition can still be sharp. However, the linear shift ensures that a small change in strain does not cause a completely different constitutive model to be used—the stress remains continuous with respect to the strain. Alternatively, we can use a step function, directly switching between the models when the uncertainty reaches the threshold:

$$\phi = \begin{cases} 0 & \text{if } \mathcal{U} < b, \\ 1 & \text{if } \mathcal{U} \geq b. \end{cases} \quad (9)$$

Note that both versions require the high-fidelity model when $\mathcal{U} \geq b$. Similar to the phase-field approach, this method suffers from time-step dependency because the mixture uses the uncertainty of the previous step. Therefore, we also perform several iterations of the staggered approach, with $k_{\max} = 3$.

In Figure 20, we show how b influences the results for these two local approaches compared to the phase-field approach. Across all simulations, the errors are nearly identical between the models. Because the local methods do not have an interface, they always have fewer IPs that require high-fidelity model evaluations for a given uncertainty. The number of Newton–Raphson iterations is similar when b is low. However, for larger b values, the local approaches require significantly more iterations with *GP30* and *GP100*. In addition, the local approaches start failing for these models as b increases. Removing the interface can thus be beneficial in reducing the number of high-fidelity evaluations for specific settings, yet generally makes the model less stable and less robust.

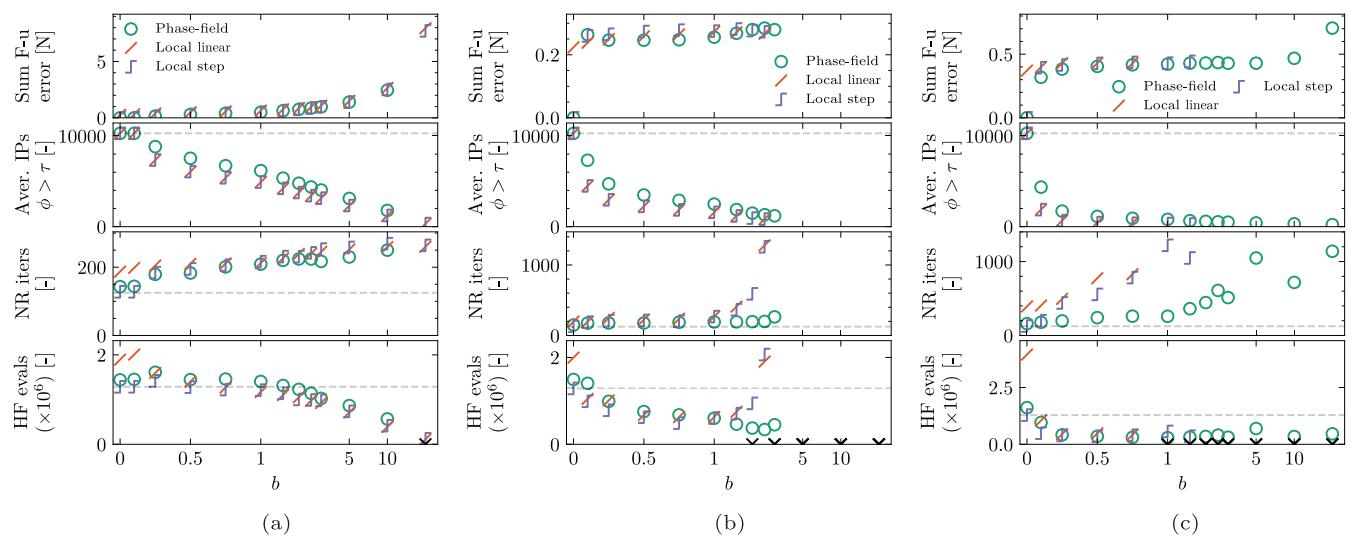


FIGURE 20 | Comparison of the opposing force parameter b for the various approaches. For the phase field we set $\epsilon = 10^{-2}$ and $\omega = 10^{-3}$. Note that the x-axis is linear from zero to one, and then scales logarithmically. All simulations that fail to converge fail due to the mechanical problem. (a) GP 10, (b) GP 30, and (c) GP 100.

4.5 | Surrogate Data

So far, we have shown the influence of the various parameters on the behavior of the mixture of models. We now study how the training dataset size affects the reduction in high-fidelity samples in more detail. To investigate whether this framework can handle complex phase fields, we consider a plate with holes to create numerous local regions with high stresses that challenge the GP surrogates and induce a switch. We constrain the left side of the domain and incrementally prescribe a displacement to the right side of the domain. We use $\epsilon = 10^{-2}$, $\omega = 10^{-3}$, and $b = 1$ for the phase field, and do not consider the local approach here.

Figure 21 shows the number of high-fidelity evaluations during the simulation as we vary the amount of training data in the GP surrogate. For three of these cases, we plot the full-field phase field at the end of the simulation in Figure 22. These figures show the number of high-fidelity evaluations quickly decreasing as the GP is trained with more training data. Still, at a certain point, the downward trend appears to converge, and the number of high-fidelity evaluations is not further reduced. This is because the GP fails to capture localization with strains outside its training data range, which go up to a strain norm of 10%, a choice that must be made prior to running the simulation. This shows the difficulty of creating an appropriate dataset in advance. Different techniques for generating datasets are appropriate for different surrogates. Generally, the dataset should be generated by sampling the input (strain) space such that it covers the expected inputs that occur in the simulation of interest. For path-dependent materials, the input space should also include the internal variables, which is commonly achieved by sampling random, non-monotonic load paths. A high-quality dataset leads to a high-quality surrogate, increasing the efficiency of the approach. However, adequately covering the input space a priori is difficult, which is why our proposed method is advantageous.

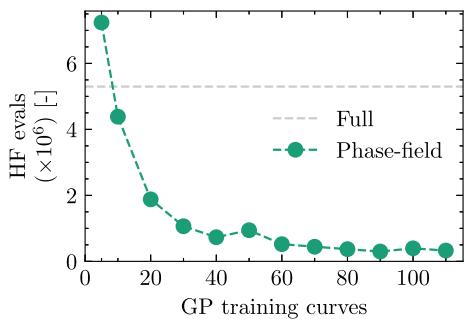


FIGURE 21 | GP training dataset size influence on the number of high-fidelity evaluations. The plot shows the average of 5 runs with different datasets. The full model is used as a reference. For very small datasets, our approach can use more evaluations than the full model due to the staggered iterations ($k_{\max} = 3$).

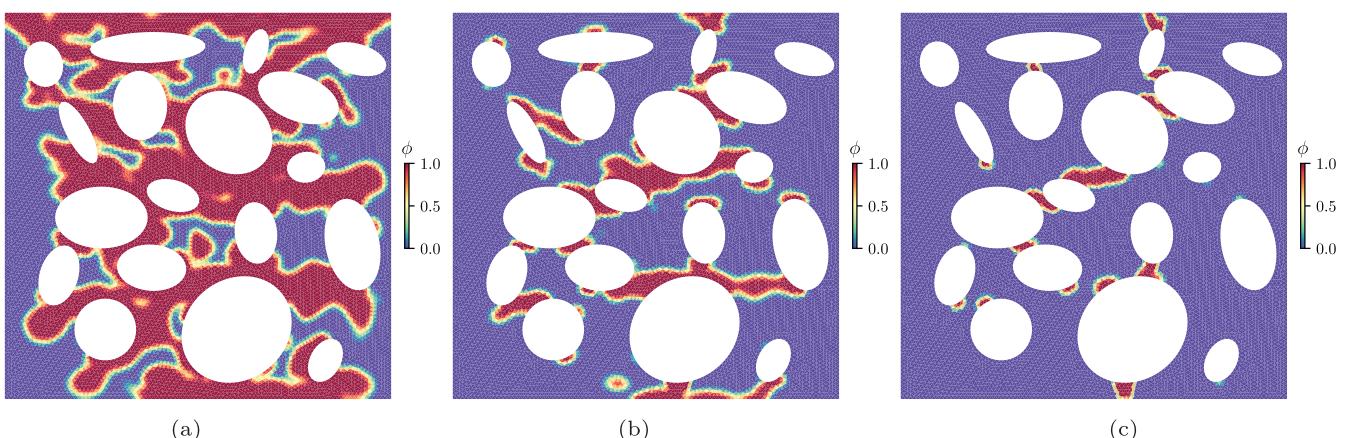


FIGURE 22 | Phase-field variable ϕ at the maximum prescribed displacement for surrogates with varying amounts of training data. (a) GP10, (b) GP30, and (c) GP100.

5 | Conclusion

This work introduces an approach for mixing constitutive models in finite element analysis and tests its performance by combining a physics-based constitutive model with a data-driven surrogate counterpart. The method dynamically determines where to use each model based on the surrogate uncertainty, creating a spatially varying mixture that preserves accuracy while reducing computational cost. This is achieved by solving a phase field based on the surrogate model uncertainty that determines which constitutive model is used. Then, given the mixture of constitutive models, the mechanical problem of interest is solved. Solving the problem in a staggered way introduces time-step dependence, which can be resolved by performing multiple staggered iterations per load step. Running several iterations brings additional costs but is necessary for the robustness of the method.

For computational feasibility, the results have been obtained for single-scale problems, where an analytical constitutive model acts as the high-fidelity model. A Gaussian Process surrogate is used, as it naturally provides an estimate of its uncertainty. Because this surrogate cannot capture history-dependent effects, the experiments consider only monotonic loading scenarios. However, our approach is model-agnostic and does not depend on the specific surrogate or high-fidelity model, but on the accuracy of the surrogate uncertainty, and therefore the conclusions hold for more general multiscale models and history-dependent surrogates. Future work should compare the approach for other material behaviors and loading scenarios, such as cyclic loading.

The use of a phase field was motivated by the hypothesis that the spatial smoothness would improve the numerical stability of the mechanical problem. To test this, we studied the effect of decreasing the size of the interface. As there is a limit to the minimum interface width for the phase field itself, we compared the phase-field approach to an alternative local approach with no spatial smoothing. There, we used the surrogate model if its uncertainty is below a threshold, and the high-fidelity model otherwise, creating a sharp transition between the models. We demonstrated on a notched plate problem that while the local approach can work and reduce the high-fidelity evaluations further, it is less robust than using a phase field. For this reason, the spatial smoothness introduced by the phase field is beneficial.

The results demonstrate that the mixing approach significantly reduces the number of high-fidelity model evaluations compared to the full model. The quality of the surrogate model directly impacts performance—better-trained surrogate models with more data allow for increased computational savings. If a surrogate model can be curated such that it accurately captures the behavior during the full simulation, then the uncertainty remains low, and no additional high-fidelity model evaluations are required. This framework then adds the extra cost of computing the surrogate uncertainty. However, creating an all-encompassing surrogate model is very challenging, and it is generally difficult to know in advance whether the surrogate model is sufficient. Therefore, this mixture of models uses the surrogate where possible while still enabling an accurate simulation. If the surrogate is poorly trained and can only capture a small part of the behavior, the additional Newton–Raphson iterations to obtain a consistent solution can result in more high-fidelity model evaluations than simply using the high-fidelity model without this approach. It is, however, unlikely that this would be the case in practice. This work thus provides a framework for accelerating multiscale simulations where micromodel evaluations are computationally expensive. The approach works independently of the specific surrogate model and can, therefore, be combined with the latest advances in developing surrogate models to enable multiscale simulations.

Acknowledgments

This work is supported by the TU Delft AI Labs programme. We gratefully acknowledge the TU Delft AI Lab Travel grant for enabling this work.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in GitHub at: <https://github.com/SLIMM-Lab/phase-field-mixture>, reference number [10.5281/zenodo.15267978](https://zenodo.org/record/15267978).

References

1. X. W. J. Ghaboussi and J. H. Garrett, Jr., "Knowledge-Based Modeling of Material Behavior With Neural Networks," *Journal of Engineering Mechanics* 117, no. 1 (1991): 132–153.
2. F. Ghavamian and A. Simone, "Accelerating Multiscale Finite Element Simulations of History-Dependent Materials Using a Recurrent Neural Network," *Computer Methods in Applied Mechanics and Engineering* 357 (2019): 112594.
3. M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations," *Journal of Computational Physics* 378 (2019): 686–707.
4. L. Lu, P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis, "Learning Nonlinear Operators via Deepenet Based on the Universal Approximation Theorem of Operators," *Nature Machine Intelligence* 3, no. 3 (2021): 218–229.
5. N. N. Vlassis and W. Sun, "Geometric Learning for Computational Mechanics Part II: Graph Embedding for Interpretable Multiscale Plasticity," *Computer Methods in Applied Mechanics and Engineering* 404 (2023): 115768.
6. J. N. Fuhr, G. A. Padmanabha, N. Bouklas, et al., "A Review on Data-Driven Constitutive Laws for Solids," *Archives of Computational Methods in Engineering* 32, no. 3 (2024): 1841–1883.
7. X. Liu, S. Tian, F. Tao, and W. Yu, "A Review of Artificial Neural Networks in the Constitutive Modeling of Composite Materials," *Composites Part B: Engineering* 224 (2021): 109152.
8. M. Maia, I. Rocha, P. Kerfriden, and F. van der Meer, "Physically Recurrent Neural Networks for Path-Dependent Heterogeneous Materials: Embedding Constitutive Models in a Data-Driven Surrogate," arXiv Preprint arXiv:2209.07320 (2022).
9. D. K. Klein, R. Ortigosa, J. Martínez-Frutos, and O. Weeger, "Nonlinear Electro-Elastic Finite Element Analysis With Neural Network Constitutive Models," arXiv Preprint arXiv:2402.07007 (2024).
10. J. Dornheim, L. Morand, H. J. Nallani, and D. Helm, "Neural Networks for Constitutive Modeling: From Universal Function Approximators to Advanced Models and the Integration of Physics," *Archives of Computational Methods in Engineering* 31, no. 2 (2024): 1097–1127.
11. I. Rocha, P. Kerfriden, and F. van der Meer, "On-The-Fly Construction of Surrogate Constitutive Models for Concurrent Multiscale Mechanical Analysis Through Probabilistic Machine Learning," *Journal of Computational Physics X* 9 (2021): 100083.
12. K. A. Kalina, L. Linden, J. Brummund, and M. Kästner, "Fe ANN: An Efficient Data-Driven Multiscale Approach Based on Physics-Constrained Neural Networks and Automated Data Mining," *Computational Mechanics* 71, no. 5 (2023): 827–851.
13. M.-J. Li, Y. Lian, Z. Cheng, et al., "A Clustering Adaptive Gaussian Process Regression Method: Response Patterns Based Real-Time Prediction for Nonlinear Solid Mechanics Problems," *Computer Methods in Applied Mechanics and Engineering* 436 (2025): 117669.
14. F. Fritzen, M. Fernández, and F. Larsson, "On-The-Fly Adaptivity for Nonlinear Twoscale Simulations Using Artificial Neural Networks and Reduced Order Modeling," *Frontiers in Materials* 6 (2019): 75.
15. I. B. Rocha, F. P. van der Meer, L. A. Mororó, and L. J. Sluys, "Accelerating Crack Growth Simulations Through Adaptive Model Order Reduction," *International Journal for Numerical Methods in Engineering* 121, no. 10 (2020): 2147–2173.
16. R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "Adaptive Mixtures of Local Experts," *Neural Computation* 3, no. 1 (1991): 79–87.
17. P. K. Rajaraman, T. Manteuffel, M. Belohlavek, and J. J. Heys, "Combining Existing Numerical Models With Data Assimilation Using Weighted Least-Squares Finite Element Methods," *International Journal for Numerical Methods in Biomedical Engineering* 33, no. 1 (2017): e02783.
18. W. J. Boettinger, J. A. Warren, C. Beckermann, and A. Karma, "Phase-Field Simulation of Solidification," *Annual Review of Materials Research* 32, no. 1 (2002): 163–194.
19. L.-Q. Chen, "Phase-Field Models for Microstructure Evolution," *Annual Review of Materials Research* 32, no. 1 (2002): 113–140.
20. J.-Y. Wu, V. P. Nguyen, C. T. Nguyen, D. Sutula, S. Sinaie, and S. P. Bordas, "Phase-Field Modeling of Fracture," *Advances in Applied Mechanics* 53 (2020): 1–183.
21. T. Gerasimov and L. De Lorenzis, "A Line Search Assisted Monolithic Approach for Phase-Field Computing of Brittle Fracture," *Computer Methods in Applied Mechanics and Engineering* 312 (2016): 276–303.
22. C. Luo, "Fast Staggered Schemes for the Phase-Field Model of Brittle Fracture Based on the Fixed-Stress Concept," *Computer Methods in Applied Mechanics and Engineering* 404 (2023): 115787.
23. Y. Gal and Z. Ghahramani, "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning," in *Proceedings of the International Conference on Machine Learning* (PMLR, 2016), 1050–1059.
24. I. Rocha, "Numerical and Experimental Investigation of Hygrothermal Aging in Laminated Composites," (2019).
25. I. A. Baratta, J. P. Dean, J. S. Dokken, et al., "DOLFINx: The Next Generation FEniCS Problem Solving Environment," Preprint (2023), <https://doi.org/10.5281/zenodo.1044766>.

26. J. Bleyer, “Dolfinx_Materials: A Python Package for Advanced Material Modelling,” (2024), <https://doi.org/10.5281/zenodo.13882183>.

27. S. Balay, S. Abhyankar, M. F. Adams, et al., “PETSc Web Page,” (2024), <https://petsc.org/>.

28. S. Goswami, C. Anitescu, and T. Rabczuk, “Adaptive Phase Field Analysis With Dual Hierarchical Meshes for Brittle Fracture,” *Engineering Fracture Mechanics* 218 (2019): 106608.

29. C. E. Rasmussen, “Gaussian Processes in Machine Learning,” in *Summer school on machine learning* (Springer, 2003), 63–71.

Appendix A

Gaussian Process Regression

A comprehensive treatment of GPs can be found in [29]. A Gaussian Process is a collection of random variables, any finite number of which follow a joint Gaussian distribution. GPs are specified by a mean $m(\mathbf{x})$ and a covariance function $k(\mathbf{x}, \mathbf{x}')$, written as:

$$f(\mathbf{x}) \sim \text{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')). \quad (\text{A1})$$

For simplicity, it is common to assume a zero mean function, $m(\mathbf{x}) = 0$. The covariance function $k(\mathbf{x}, \mathbf{x}')$ specifies the correlation between outputs at different inputs, often capturing how similar inputs should yield similar outputs. In this work, we use a squared exponential (radial basis) kernel function:

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\ell^2}\right), \quad (\text{A2})$$

where σ_f^2 is the signal variance, and ℓ is the characteristic length scale.

Given a training dataset $\mathcal{D} = \{(\mathbf{x}_i, y_i) | i = 1, \dots, n\}$ with inputs $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]^T$ and outputs $\mathbf{y} = [y_1, \dots, y_n]^T$, we aim to predict the output $f(\mathbf{x}_*)$ at a new test point \mathbf{x}_* . Assuming noisy observations $y_i = f(\mathbf{x}_i) + \epsilon_i$ where $\epsilon_i \sim \mathcal{N}(0, \sigma_n^2)$, the joint distribution of the training outputs and the test output becomes:

$$\begin{bmatrix} \mathbf{y} \\ f(\mathbf{x}_*) \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I} & K(\mathbf{X}, \mathbf{x}_*) \\ K(\mathbf{x}_*, \mathbf{X}) & K(\mathbf{x}_*, \mathbf{x}_*) \end{bmatrix}\right), \quad (\text{A3})$$

where $K(\mathbf{X}, \mathbf{X})$ denotes the $n \times n$ matrix of covariances evaluated at all pairs of training points, $K(\mathbf{X}, \mathbf{x}_*)$ is the $n \times 1$ vector of covariances between training points and the test point, and $K(\mathbf{x}_*, \mathbf{x}_*)$ is the covariance of the test point with itself. The posterior distribution of $f(\mathbf{x}_*)$ given the observations becomes:

$$p(f(\mathbf{x}_*) | \mathbf{X}, \mathbf{y}, \mathbf{x}_*) = \mathcal{N}(\mu_*, \sigma_*^2) \quad (\text{A4})$$

$$\mu_* = K(\mathbf{x}_*, \mathbf{X})[K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I}]^{-1} \mathbf{y} \quad (\text{A5})$$

$$\sigma_*^2 = K(\mathbf{x}_*, \mathbf{x}_*) - K(\mathbf{x}_*, \mathbf{X})[K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I}]^{-1} K(\mathbf{X}, \mathbf{x}_*) \quad (\text{A6})$$

The predictive mean μ_* serves as our point estimate, while the predictive variance σ_*^2 quantifies our uncertainty about the estimate. This uncertainty naturally increases as we move away from the training data.

There are several methods for optimizing the hyperparameters of a GP. We use the maximum likelihood estimation approach, which finds the hyperparameters that maximize the marginal likelihood of the observed data. The hyperparameters θ (i.e., $\theta = \{\sigma_f, \ell, \sigma_n\}$) are optimized by maximizing the log marginal likelihood:

$$\log p(\mathbf{y} | \mathbf{X}, \theta) = -\frac{1}{2} \mathbf{y}^T [K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I}]^{-1} \mathbf{y} - \frac{1}{2} \log |K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I}| - \frac{n}{2} \log(2\pi) \quad (\text{A7})$$

We perform this optimization using the L-BFGS-B algorithm, starting from 20 different initial configurations to avoid local optima.