Computers and Structures 321 (2026) 108103

Contents lists available at ScienceDirect

Computers and Structures

journal homepage: www.elsevier.com/locate/cas

Model calibration and uncertainty quantification of cellular automata-based pitting corrosion model☆

J. Ramesh Babu*, Pranav M. Karve, Sankaran Mahadevan

Department of Civil and Environmental Engineering, Vanderbilt University, Nashville, TN 37235, USA

HIGHLIGHTS

- Developed a cellular automata-based uncertainty quantification framework for corrosion.
- Developed a hierarchical surrogate modeling approach for pit evolution.
- Sensitivity analysis identified sedimentation and ion diffusion as key drivers.
- Bayesian calibration using both pit depth and aspect ratio improved model reliability.

ARTICLE INFO

Keywords:
Pitting corrosion
Cellular automata
Surrogate modeling
Bayesian inference
Uncertainty quantification
Model validation

ABSTRACT

This study presents a Bayesian framework for calibrating cellular automata (CA) models of pitting corrosion. A high-fidelity two-dimensional CA model is used to simulate pit morphology evolution under coupled electrochemical and mass transport processes, incorporating eight uncertain model parameters related to metal dissolution, hydrolysis, diffusion, and sedimentation. Key geometric indicators—maximum pit depth and pit aspect ratio—are extracted from the simulation outputs and reduced to interpretable scalar features using power-law fitting. Correlation analysis is conducted to assess linear relationships between model parameters and power-law coefficients. Gaussian Process Regression (GPR) surrogate models are constructed using Latin Hypercube Sampling (LHS) to emulate the high-fidelity CA model. Global sensitivity analysis (GSA) using Sobol' indices is performed by utilizing the trained surrogate models; it identifies sedimentation and ionic diffusion as the dominant contributors to output feature variability, with significant nonlinear interactions. The trained surrogates are integrated into a Bayesian inference framework using Metropolis-Hastings Markov Chain Monte Carlo (MH-MCMC) sampling to infer posterior distributions of the uncertain model parameters. The posterior samples are further propagated through the surrogates to quantify the output uncertainty, and the prediction is evaluated using a distance-based probabilistic model validation metric. Two numerical examples related to API-5L X65 steel pipelines are presented, demonstrating that incorporating multiple geometric features—both pit depth and aspect ratio—improves predictive accuracy. The proposed framework supports uncertainty-aware modeling and decision-making for structural health assessment and maintenance planning in corrosion-critical infrastructure.

1. Introduction

Pitting corrosion is a serious concern for metals used in aggressive environments, due to its localized and destructive nature. This form of corrosion leads to the formation of small cavities or pits on a metal surface, which can propagate over time, eventually weakening the underlying material. Unlike uniform corrosion which affects an entire area of interest, pitting corrosion can occur in isolated spots, making it difficult to detect and predict. The initiation of pitting corrosion is primarily attributed to the breakdown of the passive oxide film that protects metals from corrosion. Once this protective layer is compromised, localized electrochemical reactions commence, leading to anodic and cathodic regions within the pits.

☆ This document is the result of research that was partly funded by the US Department of Energy, Idaho National Laboratory.
* Corresponding author.
Email addresses: jay.jangala@ou.edu (J.R. Babu), pranav.m.karve@vanderbilt.edu (P.M. Karve), sankaran.mahadevan@vanderbilt.edu (S. Mahadevan).

https://doi.org/10.1016/j.compstruc.2026.108103
Received 27 August 2025; Accepted 3 January 2026
Available online 20 January 2026

The presence of aggressive ions, particularly chloride ions, plays a crucial role in exacerbating pitting corrosion. Chloride ions can penetrate the protective oxide layer and induce localized dissolution of the metal, which further enhances the development of pits. This process is particularly evident in marine environments where deicing salts are used, as these conditions can drastically increase the susceptibility of metals to localized corrosion. Stress corrosion cracking (SCC) can also be a concerning consequence of pitting, as the pits can act as stress concentrators under mechanical loads, leading to premature failure of structural components.

Moreover, localized corrosion is also influenced by other environmental and mechanical factors that can aggravate the degradation process. Factors such as temperature, pH variations, and microstructural differences within materials play significant roles in how pitting develops. Notably, the local pH can fluctuate significantly during the corrosion process, potentially reaching levels that favor further metal dissolution. This phenomenon has been observed during the corrosive challenges faced by various steel alloys, particularly in saline or acidic environments [29,48]. For instance, pitting corrosion dynamics are often linked with microstructural characteristics such as grain boundaries and alloy composition, which can dictate susceptibility to various forms of localized corrosion [20,36].

One of the complexities of pitting corrosion lies in the stochastic nature of pit formation and growth. Each pit develops randomly, varying in shape, size, depth, and location, often creating a rugged profile that complicates modeling and understanding the corrosion process [44]. Traditional prediction models tend to overlook the irregularity of pitting, leading to challenges in accurately assessing corrosion risk and creating effective prevention strategies. Current research is pursuing novel computational models and advanced detection techniques to better predict the onset and progression of pitting corrosion [12,58].

Various modeling strategies have been developed to study pitting corrosion, ranging from empirical [5,43,44,65,67] and phenomenological approaches [25,57] to advanced computational methods [3,4,10,15,38,62,64] such as finite element methods (FEM) [28,56,70], extended finite element methods(XFEM)-level set methods (LSM) [15,64], finite volume methods (FVM) [52,53] cellular automata (CA) [4,11,16,17,32,39,61], phase-field (PF) [7,37,38], and peridynamics (PD) [3,10,72]. These models differ in their ability to capture physical fidelity, computational efficiency, and the complexity of corrosion front evolution.

Among these computational frameworks, CA, PF, and PD are considered autonomous models as they have their own internal rules to propagate the corrosion front, without any external input. Among these autonomous modeling frameworks, CA, PF, and PD offer distinct trade-offs. CA models are computationally efficient as they do not require solving partial differential equations (PDEs), instead relying on discrete probabilistic rules to simulate pit nucleation and growth. This makes CA particularly suitable for simulations involving multiple pits and stochastic evolution. PF models capture interface dynamics through coupled PDEs and introduce a characteristic length scale via diffuse interface thickness, allowing high morphological fidelity but at increased computational cost. PD models use a nonlocal formulation to simulate long-range material interactions, making them suitable for coupled corrosion-mechanical analysis but computationally intensive. While PF and PD are effective for resolving detailed interface evolution and stress-damage coupling, they involve significant numerical complexity.

In contrast, CA offers a favorable balance between efficiency and morphological flexibility. It can reproduce a variety of pit shapes consistent with experimental observations, but its time and spatial units are not physically based—they are governed by model parameters that must be calibrated. This dependence on parameter tuning means CA models require rigorous validation to ensure accurate representation of pit growth behavior. Several CA models have been developed to simulate pitting corrosion under various conditions, each employing different calibration techniques. Lishchuk et al. [32] proposed a CA model for intergranular corrosion, capturing preferential propagation along grain boundaries. The model was calibrated using a differential evolution algorithm, optimizing grain, boundary, and surface layer corrosion probabilities by minimizing the mean squared error (MSE) between simulated and experimental data, ensuring agreement with the predicted time-dependent corrosion path length, which refers to the distance intergranular corrosion propagates along grain boundaries over time.

Stafiej et al. [61] developed a CA model to simulate corrosion–passivation processes, capturing the spatial separation of anodic and cathodic reactions that drive cavity formation, and calibrated it through qualitative tuning, by manually adjusting reaction probabilities, diffusion step sizes, and passivation breakdown criteria to replicate known corrosion behaviors. Fatoba et al. [17] proposed a cellular automata finite element (CAFE) model to simulate stress-assisted localized corrosion, integrating CA for corrosion evolution and finite element analysis (FEA) for mechanical stress effects, and optimized its parameters using a grid search with a Sobol sequence, by minimizing MSE between simulated and experimental pit depths under different environmental conditions. Cui et al. [11] introduced a 3D CA model to simulate the initiation and growth of corrosion pits on Q345 steel under a salt-spray environment, capturing the randomness and time-dependent evolution of pitting corrosion. The model was calibrated by tuning corrosion reaction, passivation, and downward movement probabilities, ensuring alignment between simulated pit growth trends and experimental salt-spray corrosion test data.

Fan et al. [16] proposed a CA model to predict structural steel corrosion under atmospheric conditions, calibrating it using a Genetic Algorithm (GA) to optimize key parameters by minimizing errors between simulated and experimental corrosion depths. Babu et al. [4] developed a CA model coupled with XFEM to simulate fatigue-assisted pitting corrosion in pipeline steel, capturing pit propagation, pit-to-crack transition, and crack growth until fracture under cyclic loading. The model was calibrated using experimental data, where CA parameters governing pit growth were adopted as in Fatoba et al. [17].

These traditional calibration techniques, such as grid search [17], genetic algorithms [16], and differential evolution [32], have been widely employed to optimize CA model parameters for simulating pitting and intergranular corrosion. However, these deterministic calibration methods present several limitations. They provide only point estimates of parameters without incorporating uncertainty quantification (UQ), making them less robust when applied to varying environmental conditions. Moreover, they require extensive computational resources due to the high-dimensional parameter space needed to optimize the model parameters. Another drawback is their inability to incorporate prior knowledge or systematically refine parameters based on new experimental data, limiting their adaptability to different corrosion scenarios. Furthermore, most existing CA models focus primarily on pit depth while overlooking other geometric features, such as aspect ratio, which are important for evaluating pit morphology and assessing structural degradation.

While CA models are commonly implemented in two dimensions for pitting corrosion—for example, the 2D CA frameworks used by Fatoba et al. [17] and Babu et al. [4]—it is important to note that pitting corrosion is inherently a three-dimensional process. Several fully 3D models exist in the literature, including those by Van der Weeën et al. [66], Cui et al. [11], and the studies reviewed in Larrosa et al. [27], all of which capture more realistic pit shapes and growth behavior. Moving to a 3D representation would allow pit shapes, surface effects, and growth patterns to be modeled more realistically, reduce biases introduced by 2D shape measures, and provide a more complete and accurate description of pitting corrosion, as also highlighted in pit-shape studies such as Gnefid and Akid [19], although with greater computational cost.

Given these limitations and the stochastic nature of pitting corrosion, a Bayesian framework is adopted in this work for calibrating CA model parameters in Babu et al. [4] by integrating prior knowledge with experimental observations. Recent advances in uncertainty quantification and model calibration emphasize the importance of propagating

both parameter and model-form uncertainty when predicting complex material degradation processes. In particular, modern UQ frameworks integrate surrogate modeling, Bayesian inference, and sensitivity analysis to improve predictive robustness across engineering applications. Examples include UQ of architected materials for energy absorption [34] and physics-informed machine-learning-assisted UQ for corrosion processes [6]. These studies highlight the need for UQ-focused calibration approaches capable of handling stochastic, nonlinear, and data-limited systems such as pitting corrosion.

The calibration process in this study is carried out by simultaneously comparing both the maximum pit depth and maximum aspect ratio against experimental measurements reported by Fatoba et al. [17]. Such a calibration approach considering both features is absent in previous models, such as Fatoba et al. [17] and Babu et al. [4], which considered only pit depth for calibration without accounting for other geometric features like aspect ratio. Maximum pit depth represents the vertical penetration of corrosion, indicating the extent of material loss, but does not fully characterize pit morphology or lateral expansion. The aspect ratio, defined as the ratio of pit depth to half of the pit width, accounts for both depth and width variations that are critical in determining the progression of localized corrosion damage. Model calibration considering both pit depth and pit width enhances the CA model's reliability in predicting pitting corrosion by ensuring accurate reproduction of both pit depth progression and shape evolution.

The motivation for developing this framework is to enable uncertainty-aware prediction of pitting corrosion for engineering decision-making. In practical applications—such as pipeline integrity management, offshore structures, or industrial vessels—engineers require probabilistic predictions of pit depth and morphology to assess remaining life, schedule inspections, and plan maintenance activities. Because CA models are stochastic and contain several uncertain kinetic and transport parameters, a calibration framework capable of updating these parameters with new data is essential. The proposed Bayesian-surrogate framework provides this capability: it allows the model to be updated during operation when inspection or monitoring data become available, and it delivers calibrated, uncertainty-quantified predictions suitable for integration into structural reliability assessments and maintenance planning workflows.

The Bayesian uncertainty quantification framework developed here incorporates several advanced features to improve the accuracy and efficiency of simulating pitting corrosion. (i) Correlation analysis (typically linear, e.g., Pearson correlation) is first used to identify relationships between different corrosion parameters. (ii) This is followed by global sensitivity analysis to determine the most influential factors affecting corrosion, such as sedimentation and ion diffusion. (iii) Gaussian process regression surrogates are then trained using pit depth and aspect ratio data to significantly reduce the computational cost of the full simulations. (iv) Based on the results of the correlation and sensitivity analyses, the input space is reduced specifically for the Bayesian calibration step, ensuring that only the most impactful parameters are considered. (v) Combined with multi-feature calibration, this framework enhances the overall accuracy and reliability of the cellular automata model in capturing pitting corrosion behavior.

The remainder of the paper is organized as follows. Section 2 reviews the modeling methodology, beginning with the CA framework for localized corrosion, developed following the approaches described in Fatoba et al. [17] and Babu et al. [4]. The framework includes a detailed representation of the electrochemical system, rules governing corrosion mechanisms, and CA-based treatment of ionic and precipitate diffusion, along with an outline of the algorithmic implementation. Section 2.3 introduces a reduced-dimension surrogate modeling framework, where time-series outputs of geometric features such as maximum pit depth and maximum aspect ratio are transformed into low-dimensional quantities via power-law fitting. Section 2.1 describes correlation analysis for input screening, and Section 2.3.1 describes the construction of Gaussian Process Regression (GPR) models for efficient simulation of CA outputs. Section 2.2 describes the global sensitivity analysis (GSA) based on Sobol' indices—both first-order and total effects—to identify key parameters that influence the variability of model predictions. Section 2.4 presents a Bayesian inference framework for model calibration, where experimental observations are used to obtain the posterior distributions of uncertain model parameters. In addition, a probabilistic model validation metric is introduced in Section 2.5 to assess the predictive accuracy of the calibrated models by quantifying the probability that model outputs fall within an acceptable tolerance of experimental measurements over time. Section 3 presents the results and discussion, using two numerical examples that assess the framework's accuracy: one based solely on pit depth features and another incorporating both depth and aspect ratio. These examples demonstrate the added value of geometric detail in enhancing predictive accuracy and UQ. Finally, Section 4 summarizes the key findings and outlines recommendations for future research in UQ of corrosion modeling.

Overall, the intended use of the proposed framework is to serve as a predictive and updating tool for existing structures in service. By combining physics-based CA modeling with hierarchical surrogates and Bayesian inference, the framework enables life-cycle evaluation of corrosion progression, supports risk-informed maintenance decisions, and can be updated whenever new observations are collected. This clarifies the engineering utility of the approach and its relevance to structural integrity management.

2. Methodology

This section presents the models and analysis tools used in this study, including the CA model for simulating pitting corrosion and a Bayesian framework for model calibration, uncertainty quantification, and propagation. The proposed methodology integrates surrogate model construction for computational efficiency, correlation analysis for initial screening of input–output relationships, GSA to quantify the relative influence of parameters, Bayesian inference to obtain updated (posterior) distributions of model parameters, and uncertainty propagation (UP) to evaluate the impact of input parameter uncertainties on model predictions over time. To further assess model performance, a reliability-based model validation metric is employed to quantify the probability that model predictions remain within an acceptable tolerance of experimental data across different time instances.

In this study, we employ a two-dimensional CA model adapted from Lishchuk et al. [32] and Fatoba et al. [17] and further developed in Babu et al. [4] to simulate pitting corrosion at the mesoscopic scale. The model captures the key electrochemical mechanisms driving pit evolution, including metal dissolution, ion hydrolysis, diffusion of species, and sedimentation of precipitates. The model parameters targeted for calibration in this study include those governing the electrochemical and diffusion processes that drive the evolution of corrosion pits. These parameters include the metal dissolution rate (P_{corr}), the oxidation rate (P_{oxd}), and the hydrolysis rates for Fe^{2+} and Fe^{3+} ions (P_{hyd_1}, P_{hyd_2}). Additionally, diffusion probabilities for ionic species ($P_{\text{diff}}^{\text{ion}}$) and corrosion products such as $Fe(OH)_2$ and $Fe(OH)_3$ ($P_{\text{diff}}^{\text{Fe(OH)}_2}$, $P_{\text{diff}}^{\text{Fe(OH)}_3}$) are calibrated, along with the sedimentation factor (`Sed`) that governs the downward movement of precipitates. These parameters are calibrated using experimental data to ensure the model accurately reflects corrosion behavior and pit development. Detailed implementation of the CA framework, including electrochemical reaction mechanisms, diffusion rules, and the algorithmic procedure, is provided in Babu et al. [4].

The uncertainty quantification framework built in this paper around the traditional CA-based modeling approach consists of the following elements:

- **Input screening and sensitivity analysis:** Correlation analysis is used for initial input screening to identify inputs that significantly influence the output. This is followed by global sensitivity analysis (GSA) using Sobol' indices to quantify the relative influence of model

parameters on pit depth and aspect ratio variability. These steps guide dimensionality reduction for efficient and targeted calibration.

- **Surrogate modeling:** To enable efficient uncertainty quantification and parameter sensitivity analyses, a surrogate model is needed to reduce the computational burden of repeatedly running the expensive cellular automata (CA) simulations—particularly within the Bayesian calibration framework, where Markov Chain Monte Carlo (MCMC) sampling requires thousands of model evaluations to infer posterior distributions of the most influential input parameters. Performing these runs using the high-fidelity CA model would be prohibitively expensive, making surrogate modeling essential. A reduced-dimension, hierarchical surrogate modeling strategy is implemented to alleviate the computational cost associated with repeated CA simulations. First, (i) time-series outputs of pit depth and aspect ratio are fit using power-law functions to parameterize the output. Then, (ii) Gaussian Process Regression (GPR) surrogate models are constructed to relate the fitting coefficients of the power-law functions to the CA model parameters, enabling fast and accurate approximation of the CA model outputs.
- **Bayesian calibration:** Bayesian inference is applied to calibrate uncertain CA model parameters using both maximum pit depth and maximum aspect ratio as calibration targets. This multi-feature calibration strategy enhances the model's ability to capture both vertical and lateral aspects of pit morphology, thereby improving predictive reliability over prior studies that used only depth-based metrics.
- **Uncertainty propagation:** Calibrated posterior distributions of CA model parameters are used to predict the corresponding distributions of power law coefficients using the GPR surrogate models. The predicted distributions of these power law coefficients are sampled and propagated through the power law relationships, thus producing probabilistic estimates of pit depth and aspect ratio over time.
- **Model validation:** The predictive performance of the calibrated CA model is assessed using a probabilistic model validation metric by comparing the predicted pit depth and aspect ratio against experimental data not used in calibration.

This comprehensive uncertainty quantification (UQ) framework incorporates correlation analysis for input screening, global sensitivity analysis for identifying influential parameters, hierarchical surrogate modeling for computational efficiency, Bayesian inference for parameter calibration, uncertainty propagation for probabilistic prediction of pit morphology, and model validation to assess predictive performance against experimental observations. Together, these components enable the prediction of probabilistic model responses under varying input conditions and quantify how parameter uncertainty influences the temporal evolution of output features such as pit depth and aspect ratio. The proposed UQ framework consists of both inverse analysis (Bayesian inference) and forward analysis (uncertainty propagation in predicting the pit evolution), and the computations are made efficient by incorporating correlation and sensitivity analyses.

The following subsections describe each of these steps in detail.

2.1. Preliminary correlation analysis for input screening

Before constructing the surrogate model, a preliminary correlation analysis was conducted to gain early insight into the relationships between uncertain physics model parameters and selected scalar quantities of interest (QoIs) [9,24]. This analysis was based on a small ensemble of simulation runs generated using a space-filling Design of Experiments (DoE) constructed via Latin Hypercube Sampling (LHS) [22,35]. The objective of this step was to identify dominant physics model parameters that have a noticeable influence on the model outputs and to support input screening prior to surrogate modeling. This is particularly useful for reducing model complexity, improving computational efficiency, and guiding the selection of training points in the surrogate construction phase.

Since the model outputs multiple scalar QoIs—denoted as $z^{(j)}_{\text{model}}$ for $j = 1, \dots, m$ (e.g., pit depth, pit aspect ratio, etc.)—correlation coefficients were computed between each input parameter $\theta_i \in \Theta = \{\theta_1, \theta_2, \dots, \theta_p\}$ and each QoI. The resulting correlation matrix $\mathbf{R} \in \mathbb{R}^{p\times m}$ summarizes these pairwise relationships. The resulting correlation matrix $\mathbf{R} \in \mathbb{R}^{p\times m}$ summarizes the pairwise linear relationships between each of the p uncertain model parameters and the m scalar quantities of interest extracted from the model responses. Each element of the matrix is defined using the Pearson correlation coefficient:

$$R_{ij} = \frac{\text{Cov}(\theta_i, z^{(j)}_{\text{model}})}{\sigma_{\theta_i}\, \sigma_{z^{(j)}_{\text{model}}}}, \tag{1}$$

where $\text{Cov}(\cdot,\cdot)$ denotes the sample covariance, and σ_{θ_i}, $\sigma_{z^{(j)}_{\text{model}}}$ are the standard deviations of input θ_i and output $z^{(j)}_{\text{model}}$, respectively. Values of R_{ij} close to ± 1 indicate a strong linear correlation (positive or negative), while values near zero suggest a weak or no linear relationship.

Due to the limited number of samples in this preliminary stage, the correlation values are treated as qualitative indicators rather than definitive metrics. Nonetheless, they provide initial screening for identifying parameters with negligible effect on the output, which may be fixed at chosen values (e.g., mean value) to reduce the dimensionality of the surrogate or Bayesian model parameters without compromising fidelity. A more comprehensive and quantitative sensitivity analysis was later carried out as described in Section 2.2.

2.2. Global sensitivity analysis

In this study, GSA is performed using outputs from the high-fidelity CA model. The objective of GSA is to quantify the relative contributions of uncertain model inputs and parameters to the uncertainty in the output QoIs $\mathbf{z}_{\text{model}}$. This is particularly valuable for identifying dominant inputs and model parameters, thus guiding experimental design and prioritizing further model refinement.

In this work, we adopt a variance-based GSA method using Sobol' indices [59,60]. For each constructed surrogate model output feature $\hat{\mathbf{z}}_{\text{model}}(\Theta)$, the total output variance $\mathbb{V}[\mathbf{z}_{\text{model}}]$ can be decomposed into contributions from individual parameters and their interactions as:

$$\mathbb{V}[z_{\text{model}}] = \sum_{i=1}^{p} \mathbb{V}_i + \sum_{i<j} \mathbb{V}_{ij} + \cdots + \mathbb{V}_{1,2,\dots,p}, \tag{2}$$

where $\mathbb{V}_i = \mathbb{V}_{\theta_i}[\mathbb{E}[\mathbf{z}_{\text{model}} \mid \theta_i]]$ is the first-order contribution of model parameter θ_i, and $\mathbb{V}_{ij}$, $\mathbb{V}_{ijk}$, etc., represent higher-order interaction terms.

The normalized *first-order Sobol' index* for parameter θ_i is defined as:

$$S_i = \frac{\mathbb{V}_i}{\mathbb{V}[\mathbf{z}_{\text{model}}]}, \tag{3}$$

which measures the fractional contribution of θ_i to the total output variance, ignoring interaction effects.

Similarly, the *total-effect index* is given by:

$$S_i^T = 1 - \frac{\mathbb{V}_{\sim i}}{\mathbb{V}[\mathbf{z}_{\text{model}}]}, \tag{4}$$

where $\mathbb{V}_{\sim i}$ denotes the variance of the output excluding all terms involving θ_i. The total-effect index captures the interaction effects of a parameter with other parameters in addition to its individual effect.

2.3. Hierarchical surrogate modeling framework

High-fidelity computational models (e.g., finite element, cellular automata, phase-field, or peridynamics models) are often computationally expensive and produce high-dimensional or structured outputs, such

as time-series data, spatial fields, or multi-variate responses. This complexity becomes particularly challenging in applications that require repeated evaluations, such as GSA, UQ, and design optimization, where the computational cost can quickly become prohibitive. A notable example is Bayesian model calibration, where the direct use of high-fidelity models remains impractical, since *Markov Chain Monte Carlo* (MCMC) sampling, in particular the Metropolis–Hastings algorithm [21,45], typically requires thousands of forward model evaluations in Bayesian calibration. When each single model run is computationally demanding, using the high-fidelity computational model makes Bayesian inference intractable.

Among the available surrogate modeling techniques, *Gaussian Process Regression* (GPR) is a powerful and flexible choice due to its non-parametric and Bayesian nature. Its ability to capture nonlinear input–output relationships and quantify prediction confidence makes it particularly suitable for scientific applications where data may be limited or expensive to generate [49,54].

GPR has found successful application in a variety of disciplines, including materials modeling [8,13], geostatistics [47], and deep neural networks [14]. In multiscale simulations, for example, GPR has been used to simulate computationally intensive crystal plasticity models. Moreover, its probabilistic nature makes it well suited for integration within Bayesian calibration and UQ frameworks [2,73].

In this work, the dimension reduction-based surrogate modeling framework utilizes GPR as a statistical surrogate for the full computational model. Rather than approximating the entire output field, the GPR is trained on a reduced-order representation of the output, specifically, scalar features extracted from the high-dimensional simulation response. The surrogate model reduces the computational cost of repeated evaluations required during MCMC while retaining the capacity to quantify uncertainty in both model outputs and parameter estimates. Furthermore, the probabilistic formulation of GPR integrates naturally with Bayesian inference and provides both the mean and variance of the predictions.

2.3.1. Gaussian process regression

In the proposed framework, a hierarchical surrogate modeling strategy is employed to approximate the behavior of a high-fidelity computational model that includes both a deterministic simulation and a model discrepancy term. The true system response $\mathbf{y} \in \mathbb{R}^q$ is assumed to be governed by:

$$\mathbf{y} = G(\Theta, \mathbf{x}) + \delta(\mathbf{x}), \tag{5}$$

where $G(\Theta, \mathbf{x})$ is a deterministic computational model, $\Theta \in \mathbb{R}^p$ are uncertain model parameters, $\mathbf{x} \in \mathbb{R}^d$ are controllable input variables, and $\delta(\mathbf{x}) \in \mathbb{R}^q$ captures the discrepancy between the computational model and the physical system, where q denotes the number of discrepancy terms.

A set of scalar features is extracted from the full model response using a feature transformation operator $\mathcal{F}$:

$$\mathbf{z}_{\text{model}} = \mathcal{F}(\mathbf{y}) \in \mathbb{R}^m, \tag{6}$$

which defines the target quantities for the surrogate model.

This transformation serves two primary purposes. First, the output $\mathbf{y}$ of high-fidelity simulations is often high-dimensional or structured (e.g., time-dependent curves, spatial fields, or multivariate responses), making direct simulation with regression models inefficient or impractical. Typical examples include stress–strain response, pit growth as a function of time, pit aspect ratio evolution, or crack front propagation—all of which involve temporally or spatially varying outputs. Second, many downstream tasks—such as calibration, optimization, or decision-making—depend on specific quantities of interest (QoIs) derived from the raw model output, such as peak stress, yield strength, and strain energy in case of mechanical response, geometric descriptors such as pit depth, pit aspect ratio, etc. in case of pitting or other scalar performance metrics. By mapping the complex output $\mathbf{y}$ to a reduced set of informative scalar features $\mathbf{z}_{\text{model}}$, the surrogate modeling problem becomes both computationally tractable and focused on predicting outputs that are most relevant to the application, such as pit depth and aspect ratio.

The transformation $\mathcal{F}$ can be based on physical knowledge (e.g., stress–strain offsets, extrema, integration over domains, or geometric measurements) or mathematical dimension reduction techniques (e.g., principal component analysis), depending on the context and goals of the modeling task [23,33,55,68,69]. Instead of approximating the full-field model response or time-series outputs, the surrogate focuses on the application-relevant output. In this paper, the controllable input variables $\mathbf{x}$ are held fixed, and the surrogate is trained to model the dependence of the extracted QoIs $\mathbf{z}_{\text{model}}$ on the uncertain model parameters Θ only. This is particularly relevant for forward UQ or GSA under fixed boundary, loading, or environmental conditions.

In this work, GPR is employed to build a probabilistic surrogate model, offering both predictive accuracy and quantified uncertainty. Although the simulator produces multiple outputs $\mathbf{y} \in \mathbb{R}^q$, we construct the surrogate model on the reduced feature set $\mathbf{z}_{\text{model}} \in \mathbb{R}^m$. Each component $z^{(j)}_{\text{model}}$ is treated separately and modeled using a separate GPR model, yielding a collection of m scalar-valued Gaussian Processes (GPs). Each surrogate takes the form:

$$\hat{\mathbf{z}}_{\text{model}}(\Theta) = \mathcal{GP}(m(\Theta), k(\Theta, \Theta')) + \epsilon_{\text{surr}}, \tag{7}$$

where the mean function $m(\Theta)$ captures general trends or baseline behavior in the data and can take various forms, such as constant, linear, or nonlinear functions. The covariance function $k(\Theta, \Theta')$, also referred to as the kernel, describes the statistical correlation between the values of the GP at two input points (Θ, Θ') in parameter space and governs the smoothness and flexibility of the learned function. The choice of kernel depends on the problem characteristics and may include commonly used forms such as the *squared exponential (SE)*, *Matérn*, or *radial basis function (RBF)* kernel, given by:

$$k_{\text{SE}}(\Theta, \Theta') = \sigma_f^2 \exp\left(-\frac{1}{2}(\Theta - \Theta')^\top \mathbf{L}^{-2}(\Theta - \Theta')\right), \tag{8}$$

where:

- σ_f^2 is the signal variance, which controls the overall variability of the function,
- $\mathbf{L} = \text{diag}(\ell_1, \ell_2, \ldots, \ell_d)$ is a diagonal matrix of length-scale parameters ℓ_i, each governing the sensitivity of the function to variations in the i-th dimension of Θ.

This kernel assumes the underlying function is infinitely differentiable, leading to smooth and flexible function approximations. The length-scales ℓ_i also serve as an automatic relevance determination (ARD) mechanism, and their values are typically inferred from data during model training. Finally, the term $\epsilon_{\text{surr}} \sim \mathcal{N}(0, \sigma^2_{\text{surr}})$ accounts for the residual arising from fitting of the regression model to the training data.

Given training data $\mathcal{D} = \{(\Theta_i, z_i)\}_{i=1}^N$, where $z_i = \mathcal{F}(G(\Theta_i, \mathbf{x}) + \delta(\mathbf{x}))$, the GPR posterior at a test input Θ^* is a Gaussian distribution:

$$\hat{\mathbf{z}}_{\text{model}}(\Theta^*) \mid \mathcal{D} \sim \mathcal{N}(\mu^*, \sigma^{*2}), \tag{9}$$

with predictive mean and variance:

$$\mu^* = m(\Theta^*) + \mathbf{k}^T (K + \sigma_n^2 I)^{-1} (\mathbf{z} - m(\Theta)), \tag{10}$$

$$\sigma^{*2} = k(\Theta^*, \Theta^*) - \mathbf{k}^T (K + \sigma_n^2 I)^{-1} \mathbf{k}, \tag{11}$$

In this formulation, N denotes the number of training samples used to construct the surrogate model, $\boldsymbol{\Theta} = [\Theta_1, \ldots, \Theta_N]^T$ denotes the matrix of training values of physics model parameters, and $\mathbf{z} = [z_1, \ldots, z_N]^T$ is the corresponding vector of extracted feature responses used to train the surrogate. The vector $\mathbf{k} = [k(\Theta^*, \Theta_1), \ldots, k(\Theta^*, \Theta_N)]^T$ represents the

covariance between a new test point Θ^* and each of the training inputs. The kernel matrix $K \in \mathbb{R}^{N\times N}$ is constructed such that each entry is given by $K_{ij} = k(\Theta_i, \Theta_j)$, where $k(\cdot,\cdot)$ is the covariance function. The mean vector $m(\Theta)$ is evaluated at the training physics model parameters and represents the baseline trend of the response before accounting for any correlation structure captured by the kernel. The term σ_n^2 accounts for the surrogate residual variance and any observational noise in the training data.

The hyperparameters of the GPR surrogate model, including kernel parameters and noise variance σ_n^2, are typically learned by maximizing the marginal likelihood of the observed data. This results in a predictive model that provides both the mean and variance of the prediction, thus making GPR well-suited for uncertainty-aware modeling tasks.

2.4. Bayesian model calibration

Following correlation analysis (in Section 2.1), GSA (in Section 2.2), and surrogate model construction (in Section 2.3.1), Bayesian inference is used to calibrate the uncertain model parameters $\Theta \in \mathbb{R}^p$ and the model discrepancy term $\delta \in \mathbb{R}^q$ using available observational data. The goal is to infer the posterior distribution of (Θ, δ) given the observed QoIs while accounting for model discrepancy, data noise, and prior knowledge. The Bayesian framework provides a unified probabilistic approach that assigns prior distributions to unknown parameters and computes their posterior distributions in light of the observed data [26,46,63]. Unlike deterministic or point-estimate methods [16,17,32], it is well suited for cases where both epistemic and aleatory uncertainty must be represented explicitly.

In Bayesian model calibration, the discrepancy function $\delta(\cdot)$ captures systematic bias between model predictions and experimental observations. Several functional forms of $\delta(\cdot)$ have been proposed, ranging from constant offsets to parametric expressions to Gaussian Process (GP) structures with correlation across the input domain [31]. These forms introduce an increasing number of hyperparameters that describe variance, correlation lengths, and smoothness. Because only a single experimental dataset is available in this study, a constant discrepancy is adopted to avoid overfitting and non-identifiability, while still allowing for the correction of structural bias in the CA model.

The overall workflow consists of surrogate-based prediction, prior specification, and posterior sampling via the Metropolis–Hastings MCMC algorithm, as summarized in Fig. 1.

Let $\mathbf{z}_{\text{obs}} \in \mathbb{R}^m$ denote the experimental feature vector and $\hat{\mathbf{z}}_{\text{model}}$ the surrogate-based model prediction. Bayes' theorem yields

$$p(\Theta, \delta \mid \mathbf{z}_{\text{obs}}) = \frac{p(\mathbf{z}_{\text{obs}} \mid \Theta, \delta)\, p(\Theta)\, p(\delta)}{p(\mathbf{z}_{\text{obs}})}. \tag{12}$$

The likelihood is modeled as a multivariate Gaussian:

$$P(\mathbf{z}_{\text{obs}} \mid \boldsymbol{\Theta}, \boldsymbol{\delta}) = \frac{1}{(2\pi)^{m/2}|\boldsymbol{\Sigma}_{\text{total}}|^{1/2}} \times \exp\left(-\frac{1}{2}(\mathbf{z}_{\text{obs}} - \hat{\mathbf{z}}_{\text{model}})^T \boldsymbol{\Sigma}_{\text{total}}^{-1}(\mathbf{z}_{\text{obs}} - \hat{\mathbf{z}}_{\text{model}})\right), \tag{13}$$

where

$$\boldsymbol{\Sigma}_{\text{total}} = \boldsymbol{\Sigma}_{\text{obs}} + \boldsymbol{\Sigma}_{\text{surr}} + \boldsymbol{\Sigma}_{\delta}. \tag{14}$$

- $\boldsymbol{\Sigma}_{\text{obs}}$ represents measurement uncertainty,
- $\boldsymbol{\Sigma}_{\text{surr}}$ is the posterior predictive covariance from the GP surrogate, and
- $\boldsymbol{\Sigma}_{\delta}$ represents uncertainty associated with the constant discrepancy.

Strong empirical correlations among the features (e.g., 0.97 between C_1–C_2, 0.99 between m_1–m_2, and −0.86 between C_1–m_1) motivated the use of a full covariance structure. All three covariance components are specified a priori and are *not inferred* during MCMC; their values are fixed based on experimental uncertainty quantification and GP surrogate outputs. This ensures reproducibility of the likelihood specification and avoids mixing covariance estimation with parameter calibration.

Because the marginal likelihood $p(\mathbf{z}_{\text{obs}})$ is intractable, the posterior is sampled using a Metropolis–Hastings MCMC scheme based on the unnormalized density

$$p(\Theta, \delta \mid \mathbf{z}_{\text{obs}}) \propto P(\mathbf{z}_{\text{obs}} \mid \Theta, \delta)\, p(\Theta)\, p(\delta). \tag{15}$$

At iteration t, a proposal is generated via Gaussian random-walk:

$$\Theta^* = \Theta^{(t)} + \sigma_\Theta\, \mathcal{N}(0, I),$$

$$\delta^* = \delta^{(t)} + \sigma_\delta\, \mathcal{N}(0, I),$$

where the proposal variances σ_Θ and σ_δ control mixing. The acceptance probability is

$$\alpha = \min\left(1,\ \frac{P(\Theta^*, \delta^* \mid \mathbf{z}_{\text{exp}})}{P(\Theta^{(t)}, \delta^{(t)} \mid \mathbf{z}_{\text{exp}})}\right) \tag{16}$$

The chain is updated according to

$$(\Theta^{(t+1)}, \delta^{(t+1)}) = \begin{cases} (\Theta^*, \delta^*), & \text{with prob. } \alpha, \\ (\Theta^{(t)}, \delta^{(t)}), & \text{otherwise.} \end{cases}$$

After burn-in removal and autocorrelation diagnostics [18], the remaining samples form an empirical approximation of the joint posterior. These samples are subsequently used for parameter inference, uncertainty propagation, and validation analysis.

2.5. Model validation

Model validation evaluates how well model predictions agree with experimental measurements while accounting for observation uncertainty. To achieve this, a quantitative validation metric is employed to provide a statistically grounded assessment of model fidelity, especially in complex applications such as corrosion prediction. Among existing quantitative validation approaches reviewed by Ling and Mahadevan [30], probabilistic distance-based metrics are particularly suitable for cases with limited data and nonlinear model responses, and this motivates their use in the present work.

In this work, the probabilistic distance-based metric—referred to as the model reliability metric—is employed to quantify the probability that model predictions fall within a specified tolerance band around the experimental observations. This metric, denoted as $P(H_0)(t)$, quantifies the probability that the predicted QoIs at time t fall within an acceptable tolerance δ_{tol} from the corresponding experimental values [1,30,50,71]:

$$P(H_0)(t) = \mathbb{P}\left(\left|\rho_{\text{model}}(t) - \rho_{\text{exp}}(t)\right| \le \delta_{tol}\right), \tag{17}$$

where:

- $\rho_{\text{model}}(t)$ is the model-predicted value of the quantity of interest at time t,
- $\rho_{\text{exp}}(t)$ is the experimentally measured value at time t,
- δ_{tol} is the specified tolerance threshold accounting for experimental uncertainty,
- $P(H_0)(t)$ is the probability that the model prediction is within δ_{tol} of the experimental observation.

Numerically, this metric is computed using a sample-based indicator function approach [1]:

$$P(H_0)(t) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\left(\left|\rho_{\text{model}}^{(i)}(t) - \rho_{\text{exp}}(t)\right| \le \delta_{tol}\right), \tag{18}$$

where $\mathbb{1}(\cdot)$ is the binary indicator function used to count how many model predictions fall within a tolerance band around the experimental value.:

$$\mathbb{1}\left(\left|\rho_{\text{model}}^{(i)}(t) - \rho_{\text{exp}}(t)\right| \le \delta_{tol}\right) = \begin{cases} 1, & \text{if } \left|\rho_{\text{model}}^{(i)}(t) - \rho_{\text{exp}}(t)\right| \le \delta_{tol}, \\ 0, & \text{otherwise,} \end{cases}$$

and N is the total number of posterior samples.

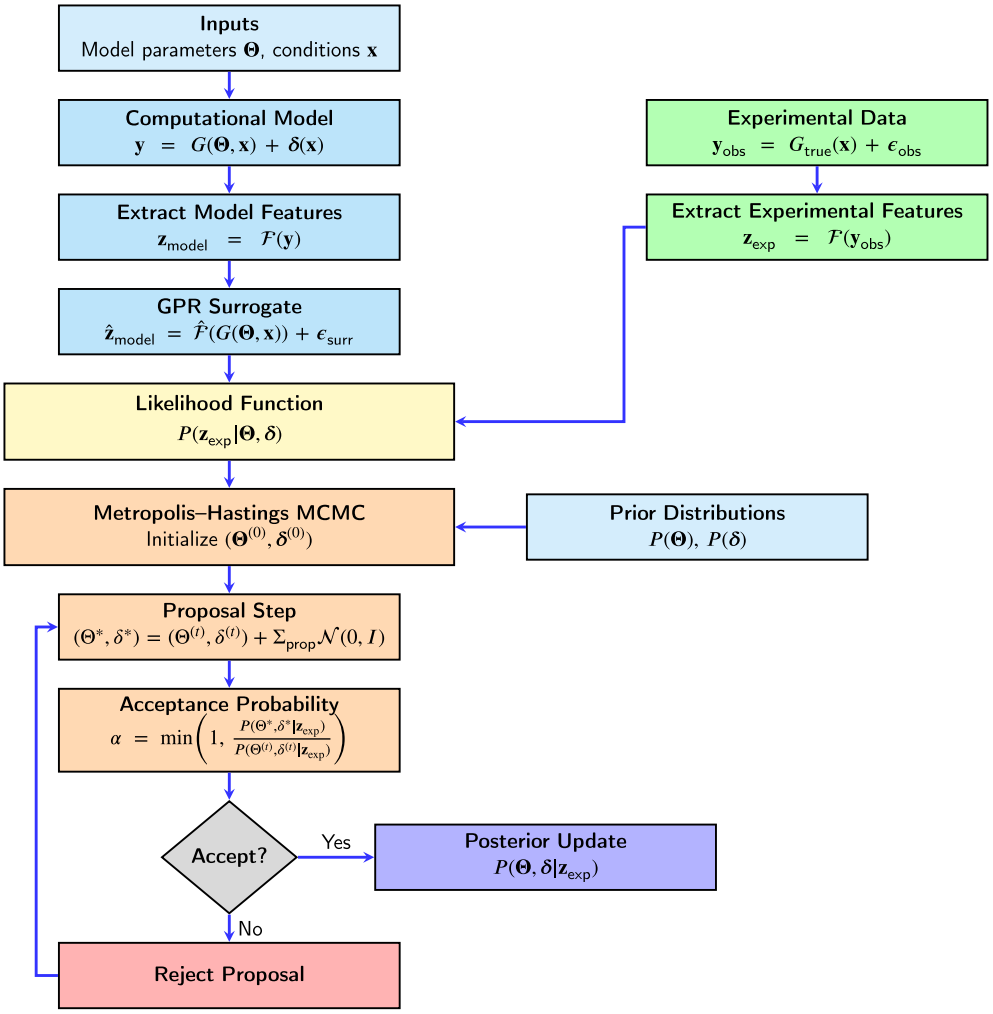


Fig. 1. Bayesian calibration framework with CA model simulation, feature extraction, surrogate modeling, likelihood evaluation, and Metropolis–Hastings MCMC sampling.

This metric provides a probabilistic measure of model accuracy over time and can be used to compare competing models or assess the calibrated model [1]. The choice of δ_{tol} should reflect realistic measurement uncertainty to avoid penalizing models for noise-driven deviations.

3. Numerical example

This section presents the results of the Bayesian parameter estimation and UQ framework applied to a CA model for pitting corrosion. The overall methodology, summarized in the flowchart in Fig. 1 earlier, includes extraction of scalar QoIs, preliminary correlation analysis, GSA to assess parameter importance, construction of a GPR surrogate model, Bayesian inference via MH-MCMC sampling for model calibration, and computation of a reliability-based validation metric to compare model predictions with experimental data. The methodology accounts for various sources of uncertainty, including measurement noise, surrogate prediction error, and physics model parameters and discrepancies.

Two cases are presented to demonstrate the application of the proposed framework under different levels of geometric detail. The first focuses on scalar features extracted from the evolution of maximum pit depth. The second incorporates both maximum pit depth and maximum pit aspect ratio (i.e., ratio of depth to half width), enabling simultaneous assessment of depth and lateral growth. This multi-feature representation supports the comprehensive calibration of model parameters that govern both the extent and morphology of corrosion, thereby improving predictive performance. The study highlights how physically meaningful, time-dependent feature selection—coupled with UQ—can improve the accuracy of model calibration in complex corrosion simulations.

3.1. CA model setup

The CA model simulates iron corrosion in an aerated aqueous environment at ambient temperature, replicating experimental pitting conditions reported in Fatoba et al. [17]. A two-dimensional lattice space consisting of $N_x = 500$ elements in the x-direction and $N_y = 1000$ elements in the y-direction was employed in the simulations, as shown in Fig. 2(a). The domain is vertically divided into two equal halves, with the upper half representing the aqueous electrolyte and the lower half corresponding to the metallic iron (Fe) substrate. To simulate the laboratory electrolyte chemistry, 0.6 moles of NaCl at pH 6.2 were introduced into the water phase to promote stable pitting under the assumption that the lacy cover had already broken down to allow pit propagation. The salt concentration and pH typically reflect that of a seawater environment. Although NaCl is a neutral salt, its dissociation provides Na^+ and Cl^- ions that increase the ionic strength and indirectly promote H^+ formation through hydrolysis and local charge balance, thereby mimicking the acidic microenvironment often observed during pit propagation. A Moore neighborhood configuration was used to simulate electrochemical interactions by considering all eight neighboring cells around a given lattice site, providing a more realistic representation of mass transport and reaction pathways.

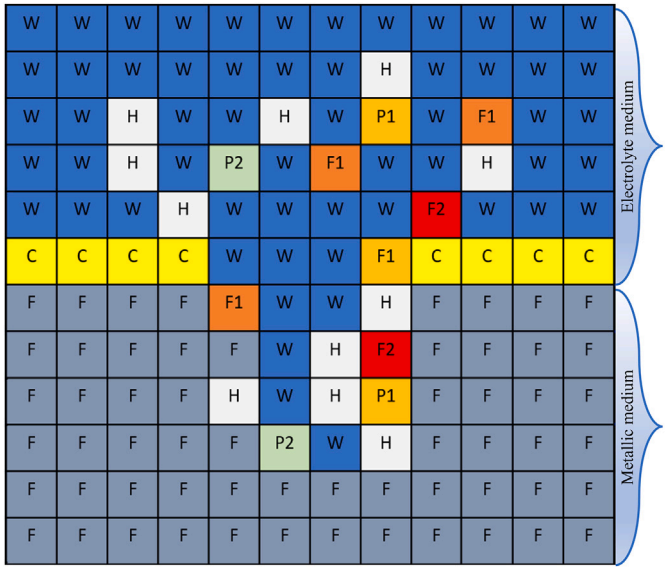


Fig. 2. Cellular automata grid representation of corrosion modeling, where different species (e.g., water, protons, iron ions, hydroxides) are spatially mapped in electrolyte and metal regions.

Capillary wall sites C were introduced at the metal–electrolyte interface to regulate electrolyte access and spatially restrict water–iron interaction. Acting as a semi-permeable barrier, these sites enable localized transport and reaction, allowing the model to define an initial single pit geometry with a prescribed width $\rho_{w,\mathrm{ini}}$. To replicate the capillary tip size used in experiments by Fatoba et al. [17], the initial pit width was set to 500 μm (see Fig. 2). Each CA lattice site was scaled accordingly, resulting in a spatial resolution of $2.5\,\mu\mathrm{m} \times 2.5\,\mu\mathrm{m}$ per cell, which was fixed for all simulations. Boundary cells along the left and right edges of the metallic domain were excluded from electrochemical reactions to prevent edge effects and ensure proper enforcement of electrochemical boundary conditions. This setup maintains localized, physically realistic pit formation and ensures numerical stability consistent with experimental conditions. Together, the water chemistry, domain structure, and capillary wall configuration establish the initial conditions for simulating localized corrosion.

Eight uncertain model parameters $\Theta \in \mathbb{R}^8$ are considered in this study; these relate to key physical processes involved in the pitting corrosion simulation. These include the corrosion probability P_{corr}, which governs the transformation of metallic iron (Fe) into ferrous ions (Fe^{2+}), and the oxidation probability P_{dis}, which controls the subsequent conversion of Fe^{2+} to Fe^{3+}. The hydrolysis behavior is captured by two probabilities, P_{hyd1} and P_{hyd2}, associated with the formation of $Fe(OH)_2$ and $Fe(OH)_3$, respectively. Ion mobility is governed by a shared diffusion probability $P_{\mathrm{diff}}^{\mathrm{ion}}$ for H^+, Fe^{2+}, and Fe^{3+}, while the transport of hydrolysis products is described separately via $P_{\mathrm{diff}}^{\mathrm{Fe(OH)_2}}$ and $P_{\mathrm{diff}}^{\mathrm{Fe(OH)_3}}$. Finally, the sedimentation factor Sed introduces a gravity-like bias that influences the directional diffusion of heavier species within the simulation domain. Different subsets of these eight model parameters will be considered in this numerical example, based on correlation and sensitivity analyses, as discussed later.

LHS was employed to generate 1000 high-fidelity CA model runs across the uncertain model parameter space. The simulations are executed on a high-performance workstation equipped with a 13th Gen Intel® Core™ i9-13900K CPU (24 cores, 32 threads) and 128 GB of RAM. Each individual simulation requires approximately 2.5 hours to complete. Given a batch of 1000 simulations, the total computational time amounts to around 2500 hours, which corresponds to approximately 3.5 months of continuous processing time if run sequentially. The model parameters are sampled (for training the surrogate model) from independent uniform distributions, as summarized in Table 1. These parameters are grouped into corrosion kinetics (e.g., metal dissolution and hydrolysis), ionic and precipitate diffusion, and sedimentation effects. The selected ranges reflect physically plausible regimes informed by prior modeling studies [4,17]. Each simulation produced evolving pit morphologies, from which time-dependent quantities of interest, such as maximum pit depth, maximum pit width, and maximum aspect ratio, were extracted. These geometric features, paired with the corresponding LHS-sampled model parameters, were then used to train GPR surrogate models and support subsequent tasks such as GSA and Bayesian calibration. To demonstrate the effectiveness of the proposed framework, two cases are presented in the following sections: one considering pit depth features alone and the other incorporating both pit depth and aspect ratio to capture more detailed pit morphology.

Table 1
Parameter ranges for uncertain CA model parameters used in LHS.

Group	CA Model Parameter	Range
Corrosion	P_{corr} (corrosion probability)	0.1 – 0.9
	P_{oxd} (oxidation probability)	0.1 – 0.9
	P_{hyd1} (hydrolysis of Fe^{2+})	0.1 – 0.6
	P_{hyd2} (hydrolysis of Fe^{3+})	0.1 – 0.6
Diffusion	$P_{\mathrm{diff}}^{\mathrm{ion}}$ (for H^+, Fe^{2+}, Fe^{3+})	0.1 – 0.9
	$P_{\mathrm{diff}}^{\mathrm{Fe(OH)_2}}$ (diffusion of $Fe(OH)_2$)	0.02 – 0.09
	$P_{\mathrm{diff}}^{\mathrm{Fe(OH)_3}}$ (diffusion of $Fe(OH)_3$)	0.02 – 0.09
	Sed (sedimentation factor)	1 – 8

3.1.1. Feature extraction

To facilitate surrogate modeling, global sensitivity analysis (GSA), and uncertainty quantification (UQ), scalar features were extracted from the time-dependent outputs of 1000 high-fidelity cellular automata (CA) simulations. These simulations were executed on a workstation equipped with an Intel® Core™ i9-13900K CPU (24 cores, 32 threads) and 128 GB of RAM. On average, each simulation required approximately 2.5 hours to complete. Two cases were considered to examine different levels of geometric detail in pit evolution.

In **Case 1**, the analysis focused solely on the temporal evolution of maximum pit depth. This approach is motivated by the fact that pit depth is one of the most commonly reported and experimentally accessible indicators of localized corrosion severity. From each of the 1000 high-fidelity CA simulations, the temporal evolution of maximum pit depth was recorded. To convert these time-series outputs into low-dimensional, interpretable QoIs, each time series was fit with a power-law relation of the form:

$$\rho_{d_{\max}}(t) = C_1 t^{m_1}, \tag{19}$$

where $\rho_{d_{\max}}(t)$ is the maximum pit depth at time t, and C_1 and m_1 are fitting parameters that serve as scalar features describing the corrosion rate and growth kinetics. The feature C_1 reflects the initial corrosion intensity, while m_1 captures the time exponent that governs the pit propagation behavior. The use of power-law representations follows established practice in localized corrosion modeling (e.g., Fatoba et al. [17], Babu et al. [4]), where both experimental pit-growth measurements and CA-based simulations consistently exhibit power-law behavior over time. Because the experimental dataset in this study displays the same trend, fitting a power-law model does not impose an artificial structure but instead captures the intrinsic geometric evolution of pitting.

The fitting procedure was applied across all simulation outputs to ensure consistency in feature extraction.

The output of each simulation was thus reduced to a feature vector:

$$\mathbf{z}_{\mathrm{model}} = (C_1, m_1), \tag{20}$$

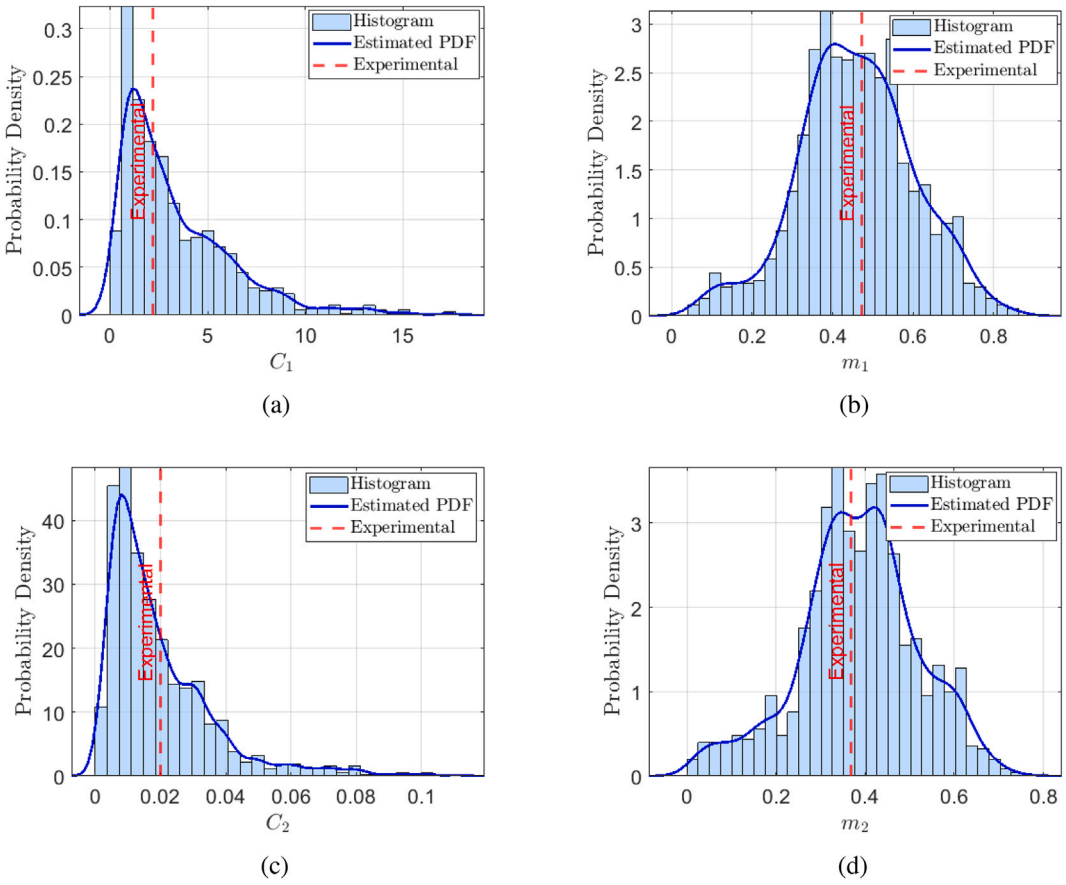


Fig. 3. Statistical distributions of four power-law coefficients from 1000 high-fidelity CA simulations. Subplots (a) and (b) show the distributions of maximum pit depth features (C_1, m_1), while (c) and (d) correspond to pit aspect ratio features (C_2, m_2) [17].

which serves as the output QoI for surrogate modeling and subsequent inference.

In **Case 2**, the framework was extended to include the evolution of the maximum pit aspect ratio, which characterizes the shape of the pit as it grows. Similar to depth, the aspect ratio was observed as a function of time and fitted with a power-law relation:

$$AR_{\max}(t) = C_2\, t^{m_2}, \tag{21}$$

where $AR_{\max}(t)$ denotes the maximum pit aspect ratio over time, and C_2, m_2 represent the scaling and growth exponent parameters, respectively. The reduced output for each simulation in this case was:

$$\mathbf{z}_{\text{model}} = (C_1, m_1, C_2, m_2), \tag{22}$$

enabling simultaneous modeling of both the extent and morphology of corrosion evolution.

Fig. 3(a)–(d) illustrates the extraction of four geometric features—C_1, m_1, C_2, and m_2—from 1000 cellular automata (CA) simulations by fitting power-law curves to time-series pit morphology data. To evaluate surrogate accuracy in the physical output space, the surrogate-predicted power-law coefficients (C_1, m_1, C_2, m_2) were inserted into the corresponding power-law relations to reconstruct the full pit-depth and aspect-ratio trajectories. These reconstructed profiles were then compared against those obtained from the high-fidelity CA simulations across a large ensemble of test inputs. The resulting error distributions demonstrate that the surrogate models retain high predictive fidelity after reconstruction: the mean absolute percentage error (MAPE) exhibits a narrow distribution with a mean value of approximately 6.75%, while the coefficient of determination shows a strong skew toward unity with a mean of $R^2 = 0.8477$. Together, these metrics confirm that surrogate-induced errors remain small when propagated into the physical quantities of interest, and that the surrogate models accurately preserve the high-fidelity CA model behavior in both pit-depth and aspect-ratio evolution. The blue, green, cyan, and magenta markers denote CA-predicted feature values across different realizations. Each feature is compared against a single experimental data point, shown respectively as a red, black, yellow, or orange marker.

These experimental values were obtained from an electrochemical–mechanical pit-growth experiment on API 5L X65 steel, in which pits were generated using a three-electrode micro-capillary setup containing 0.6 M NaCl electrolyte. A 500 μm capillary tip delivered localized exposure to the specimen surface, while a Pt counter electrode and Ag/AgCl reference electrode controlled the electrochemical environment. Cyclic loading ($\sigma_{\max}$ = 200 MPa, R = 0.1, 2 Hz) was applied simultaneously to replicate pipeline operating conditions. Under these coupled conditions, a single pit was allowed to evolve over time. Pit morphology was captured at selected intervals using a Keyence VK-X200K 3D confocal laser microscope, and pit depth and pit aspect ratio were extracted from the resulting surface profiles. These two geometric descriptors form the experimental dataset used as calibration targets for the power-law features shown in Fig. 3, based on measurements reported by Fatoba et al. [17].

This approach enables the reduction of time-series output into scalar features, which are essential for efficient surrogate modeling, global sensitivity analysis, and Bayesian calibration.

3.1.2. Correlation analysis

Prior to surrogate modeling, Pearson correlation analysis was performed (as detailed in Section 2.1) to investigate linear trends between

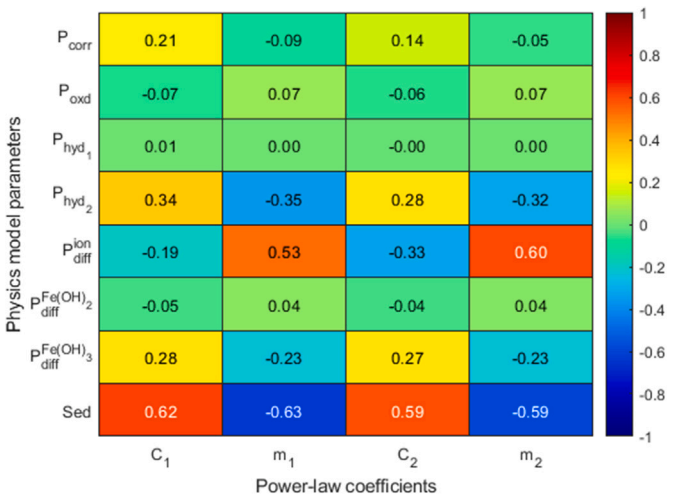


Fig. 4. Correlation matrix showing Pearson correlation coefficients between physics model parameters and power-law coefficients(C_1, m_1, C_2, m_2).

the uncertain model parameters and the extracted scalar features C_1, m_1 (from maximum pit depth) and C_2, m_2 (from maximum pit aspect ratio). The full correlation matrix summarizes all pairwise correlations in a compact form in Fig. 4.

Pit depth features (C_1, m_1). The initial corrosion intensity coefficient C_1 shows strong positive correlation with the sedimentation factor (Sed = 0.62) and moderate correlations with the hydrolysis probability of Fe^{3+} ($P_{\text{hyd2}} = 0.34$) and diffusion probability of $Fe(OH)_3$ ($P_{\text{diff}}^{Fe(OH)_3} = 0.28$). These results indicate that increased downward mass transport and precipitate mobility are associated with faster pit initiation. In contrast, C_1 is negatively correlated with ionic diffusion ($P_{\text{diff}}^{\text{ion}} = -0.19$), suggesting that rapid dispersion of ions from the pit may slow down the initial corrosion rate; however, this correlation is weak.

The exponent m_1, which governs the time-dependent pit growth rate, exhibits a strong negative correlation with Sed = −0.63, a moderate negative correlation with $P_{\text{hyd2}} = -0.35$, and a moderate positive correlation with $P_{\text{diff}}^{\text{ion}} = 0.53$. These trends imply that faster ion mobility supports sustained pit growth over time, while sedimentation and hydrolysis may inhibit long-term propagation.

Pit aspect ratio features (C_2, m_2). : The coefficient C_2, which characterizes the initial rate of aspect ratio change, exhibits moderate correlation with Sed = 0.59 and weak positive correlations with $P_{\text{hyd2}} = 0.28$, and $P_{\text{diff}}^{Fe(OH)_3} = 0.27$. These findings suggest that the early geometry of the pit is sensitive to vertical transport effects, which enhance the formation of deep, narrow features in the initial stages.

For the exponent m_2, which describes the evolution of the pit shape over time, the trends closely mirror those of m_1. The strongest negative correlation is again observed with Sed = −0.59, while the strongest positive correlation is with $P_{\text{diff}}^{\text{ion}} = 0.60$. This reaffirms the importance of balanced ionic mobility over gravitational settling to maintain pit geometry evolution and lateral growth.

Across all four features, sedimentation (Sed) emerges as the most influential parameter, positively affecting early-stage growth rates (C_1, C_2) while negatively impacting the long-term growth exponents (m_1, m_2). This dual role reflects its physical interpretation: while gravity-driven species transport initially enhances localized attack, it can hinder sustained transport and growth by causing stagnation or clogging near the pit base. Ionic diffusion ($P_{\text{diff}}^{\text{ion}}$) consistently shows a positive influence on time-dependent exponents m_1 and m_2, highlighting its role in facilitating continued corrosion propagation through efficient species exchange.

3.1.3. Global sensitivity analysis

The GSA, detailed in Section 2.2, was conducted using the variance-based Sobol' method to evaluate the relative contributions of uncertain CA model inputs to the variability observed in the extracted feature set $\mathbf{z}_{\text{model}} = (C_1, m_1, C_2, m_2)$.

The prior distributions of model parameters (as listed in Table 1) are modeled as statistically independent uniform distributions across their defined ranges (see Section 3.1.5). A set of 1000 high-fidelity CA simulations, previously generated using Latin Hypercube Sampling (LHS) for surrogate model training, was also employed to perform the Sobol's sensitivity analysis. The UQLab toolbox [40,41] was employed to compute the first-order and total-effect Sobol' indices using this dataset.

Results and interpretation. : Fig. 5(a)-(d) presents the Sobol' indices for all four features. For each, both first-order (blue bars) and total-effect (red bars) indices are shown, highlighting direct effects and higher-order interactions, respectively.

The results for C_1 and m_1 are presented in Fig. 5(a) and (b), respectively. For C_1, the Sed was found to be the most dominant contributor, with a total Sobol' index of approximately 0.90 and a first-order index of 0.55, indicating both strong main effects and interaction effects. This confirms the dominant role of the diffusion-controlled dissolution process, influencing both the onset and sustained progression of pitting corrosion. Moderate total contributions were also observed from P_{hyd2} and $P_{\text{diff}}^{\text{Fe(OH)}_3}$, suggesting the role of hydrolysis and precipitate diffusion in early-stage corrosion behavior. Interestingly, ionic diffusion ($P_{\text{diff}}^{\text{ion}}$) exhibited a moderate total effect (0.34) but a lower first-order effect (0.02), indicating significant interactions with other parameters.

For m_1, which governs the time-dependent pit growth rate, the sedimentation factor remained the most influential input (total index: 0.76; first-order: 0.47). Ionic diffusion again showed strong influence, with a total Sobol' index of 0.43, followed by P_{hyd2} and P_{corr} with smaller but non-negligible contributions, suggesting that rapid ionic mobility and hydrolysis play critical roles in sustaining corrosion propagation.

For C_2, which reflects the initial lateral corrosion intensity (aspect ratio growth), the Sed again emerged as the most dominant contributor, with a total Sobol' index of approximately 0.93 and a first-order index of 0.50. This highlights the importance of a diffusion-controlled dissolution process in shaping not only vertical but also lateral corrosion features. Additionally, $P_{\text{diff}}^{\text{ion}}$ contributed significantly (total index: 0.46; negligible first-order), suggesting that mass transport effects play a critical role in the formation and expansion of lateral pit structures. Moderate influence was also observed from the P_{hyd2}, P_{corr}, and P_{oxd}, indicating that local chemistry and oxidation processes also modulate geometric pit morphology.

For m_2, which governs the time-dependent growth rate of pit aspect ratio, Sed, and $P_{\text{diff}}^{\text{ion}}$ were again the dominant parameters, with total Sobol' indices of 0.71 and 0.52, respectively, and corresponding first-order indices of 0.41 and 0.27. This indicates strong main effects from both parameters. The relatively narrow gap between total and first-order indices for m_2 (compared to m_1) suggests more additive, less interactive behavior in the lateral propagation mechanism. Nonetheless, secondary contributions from P_{hyd2} imply hydrolysis processes also influence the geometric evolution over time.

The GSA reveals that Sed and $P_{\text{diff}}^{\text{ion}}$ are consistently dominant contributors to the variability in all four extracted features—C_1, m_1, C_2, and m_2. For C_1 and C_2, which reflect the intensity of pit depth and aspect ratio growth, respectively, Sed exhibits the highest total Sobol' indices (approximately 0.90 and 0.93), with corresponding first-order effects also being substantial. This suggests that the diffusion-controlled dissolution process plays a central role in the onset and development of both vertical and lateral pit geometries. Moderate contributions from P_{hyd2} highlight the role of chemical speciation in shaping early-stage corrosion. For the time exponents m_1 and m_2, which control the growth kinetics, both Sed and $P\text{diff}^{\text{ion}}$ again show strong influence, indicating that ion mobility and vertical mass flux significantly govern the sustained propagation of

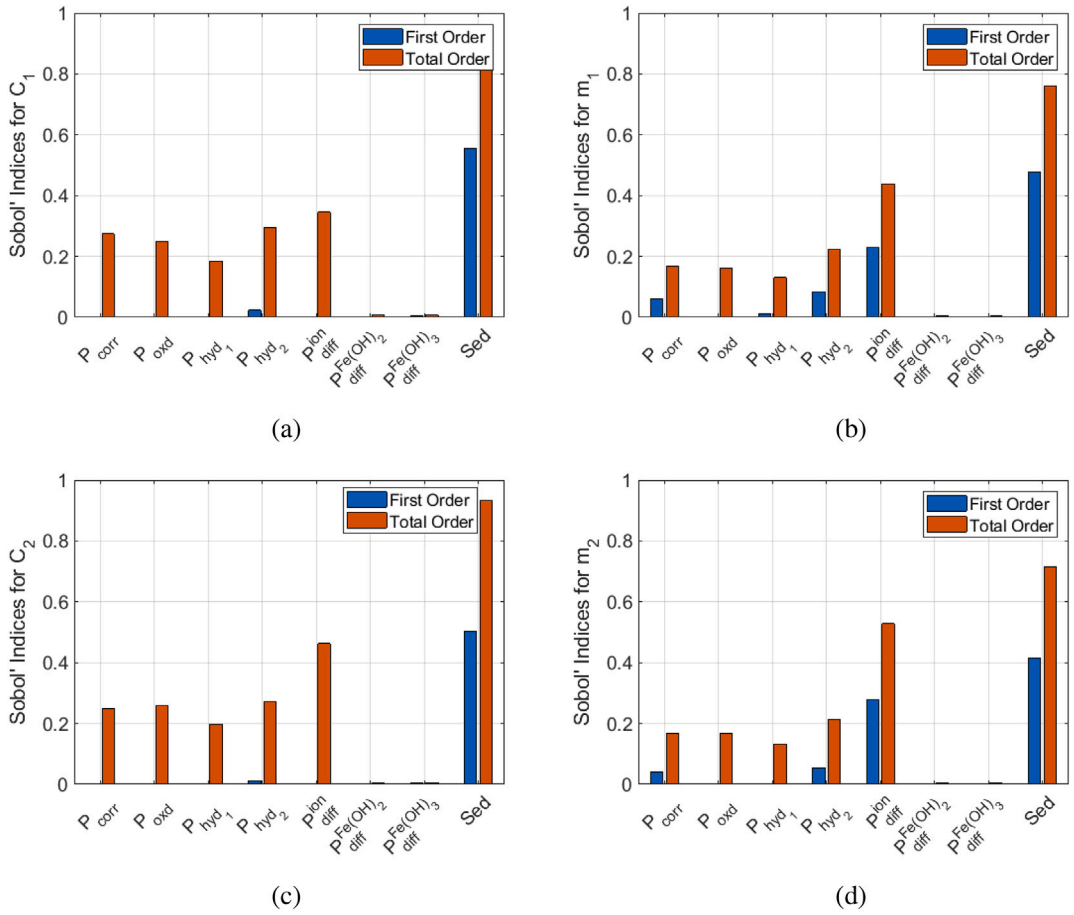


Fig. 5. First-order (blue) and total-effect (red) Sobol' sensitivity indices for all four power-law coefficients: (a) C_1, (b) m_1, (c) C_2, and (d) m_2.

pits. Notably, in all cases, the differences between total and first-order indices—particularly for `Sed` and $P_{\text{diff}}^{\text{ion}}$—underscore the importance of higher-order, nonlinear interactions among inputs. This highlights the complexity of the underlying corrosion process, where the effects of physics model parameters are not purely additive but coupled through interdependent transport and reaction pathways.

The GSA results are broadly consistent with the findings from the earlier Pearson correlation analysis. In both analyses, the sedimentation factor (`Sed`) and ionic diffusion probability ($P_{\text{diff}}^{\text{ion}}$) consistently emerged as the most influential parameters affecting all four extracted features (C_1, m_1, C_2, m_2). Similarly, $P_{\text{diff}}^{\text{ion}}$, which showed moderate positive correlations with the exponents m_1 and m_2, was also identified in the GSA as a dominant contributor to their variability. Secondary parameters such as P_{hyd2} and $P_{\text{diff}}^{\text{Fe(OH)}_3}$, which showed weak to moderate correlations in the Pearson analysis, also appeared with modest influence in the Sobol' indices. This consistency reinforces confidence in the physical interpretations drawn from both analyses, while also validating the GSA as a robust tool for understanding parameter influence. Table 2 summarizes the selection of physics model parameters for use in Bayesian calibration and uncertainty quantification, based on a combination of correlation analysis and Sobol' sensitivity indices. The inputs are categorized by physical processes—*corrosion*, *diffusion*, and *other*—and evaluated according to their maximum absolute correlation and their highest total-effect Sobol' indices across all the power-law coefficients. Parameters such as `Sed`, $P_{\text{diff}}^{\text{ion}}$, and P_{hyd_2} were chosen for calibration due to their consistently high influence across both metrics, indicating their significant contribution to output variability and morphological evolution. In contrast, parameters like P_{hyd_1} and $P_{\text{diff}}^{\text{Fe(OH)}_2}$, which exhibited low correlation and negligible sensitivity, were excluded from the calibration process and held fixed at nominal values. Moderately influential inputs—such as P_{corr}, P_{oxd}, and $P_{\text{diff}}^{\text{Fe(OH)}_3}$—were marked as optional, depending on the desired trade-off between model simplicity and accuracy. This combined use of correlation and global sensitivity analyses ensures that the Bayesian calibration focuses on the most influential parameters, improving computational tractability while preserving the precision of the calibration results [31,51].

3.1.4. Surrogate model construction

In this section, GPR surrogate models were constructed (as described in Section 2.3) to simulate the relationship between the uncertain model parameters Θ and the extracted geometric features $\mathbf{z}_{\text{model}} = (C_1, m_1, C_2, m_2)$ from the high-fidelity CA simulations. These surrogate models serve as computationally efficient approximations of the CA simulator under fixed input variables $\mathbf{x}$ and specified boundary and environmental conditions.

Separate GPR models were trained for each scalar feature—C_1 and m_1 from maximum pit depth evolution, and C_2 and m_2 from maximum aspect ratio evolution—using data from 1000 LHS runs. The input to each model was the eight-dimensional vector of uncertain model parameters $\Theta \in \mathbb{R}^8$, and the corresponding output was each one of the extracted power-law features. To ensure generalization and avoid overfitting, the dataset was randomly split: 80% of samples were used for training and 20% for testing, based on MATLAB's `cvpartition` function. Each GPR model employed a linear basis function to represent global trends, and an automatic relevance determination squared exponential (ARD-SE) kernel to capture smooth nonlinear interactions among inputs. Specifically, the kernel function was set to `ardsquaredexponential`, and the basis function was set to `linear`. Model construction and

Table 2
Parameter selection for Bayesian calibration using correlation and Sobol' sensitivity measures.

Group	Input Parameter	Corr (max $\|R\|$)	Sobol Max Total Index	Recommended for calibration	Reason for Inclusion
Corrosion	P_{corr}	0.2130	0.27378	Optional	Moderate correlation and sensitivity; captures some variation.
	P_{oxd}	0.0683	0.25876	Optional	Low correlation, modest Sobol index; possible nonlinear effects.
	P_{hyd_1}	0.0070	0.19743	No	Negligible influence across outputs.
	P_{hyd_2}	0.3520	0.29485	Yes	Moderate influence across outputs; consistent effect.
Diffusion	$P^{\text{ion}}_{\text{diff}}$	0.5955	0.52843	Yes	Strongest combined influence; dominant across all outputs.
	$P^{\text{Fe(OH)}_2}_{\text{diff}}$	0.0450	0.00887	No	Minimal contribution; both metrics near zero.
	$P^{\text{Fe(OH)}_3}_{\text{diff}}$	0.2791	0.00769	Optional	Moderate correlation but negligible sensitivity; mostly linear effect.
Other	Sed	0.6290	0.93167	Yes	Highest influence across all metrics; essential for calibration.

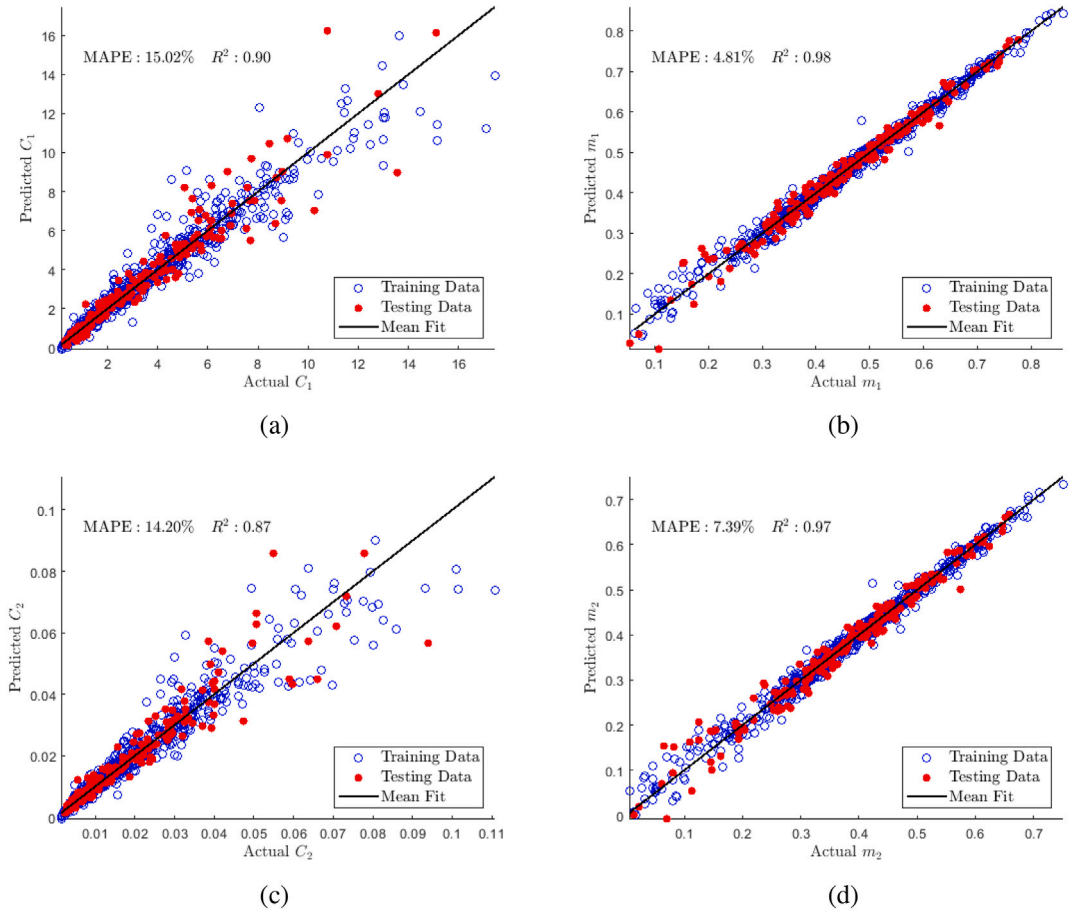


Fig. 6. GPR surrogate model predictions versus CA simulation outputs for extracted features: (a) C_1, (b) m_1, (c) C_2, and (d) m_2. The mean fit line is shown in black for all plots. Blue circles represent training data, and red circles represent testing data.

training were implemented using the `fitrgp` function from MATLAB's Statistics and Machine Learning Toolbox [42].

Model testing and performance evaluation. The predictive performance of the trained GPR surrogate models was evaluated using both training and testing datasets. Fig. 6(a)–(d) presents scatter plots comparing the predicted versus actual values for C_1, m_1, C_2, and m_2, respectively, on both training and testing datasets. The predictive accuracy of each GPR model was quantified using the coefficient of determination (R^2) and the mean absolute percentage error (MAPE) for both training and testing datasets. Fig. 7 summarizes the performance metrics across all four power-law coefficients.

The trained surrogate models were verified using the testing dataset. The results show that all models achieve strong predictive capabilities, particularly for the power law model exponents m_1 and m_2, which yield testing R^2 values of 0.98 and 0.97, respectively, and MAPE values under 7.5%. The power law model coefficients C_1 and C_2 exhibit slightly lower but still satisfactory predictive performance, with R^2 values of 0.90 and 0.87 and MAPE values of 15.02% and 14.20%, respectively. For completeness, surrogate accuracy was also quantified directly against high-fidelity CA simulations. Using surrogate-reconstructed trajectories, the pit-depth model attains $R^2 = 0.8934$ with MAPE = 9.93%, while the aspect-ratio model attains $R^2 = 0.8997$ with MAPE = 8.35%, indicating strong agreement with the corresponding CA-predicted physical outputs. These outcomes confirm that the surrogate models are sufficiently accurate to support subsequent GSA and Bayesian inference tasks.

3.1.5. Bayesian calibration: setup and results

The mathematical formulation and implementation details of the Bayesian calibration framework have been described earlier

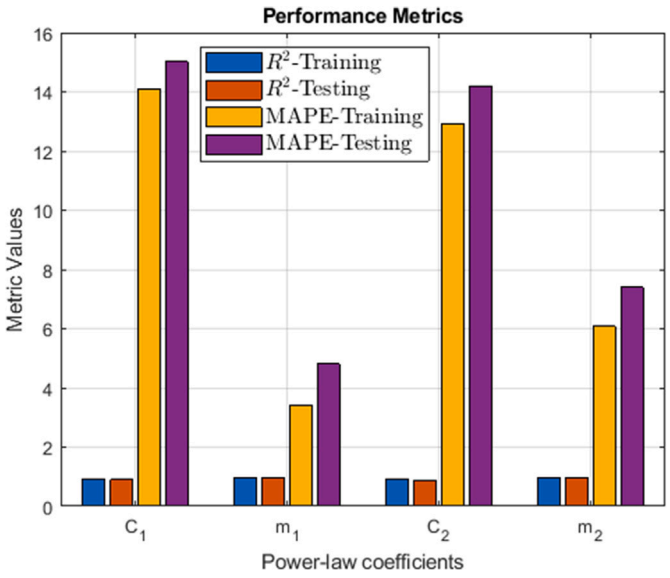


Fig. 7. Performance metrics for GPR models across all four power-law coefficients (C_1, m_1, C_2, m_2). Bar plots represent training and testing values for R^2 and MAPE. Blue and orange bars correspond to R^2 for training and testing, while yellow and purple bars denote MAPE for training and testing, respectively.

in Section 2.4. In this section, we present the numerical results of the Bayesian calibration for the two cases introduced previously.

The goal is to infer the posterior distributions of the uncertain physics model parameters $\boldsymbol{\Theta} \in \mathbb{R}^8$ by comparing the surrogate model predictions of extracted geometric features against experimentally observed values. Case 1 uses (C_1, m_1) as calibration targets derived from maximum pit depth evolution, while Case 2 extends the analysis to the full feature set (C_1, m_1, C_2, m_2) by including pit aspect ratio dynamics.

The MH-MCMC algorithm was employed to construct the posterior distribution, using the trained GPR surrogate models (presented in Section 3.1.4) to evaluate the likelihood at each iteration. All prior distributions were assumed to be uniform over the ranges in Table 1, and the total number of MCMC iterations was set to 5 million across all cases discussed below. Convergence of the MCMC chains was assessed using autocorrelation diagnostics and visual inspection of trace plots, following the guidelines in [18]. Chains showed decay in autocorrelation and reached stable regions after initial transients, supporting the decision to discard the first half of the samples as burn-in to ensure reliable posterior estimation. This allowed for the estimation of the joint posterior distribution of both the uncertain model parameters (Θ) and the discrepancy terms (δ) introduced to account for the structural limitations of the computational model.

3.1.6. Case 1: Calibration using pit depth alone

In Case 1, Bayesian calibration was performed using only pit depth evolution time series; from this time series data, two features are extracted—C_1 and m_1—which quantify the initial magnitude and temporal evolution of the *maximum pit depth*. These outputs are physically meaningful and readily extractable from experiments, making them ideal targets for calibration.

To investigate the effect of physics model parameter dimensionality on identifiability and predictive accuracy, three calibration cases were defined:

- **Case 1a (8 parameters):** All eight uncertain model parameters were calibrated simultaneously. While this setup allows for comprehensive exploration of parameter interactions, it also presents challenges related to overparameterization and reduced identifiability.
- **Case 1b (6 parameters):** This reduced case excluded P_{hyd1} and $P_{\text{diff}}^{\text{Fe(OH)}_2}$, which were previously shown to have negligible influence on C_1 and m_1 through Sobol' and Pearson correlation analyses. The reduced parameter space aimed to improve inference quality and computational efficiency.
- **Case 1c (3 parameters):** This minimalist setup included only the three most influential parameters: P_{hyd2}, $P_{\text{diff}}^{\text{ion}}$ and Sed. It served as a test of whether reliable calibration could be achieved with the benefit of an efficient, reduced-parameter calibration approach.

In all cases, samples from the prior distribution were generated using LHS. Posterior inference was performed using MCMC, and kernel density estimation (KDE) was used to visualize posterior contraction and central tendency. An acceptance rate of approximately 45.01% was achieved for Case 1a. For Case 1b, the acceptance rate improved to 53.58%, and for Case 1c, it reached 67.70%. This improvement is attributed to the progressive reduction in the number of calibrated parameters—from 8 in Case 1a to 6 in Case 1b and 3 in Case 1c—which reduces the dimensionality of the parameter space and simplifies the posterior distribution, allowing the sampler to explore it more efficiently and accept proposed samples more frequently.

Formulation of the discrepancy term

To enhance the reliability of Bayesian inference and acknowledge potential model-form inadequacies, additive discrepancy terms were introduced into the calibration framework. The decision to include specific discrepancy terms was guided by comparing the predicted distributions (from 1000 high-fidelity CA simulations) of the power-law coefficients C_1 and m_1 against experimentally observed values of C_1 and m_1.

As shown in Fig. 3(a)–(b), the distribution of C_1 is positively skewed, with a large proportion of simulated outcomes underpredicting the experimental reference value. This asymmetry suggests a persistent structural bias in the surrogate model's prediction of C_1, which cannot be resolved through parameter adjustment alone. To account for this, a mean discrepancy term $\mu_{\delta_{C_1}}$ was introduced to absorb this systematic offset. Additionally, to reflect uncertainty in the magnitude of C_1 across the design space, especially given its widespread nature, an independent variance term $\sigma_{\delta_{C_1}}$ was also introduced. This allows the inference process to recognize not only bias (via $\mu_{\delta_{C_1}}$) but also stochastic variability in the observed response.

In contrast, the distribution of m_1 appears approximately symmetric and centered near the experimental observation, indicating that the surrogate model captures its central tendency reasonably well. However, noticeable spread and tail behavior suggest the presence of residual variability not captured by the model. To account for this uncertainty, a variance discrepancy term $\sigma_{\delta_{m_1}}$ was included. The discrepancy terms function like regularization terms that prevent overfitting, i.e., avoid ascribing all of the difference between prediction and observation to the model parameters.

Posterior distributions and parameter insights

Posterior distributions revealed clear trends in parameter identifiability across the calibrated cases, as shown in Fig. 8. The Sed consistently exhibited sharp contraction from the prior across all three cases, underscoring its dominant role in controlling the rate of pit deepening. The posterior mean shifted significantly away from the prior center, suggesting strong data influence. A similar trend was observed for the ionic diffusion probability ($P_{\text{diff}}^{\text{ion}}$), particularly in Cases 1b and 1c, where its posterior became sharply peaked.

In contrast, P_{hyd_1} and $P_{\text{diff}}^{\text{Fe(OH)}_2}$ showed little to no posterior update in Case 1a, validating their exclusion in reduced-dimensional Cases 1b and 1c. Their posteriors remained nearly uniform, indicating weak sensitivity to the selected power-law coefficients and poor identifiability under the given data.

Intermediate behaviors were seen for P_{hyd2} and $P_{\text{diff}}^{\text{Fe(OH)}_3}$, which exhibited broader posterior contraction in Case 1b than in Case 1a. These

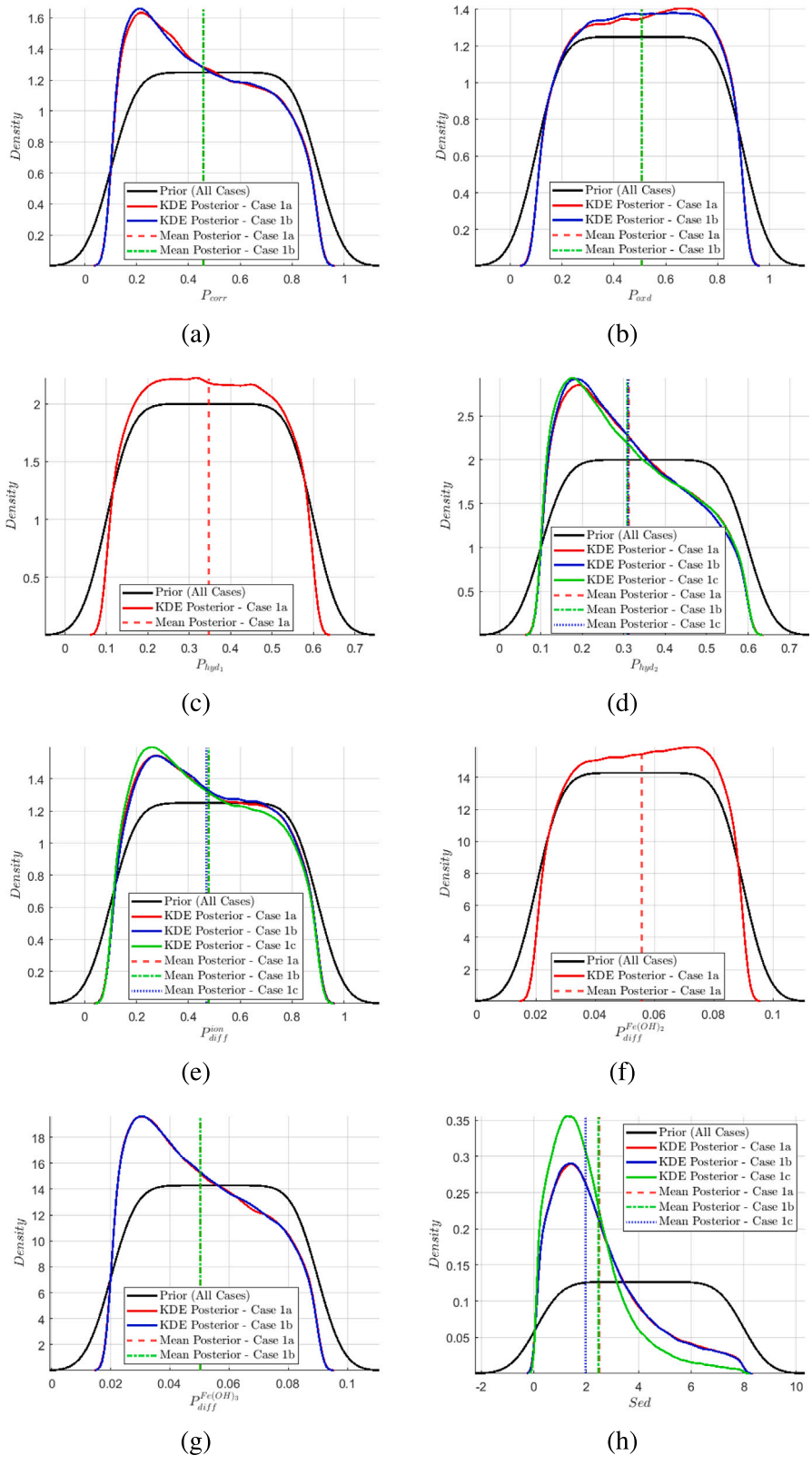


Fig. 8. Posterior distributions of all physics model parameters across the three cases of two-feature calibration. Each subplot corresponds to one model parameter: (a) P_{corr}, (b) P_{oxd}, (c) P_{hyd_1}, (d) P_{hyd_2}, (e) $P_{\text{diff}}^{\text{ion}}$, (f) $P_{\text{diff}}^{\text{Fe(OH)}_2}$, (g) $P_{\text{diff}}^{\text{Fe(OH)}_3}$, and (h) Sed.

parameters, though not dominant individually, were more identifiable when the confounding influence of low-sensitivity parameters was removed.

Model discrepancy terms

Discrepancy modeling provided additional insight into the structural limitations of the model. The mean discrepancy for C_1 ($\mu_{\delta_{C_1}}$) was

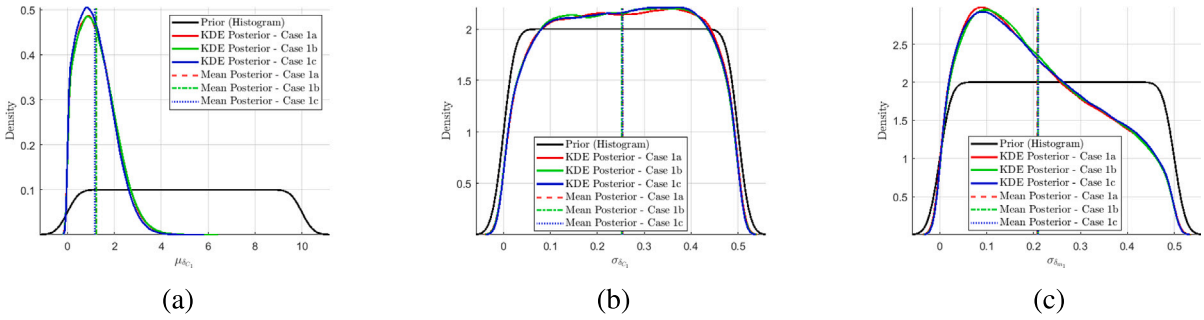


Fig. 9. Posterior distributions of model discrepancy hyperparameters inferred during two-feature calibration. Each subplot corresponds to a different hyperparameter: (a) $\mu_{\delta_{C_1}}$, (b) $\sigma_{\delta_{C_1}}$, and (c) $\sigma_{\delta_{m_1}}$.

significantly updated across all three cases, consistently indicating the presence of a systematic underprediction by the computational model. This finding supports the inclusion of additive discrepancy terms to account for structural bias in the simulator.

The variance components of the discrepancy terms exhibited distinct behaviors across the two calibrated power-law coefficients. The posterior distribution of $\sigma_{\delta_{C_1}}$ remained broadly similar to its uniform prior, indicating that this term was weakly constrained by the calibration data. This suggests that while the model exhibited a systematic bias in predicting C_1—captured effectively by the mean discrepancy term $\mu_{\delta_{C_1}}$—the available observations provided limited information to resolve the extent of its residual variability.

In contrast, the posterior of $\sigma_{\delta_{m_1}}$ showed pronounced contraction from the prior across all three calibration cases, with consistent and well-defined peaks. This reflects that the data were sufficiently informative to constrain the residual uncertainty in the growth exponent m_1. The inclusion of $\sigma_{\delta_{m_1}}$ therefore plays a key role in quantifying epistemic uncertainty, enabling the model to represent the spread in m_1 predictions without attributing it to structural bias. Together, these results underscore the benefit of selectively incorporating feature-specific discrepancy terms based on empirical evidence of residual behavior.

3.1.7. Case 2: Calibration using both pit depth and aspect ratio

In Case 2, Bayesian calibration was implemented to incorporate not only pit depth but also pit aspect ratio. Four experimentally derived features—C_1, m_1, C_2, and m_2—which jointly characterize both the initial magnitude and temporal evolution of the *maximum pit depth* and *maximum pit aspect ratio* were utilized in this calibration. It was investigated whether this expanded set of outputs allows for a more comprehensive calibration and enables improved predictive performance of the model.

To systematically evaluate the effect of parameter dimensionality on calibration quality, the same three calibration configurations described in the two-feature case were reused: the full 8-parameter model (Case 2a), a reduced 6-parameter model (Case 2b) excluding parameters with minimal influence, and a 3-parameter model (Case 2c) focused on the most influential factors based on global sensitivity and correlation analyses.

All Bayesian inference components—prior samples using LHS, discrepancy modeling, likelihood formulation, and posterior estimation via Metropolis-Hastings MCMC—were implemented consistently with the approach used for the two-feature case. KDE was employed to visualize posterior distributions and assess contraction around plausible parameter regions. An acceptance rate of approximately 40.08% was achieved for Case 2a. For Case 2b, the acceptance rate increased to 47.43%, and for Case 2c, it further improved to 61.03%. These trends closely mirror those observed in the two-feature calibration (Case 1), where acceptance rates increased from 45.01% in Case 1a to 67.70% in Case 1c. The slightly lower acceptance rates observed in Case 2 are likely due to the additional constraint introduced by incorporating aspect ratio as a calibration target, which tightens the posterior distribution and reduces the probability of accepting proposed samples.

Formulation of the discrepancy term

To prevent the overfitting in Bayesian calibration and account for potential structural inadequacies, the discrepancy term adopted in the two-feature calibration—comprising $\mu_{\delta_{C_1}}$, $\sigma_{\delta_{C_1}}$, and $\sigma_{\delta_{m_1}}$—is again pursued here to include additional terms targeting the newly introduced power-law coefficients, C_2 and m_2.

- **For** C_2, a similarly skewed distribution was observed (as with C_1, shown in Fig. 3(c)), necessitating both a mean discrepancy term $\mu_{\delta_{C_2}}$ to capture systematic bias and a variance term $\sigma_{\delta_{C_2}}$ to model variability in prediction uncertainty.
- **For** m_2, while the distribution remained centered near the experimental value (shown in Fig. 3(d)), its visible spread justified the inclusion of a variance discrepancy term $\sigma_{\delta_{m_2}}$, analogous to the approach used for m_1.

Posterior distributions and parameter insights

Posterior distributions obtained from the 4-feature calibration (C_1, m_1, C_2, m_2) further illuminate trends in parameter identifiability under increased observational features, as shown in Fig. 10. Note that the additional consideration of aspect ratio serves as a constraint on the pit geometry and makes the calibration information more realistic; not considering the aspect ratio will make the calibrated model unrealistic in predicting the pit geometry evolution.

Across all three calibration cases, the `Sed` parameter continued to exhibit strong posterior contraction, with even sharper peaks and further shifts in the posterior mean compared to the 2-feature model. This reaffirms its critical role in governing both pit depth and pit aspect ratio dynamics. A similar pattern was observed for $P_{\text{diff}}^{\text{ion}}$, whose posteriors remained tightly concentrated, particularly in Cases 2 and 3, where dimensionality was reduced. The consistency in sharp contraction across both models emphasizes the parameter's dominant influence.

Notably, with the inclusion of C_2 and m_2 as calibration output QoIs, $P_{\text{diff}}^{\text{Fe(OH)}_3}$ and P_{hyd2} exhibited clearer identifiability than in the 2-feature case. Their posterior distributions showed greater narrowing and shifted means, especially in Case 2, indicating enhanced sensitivity under richer observational data.

In contrast, P_{hyd1} and $P_{\text{diff}}^{\text{Fe(OH)}_2}$ continued to show minimal contraction from their priors in Case 1, confirming their limited influence on all four power-law coefficients. Their exclusion in reduced Cases 2 and 3 remains well-justified.

Interestingly, P_{corr} and P_{oxd}, which had limited identifiability under the 2-feature model, showed improved but modest contraction in the 4-feature setup, particularly in Case 2. This suggests a weak but non-negligible correlation with the additional outputs, though their role

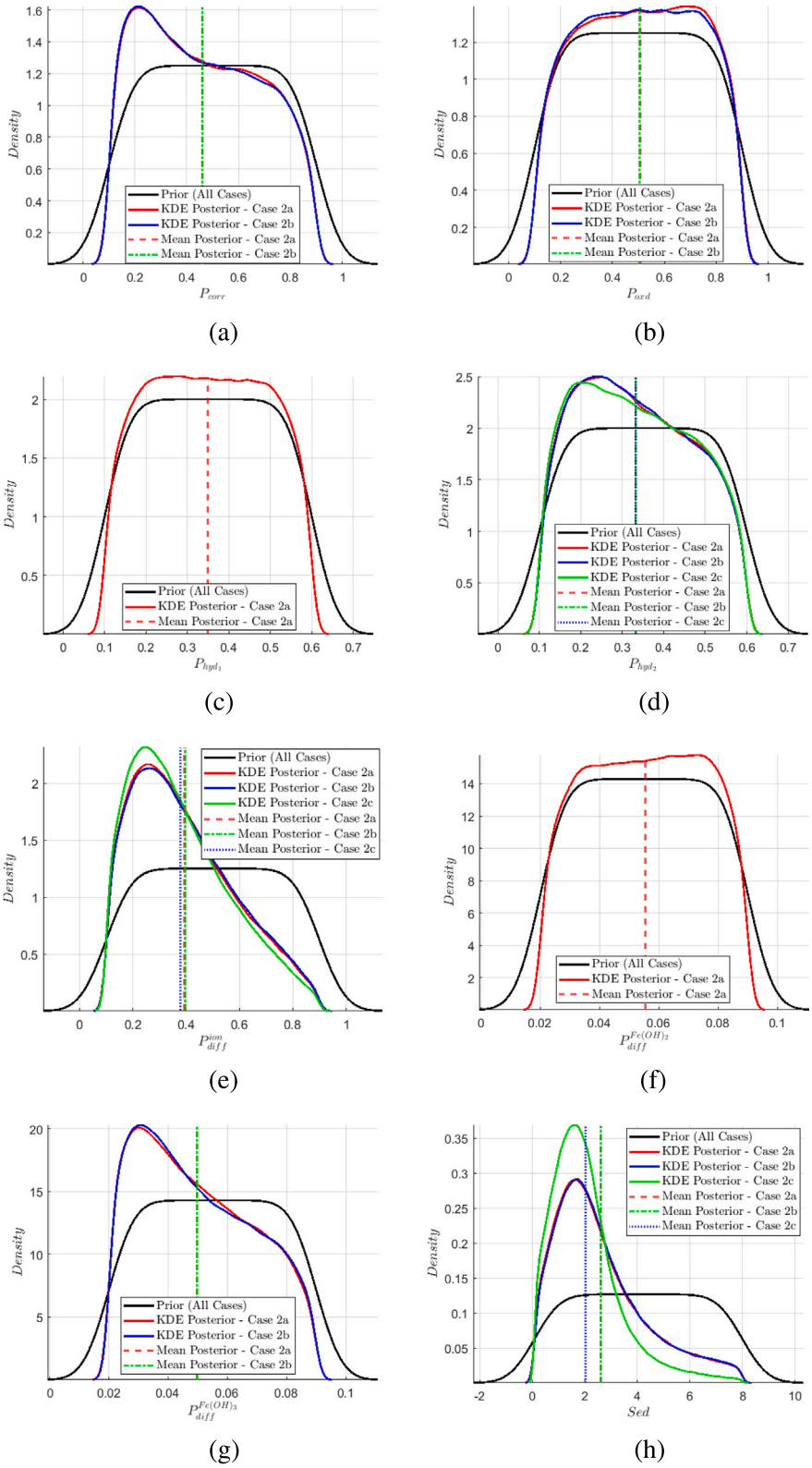


Fig. 10. Posterior distributions of all physics model parameters across the three cases of four-feature calibration. Each subplot corresponds to one model parameter: (a) P_{corr}, (b) P_{oxd}, (c) P_{hyd_1}, (d) P_{hyd_2}, (e) $P_{\text{diff}}^{\text{ion}}$, (f) $P_{\text{diff}}^{\text{Fe(OH)}_2}$, (g) $P_{\text{diff}}^{\text{Fe(OH)}_3}$, and (h) Sed.

remains secondary to the more influential transport and sedimentation parameters.

Overall, the 4-feature calibration not only reinforced conclusions drawn from the 2-feature case but also sharpened inference on a broader subset of parameters. The added observational richness led to clearer posterior contraction and improved distinction between influential and non-influential parameters across all calibration cases.

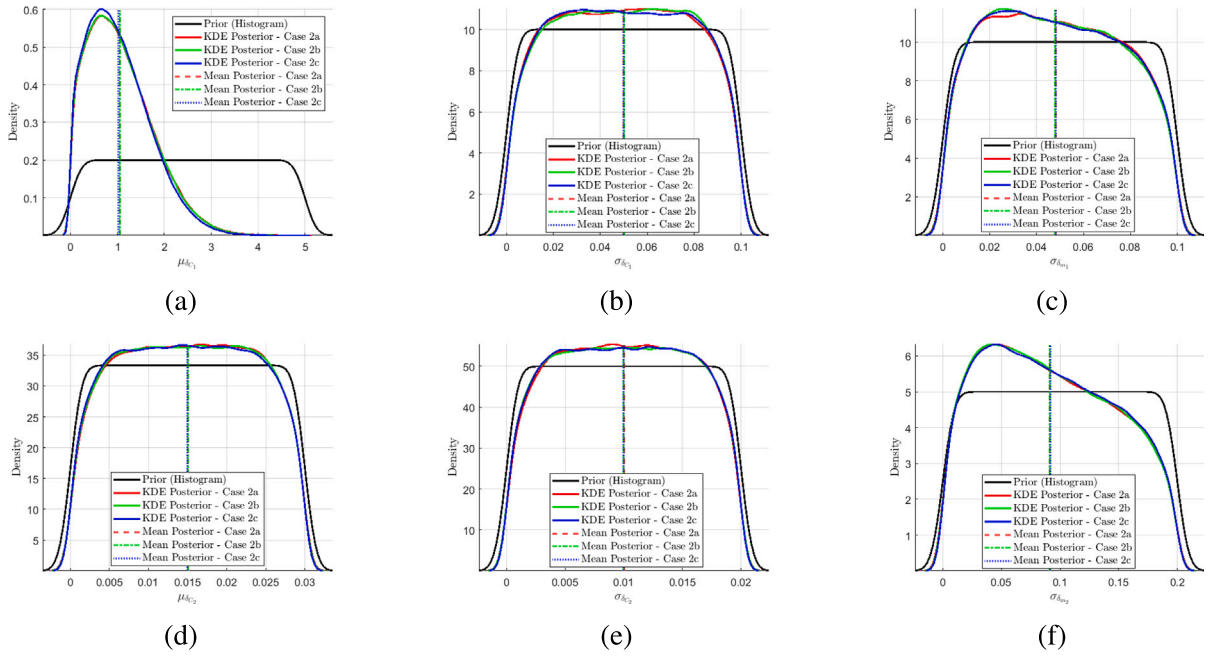


Fig. 11. Posterior distributions of model discrepancy hyperparameters inferred during four-feature calibration. Each subplot corresponds to a different hyperparameter: (a) $\mu_{\delta_{C_1}}$, (b) $\sigma_{\delta_{C_1}}$, (c) $\sigma_{\delta_{m_1}}$, (d) $\mu_{\delta_{C_2}}$, (e) $\sigma_{\delta_{C_2}}$, and (f) $\sigma_{\delta_{m_2}}$.

Model discrepancy terms

As in the 2-feature model, the mean discrepancy for C_1 ($\mu_{\delta_{C_1}}$) showed substantial posterior contraction in all three calibration cases, reinforcing the earlier finding of systematic underprediction in maximum pit depth by the surrogate. For C_1, the variance component $\sigma_{\delta_{C_1}}$ again displayed weak identifiability, with posteriors closely resembling the uniform prior. In contrast, the discrepancy variance for m_1 ($\sigma_{\delta_{m_1}}$) continued to exhibit small posterior contraction and shift toward the left, indicating that the calibration data were sufficient to resolve its stochastic uncertainty.

For the additional power-law coefficients, new discrepancy terms were introduced, but these exhibited limited posterior updates (shown in Fig. 11):

- The mean discrepancy for C_2 ($\mu_{\delta_{C_2}}$) remained nearly uniform across all cases, suggesting that the posterior distribution was not much different from the prior. This indicates a lack of significant evidence for systematic bias in the surrogate prediction of C_2. There was no significant evidence of systematic bias in the surrogate prediction of C_2.
- Similarly, the variance discrepancy term $\sigma_{\delta_{C_2}}$ showed minimal departure from the prior, indicating that the observed spread in C_2 could not be strongly constrained by the available data.
- The variance discrepancy for m_2 ($\sigma_{\delta_{m_2}}$) showed modest posterior contraction, reflecting some sensitivity to the observed variability, though less than that observed for m_1.

These results suggest that while discrepancy terms for C_1 and m_1 were critical and well-informed by the data, the added discrepancy terms for C_2 and m_2 contributed less to model refinement. Nonetheless, their inclusion is necessary to prevent overfitting as discussed earlier.

3.1.8. Posterior prediction of pit geometry evolution

The posterior samples of model parameters from the Bayesian calibration are used to predict the pit morphology under two configurations: (i) the 2-feature model, calibrated solely using maximum pit depth evolution (C_1, m_1), and (ii) the 4-feature model, which incorporates both maximum pit depth and maximum pit aspect ratio evolution (C_1, m_1, C_2, m_2). The posterior samples were propagated through the GPR surrogates to generate forward predictions, capturing the impact of parameter uncertainty on the model's output. In the following, we first present the posterior predictive distributions for these two configurations before reconstructing how the maximum pit depth and maximum pit aspect ratio propagate.

Posterior predictions for two power-law coefficients

To better understand the model's performance, we examine the posterior predictive distributions for the two power-law coefficients, C_1 and m_1. These distributions are analyzed across three calibration cases, each corresponding to different parameter dimensionalities. The objective is to assess how the inclusion of additional parameters influences the model's ability to predict these two power-law coefficients. Case 1, which involves 8 parameters, provides a broad perspective, while Cases 2 (6 parameters) and 3 (3 parameters) progressively reduce dimensionality, enabling a more focused evaluation of the parameter interactions and their impact on predictive accuracy.

The posterior predictive distributions for the two power-law coefficients, C_1 and m_1, reveal distinct sensitivities to parameter dimensionality. For C_1, the predictive distributions across all three calibration cases exhibit near-identical shapes, widths, and alignment with the experimental value, as shown in the Fig. 12. All posteriors consistently peak around the observed data, and show a rightward shift from the prior distribution, regardless of whether 8, 6, or 3 parameters are inferred. This consistency indicates that C_1 is robustly identifiable even in higher-dimensional spaces, and its predictive behavior is largely insensitive to the inclusion of low-influence parameters.

In contrast, the predictive distributions for m_1 show a marked improvement in sharpness and accuracy with progressive parameter reduction. In all cases, the distribution is slightly offset from the experimental data, with Case 3 (3 parameters) producing a sharply peaked distribution. This trend suggests that m_1 is more sensitive to parameter interaction and benefits from a reduced-dimensional calibration setup. These findings highlight the value of sensitivity-informed dimensionality reduction: while it preserves the predictive fidelity of well-identified

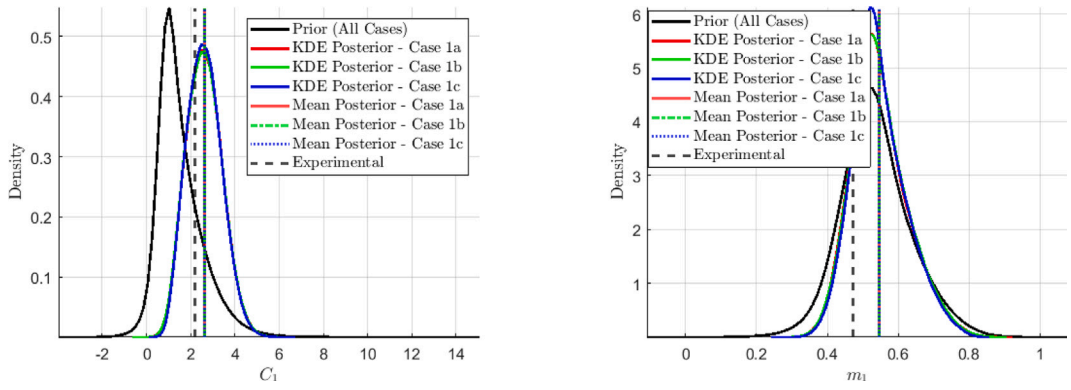


Fig. 12. Posterior predictions for the two power-law coefficients: (a) C_1, and (b) m_1, across the three calibration cases.

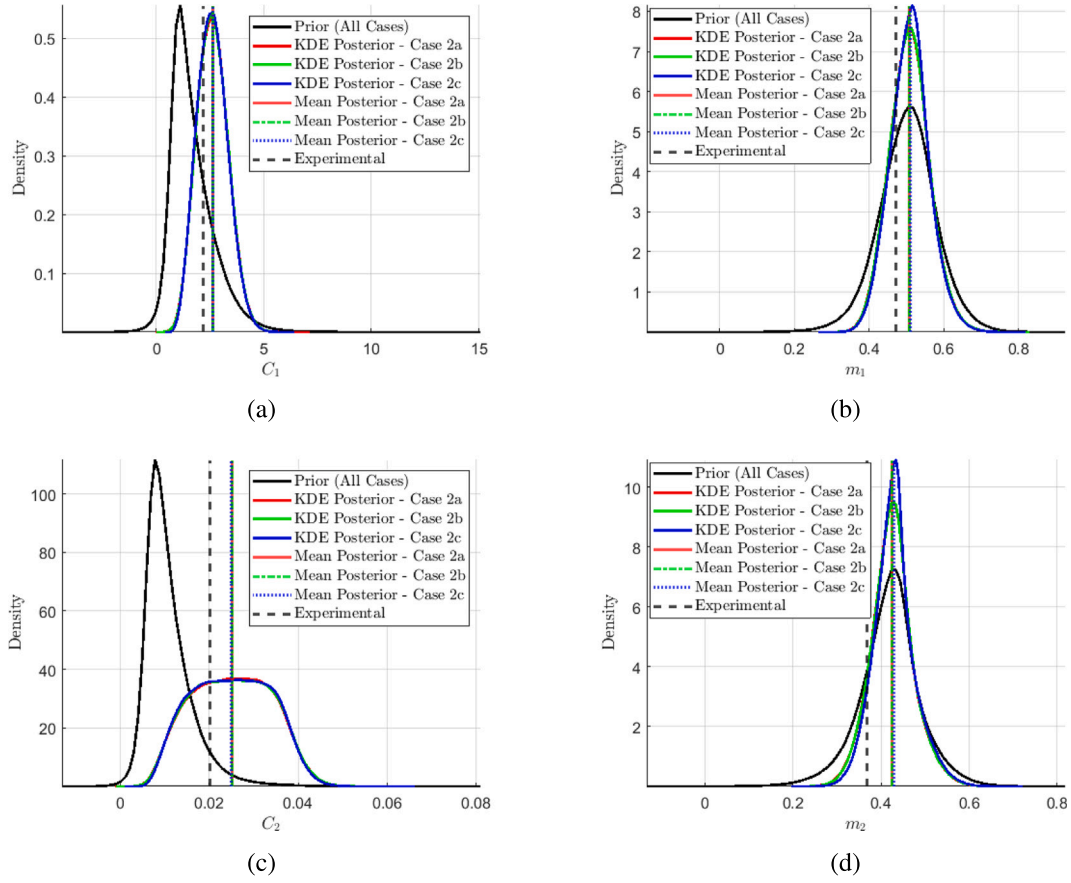


Fig. 13. Posterior predictive distributions for the four power-law coefficients, (a) C_1, (b) m_1, (c) C_1, (d) m_1, across the three calibration cases.

features like C_1, it significantly enhances the precision and identifiability of features like m_1 (Fig. 12).

Posterior predictions for four power-law coefficients

Here, we expand the analysis to the four power-law coefficients—C_1, m_1, C_2, and m_2—to explore how the inclusion of C_2 and m_2 influences the predictive distributions and model uncertainty. The posterior predictive distributions provide insight into the model's predictive capabilities under varying parameter dimensionalities across all three calibration cases (see Fig. 13).

As discussed in the 2-feature model, the predictive distribution for C_1 showed strong contraction and close alignment with the experimental value. In the 4-feature model, this behavior is retained and further sharpened—posteriors across all three cases remain consistent and even more concentrated. This indicates that incorporating additional outputs in the calibration enhances the precision of predictions for C_1 without degrading robustness, reinforcing its reliability as a model output.

Across both the two-feature and four-feature calibration settings, the posterior predictive distributions for m_1 show consistent narrowing relative to the prior, indicating effective constraint by the data. In all three calibration cases, the posteriors are slightly to the right of the experimental value, with progressively sharper peaks observed from Case 1 through Case 3. Comparing models, the four-feature calibration achieves even more pronounced posterior contraction than the

two-feature setup, particularly in Cases 2 and 3. This improvement suggests that incorporating additional outputs (such as C_2 and m_2) enhances the overall information content available to constrain m_1, thereby refining its predictive resolution.

The posterior predictions for C_2—evaluated only in the four-feature calibration model—exhibit a noticeable shift toward the experimental value, indicating successful bias correction through calibration. However, unlike the prior, which is sharply peaked and underpredicts the experimental observation, the posteriors are broader and less peaked, suggesting that while calibration improves accuracy, it also acknowledges greater epistemic uncertainty in this output. This behavior highlights that C_2 is moderately identifiable: informative enough to shift the prediction toward experimental reality, yet not sharply constrained across the parameter space. The consistency of the posterior shapes across Cases 1, 2, and 3 further suggests robustness of the inference, albeit with limited precision.

The posterior predictive distributions for the output feature m_2 exhibit clear narrowing compared to the broad prior, indicating improved precision due to calibration. However, the posterior modes remain closely aligned with the prior mean rather than shifting toward the experimental value. This suggests that, despite sharper predictions, the model has limited ability to adjust m_2 toward the observed data. The mean of the posterior distribution remains nearly unchanged from the prior mean, highlighting a persistent mismatch that calibration fails to resolve. This behavior implies that m_2 is only weakly sensitive to the calibrated parameters, and that further improvements may require enhancing model structure or introducing additional influential inputs. The feature's predictive spread is reduced, but its central tendency still deviates from experimental observations.

Fig. 14(a) compares the maximum pit-depth prediction intervals from the 2-feature and 4-feature calibrations, while Fig. 14(b) presents the corresponding uncertainty bounds for the aspect ratio from the 4-feature model.

The 2-feature calibration leads to noticeably wide prediction intervals and consistent overestimation of pit depth, indicating that the model remains weakly constrained when only vertical growth kinetics (C_1, m_1) are used. These broad intervals reflect insufficient information about lateral pit evolution, limiting the model's reliability for applications such as structural integrity assessment, where overly conservative depth predictions may distort risk estimates.

In contrast, the 4-feature calibration produces substantially narrower uncertainty bounds and mean predictions that closely follow the experimental trajectory. By jointly constraining both depth and aspect ratio kinetics, the posterior distributions become more concentrated, reducing overall predictive uncertainty. As shown in Fig. 14(b), the 4-feature model also captures lateral spreading behavior—an essential aspect of pit severity that the 2-feature model cannot represent.

Overall, these results demonstrate that incorporating multiple geometric descriptors leads to more realistic and practically usable uncertainty bounds. The 4-feature calibration provides a more reliable foundation for downstream tasks such as remaining-life estimation, inspection planning, and quantifying the likelihood of extreme pit morphologies, whereas the 2-feature calibration is underspecified for such applications.

3.1.9. Model validation

As mentioned in Section 2.5, a probabilistic model validation metric, namely the model reliability metric, was computed using the predicted maximum pit depth $\rho_{d_{\max}}(t)$ as the QoI, and measurements were taken using a Keyence VK-X200K 3D Confocal Microscope [17], with a measurement error of approximately 5–10 μm. To account for both measurement and modeling uncertainty, the tolerance was conservatively set to at least 3 times the measurement error ($\delta_{tol} =\geq 30\ \mu$m). The computed $P(H_0)(t)$ (presented in Section 2.5) values over time were then averaged to obtain the reliability score for each model. To validate the predictive consistency of the calibrated models, the model reliability metric $P(H_0)(t)$ was computed for both the 2-feature and 4-feature models.

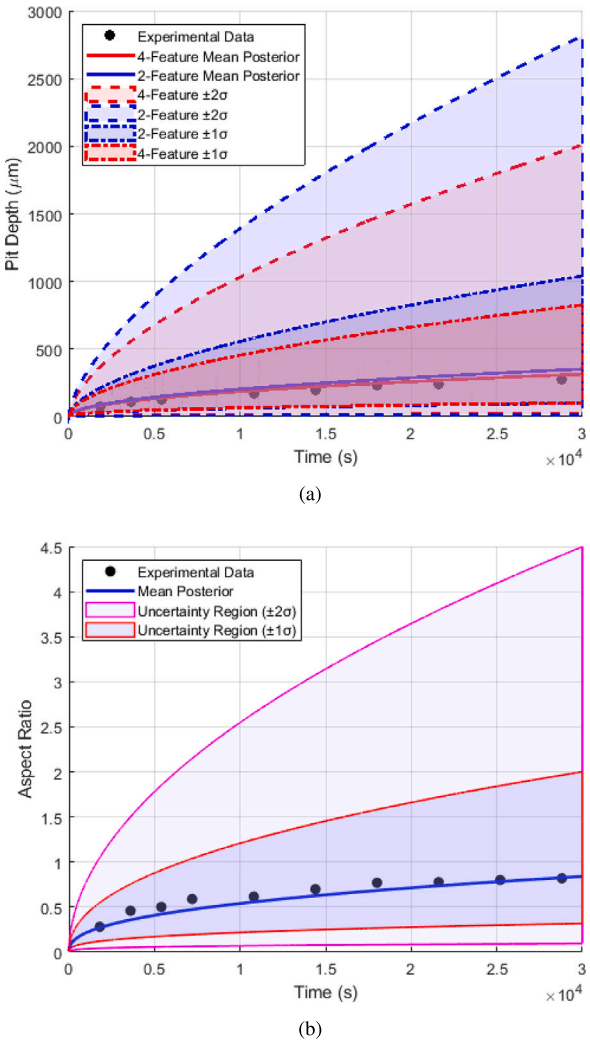
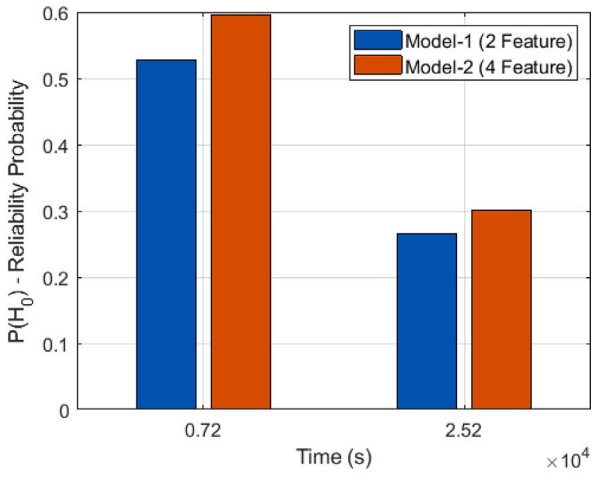


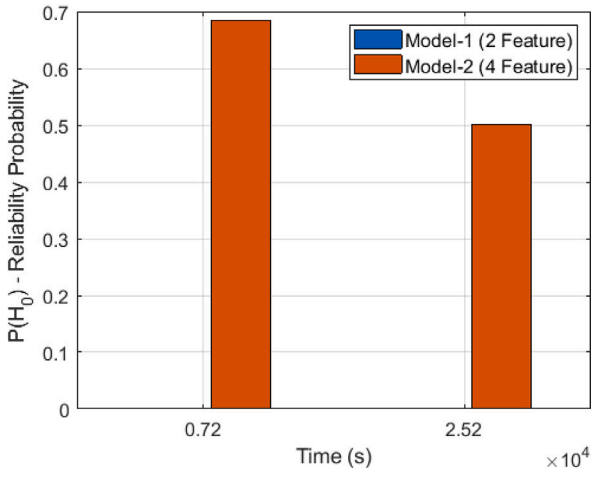
Fig. 14. UQ results using posterior distributions from Bayesian calibration. (a) Maximum pit-depth predictions comparing 2-feature and 4-feature models. (b) Maximum aspect ratio predictions with $\pm1\sigma$ and $\pm2\sigma$ bands from the 4-feature model. Experimental data are shown as black dots; shaded regions indicate predictive uncertainty.

Fig. 15(a) compares the computed model reliability $P(H_0)(t)$ for two surrogate models—one constructed using only pit depth features (2-feature model) and the other using both pit depth and aspect ratio features (4-feature model)—at two representative time instances, $t = 7200$ s and $t = 25200$ s. The metric $P(H_0)$ quantifies the likelihood that a model prediction lies within a predefined tolerance $\delta_{tol} = 30\mu$m of the experimental observations. The results show that, at both time points, the 4-feature model consistently exhibits higher reliability than the 2-feature model. Specifically, the 2-feature model yields a mean reliability of $\bar{P}_{H_0} = 0.39$, whereas the 4-feature model achieves a higher average of $\bar{P}_{H_0} = 0.45$.

In the case of aspect ratio prediction (Fig. 15(b)), the reliability metric $P(H_0)$ was computed using a relative tolerance defined as $\delta_{tol} = 0.1 \times \max(\text{Aspect Ratio})$, corresponding to 10% of the maximum experimental aspect ratio value. Under this criterion, the difference between the two models is even more pronounced. The 2-feature model, which does not include aspect ratio in its calibration, shows near-zero reliability across both time points. This is visually evident from the absence of blue bars in the plot. In contrast, the 4-feature model demonstrates



(a)



(b)

Fig. 15. Model reliability metric $P(H_0)$ computed at two time instants (t = 7200 s and t = 25200 s). (a) Reliability for pit depth prediction. (b) Reliability for pit aspect ratio prediction. In both cases, results are shown for models calibrated using 2 features (blue bars: pit depth only) and 4 features (red bars: both pit depth and aspect ratio). **Note:** In (b), the blue bars are barely visible due to near-zero values. For clarity, the actual $P(H_0)$ values for Model 1 are 0.0001 at t = 7200 s and 0.005 at t = 25200 s.

substantial reliability, highlighting the importance of incorporating aspect ratio in the calibration process when aspect ratio prediction is of interest.

This improvement demonstrates that incorporating multiple geometric features—namely, both maximum pit depth and aspect ratio—enhances the predictive accuracy of the model when calibrated against experimental data. As time advances, both models show a decline in $P(H_0)$, highlighting the increased challenge of maintaining predictive accuracy during later stages of corrosion where morphological complexity becomes more pronounced. This trend also suggests that more extensive experimental datasets may be necessary for improving long-term reliability, as the present analysis is limited to a single set of reference data.

4. Conclusion

This study developed a Bayesian framework for calibrating a high-fidelity cellular automata (CA) model of pitting corrosion in API-5L X65 steel and quantifying uncertainty in its predictions. To address the computational cost of repeated CA simulations, a feature-based surrogate modeling approach was employed in which the time evolution of pit depth and aspect ratio was reduced to four power-law coefficients (C_1, m_1, C_2, m_2), enabling efficient correlation analysis, global sensitivity analysis, surrogate construction, Bayesian inference, and uncertainty propagation.

Pearson correlation and Sobol' sensitivity analyses identified sedimentation (Sed) and ionic diffusion ($P_{\text{diff}}^{\text{ion}}$) as the dominant parameters influencing pit morphology, with differences between total and first-order indices revealing significant nonlinear interaction effects among dissolution, hydrolysis, diffusion, and sedimentation. Gaussian Process Regression (GPR) surrogate models achieved strong predictive performance across all four power-law coefficients ($R^2 > 0.87$; MAPE < 8%), effectively eliminating the need for repeated high-fidelity simulations during calibration and uncertainty propagation. Sensitivity-informed parameter reduction further improved calibration efficiency, as low-impact inputs such as P_{hyd1} and $P_{\text{diff}}^{\text{Fe(OH)}_2}$ were excluded from Cases 1b and 1c without degrading predictive accuracy.

Influential parameters including Sed, $P_{\text{diff}}^{\text{ion}}$, and P_{hyd_2} exhibited the strongest updates in calibrated values, with the most compact posteriors obtained for Case 1c. Acceptance rates in the Metropolis–Hastings sampler increased as the number of calibrated parameters decreased (8→6→3), consistent with a lower-dimensional posterior space, while slightly lower acceptance rates in Case 2 reflected constraints introduced by aspect-ratio features. Incorporating all four features (C_1, m_1, C_2, m_2) produced narrower posterior distributions, improved bias correction for C_2, and yielded more consistent pit-morphology predictions; intermediate-sensitivity parameters such as P_{hyd2} and $P_{\text{diff}}^{\text{Fe(OH)}_3}$ also exhibited increased contraction, although m_2 remained weakly constrained. Bayesian calibration and subsequent uncertainty propagation demonstrated improved predictive behavior in the four-feature case, and reliability-based validation using the metric $P(H_0)$ confirmed this improvement, with the four-feature model achieving a higher average reliability score ($\bar{P}(H_0) = 0.4481$) than the two-feature model (0.3966).

A key limitation of the study is the reliance on a single experimental dataset, which restricts the diversity of observable pit morphologies and limits generalizability across environmental conditions; future work should incorporate multiple datasets or time-resolved measurements to enhance parameter identifiability and predictive robustness. Additionally, the current framework employs a two-dimensional CA model, whereas pitting corrosion is inherently three-dimensional. Extending the framework to a fully 3D representation would enable more realistic pit geometries and evolution patterns while necessitating revised sensitivity analysis and surrogate-modeling strategies due to increased geometric complexity.

In summary, the proposed methodology provides an integrated uncertainty quantification framework for localized corrosion modeling. The approach is generalizable and can be extended to stress-assisted [17] and fatigue-assisted pitting corrosion [4], supporting quantitative assessment of pit-to-crack transition and full corrosion-life estimation. By combining physics-based modeling, surrogate construction, and Bayesian inference, the framework offers a structured basis for reliability assessment and maintenance planning in corrosion-critical infrastructure.

CRediT authorship contribution statement

J. Ramesh Babu: Writing – review & editing, Conceptualization of this study, Methodology, Software. **Pranav M. Karve:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Sankaran Mahadevan:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was partially funded by the US Department of Energy, Idaho National Laboratory (Contract DE-AC07-05ID14517, Project Monitor: Svetlana Lawrence). The support is gratefully acknowledged.

Data availability

Data will be made available on request.

References

[1] Ao D, Hu Z, Mahadevan S. Dynamics model validation using time-domain metrics. J Verif Valid Uncertain Quant Mar 2017;2(1):011004. https://doi.org/10.1115/1.4036182
[2] Aziz I, Soltanaghai E, Watts A, Alipour M. Bayesian inversion of gpr waveforms for sub-surface material characterization: an uncertainty-aware retrieval of soil moisture and overlaying biomass properties. Remote Sens Environ 2024;313:114351. https://doi.org/10.1016/j.rse.2024.114351
[3] Babu JR, Gopalakrishnan S. Synergistic approach: peridynamics and machine learning regression for efficient pitting corrosion simulation. Computers Struct Dec 2024;305. https://doi.org/10.1016/j.compstruc.2024.107588
[4] Babu JR, Gopalakrishnan S, Fatoba OO, Leiva-Garcia R, Akid R. Modeling fatigue-assisted pitting in pipeline steel with pca-xfem. Corros Sci Mar 2024;228. https://doi.org/10.1016/j.corsci.2023.111772
[5] Babu JR, Karve PM, Mahadevan S. Probabilistic prediction of external corrosion fatigue life in buried steel pipes. Int J Press Vessels Piping Apr 2025;214. https://doi.org/10.1016/j.ijpvp.2024.105415
[6] Bansal P, Zheng Z, Shao C, Li J, Banu M, Carlson BE, Li Y. Physics-informed machine learning assisted uncertainty quantification for the corrosion of dissimilar material joints. Reliab Eng Syst Saf Nov 2022;227:108711. https://doi.org/10.1016/j.ress.2022.108711
[7] Borden MJ, Hughes TJR, Landis CM, Anvari A, Lee IJ. A phase-field formulation for fracture in ductile materials: finite deformation balance law derivation, plastic degradation, and stress triaxiality effects. Comput Methods Appl Mech Eng Dec 2016;312:130–66. https://doi.org/10.1016/j.cma.2016.09.005
[8] Chen B, Shen L, Zhang H. Gaussian process regression-based material model for stochastic structural analysis. ASCE-ASME J Risk Uncertain Eng Syst Part A Civ Eng 2021;7(3):04021025. https://ascelibrary.org/doi/10.1061/AJRUA6.0001145.
[9] Chen PY, Popovich PM. Correlation: parametric and nonparametric measures. Thousand Oaks, CA: Sage Publications; 2002.
[10] Chen Z, Bobaru F. Peridynamic modeling of pitting corrosion damage. J Mech Phys Solids May 2015;78:352–81. https://doi.org/10.1016/j.jmps.2015.02.015
[11] Cui C, Ma R, Chen A, Pan Z, Tian H. Experimental study and 3d cellular automata simulation of corrosion pits on q345 steel surface under salt-spray environment. Corros Sci 2019;154:80–9. https://doi.org/10.1016/j.corsci.2019.03.011
[12] Daviddi E, Shkirskiy V, Kirkman PM, Robin MP, Bentley CL, Unwin PR. Screening the surface structure-dependent action of a benzotriazole derivative on copper electrochemistry in a triple-phase nanoscale environment. J Phys Chem C 2022;126(35):14897–907. https://doi.org/10.1021/acs.jpcc.2c04494
[13] Deringer VL, Bartók AP, Bernstein N, Wilkins DM, Ceriotti M, Csányi G. Gaussian process regression for materials and molecules. Chem Rev 2021;121(16):10073–141. https://pubs.acs.org/doi/10.1021/acs.chemrev.1c00022.
[14] Deshpande S, Rappel H, Hobbs M, Bordas SPA, Lengiewicz J. Gaussian process regression + deep neural network autoencoder for probabilistic surrogate modeling in nonlinear mechanics of solids. Comput Methods Appl Mech Eng 2025;437:117790. https://doi.org/10.1016/j.cma.2025.117790
[15] Duddu R. Numerical modeling of corrosion pit propagation using the combined extended finite element and level set method. Computational Mech 2014;54:613–27. https://doi.org/10.1007/s00466-014-1010-8
[16] Fan Y, Pan J, Wu Z, Li B, Wang Z. A method for predicting the corrosion behavior of structural steel under atmosphere. Buildings 2023;13(253). https://doi.org/10.3390/buildings13010253
[17] Fatoba OO, Leiva-Garcia R, Lishchuk SV, Larrosa NO, Akid R. Simulation of stress-assisted localised corrosion using a cellular automaton finite element approach. Corros Sci 2018;137:83–97. https://doi.org/10.1016/j.corsci.2018.03.029
[18] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis. 3rd ed. Boca Raton, FL: Chapman and Hall/CRC; 2013.
[19] Gnefid S, Akid R. The effects of pits size and shape on fatigue crack initiation. J Pure Appl Sci 2021;20(2):195–203. https://doi.org/10.51984/jopas.v20i2.1176
[20] Guo X, Gin S, Liu H, Ngo D, Luo J, Kim SH, Frankel GS. Near-field corrosion interactions between glass and corrosion resistant alloys. NPJ Mater Degrad 2020;4(1). https://doi.org/10.1038/s41529-020-0114-1
[21] Hastings WK. Monte carlo sampling methods using markov chains and their applications. Biometrika 1970;57(1):97–109. https://doi.org/10.2307/2334940. https://www.jstor.org/stable/2334940.
[22] Helton JC, Davis FJ. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Technical Report SAND2001-0417, Albuquerque, NM and Livermore, CA: Sandia National Laboratories; 2002.
[23] Hou CKJ, Behdinan K. Dimensionality reduction in surrogate modeling: a review of combined methods. Data Sci Eng 2022;7:402–27. https://doi.org/10.1007/s41019-022-00193-5
[24] Hüsser AP. Correlation analysis. In: The international encyclopedia of communication research methods. Wiley-Blackwell; 2017. https://doi.org/10.1002/9781118901731.iecrm0048
[25] Kruger J, Hardman VK. Current understanding of pitting and crevice corrosion and its application to test methods for determining the susceptibility to such corrosion of nuclear waste metallic containers. Technical Report NBSIR 82-2477, Washington, DC, USA: National Bureau of Standards, U.S. Department of Commerce. Mar 1982.
[26] Kapusuzoglu B, Mahadevan S, Matsumoto S, Miyagi Y, Watanabe D. Multi-level Bayesian calibration of a multi-component dynamic system model. J Comput Inf Sci Eng Feb 2023;23(1):011006. https://doi.org/10.1115/1.4055315
[27] Larrosa NO, Akid R, Ainsworth RA. Corrosion-fatigue: a review of damage tolerance models. Int Mater Rev 2018;63(5):283–308. https://doi.org/10.1080/09506608.2017.1375644
[28] Laycock NJ, White SP, Noh JS, Wilson PT, Newman RC. Perforated covers for propagating pits. J Electrochem Soc 1998;145(4):1101. https://doi.org/10.1149/1.1838423
[29] Li X, Shao Y, Miao W, Liu Y, Gao Z, Liu C. Galvanic corrosion behaviors of the low-carbon ferritic stainless steel erw (electrical resistance welding) joint in the simulated seawater. Anti-Corros Method Mater 2020;67(3):269–79. https://doi.org/10.1108/acmm-01-2020-2247
[30] Ling Y, Mahadevan S. Quantitative model validation techniques: new insights. Reliab Eng Syst Saf Mar 2013;111:217–31. https://doi.org/10.1016/j.ress.2012.11.011
[31] Ling Y, Mullins J, Mahadevan S. Selection of model discrepancy priors in Bayesian calibration. J Comput Phys Nov 2014;276:665–80. https://doi.org/10.1016/j.jcp.2014.08.005
[32] Lishchuk SV, Akid R, Worden K, Michalski J. A cellular automaton model for predicting intergranular corrosion. Corros Sci 2011;53:2518–26. https://doi.org/10.1016/j.corsci.2011.04.027
[33] Liu Y, Li L, Zhao S, Song S. A global surrogate model technique based on principal component analysis and kriging for uncertainty propagation of dynamic systems. Reliab Eng Syst Saf Mar 2021;207:107365. https://doi.org/10.1016/j.ress.2020.107365
[34] Liu Z, Xu Y, Jiang Y, Renteria A, Bansal P, Xu C, Wang P, Li Y. Uncertainty quantification of additively manufactured architected cellular materials for energy absorption applications. ASME J Risk Uncertain Eng Syst Part B Mech Eng Sep 2025;11(3):031204. https://doi.org/10.1115/1.4066933
[35] Loh W-L. On latin hypercube sampling. Ann Stat 1996;24(5):2058–80. Research supported in part by NSA Grant MDA 904-93-H-3011.
[36] Lopes BLS, Rodrigues SF, Silva ES, Reis GS, Martins WDP, Damascena JC, Leal VS. Influence of mig/mag welding process on mechanical and pitting corrosion behaviors on the super-duplex stainless steel SAF 2507 welded joints. Mater Sci Appl 2018;9(2):228–45. https://doi.org/10.4236/msa.2018.92015
[37] Mai W, Soghrati S. New phase field model for simulating galvanic and pitting corrosion processes. Electrochim Acta Jan 2018;260:290–304. https://doi.org/10.1016/j.electacta.2017.12.086
[38] Mai W, Soghrati S, Buchheit RG. A phase field model for simulating the pitting corrosion. Corros Sci Sep 2016;110:157–66. https://doi.org/10.1016/j.corsci.2016.04.001
[39] Malki B, Baroux B. Computer simulation of the corrosion pit growth. Corros Sci 2004;46(8):1715–30. https://doi.org/10.1016/j.corsci.2004.05.004
[40] Marelli S, Sudret B. UQLab: a framework for uncertainty quantification in MATLAB. In: 2nd international conference on vulnerability and risk analysis and management (ICVRAM 2014); University of Liverpool, United Kingdom. Jul 13–16 2014. p. 2554–63. https://www.scopus.com/record/display.uri?eid=2-s2.0-84920643339 [Accessed: 27 Mar 2025].
[41] Marelli S, Lamas C, Konakli K, Mylonas C, Wiederkehr P, Sudret B. UQLab user manual – sensitivity analysis. Technical report, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland; 2024. Report UQLab-V2.1-106.
[42] MathWorks. Statistics and machine learning toolbox: fit Gaussian process regression (GPR) models. The MathWorks, Inc.; 2023. https://www.mathworks.com/help/stats/fitrgp.html.
[43] Melchers RE. Pitting corrosion of mild steel in marine immersion environment—part 2: variability of maximum pit depth. Corrosion 2004;60(10):937–44. https://doi.org/10.5006/1.3287827
[44] Melchers RE. Statistical characterization of pitting corrosion—part 1: data analysis. Corrosion 2005;61(7):655–64. https://doi.org/10.5006/1.3278201
[45] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. J Chem Phys Jun 1953;21(6):1087–92. https://doi.org/10.1063/1.1699114
[46] Muehleisen RT, Bergerson J. Bayesian calibration – what, why and how. In: Proceedings of the 4th international high performance buildings conference. Purdue University; Jul 2016. p. 167. https://docs.lib.purdue.edu/ihpbc/167/.
[47] Murakami D, Yamagata Y, Hirano T. Chapter four - geostatistics and Gaussian process models. In: Spatial analysis using big data: methods and urban applications. Elsevier; 2020. pp. 57–112. https://doi.org/10.1016/B978-0-12-813127-5.00004-7
[48] Parangusan H, Bhadra J, Al-Thani N. A review of passivity breakdown on metal surfaces: influence of chloride- and sulfide-ion concentrations, temperature, and ph. Emergent Mater 2021;4(5):1187–203. https://doi.org/10.1007/s42247-021-00194-6
[49] Rasmussen CE, Williams CKI. Gaussian processes for machine learning. Cambridge, MA: MIT Press; 2006.

[50] Rebba R, Mahadevan S. Computational methods for model reliability assessment. Reliab Eng Syst Saf Aug 2008;93(8):1197–207. https://doi.org/10.1016/j.ress.2007.08.001
[51] Sankararaman S, Mahadevan S. Separating the contributions of variability and parameter uncertainty in probability distributions. Reliab Eng Syst Saf Apr 2013;112:187–99. https://doi.org/10.1016/j.ress.2012.11.024
[52] Scheiner S, Hellmich C. Stable pitting corrosion of stainless steel as diffusion-controlled dissolution process with a sharp moving electrode boundary. Corros Sci Feb 2007;49(2):319–46. https://doi.org/10.1016/j.corsci.2006.03.019
[53] Scheiner S, Hellmich C. Finite volume model for diffusion- and activation-controlled pitting corrosion of stainless steel. Comput Methods Appl Mech Eng 2009;198(37–40):2898–910. https://doi.org/10.1016/j.cma.2009.04.012
[54] Schulz E, Speekenbrink M, Krause A. A tutorial on Gaussian process regression: modelling, exploring, and exploiting functions. bioRxiv preprint, 2016. https://doi.org/10.1101/095190
[55] Senthilnathan A, Nath P, Mahadevan S, Witherell P. Surrogate modeling of microstructure prediction in additive manufacturing. Comput Mater Sci Jan 2025;247:113536. https://doi.org/10.1016/j.commatsci.2024.113536
[56] Sharland SM, Jackson CP, Diver AJ. A finite-element model of the propagation of corrosion crevices and pits. Corros Sci 1989;29(9):1149–66. https://doi.org/10.1016/0010-938X(89)90051-6
[57] Shi X, Nguyen TA, Kumar P, Liu Y. A phenomenological model for the chloride threshold of pitting corrosion of steel in simulated concrete pore solutions. Anti-Corros Method Mater 2011;58(4):179–89. https://www.emerald.com/insight/content/doi/10.1108/00035591111148894/full/html.
[58] Siegel MD, Liu L, Pfaffenberger Z, Kisley L. Quantitative advantages of corrosion sensing using fluorescence, microscopy, and single-molecule detection. ACS Appl Mater Interfaces 2024;16(42):56481–96. https://doi.org/10.1021/acsami.4c07800
[59] Sobol' IM. On sensitivity estimation for nonlinear mathematical models. Mat. Model. 1990;2(1):112–8. http://mi.mathnet.ru/mm2320.
[60] Sobol' IM. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. Math Comput Simul Feb 2001;55(1–3):271–80. https://doi.org/10.1016/S0378-4754(00)00270-6
[61] Stafiej J, di Caprio D, Bartosik Ł. Corrosion-passivation processes in a cellular automata based simulation study. J Supercomput 2013;65:697–709. https://doi.org/10.1007/s11227-013-0933-8
[62] Sun X, Duddu R. A sequential non-iterative approach for modeling multi-ionic species reactive transport during localized corrosion. Finite Elem Anal Des 2019;166. https://doi.org/10.1016/j.finel.2019.103318
[63] Tohme T, Vanslette K, Youcef-Toumi K. A generalized Bayesian approach to model calibration. Reliab Eng Syst Saf Dec 2020;204:107141. https://doi.org/10.1016/j.ress.2020.107141
[64] Vagbharathi AS, Gopalakrishnan S. An extended finite-element model coupled with level set method for analysis of growth of corrosion pits in metallic structures. Proc R Soc A 2014;470(2168). https://doi.org/10.1098/rspa.2014.0001
[65] Valor A, Caleyo F, Alfonso L, Rivas D, Hallen JM. Stochastic modeling of pitting corrosion: a new model for initiation and growth of multiple corrosion pits. Corros Sci 2007;49(2):559–79. https://doi.org/10.1016/j.corsci.2006.05.049
[66] Van der Weeën P, Zimer AM, Pereira EC, Mascaro LH, Bruno OM, De Baets B. Modeling pitting corrosion by means of a 3d discrete stochastic model. Corrosion Sci 2014;80:21–31. https://doi.org/10.1016/j.corsci.2014.01.010
[67] Velázquez JC, Caleyo F, Valor A, Hallen JM. Predictive model for pitting corrosion in buried oil and gas pipelines. Corrosion 2009;65(5):332–42. https://doi.org/10.5006/1.3319138
[68] Venkatraman A, Wilson MA, Zapiain DMDO. Accelerating charge estimation in molecular dynamics simulations using physics-informed neural networks: corrosion applications. npj Comput Mater 2025;11:24. https://www.nature.com/articles/s41524-024-01495-0.
[69] Vohra M, Nath P, Mahadevan S, Lee YT. Fast surrogate modeling using dimensionality reduction in model inputs and field output: application to additive manufacturing. Reliab Eng Syst Saf Sep 2020;201:106986. https://doi.org/10.1016/j.ress.2020.106986
[70] Walton JC. Mathematical modeling of mass transport and chemical reaction in crevice and pitting corrosion. Corrosion Sci 1990;30(8–9):915–28. https://doi.org/10.1016/0010-938X(90)90013-U
[71] White A, Mahadevan S, Schmucker J, Karl A. Multi-metric validation under uncertainty for multivariate model outputs and limited measurements. J Verif Valid Uncertain Quant 2022;7(4):041004. https://doi.org/10.1115/1.4056548
[72] Xu Z, Zhang G, Chen Z, et al. Elastic vortices and thermally-driven cracks in brittle materials with peridynamics. Int J Fract 2018;209:203–22. https://doi.org/10.1007/s10704-017-0256-5
[73] Zhang J, Xiao C, Yang W, Liang X, Zhang L, Wang X. Probabilistic geothermal resources assessment using machine learning: Bayesian correction framework based on Gaussian process regression. Geothermics 2023;114:102787. https://doi.org/10.1016/j.geothermics.2023.102787