



A regularized ensemble Kalman filter for stochastic phase field models of brittle fracture

Downloaded from: <https://research.chalmers.se>, 2026-08-30 01:46 UTC

Citation for the original published paper (version of record):

Hermann, L., Janicke, R., Meyer, K. et al (2026). A regularized ensemble Kalman filter for stochastic phase field models of brittle fracture. *Computer Methods in Applied Mechanics and Engineering*, 461. <http://dx.doi.org/10.1016/j.cma.2026.119182>

N.B. When citing this work, cite the original published paper.



A regularized ensemble Kalman filter for stochastic phase field models of brittle fracture

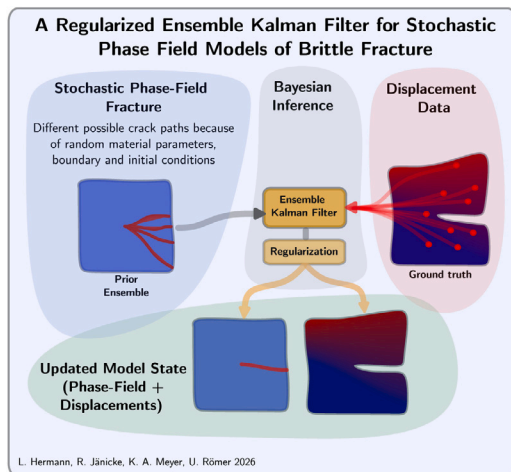
Lucas Hermann ^a,^{*} Ralf Jänicke ^b, Knut Andreas Meyer ^c, Ulrich Römer ^a

^a TU Braunschweig, Institute for Acoustics and Dynamics, Langer Kamp 19, Braunschweig, 38106, Germany

^b TU Braunschweig, Institute for Applied Mechanics, Pockelsstrasse 3, Braunschweig, 38106, Germany

^c Chalmers University of Technology, Department of Mechanical Engineering, Maskingränd 1, Göteborg, 412 96, Sweden

GRAPHICAL ABSTRACT



ARTICLE INFO

Link to source code: https://github.com/herluc/EnKF_Phase_Field, <https://doi.org/10.5281/zenodo.20828230>

Keywords:

Phase field fracture
Ensemble Kalman filter
Uncertainty quantification
Data assimilation

ABSTRACT

The phase-field approach to brittle fracture provides a continuum framework for modeling crack initiation and propagation without explicit representation of discrete crack surfaces, provided the spatial discretization is fine enough to resolve the regularization length scale. However, uncertain local material parameters due to material defects can strongly influence simulation results, such as crack paths and remaining structural strength. At the same time, the ability to continuously monitor structures using sensors allows complementing modeling predictions with displacement measurements for example.

In this contribution, we connect these two complementary sources of information and present a Bayesian inference procedure that allows updating the current model state with

^{*} Corresponding author.

E-mail address: l.hermann@tu-braunschweig.de (L. Hermann).

<https://doi.org/10.1016/j.cma.2026.119182>

Received 9 March 2026; Received in revised form 8 June 2026; Accepted 18 June 2026

Available online 4 July 2026

0045-7825/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

incoming sensor data. We construct a Bayesian prior for the model state (both displacements and phase-field) and employ an ensemble Kalman filter (EnKF) to perform the update. In the EnKF, the update is computed by performing a Kalman shift on each ensemble member. Since the standard EnKF may produce assimilated states that violate common modeling assumptions, we present a phase field-based regularization technique as a proximal step correction towards model-consistent updates. 1D and 2D numerical examples demonstrate the performance and accuracy of the proposed method and show that the updated state matches the ground truth reasonably well. Unlike traditional Bayesian inversion techniques, which have already been applied to brittle fracture, we infer not the model parameters but the model state, i.e., the displacement field and the phase-field. Although only displacements are observed, the strong correlation between both fields also allows inference of the posterior phase-field.

1. Introduction

Phase field fracture modeling was originally proposed to simulate crack propagation in brittle materials [1,2], and has later been extended to ductile fracture [3] and fatigue crack growth [4]. The gradient-regularization ensures mesh-independent results in the case of a sufficiently fine spatial discretization. However, crack propagation is highly dependent on variations in boundary conditions and local material defects, both in physical experiments and models. Given that material parameters and initial conditions are often only partially known, Uncertainty Quantification (UQ) methods are employed to establish a probabilistic forward problem [5,6].

In this context, the forward problem entails propagating uncertainty in the input parameters to model outputs. Input parameters can be related to material properties, initial conditions, and boundary conditions. The goal is then to determine the induced probability distribution for model outputs, such as displacement and the phase field and the corresponding crack paths. In this context, multivariate Gaussian approximations for the output densities are appealing to reduce the computational cost. These Gaussian distributions are defined via a mean and variance for each point of the domain, as well as two-point correlation functions and the correlations across fields.

For non-linear phase field problems, the uncertainty in the output can be substantial. To mitigate this uncertainty, real-world data can be collected, for instance, through Digital Image Correlation (DIC) [7]. In a Bayesian context, prior distributions can be conditioned on this data, see, e.g. [8], effectively assimilating the data into the model. For a specific structure of the data model, this results in a new Gaussian posterior that has a reduced variance, particularly around sensor locations.

The phase field fracture model operates using a quasi-static time-stepping regime, incrementally increasing the load to trace the evolving damage. In some cases, measurement data may be available at various time steps or loads during the simulation, requiring conditioning of the time-evolving probability distribution on incoming data. This process can be managed through Kalman filtering, where the model is advanced until the time step with available data. The solution is then conditioned on the measured data, and the resulting posterior distribution serves as the initial condition for subsequent uncertainty propagation. For non-linear models, the Extended Kalman Filter (EKF) is typically employed to linearize the problem. Alternatively, the Ensemble Kalman Filter (EnKF) can be used for non-linear problems, operating in a Monte Carlo fashion, which reduces uncertainty over time as different members of the ensemble converge towards a common mean. The reader is referred to [9] and the references therein for a discussion.

Applying Kalman filtering to the phase field brittle fracture model presents several challenges. Since the phase field variable attempts to model a smeared but localized crack, fine discretization and the positioning of sensors around the crack path are needed. Additionally, the phase field variable is history-dependent, requiring measures to prevent unphysical solutions during the conditioning process. The aforementioned non-linearities further complicate the filtering approach.

This paper aims to provide a data-corrected solution to the phase field fracture problem, particularly when dealing with non-unique crack path solutions. These may occur if the phase field is highly sensitive to mesh parameters, initial conditions or material configurations [5]. Our goal is to infer the phase field solution based on data for the displacement, leveraging the model-based correlation between both fields.

It is important to note that there exist other works on data assimilation procedures, applied to numerical fracture problems. For instance, [10] derived a Bayesian approach to update scalar parameters of the phase field model, albeit not the complete high-dimensional state as targeted in this publication. The authors in [11] employ a linear elastic fracture mechanics approach and use it in conjunction with an EKF. The method parametrizes the crack, hence e.g. the length or the angle of the crack surface take a certain value, given a load increment. These parameters are subject to uncertainty and updated throughout time with the EKF. The authors state that, so far, only cracks in form of a straight line can be parametrized and updated in the given framework. In [12], the authors make use of an hydraulic fracture model in 1D, the parameters of which are updated using, again, an EKF. Here, the corresponding crack surface, i.e. the shape of the crack path, is parametrized with few parameters which are updated with the EKF. Using an implicit level set algorithm, that approach was adapted to infer the crack front of a planar hydraulic fracture in [13]. See [14] for an application thereof with real measurement data. The authors of [15] focus on fatigue fracture. The model is based on XFEM and three crack parameters, stemming from Paris' law for fatigue crack growth, are updated with an EKF. In [16], the authors use an EnKF to update fracture length and angle parameters of a 2D fracture field featuring multiple straight cracks. The mentioned references have in common that they use a Kalman Filter and parametric fracture models. The downside of these parametric fracture models, albeit being rather computationally cheap, is a lack in flexibility concerning the crack path compared to phase field models. In many



a) A crack is represented by a sharp fracture surface, which is challenging to handle numerically.

b) Through phase field regularization, the fracture surface is effectively approximated by a smooth field.

Fig. 1. Phase field regularization of a fracture in a continuum, based on [18]. On the displacement fields, Dirichlet and Neumann boundary conditions Γ_D^u and Γ_N^u can be imposed. The phase field is left unconstrained.

cases, the overall shape and rough path of the expected crack need to be known beforehand. Geometrically complex crack surfaces are very hard to model. For XFEM, specifically, tracking a crack in three dimensions is challenging. The phase field approach to brittle fracture, as adopted in this paper, however, does not assume a certain geometry of the crack. An extension to three dimensions is rather simple. Instead of updating only few parameters, the complete state of the phase field and displacement is updated using an EnKF. This makes it possible to gather real world data from complex crack paths and, in theory, end up with a model solution much closer to reality compared to the parametric models. The differences of phase field fracture models to a parametric one (in that case XFEM) are explained in great detail in [17].

The paper is structured as follows: Section 2 introduces phase field brittle fracture modeling, with a focus on the micromorphic phase field model [18]. Following this, the stochastic forward problem is presented, resulting in the prior state suitable for data assimilation. Section 3 addresses the stochastic data model as well as the filtering. We put a special focus on the regularization of the posterior which is necessary in order to enforce physically relevant properties. A 1D numerical example is developed alongside the theory to illustrate the procedure. In Section 4, a 2D example is used to demonstrate the ideas in a more relevant setting.

1.1. Notation

Before proceeding to the phase field model and Kalman filtering techniques, we first comment on the notation used throughout this study. Scalars (scalar fields) are denoted by Roman type, while vectors (vector fields) are represented with bold-italic characters. Additionally, blackboard bold characters are reserved for function spaces and sets (e.g. $\mathbb{R}$). After discretization, bold-italic characters are used again for vectors and matrices. A Gaussian random vector is denoted as $\mathbf{a} \sim \mathcal{N}(\bar{\mathbf{a}}, \mathbf{C}_a)$, with mean vector $\bar{\mathbf{a}}$ and covariance matrix $\mathbf{C}_a$. We use $p(\mathbf{a})$ to denote the (multivariate) probability density function of $\mathbf{a}$ and $p(\mathbf{a}|\mathbf{b})$ to denote the density, conditional on $\mathbf{b}$. We use the same symbol for a random variable/vector and the respective realization, since the meaning is clear from the context. The quantity $\|\mathbf{a}\|_{\mathbf{C}}^2 = \mathbf{a}^T \mathbf{C}^{-1} \mathbf{a}$ denotes a weighted Euclidean norm. We will employ indices as follows: $\mathbf{a}_{n,k}$ denotes the k th iterate in the Newton–Raphson iteration at displacement increment n and $\mathbf{a}_n$ refers to the final iterate. During Kalman filtering we distinguish between the forecast state $\mathbf{a}^F$ and the assimilated state $\mathbf{a}^A$ from data. In the context of the ensemble Kalman filter, $\mathbf{a}_{n,i}^{F/A}$ denote the forecast/assimilated state at increment n of the i th ensemble member. Finally, we distinguish between a function and the value it takes, e.g., $\hat{E}$ refers to a function, whereas E is the value it takes given specific arguments.

2. Phase field brittle fracture

The phase field modeling approach summarized in this chapter has been presented in [18–20]. Let $\Omega \subset \mathbb{R}^d$, $d = 1, 2$, be an open and bounded domain, occupied by a linear elastic continuum. The boundary is separated into an inhomogeneous essential (Dirichlet) boundary Γ_D^u , an inhomogeneous natural (Neumann) boundary Γ_N^u , such that $\Gamma_D^u \cap \Gamma_N^u = \emptyset$, and the homogeneous natural Neumann boundary Γ_N^0 . The homogeneous Neumann boundary conditions appear naturally on the complement of $\Gamma_D^u \cup \Gamma_N^u$. In principle, mixed boundary conditions are possible, but these are not considered here. For the phase field itself, no boundary conditions are imposed, hence only the natural homogeneous Neumann boundary conditions are involved on the phase field boundary. Within the continuum, cracks can develop, as visible in Fig. 1(a).

A sharp crack results in a discontinuity within the continuum along the time-evolving crack surface Γ_c . The phase field method for brittle fracture aims to regularize the discontinuity in terms of the phase field variable $\varphi(\mathbf{x}, t) \in [0, 1]$, where 0 signifies an undamaged material and 1 a fully fractured one. The length scale parameter $\ell \in \mathbb{R}^+$ controls the width of the resulting regularized crack, as visible in Fig. 1(b). The phase field method for brittle fracture is based on minimizing an energy functional

$$E(\mathbf{u}, \varphi) = \int_{\Omega} g(\varphi) \Psi^+(\hat{\epsilon}(\mathbf{u})) \, d\Omega + \int_{\Omega} \Psi^-(\hat{\epsilon}(\mathbf{u})) \, d\Omega - \int_{\Gamma_N^u} \mathbf{t}_N \cdot \mathbf{u} \, d\Gamma + \int_{\Omega} \frac{G_c}{c_w \ell} (w(\varphi) + \ell^2 \|\nabla \varphi\|^2) \, d\Omega, \quad (1)$$

in which $\Psi^\pm(\hat{\epsilon}(\mathbf{u}))$ is the split elastic strain energy density, where Ψ^+ is the fracture driving part and Ψ^- is the residual part. Moreover, $\hat{\epsilon}(\mathbf{u})$ is the symmetric part of the displacement gradient and $\mathbf{u}(\mathbf{x}, t)$ the displacement. In the numerical examples, the volumetric–deviatoric split is used and we have for $\Psi^\pm$

$$\Psi^+ = \frac{1}{2} K \langle \text{tr}(\epsilon) \rangle^2 + \mu (\epsilon_{\text{dev}} : \epsilon_{\text{dev}}), \quad \Psi^- = \frac{1}{2} K \langle -\text{tr}(\epsilon) \rangle^2, \quad (2)$$

where K and μ are the bulk and shear modulus, respectively. The boundary term over the Neumann boundary Γ_N^u models an applied traction. Finally, we use the so-called AT2 fracture model with

$$g(\varphi) = (1 - \varphi)^2, \quad c_w = 2, \quad w(\varphi) = \varphi^2 \quad (3)$$

where g is the degradation function, w the local dissipation function, and c_w the associated scaling to make G_c the fracture energy. Notice that this is a quasi-static problem, hence the numerical problem will be treated as being stationary. Therefore, we omit time indices at this point of the derivation and will introduce these later when they become relevant for the EnKF. For the moment it is sufficient to introduce φ' as the phase field of the previous time/loading step.

In this paper, a modified version of the phase field problem will be used, namely the micromorphic approach according to [18]. The energy functional parametrization is then extended by the micromorphic variable, d , and a term penalizing the difference between the phase field φ and d is added, resulting in

$$\begin{aligned} E = \hat{E}(\mathbf{u}, \varphi, d) = & \int_{\Omega} g(\varphi) \Psi^+(\hat{\epsilon}(\mathbf{u})) \, d\Omega + \int_{\Omega} \Psi^-(\hat{\epsilon}(\mathbf{u})) \, d\Omega - \int_{\Gamma_N^u} \mathbf{t}_N \cdot \mathbf{u} \, d\Gamma \\ & + \int_{\Omega} \frac{G_c}{c_w \ell} (w(\varphi) + \ell^2 \|\nabla d\|^2) \, d\Omega + \int_{\Omega} \frac{\alpha}{2} (\varphi - d)^2 \, d\Omega. \end{aligned} \quad (4)$$

Choosing the interaction parameter α well leads to $d \approx \varphi$ [18], hence the micromorphic field approximates the phase field. In practice, α is defined as a function of G_c , ℓ and a scaling constant β . By introducing the function spaces

$$\mathbb{U} = \{\mathbf{u} \in (H^1(\Omega))^2 \mid \mathbf{u} = \mathbf{u}_D \text{ on } \Gamma_D^u\}, \quad (5a)$$

$$\mathbb{U}_0 = \{\mathbf{u} \in (H^1(\Omega))^2 \mid \mathbf{u} = \mathbf{0} \text{ on } \Gamma_D^u\}, \quad (5b)$$

$$\mathbb{D} = H^1(\Omega), \quad (5c)$$

the Euler–Lagrange equations, related to (4), read: find $(\mathbf{u}, d, \varphi) \in \mathbb{U} \times \mathbb{D} \times L^2(\Omega)$ such that

$$\int_{\Omega} g(\varphi) \boldsymbol{\sigma}^+ : \hat{\epsilon}(\delta \mathbf{u}) \, d\Omega + \int_{\Omega} \boldsymbol{\sigma}^- : \hat{\epsilon}(\delta \mathbf{u}) \, d\Omega - \int_{\Gamma_N^u} \mathbf{t}_N \cdot \delta \mathbf{u} \, d\Gamma = 0 \quad \forall \delta \mathbf{u} \in \mathbb{U}_0, \quad (6)$$

$$\int_{\Omega} \frac{2G_c \ell}{c_w} \nabla d \cdot \nabla \delta d \, d\Omega - \int_{\Omega} \alpha (\varphi - d) \delta d \, d\Omega = 0 \quad \forall \delta d \in \mathbb{D}, \quad (7)$$

$$\int_{\Omega} \delta \varphi \left(g'(\varphi) \Psi^+(\hat{\epsilon}(\mathbf{u})) + \frac{G_c}{c_w \ell} w'(\varphi) + \alpha (\varphi - d) \right) \, d\Omega = 0 \quad \forall \delta \varphi \in L^2(\Omega). \quad (8)$$

Eq. (8) implies that

$$g'(\varphi) \Psi^+(\hat{\epsilon}(\mathbf{u})) + \frac{G_c}{c_w \ell} w'(\varphi) + \alpha (\varphi - d) = 0, \quad (9)$$

almost everywhere¹ in Ω , however, we will assume additional regularity so that (9) holds pointwise in Ω . The constraints on φ , irreversibility, $\varphi \geq 0$, and bounds, $0 \leq \varphi \leq 1$, can then be enforced locally. With the chosen AT2 fracture model, Eq. (9) can then be solved analytically as

$$\varphi = \hat{\varphi}(\mathbf{u}, d, \varphi', \ell) = \min \left(\max \left(\frac{2\Psi^+(\hat{\epsilon}(\mathbf{u})) + \alpha d}{2\Psi^+(\hat{\epsilon}(\mathbf{u})) + \alpha + \frac{G_c}{\ell}}, \varphi' \right), 1 \right), \quad (10)$$

where the dependence on quantities $(\mathbf{u}, d, \varphi', \ell)$ is highlighted, since these will be modified later on as part of the Kalman filter.

FEM for PF fracture

The phase-field fracture problem is solved with the Finite Element Method (FEM), introducing the approximations

$$\mathbf{u}(\mathbf{x}) \approx \mathbf{u}_h(\mathbf{x}) = \sum_i \boldsymbol{\phi}_i^u(\mathbf{x}) a_i^u, \quad d(\mathbf{x}) \approx d_h(\mathbf{x}) = \sum_i \phi_i^d(\mathbf{x}) a_i^d \quad (11)$$

such that the energy, $E \approx \hat{E}(\mathbf{u}_h, d_h; \varphi'_h, \ell) = \hat{E}_h(\mathbf{a}^u, \mathbf{a}^d; \boldsymbol{\varphi}', \ell)$. Note that because of $\varphi_h = \hat{\varphi}(\mathbf{u}_h, d_h, \varphi'_h; \ell)$ we have rewritten $\hat{E}(\mathbf{u}_h, \varphi_h, d_h) = \hat{E}(\mathbf{u}_h, d_h; \varphi'_h, \ell)$, where we now explicitly state the dependence on φ'_h, ℓ . This notation will be required to introduce

¹ That is except on sets of Lebesgue measure zero.

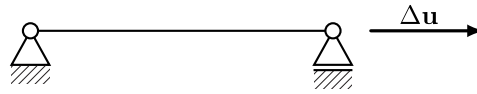


Fig. 2. 1D numerical experiment.

Table 1

Model parameters for the 1D tension rod example.

Parameter	Value
Fracture Model	AT2
E	210000 [N/mm ²]
ν	0.3 [-]
G_c	2.7 [N/mm]
l	2.5e-2 [mm]
β	200 [-]
α	$\beta G_c / \ell$

the Kalman filter regularization. Minimizing $\hat{E}_h$ leads to (cf. the original, not FE approximated Euler–Lagrange Eqs. (6), (7))

$$R_i^u(\mathbf{a}^u, \mathbf{a}^d; \boldsymbol{\varphi}', \ell) = \int_{\Omega} \hat{\mathbf{e}}(\boldsymbol{\varphi}_i^u) : [g(\varphi_h) \boldsymbol{\sigma}^+ + \boldsymbol{\sigma}^-] \, d\Omega - \int_{\Gamma_N^u} \boldsymbol{\varphi}_i^u \cdot \mathbf{t}_N \, d\Gamma = 0, \quad (12)$$

$$R_i^d(\mathbf{a}^u, \mathbf{a}^d; \boldsymbol{\varphi}', \ell) = \int_{\Omega} \nabla \boldsymbol{\varphi}_i^d \cdot \left[\frac{2G_c \ell}{c_w} \nabla d_h \right] \, d\Omega - \int_{\Omega} \boldsymbol{\varphi}_i^d \alpha (\varphi_h - d_h) \, d\Omega = 0. \quad (13)$$

Following [18], we adopt convexification via extrapolation [21], using the extrapolated micromorphic field,

$$\check{\mathbf{a}}^d = \mathbf{a}_{n-1}^d + \Delta_{t_n} \frac{(\mathbf{a}_{n-1}^d - \mathbf{a}_{n-2}^d)}{\Delta_{t_{n-1}}}, \quad \check{d}_h = \sum_i \boldsymbol{\varphi}_i^d \check{\mathbf{a}}_i^d, \quad (14)$$

and $\varphi \approx \check{\varphi} = \hat{\varphi}_h(\mathbf{u}_h, \check{d}_h; \boldsymbol{\varphi}_{n-1}, \ell)$ in (12) and (13). Note that we only indicate past indices $n-1, n-2$ and not the current one n , to keep the notation lighter. We also use φ instead of $\check{\varphi}$ in the following, keeping in mind that always $\check{d}_h$ is used here as an argument. With these definitions, we now collect the state variables (per quadrature point), $\boldsymbol{\varphi}$, unknowns, $\mathbf{a} = (\mathbf{a}^u, \mathbf{a}^d)$, and residuals, $\mathbf{R} = (\mathbf{R}^u, \mathbf{R}^d)$, such that we have the complete equation system

$$\mathbf{R}^u(\mathbf{a}^u, \check{\mathbf{a}}^d; \boldsymbol{\varphi}_{n-1}, \ell) = 0, \quad (15)$$

$$\mathbf{R}^d(\mathbf{a}^u, \mathbf{a}^d; \boldsymbol{\varphi}_{n-1}, \ell) = 0, \quad (16)$$

that we solve monolithically in each time step using the standard Newton–Raphson method. Note that (16) is linear in $\mathbf{a}^d$, as $\check{\mathbf{a}}^d$ is used to calculate φ_h .

Remark. Solving the FE system of equations, which is defined in this section, is computationally very demanding. In [22], the CPU hours were compared for different solution schemes. As we not only aim to solve a deterministic phase field problem, but a stochastic one based on an ensemble strategy, which potentially involves hundreds of parallel model evaluations, some precautions have to be taken to keep the CPU hours per ensemble member low and hence the stochastic evaluation feasible.

Firstly, a coarser mesh is applied in the regions where a crack is not expected in order to reduce the number of degrees of freedom of the resulting FE system of equations. Secondly, the monolithic solution strategy has been chosen deliberately over a staggered one, because it requires significantly less CPU hours [22]. Lastly, the Volumetric–Deviatoric energy split is chosen over the spectral decomposition, as that split is computationally more demanding than the Volumetric–Deviatoric split.

2.1. Deterministic solution of the phase field problem/true underlying solution

Solving the phase field problem for a fixed set of parameters yields a deterministic solution. This solution can in the further process serve as a reference solution from which data can be observed. We want to explain the concepts developed in this paper along with a simple example. To this end, we introduce a solution to the phase field problem on a 1D domain. Fig. 2 illustrates the mechanical setup: A tension rod on $\Omega = [-1, 1]$ in [mm] is pulled in a displacement controlled manner until failure.

With $N = 200$ degrees of freedom for each u and d , i.e. a mesh size of $h = 1e-2$ mm, the length scale is fixed to $l = 2.5e-2$. We use piecewise linear approximations for u and d . While the left boundary is fixed, a Dirichlet boundary condition on the right boundary is used to prescribe a displacement increment of $\Delta u = 1e-4$ [mm] per pseudo time step. As the problem is 1D, the energy split does not apply here. Table 1 lists all chosen parameters for the model. An initial value of $\varphi = 0.7$ for the phase field at a certain position, in this case $x_0 = -0.285$, i.e. an initial damage, serves as a nucleus for the development of a crack. In Fig. 4, the deterministic solution for both phase field and displacements is visible in blue for a single time step. The whole time series can be condensed to a force displacement diagram as visible in Fig. 3. The time step marked in red is the one at which the displacement and phase field is shown in the plots within Section 3.

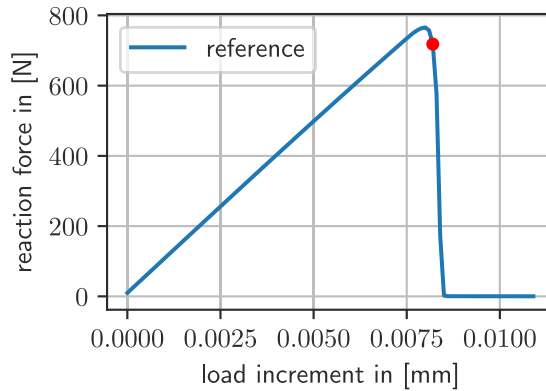


Fig. 3. Reaction forces over pseudo time for the 1D numerical example. A linear elastic interval is followed by an abrupt crack and loss of remaining stiffness. The red dot marks the first load increment and reaction force at which data assimilation will take place in the later section.

3. Data assimilation

If both a predictive model and measured data exist, the process of connecting both, to ideally obtain improved predictions, is called data assimilation [23]. Assimilating a normally distributed output from a linear model with measurement data that also follows a normal distribution can be achieved in closed-form via conditioning of multivariate Gaussian distribution, see [24]. In the time-dependent case, a Kalman filter can be used. Details on the most important flavors of the Kalman Filter and related techniques can be found in [25]. The Kalman filter is very similar to the usual conditioning of a multivariate Gaussian, with the difference that the updated distribution serves as a prior in a subsequent time step. The standard Kalman filter is derived for linear problems. If the problem is non-linear, linearization leads to the extended Kalman filter (EKF). Another method to treat the non-linearity is to use derivative-free sampling methods. Examples for sampling-based data assimilation methods are the Ensemble Kalman Filter (EnKF) [26] and the Particle Filter [27]. In any case, a prior is needed and is obtained here by propagating uncertainties through the predictive model. The following introduction and discussion is based on [9,25].

3.1. Stochastic forward problem

Many aspects of numerical modeling require making assumptions, handling a lack of knowledge and quantifying errors, originating e.g. from discretization. In these cases, instead of using deterministic quantities, accounting and propagating uncertainties is crucial. For the given phase field problem, there are many sources of uncertainty. A prominent example is parametric uncertainty, e.g. in the boundary, initial conditions and material parameters which can be either scalar or fields. Another source of uncertainty is in the chosen initial damage field. In this paper, we will focus on the latter and introduce a random initial damage field. Uncertainty in these input parameters will lead to uncertainty in the solution fields, as well. For a linear model, Gaussian distributions in the random coefficients will lead to a Gaussian solution. The EnKF, which is the data assimilation technique used in this work, propagates a particle ensemble through the model, which can be used to approximate the output statistics and obtain a prior. Hence, uncertainty propagation is carried out in a Monte Carlo (MC)-like fashion, where the phase field model is simulated repeatedly with varying realizations of the model inputs.

In Fig. 4, multiple samples of the stochastic phase field problem are shown. These arise from treating the initial value for the phase field as a random variable: The position and magnitude of the damage nucleus is different for each realization, hence also the position and behavior in time of the resulting crack. The position is sampled from a normal distribution $x_0 \sim \mathcal{N}(-0.25, 0.12)$, whereas the initial magnitude of the phase field is sampled from a uniform distribution $\varphi_{\text{mag}} \sim \mathcal{U}(0.73, 0.76)$. The shape of the initial damage is defined as a Gaussian with magnitude φ_{mag} , standard deviation $\sigma = 0.05$ and mean x_0 . From the ensemble, a Gaussian approximation can be computed, i.e. mean and covariance are calculated from the samples. The next step towards data assimilation is to connect the prior ensemble with data in form of a data model.

3.2. Data model

The basis for assimilating model predictions with data is a model of the observations. A data model connects the phase field FE solution to the sensor data and the derivation in this section mainly follows [8]. First, based on [28], we model the measurement data generating process as

$$\mathbf{y}_n = \rho \mathbf{H} \mathbf{a}_n + \delta_n + \boldsymbol{\eta}_n \quad (17)$$

now with an explicit iteration index n , the measured data $\mathbf{y}_n \in \mathbb{R}^N$ at time $n\Delta t$ and the linear observation map $\mathbf{H} : \mathbb{R}^M \rightarrow \mathbb{R}^N$ which maps from the high-dimensional state-space of the solution to the lower dimensional measurement space. The matrix can be

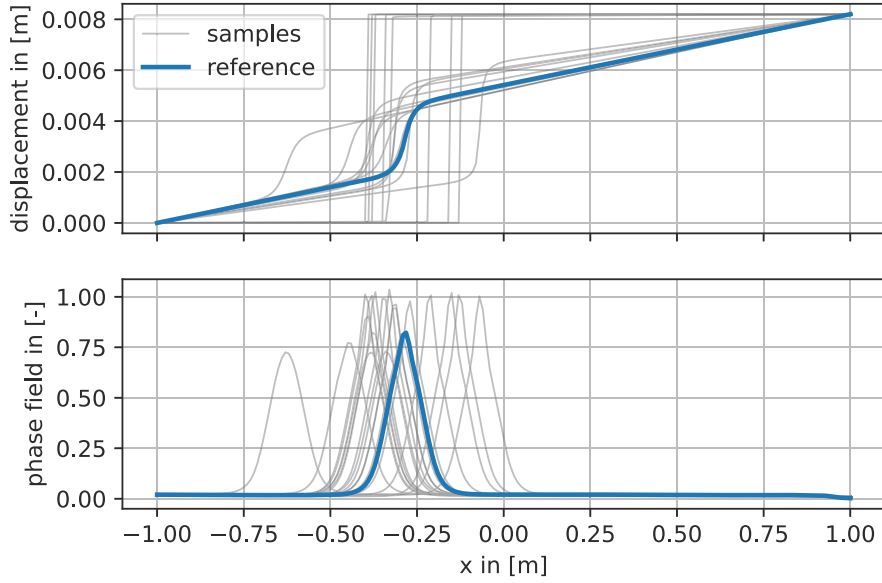


Fig. 4. Solution to the 1D phase field brittle fracture problem at a single time step. The deterministic reference solution is visible in blue, whereas the prior ensemble is visible in gray. Some ensemble members indicate a complete fracture (phase field of 1, displacement is a step function) while others only have a phase field smaller than 1, hence there is some stiffness degradation but no complete fracture. The prior ensemble is the result of row 6 in Algorithm 1.

constructed by evaluating the FE basis functions at the sensor locations. In general, $\mathbf{H}$ is constructed by discretizing the respective operator relating the FE solution to measurable quantities. Moreover, $\boldsymbol{\eta}_n \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \mathbf{I})$ models measurement noise and

$$\mathbf{a}_n \sim \mathcal{N}(\bar{\mathbf{a}}_n, \mathbf{C}_{a;n})$$

is the FE solution vector containing both displacements $\mathbf{u}$ and the micromorphic field $\mathbf{d}$. At every time/load step, data can be generated. Note that measurements will only be taken on the displacements and not on the phase field variable, hence the observation matrix is non-zero only in the part related to the displacement vector. The measurement data can be collected on arbitrary points throughout the domain which are not restricted to the nodes of the FE mesh. Finally, sensor noise is modeled with the Gaussian $\boldsymbol{\eta}_n$ whereas the model-data-mismatch is captured with another Gaussian $\boldsymbol{\delta}_n$. Note that $\boldsymbol{\delta}_n, \boldsymbol{\eta}_n$ are independent of $\mathbf{a}_n$. To do so, a kernel function, e.g. of squared exponential or Matérn type, has to be chosen and this kernel is then evaluated at the sensor points. In the numerical examples given in this manuscript, the Matérn kernel

$$c(\mathbf{x}, \mathbf{x}') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu} \|\mathbf{x} - \mathbf{x}'\|}{l} \right)^\nu K_\nu \left(\frac{\sqrt{2\nu} \|\mathbf{x} - \mathbf{x}'\|}{l} \right) \quad (18)$$

will be used throughout. It is defined with the scaling parameter $\sigma \in \mathbb{R}^+$, $\Gamma(\nu)$ the Gamma function and K_ν a modified Bessel function. Moreover, $l \in \mathbb{R}^+$ is the lengthscale parameter and $\nu \in \mathbb{R}^+$ controls the smoothness of the kernel. For $\nu \rightarrow \infty$ (18) becomes the squared exponential kernel. See [24] for a thorough discussion.

The data model, connecting model to measurements, needs to be calibrated. The calibration is accomplished by tuning the model error covariance. The corresponding hyperparameters $\boldsymbol{w} = (\nu, \sigma, l)$ can be learned from the given data by minimizing the negative log likelihood of the data model with an optimizer such as L-BFGS. A full sampling based approach, based on Markov Chain Monte Carlo (MCMC) can also be used, if uncertainties for the parameter estimation are sought. With a calibrated data model, the Kalman filter equations, to be described below, can be derived by conditioning the FE solution fields on the data. For a single observation we obtain

$$p(\mathbf{y}|\boldsymbol{w}) = \frac{1}{(2\pi)^{n_{\text{sens}}/2} |\mathbf{G}|} \exp \left(-\frac{1}{2} (\mathbf{y} - \rho \mathbf{H} \bar{\mathbf{a}})^\top \mathbf{G}^{-1} (\mathbf{y} - \rho \mathbf{H} \bar{\mathbf{a}}) \right) \quad (19)$$

with n_{sens} the number of sensors and

$$\mathbf{G} = \rho^2 \mathbf{H} \mathbf{C}_a \mathbf{H}^\top + \mathbf{C}_\delta + \mathbf{C}_e \quad (20)$$

the covariance of (17). Taking the natural logarithm yields

$$\log p(\mathbf{y}|\boldsymbol{w}) = -\frac{n_{\text{sens}}}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{G}| - \frac{1}{2} (\mathbf{y} - \rho \mathbf{H} \bar{\mathbf{a}})^\top \mathbf{G}^{-1} (\mathbf{y} - \rho \mathbf{H} \bar{\mathbf{a}}). \quad (21)$$

For multiple observations per sensor we find

$$\log p(\mathbf{Y}|\mathbf{w}) = \sum_{i=1}^{n_{\text{obs}}} \log p(y_i|\mathbf{w}), \quad (22)$$

which can be applied if at one sensor location n_{obs} multiple independent observations exist. See [8] for details on the treatment of the independence assumption in this case. After optimizing with regard to the given likelihood, the data model can be evaluated with the estimated hyperparameters.

In order to generate artificial measurement data, one can sample from the data model with a fixed set of parameters. Another possibility is misspecifying the model by changing parameters or adding complexity through different boundary terms etc. In this paper, a fixed initial phase field is chosen that is not part of the prior ensemble. Also, in order to prevent an inverse crime [29], the reference solution is computed on a different triangulation $\mathcal{T}_{\text{data}}$ with $h_{\text{data}} < h$ compared to the prior ensemble. An uncorrelated, mean-free noise term $e \sim \mathcal{N}(0, C_e)$ is added to the process. The resulting data generating process is then sampled n_{obs} times.

3.3. Ensemble Kalman filter

The EnKF is an equal-weight particle system variant of the Kalman filter [26]. This enables using non-linear models without an explicit linearization. Another sampling-based variant is the particle filter [30], where each ensemble member gets assigned a weight which is updated over time. Particle filters are known to work well for non-linear problems [31], hence they appear promising in our regard. However, the problem with the particle filters is the scaling to higher dimensions, as shown in the literature [32]: Filter breakdown is expected at only hundreds of parameters already, but a high-dimensional state-space system, as the one used here, utilizes tens of thousands or more state parameters. The named problems have also been addressed in the context of statFEM in [9], where the authors chose an EnKF instead of the particle filter. Because of the limitations in regard to high-dimensional problems, we also resort to the EnKF in this paper. Within the EnKF, each ensemble member has the same weight and in the update step, as data comes in, each member is conditioned individually on data.

The EnKF consists of two repeating steps: First, in the prediction step, the approximate solution to the PDE is evolved in time. The resulting fields can then either directly be used as initial conditions to further evolve in time or be updated. Thus, the second step is the update, or analysis, step. If at some point in time measurement data are available, the solution fields of each ensemble member are updated via a Kalman shift, which can be framed in a Bayesian updating framework. It should be noted that the EnKF converges to the true filter for an increasing particle size, only in the linear Gaussian setting [33]. However, the phase field problem is inherently non-linear, hence the accuracy of our particle approximations is less obvious. Recent works develop theory for the near-Gaussian setting, see [33,34], and bounds for the Gaussian approximation, as discussed in [25], become available. This introduction and also the implementation of the filter, presented below, follows [9,25].

It will be instructive to view the EnKF update from an optimization perspective. We also re-introduce the time index n at this point. Following [25], we consider the objective function

$$J_{n,i}(\mathbf{a}_{n,i}^A) = \frac{1}{2} \|\mathbf{y}_n - \mathbf{H} \mathbf{a}_{n,i}^A\|_r^2 + \frac{1}{2} \|\mathbf{a}_{n,i}^A - \mathbf{a}_{n,i}^F\|_{C_n}^2, \quad (23)$$

where the analysis vector $\mathbf{a}_{n,i}^A$, for each ensemble member i , is determined by minimization. The forecast $\mathbf{a}_{n,i}^F$ is obtained by propagating the particles according to the phase field model from the previous time step, to be detailed in Section 3.3.1. For simplicity, only a single time step n and the corresponding data are considered here. We can observe that the model is not strongly enforced, hence the analysis can potentially violate constraints that would be imposed by the phase field model; unless C_n tends to zero and the data are without any influence. Because the non-linear model is not part of the optimization, a closed form can be derived, i.e. (27). At the same time, we observed that the standard EnKF update can fail and we attribute this precisely to the missing explicit influence of the phase field model in the update equations, in combination with the strong nonlinearities of the phase field equations. In order to resolve this issue, a constraint approach will be introduced below.

The EnKF has to be initialized at first. As it is a sampling based approach, the forward problem is solved for n_{ens} samples of the input parameters. Uncertain input parameters could be the spatial configuration of an initial damage field, a random material parameter or probabilistic forcing with random perturbations in the initialization or at each timestep.

3.3.1. Forecast

Consider the ensemble $\{\mathbf{a}_{n-1,i}\}_i$ at time/loading step $n-1$, originating from either the initialization of the filter or a previous prediction or update. Note that we omit the dependence on $\mathbf{y}_{1:n-1}, \mathbf{w}_{1:n-1}$. For each ensemble member, the prediction $\mathbf{a}_{n,i}^F = (\mathbf{a}_{n,i}^{F,u}, \mathbf{a}_{n,i}^{F,d})$ is computed from the residual equation system (16). Subsequently, the prediction's mean and covariance can be computed as

$$\bar{\mathbf{a}}_n^F = \frac{1}{n_{\text{ens}}} \sum_{i=1}^{n_{\text{ens}}} \mathbf{a}_{n,i}^F \quad (24)$$

and

$$\mathbf{C}_n^F = \frac{1}{n_{\text{ens}} - 1} \sum_{i=1}^{n_{\text{ens}}} (\mathbf{a}_{n,i}^F - \bar{\mathbf{a}}_n^F) (\mathbf{a}_{n,i}^F - \bar{\mathbf{a}}_n^F)^T, \quad (25)$$

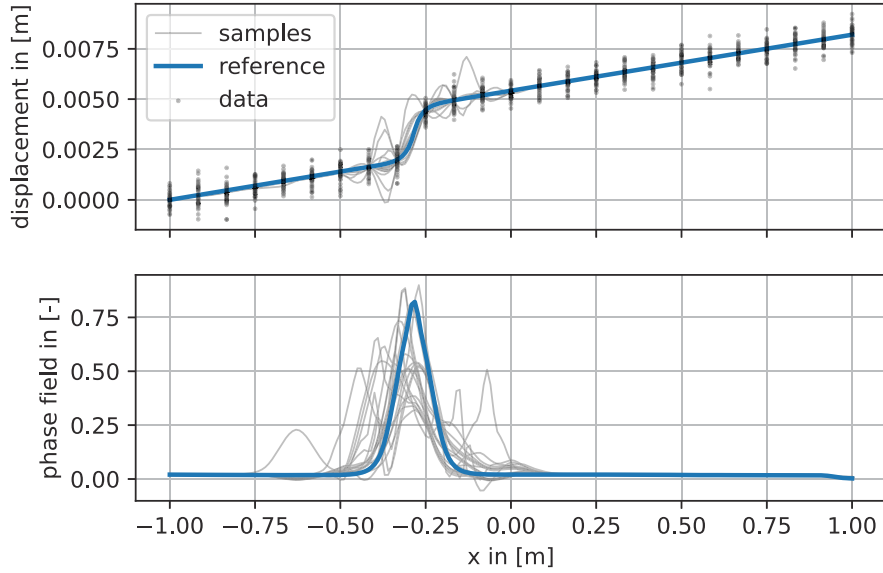


Fig. 5. The posterior ensemble in gray compared to the reference solution, from which noisy measurements are drawn, in blue. The noisy measurements of the reference solution (multiple observations per sensor) are visible as gray dots in a regular interval. They are only taken on the displacement field, not on the phase field. Compared to the prior, the ensemble now clusters more closely around the reference. However, numerical artifacts and unphysical phase field solutions appear, which can lead to the next time step not converging. This posterior ensemble, or analysis, is the result of row 10 in Algorithm 1.

respectively. To keep the notation short, from here on we drop the dependence of the forecast on previous time steps, as e.g. in $\bar{a}_{n|n-1}^F$. The prediction step is iteratively executed until reaching a time step at which data are available. Then, each ensemble member is updated with respect to the given measurement data. In the 1D numerical example, data are available from time step 82 on in an interval of 10 steps. Because a finite ensemble size is used, the covariance is usually inflated and a localization procedure is employed. Details on that can be found in the [Appendix](#).

3.3.2. Analysis

The first step towards updating the prior state, i.e. conditioning it on new data y_n , is updating the hyperparameters $w_{1:n-1}$ of the data model, in case that they change over time. If so, the negative log-marginal posterior of the data model is minimized given suitable priors for the individual parameters. With the updated hyperparameters $w_{1:n}$, each ensemble member is updated by a Kalman shift. The Kalman shift is an analytical solution to the optimization problem (23), see [25] for details. For the state of one member and a single observation per sensor we obtain

$$a_{n,i}^A = a_{n,i}^F + C_n^F H^\top (\rho^2 H C_n^F H^\top + C_\delta + C_e)^{-1} (y_n - \rho H a_{n,i}^F). \quad (26)$$

For the data model as proposed in [8], i.e. with multiple available readings per sensor, there holds for the updated state of a single ensemble member

$$a_{n,i}^A = a_{n,i}^F + C_n^F H^\top (\rho^2 n_{\text{obs}} H C_n^F H^\top + C_\delta + C_e)^{-1} \left(\sum_{j=1}^{n_{\text{obs}}} y_{n,j} - \rho n_{\text{obs}} H a_{n,i}^F \right). \quad (27)$$

The updated states now serve as an initial condition again for the next prediction step. In Fig. 5, the ensemble analysis for the running example are visible. The ensemble members now cluster closely around the reference solution, although it is clearly visible that the posterior ensemble members exhibit strong unphysical noise. Also, the phase field drops below zero, which violates the constraints encoded in the system of PDEs. Subsequent Newton iterations for the next time step might not converge due to this effect. The reason for these problems is the strong gradient in the displacement field. The problem can be overcome to some extent by smoothing the displacement field before the update, as visible in Fig. 6 (a). The resulting posterior ensemble exhibits much less noise, yet negative phase fields still occur and the displacements are still unphysical. This unphysical behavior of the posterior can be overcome by regularization within a staggered solution of the individual equations for phase field and displacement. This will be discussed in Section 3.3.3.

In Algorithm 1, the forecast time stepping as well as the EnKF analysis are summarized. T denotes the set of pseudo-time steps at which the phase field problem is solved with different loadings. T_D denotes the set of time steps at which data is available and the EnKF update is performed.

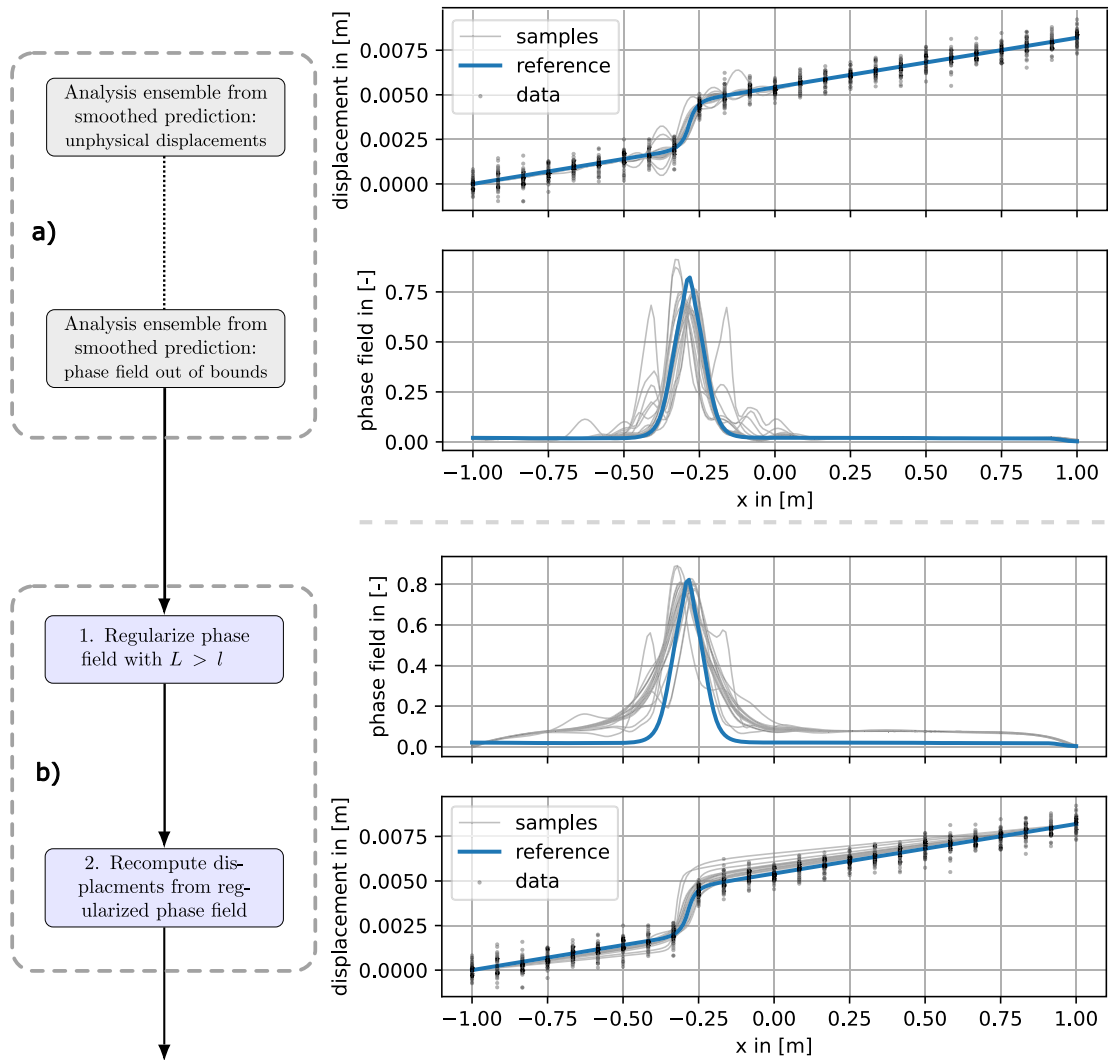


Fig. 6. (a): Posterior ensemble with prior smoothing through a Gaussian Process. The strongest artifacts are filtered out, but the unphysical phase field remains. (b): The posterior ensemble in gray after the first regularization step compared to the reference solution in blue. By the projection on a solution space with larger phase field length scale L , most of the numerical artifacts are filtered. However, the larger length scale leads to more accumulated damage in the whole domain and therefore a loss in total stiffness, which reflects in a slightly larger displacement. This intermediate result corresponds to the results of rows 2–5 in Algorithm 2. (c): Correction with the original phase field length scale. On the left in (c1), the process has been repeated once. On the right in (c2), the process has been repeated four times. It is visible that a single repetition yields results closer to the reference. This is because inference takes place on two phenomena here: fracture initiation and fracture localization. Bayesian treatment of initiation appears much harder in this context; the iterations lead to results which indicate mostly complete damage or only very little damage. However, fracture localization works as desired: the correct position of the crack is found. We are mostly interested in fracture localization at this point. This regularized ensemble corresponds to row 13 in Algorithm 2.

3.3.3. Regularization

Fig. 5 depicts the updated ensemble for the running example. It is clearly visible that the Kalman update can generate both spurious oscillations in the displacements as well as phase field variables with negative values. Several paths can be pursued to remove these effects. One could transform the phase field variable to enforce the bounds $[0, 1]$ and smooth the displacements, however, we found that in particular the smoothing is difficult to realize in higher dimensions. Another option is to enforce the phase field model as a strong constraint in a 3D-Var or 4D-Var type approach. This however, appears to be very costly to realize and is left for future research.

Here, we propose a different solution, where we regularize the Kalman shift with a few well-chosen staggered iterations using the phase field model. Instead of a standard Newton approach to solve the individual equations, we treat each Newton step as a single-step approximation to a proximal operator, which serves to balance between the purely PDE-based regularization and the data influence from the EnKF analysis. The reasoning behind this will be discussed in Section 5.

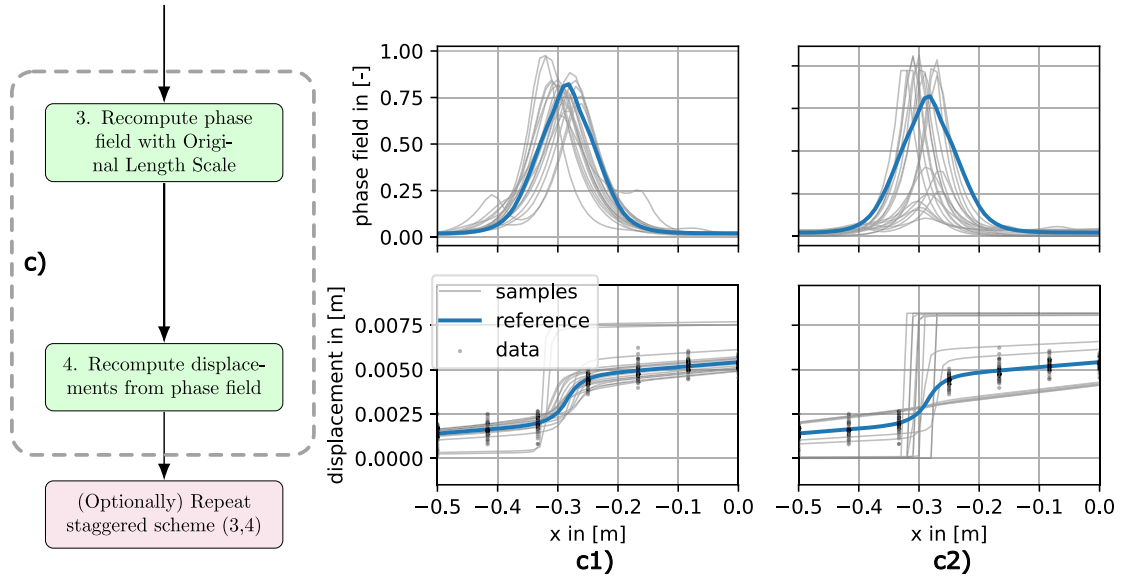


Fig. 6. (continued).

Algorithm 1 Phase Field Ensemble Kalman Filter

Require: $\mathbf{a}_0^F = \left\{ (\mathbf{a}_{0,i}^{F,u}, \mathbf{a}_{0,i}^{F,d}) \right\}_{i=1}^{n_{\text{ens}}}$, $t = t_0 = 0$

```

1: for  $t \in T$  do
2:    $t = t + \Delta t$ 
3:    $n = n + 1$ 
4:   for  $i = 1$  to  $n_{\text{ens}}$  do
5:     update  $\tilde{\mathbf{a}}^d, \varphi_{n-1}$  based on  $\mathbf{a}_{n-1,i}^F$ 
6:     solve monolithically:
7:        $\mathbf{R}^u(\mathbf{a}^u, \tilde{\mathbf{a}}^d, \varphi_{n-1}, \ell) = 0$ ,
8:        $\mathbf{R}^d(\mathbf{a}^u, \mathbf{a}^d, \varphi_{n-1}, \ell) = 0$ 
9:       set  $\mathbf{a}_{n,i}^F$  equal to  $(\mathbf{a}^u, \mathbf{a}^d)$ 
10:     $\mathbf{a}_n^F = \left\{ (\mathbf{a}_{n,i}^{F,u}, \mathbf{a}_{n,i}^{F,d}) \right\}_{i=1}^{n_{\text{ens}}}$ 
11:    if  $t \in T_D$  then
12:      for  $i = 1$  to  $n_{\text{ens}}$  do
13:        perform EnKF analysis Eq. (27)  $\mathbf{a}_{n,i}^F \rightarrow \mathbf{a}_{n,i}^A$ 
14:         $\mathbf{a}_n^A = \left\{ (\mathbf{a}_{n,i}^{A,u}, \mathbf{a}_{n,i}^{A,d}) \right\}_{i=1}^{n_{\text{ens}}}$ 
15:        for  $i = 1$  to  $n_{\text{ens}}$  do
16:          perform iterative regularization (following section)  $\mathbf{a}_{n,i}^A \rightarrow \mathbf{a}_{n,i}^R$ 
17:           $\mathbf{a}_n^R = \left\{ (\mathbf{a}_{n,i}^{R,u}, \mathbf{a}_{n,i}^{R,d}) \right\}_{i=1}^{n_{\text{ens}}}$ 
18:           $\mathbf{a}_n^F = \mathbf{a}_n^R$ 

```

▷ Initial ensemble at $t = 0$

▷ Calculate next pseudo-time forecast...

▷ ...for each ensemble member

The first step towards regularizing the posterior is to compute the displacement and phase field again at the same pseudo-time step. Instead of the monolithic approach used in the prediction, we solve for $\mathbf{u}$ and $\mathbf{d}$ individually and with a larger phase field length scale $L > \ell$.

(1) Given the result from the EnKF update, i.e. $\mathbf{a}_{n,i}^{A,u}$, solve

$$\mathbf{R}^d(\mathbf{a}_{n,i}^{A,u}, \tilde{\mathbf{a}}_{n,i}^{R,d}; \mathbf{0}, L) = 0 \quad (28)$$

for $\tilde{\mathbf{a}}_{n,i}^{R,d}$.

(2) Solve for $\tilde{\mathbf{a}}_{n,i}^{R,u}$ subject to

$$\mathbf{R}^u(\tilde{\mathbf{a}}_{n,i}^{R,u}, \tilde{\mathbf{a}}_{n,i}^{R,d}; \mathbf{0}, L) = 0. \quad (29)$$

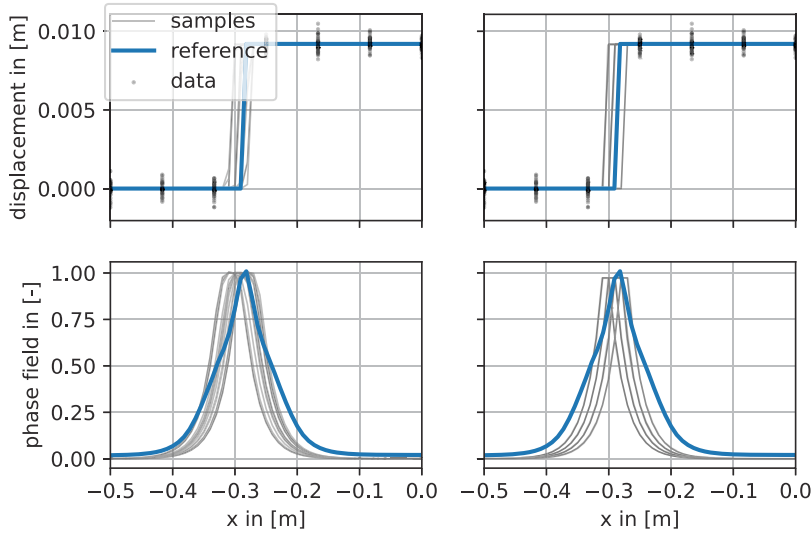


Fig. 7. Here, the EnKF analysis and regularization procedure have been performed at a later time step, where the reference solution and all ensemble members already indicate complete damage at a respective position. On the left, the process has been repeated once. On the right, the process has been repeated four times. It is visible that a single repetition yields good results already. Repeating the staggered scheme leads to similar results, but with converged ensemble member phase fields shapes. Crack nucleation does not play a role here, hence inference is simpler and the analysis exhibits a smaller standard deviation in the samples.

Note that, with $\varphi = 0$, we essentially removed the irreversibility constraint at this step. The notation $F_{1,n}$ reflects, that we are still imposing the prescribed displacement at step n .

The result is visible in Fig. 6 b). The phase field appears much wider, while still pinning the reference solution accurately. The wider phase field covers most of the noise around the position of the crack. This is also reflected in the corresponding displacement, computed with the regularized phase field. A downside is that although the displacement indeed pins the correct reference position of the crack, the remaining stiffness at the crack is lower than in the reference solution. The reason for that is the increase in the overall, cumulated, damage in the domain due to the wider phase field. In turn, that leads to a faster propagation speed of the crack in higher dimensional settings. Once a crack is fully developed, the effect is not relevant any more. Nevertheless, we consider this effect to be preferable to the noisy, unphysical posterior.

The second part consists in recomputing the phase field from the regularized displacement, this time with the original length scale:

3. Given the regularized displacements, i.e. $\tilde{\mathbf{a}}_{n,i}^{R,u}$, find $\mathbf{a}_{n,i}^{R,d}$ subject to

$$\mathbf{R}^d(\tilde{\mathbf{a}}_{n,i}^{R,u}, \mathbf{a}_{n,i}^{R,d}; \mathbf{0}, \mathcal{L}) = 0. \quad (30)$$

4. Using the phase field model without the reversibility constraint find $\mathbf{a}_{n,i}^{R,u}$ such that

$$\mathbf{R}^u(\mathbf{a}_{n,i}^{R,u}, \mathbf{a}_{n,i}^{R,d}; \mathbf{0}, \mathcal{L}) = 0. \quad (31)$$

The result is visible in Fig. 6 c). Clearly, there is still some uncertainty left, which is expected due to a limited amount of sensors and measurement noise. For comparison, the regularization scheme is applied to a completely fractured 1D domain, i.e. with a phase field maximum of 1. These results are visible in Fig. 7.

Finally, the full state vector is reconstructed as

$$\mathbf{a}_n^R = \left\{ (\mathbf{a}_{n,i}^{R,u}, \mathbf{a}_{n,i}^{R,d}) \right\}_{i=1}^{n_{\text{ens}}} \quad (32)$$

and serves as start value for the next pseudo time step. Algorithm 2 summarizes the described regularization procedure.

The regularization process is similar to the staggered approach towards solving a phase field problem [35,36] in the first place. If subsequent Newton iterations still fail to converge, the staggered scheme can be repeated. It is worth noting, however, that the stiffness can either decrease or increase further with every new iteration of the staggered scheme and hence the crack propagation speed either increases or decreases, as it can be noticed in Fig. 6(c2). The inferred position of the crack will still be inferred correctly. This effect highlights that the proposed method is able to successfully perform crack localization, which is the desired functionality for higher dimensional settings, but when repeating the staggered scheme the method is less successful at accurately inferring the magnitude of the phase field at the crack front.

Algorithm 2 Regularizing the Kalman analysis

Require: $\mathbf{a}_n^A = \left\{ (\mathbf{a}_{n,i}^{A,u}, \mathbf{a}_{n,i}^{A,d}) \right\}_{i=1}^{n_{\text{ens}}}$ ▷ Ensemble after EnKF analysis

Ensure: $L > \ell$

1: **for** $i = 1$ to n_{ens} **do** ▷ Regularize each ensemble member

2: with $\mathbf{a}_{n,i}^{A,u}$ and $\varphi_{n-1} \stackrel{!}{=} 0$

3: solve $\mathbf{R}^d(\mathbf{a}_{n,i}^{A,u}, \tilde{\mathbf{a}}_{n,i}^{R,d}, \mathbf{0}, L) = 0$ for $\tilde{\mathbf{a}}_{n,i}^{R,d}$

4: with $\tilde{\mathbf{a}}_{n,i}^{R,d}$ and $\varphi_{n-1} \stackrel{!}{=} 0$ ▷ Degrade stresses

5: solve $\mathbf{R}^u(\tilde{\mathbf{a}}_{n,i}^{R,d}, \tilde{\mathbf{a}}_{n,i}^{R,u}, \mathbf{0}, L) = 0$ for $\tilde{\mathbf{a}}_{n,i}^{R,u}$

6: with $\tilde{\mathbf{a}}_{n,i}^{R,u}$ and $\varphi_{n-1} \stackrel{!}{=} 0$

7: solve $\mathbf{R}^d(\tilde{\mathbf{a}}_{n,i}^{R,u}, \mathbf{a}_{n,i}^{R,d}, \mathbf{0}, \ell) = 0$ for $\mathbf{a}_{n,i}^{R,d}$

8: **for** $j = 1$ to n_{stagger} **do** ▷ Staggered solution scheme

9: with $\mathbf{a}_{n,i}^{R,d}$ and $\varphi_{n-1} \stackrel{!}{=} 0$

10: solve $\mathbf{R}^u(\mathbf{a}_{n,i}^{R,d}, \mathbf{a}_{n,i}^{R,u}, \mathbf{0}, \ell) = 0$ for $\mathbf{a}_{n,i}^{R,u}$

11: with $\mathbf{a}_{n,i}^{R,u}$ and $\varphi_{n-1} \stackrel{!}{=} 0$

12: solve $\mathbf{R}^d(\mathbf{a}_{n,i}^{R,u}, \mathbf{a}_{n,i}^{R,d}, \mathbf{0}, \ell) = 0$ for $\mathbf{a}_{n,i}^{R,d}$

13: $\mathbf{a}_n^R = \left\{ (\mathbf{a}_{n,i}^{R,u}, \mathbf{a}_{n,i}^{R,d}) \right\}_{i=1}^{n_{\text{ens}}}$ ▷ Ensemble after regularization

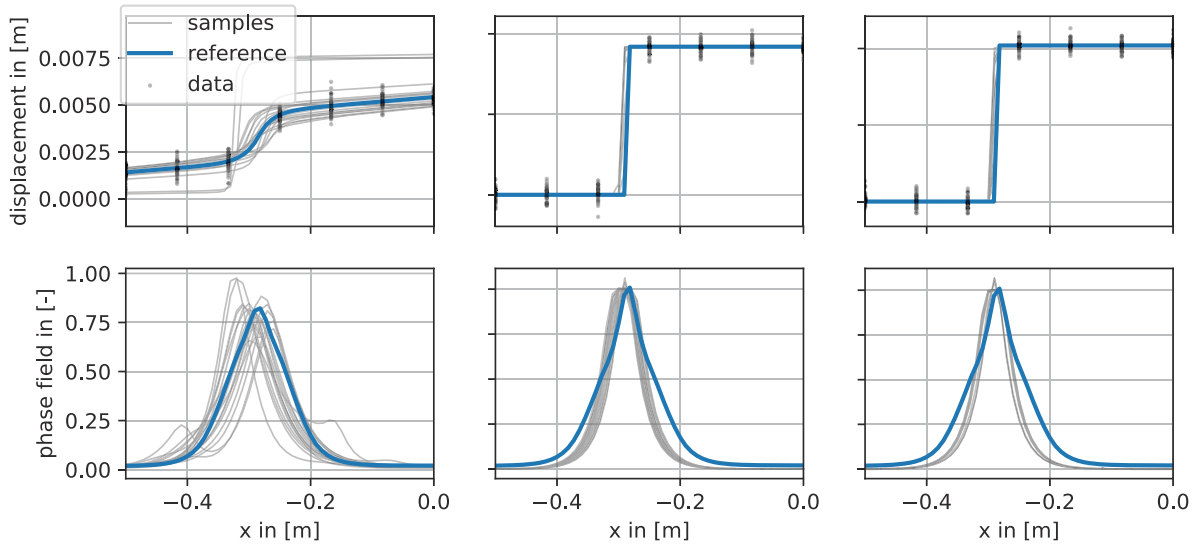


Fig. 8. The more EnKF analysis steps are included at different time steps (82, 92 and 102), the smaller the uncertainty about the crack position becomes: Displacements and phase field converge towards the reference. At each time step, the staggered scheme has been repeated once.

After regularization, the next pseudo-time prediction can be computed. Fig. 8 shows more snapshots of the analysis timeline: With more updates, the ensemble clusters more closely around the reference, hence the uncertainty decreases.

The number of sensors has an influence on the quality of the posterior ensemble. Intuitively, more sensors imply a larger weight on the data and less on the prior. Fig. 9 shows the analysis at the same time step, but with different numbers of sensors in the domain.

4. Numerical study

4.1. Single edge notched specimen under shear (SENS)

For a 2D example, the SENS benchmark is chosen. As visible in Fig. 10, the mechanical setup consists of a unit square in [mm], fixed at the bottom. No vertical displacement is permitted at the left and right edges. A notch, depicted in red, of half of the edge length is modeled within the FEM mesh. The structure is loaded in a quasi-static manner at the top boundary. The Dirichlet boundary condition on the top boundary is used to prescribe a displacement increment of $\Delta u = 1e-4$ [mm] per pseudo time step for the first

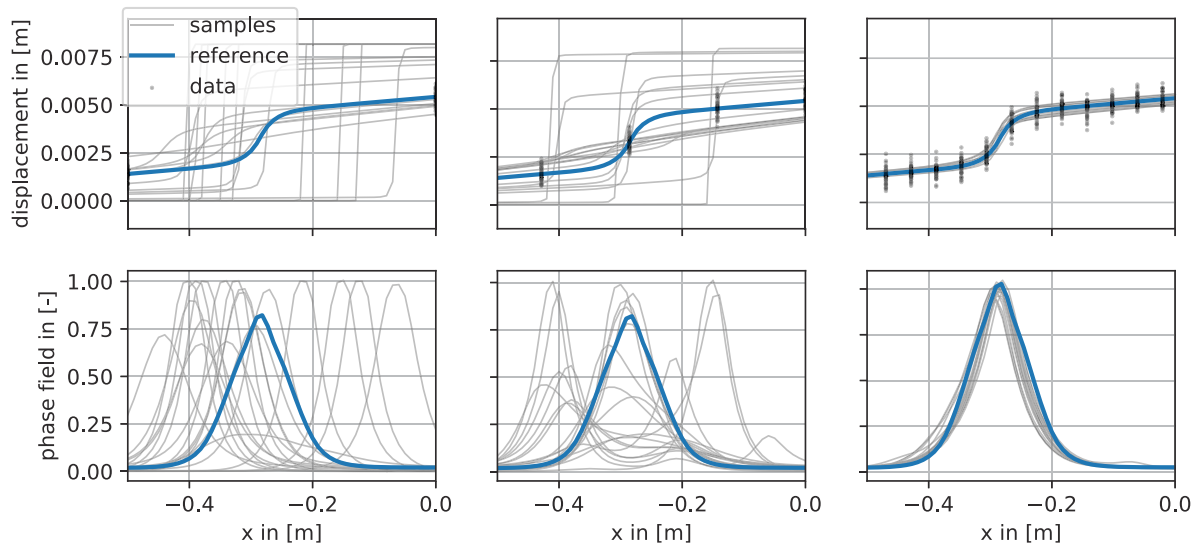


Fig. 9. The more sensor positions (5, 15, 50) are used in the data generation and analysis, the smaller the uncertainty about the crack position becomes. Compare to Fig. 8 with 25 sensor positions. All plots show the first analysis at time step 82.

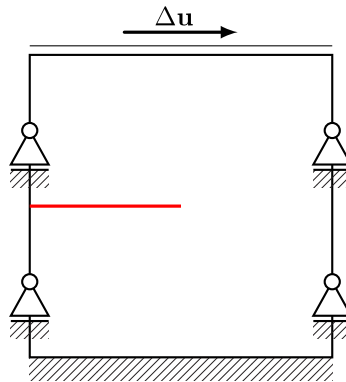


Fig. 10. SENS numerical experiment. Edge length: 1.0 mm, notch length: 0.5 mm.

Table 2

Parameters for the phase field model of the 2D numerical example.

Parameter	Value
Fracture Model	AT2
Energy Split	Volumetric-Deviatoric
E	210000 [N/mm ²]
ν	0.3 [-]
G_c	2.7 [N/mm]
l	1.5e-2 [mm]
β	200 [-]
α	$\beta G_c / l$

70 pseudo time steps. The increment decreases to $\Delta u = 1e-5$ [mm] from that time step on. The 2D solid is in a state of plane strain linear elasticity. See all model parameters in Table 2.

The ground truth, i.e. the solution fields from which noisy data are measured, is visible in Fig. 11. Data of the displacement in both spatial dimensions is collected at sparse sensor locations ($n_{\text{sens}} \ll n_{\text{dof}}$), denoted with a black cross. The parameters for the data assimilation and regularization are visible in Table 3 and Table 4, respectively.

Table 3

Parameters for the data assimilation problem. Multiple (i.e. n_{obs}) observations per sensor are assumed to be independent.

Parameter	Value
n_{ens}	100
n_{sens}	100
n_{obs}	20
σ_ϵ	$4e-4$
n_{dof}	34 071
n_{dof} ground truth	36 070
r	1.05
l_{loc}	0.45

Table 4

Regularization parameters.

Parameter	Value
l_{reg}	$4l$
n_{iter}	4

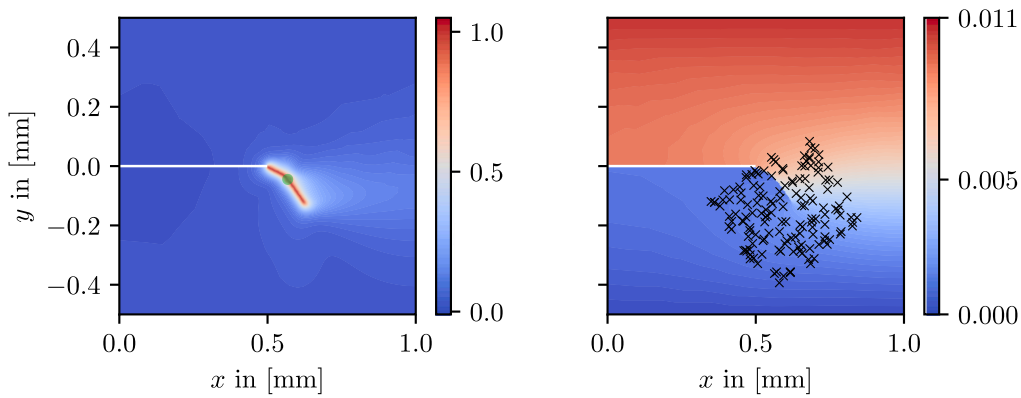


Fig. 11. Data are collected from the displacement field of the depicted phase field solution. The EnKF is supposed to restore the phase field and displacements from a prior process and the measurements only. The sensor locations are marked with black crosses. The initial pore is marked with a green dot.

The prior ensemble is constructed by randomly sampling n_{ens} different initial conditions for the phase field. Each sample is defined by a single pore in the x-y plane in proximity to the slit. The position of the pore is defined with two random variables

$$X_0 \sim 0.51 + (0.62 - 0.51) \cdot \text{Beta}(8, 8) \quad (33)$$

$$Y_0 \sim -0.11 + (0.02 - (-0.11)) \cdot \text{Beta}(8, 8). \quad (34)$$

Beta(8, 8) indicates the Beta distribution [37] with parameters $p = 8$, $q = 8$. In Fig. 12, a single prior sample of displacement in x-direction and the corresponding phase field is visible along with the absolute error. Before the EnKF analysis, for the crack tip, an error norm $\|\mathbf{x}_{\text{tip}} - \mathbf{x}_{\text{tip;ref}}\|_2$ of $3.985e^{-2}$ and an L^2 error norm $\|u_{1,i} - u_{1,\text{ref}}\|_{L^2}$ of $1.994e^{-4}$ for the displacement in x direction is observed. The second column shows the time step after an EnKF analysis step. Qualitatively, the absolute error reduced, i.e. the posterior phase field aligns more closely with the ground truth. For the crack tip, an error norm of $3.726e^{-3}$ and an L^2 norm of $9.467e^{-5}$ for the displacement in x direction is observed. The third column shows a later time step after a second analysis. The error remains small. Here, for the crack tip, an error norm of $2.244e^{-2}$ and an L^2 norm of $2.194e^{-4}$ for the displacement in x direction is observed. The chosen ensemble member exhibits a particularly strong deviation from the ground truth. Still, the analysis leads to a small error. Notice that this is also the case for all other ensemble members, i.e. the error either decreases or stays low if there is already only a small deviation. In Table 5, the corresponding hyperparameter estimates are given. Notice that the model error standard deviation is very small. This indicates a low degree of model error.

In Fig. 13, on the other hand, no analysis steps have been applied to the sample and hence the error increases over time, both in displacements and phase field. At time step 381, for the crack tip, an error norm of $3.985e^{-2}$ and an L^2 norm of $1.994e^{-4}$ for the displacement in x direction is observed. At time step 485, for the crack tip, an error norm of $5.334e^{-2}$ and an L^2 norm of $4.156e^{-4}$ for the displacement in x direction is observed. Hence, it can be stated that performing a Kalman shift on the prior ensemble indeed leads to a reduction in the error of the individual sample and therefore also in the uncertainty of the whole ensemble.

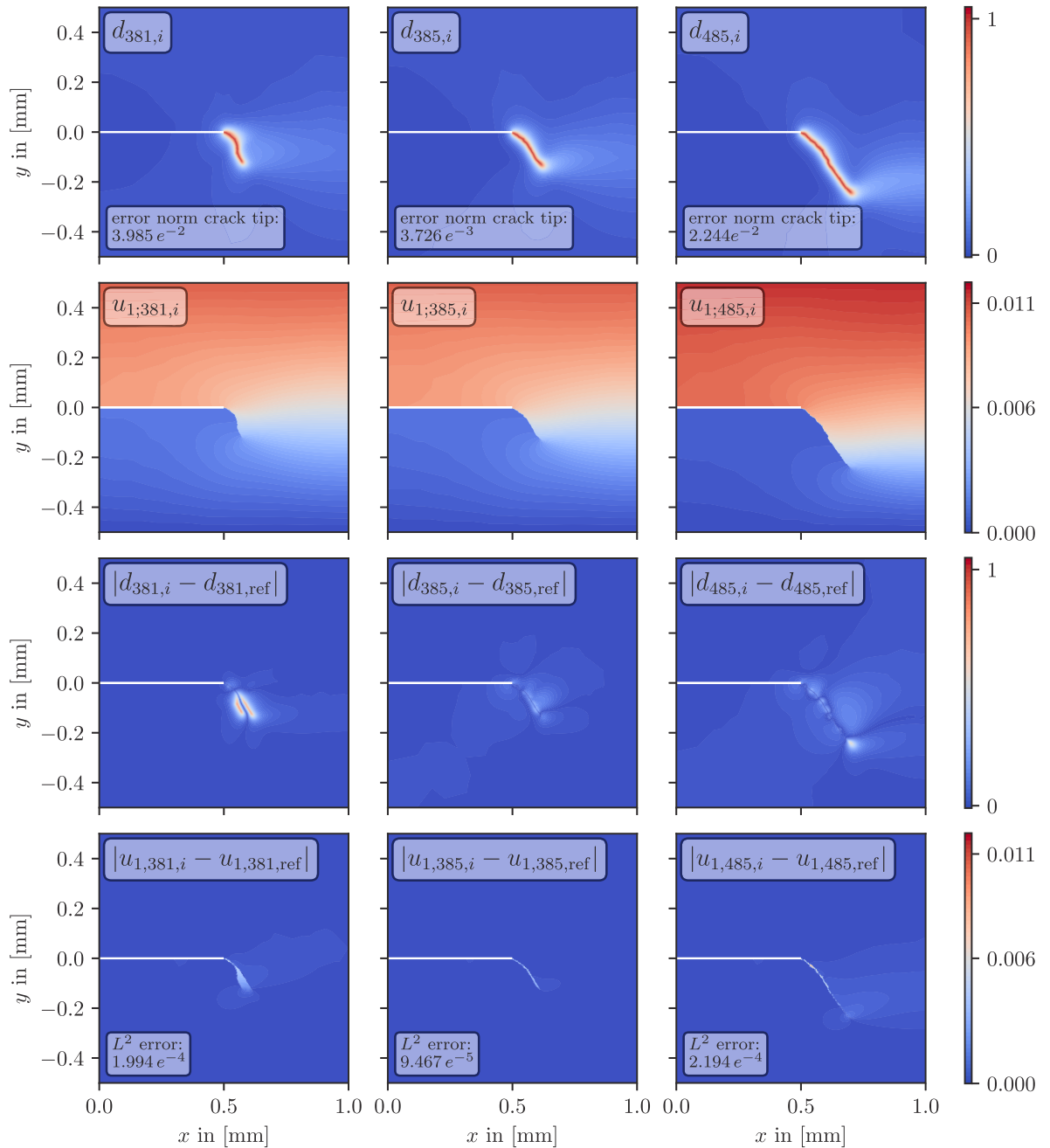


Fig. 12. SENS analysis timeline for a single ensemble member. From left to right, three different time steps are visible. At time step 381, no data assimilation has been performed yet. An error both in phase field and displacement is noticeable. At time step 382, the first data assimilation and regularization procedure has been performed. As the regularization introduced smoothing, a couple of time steps are necessary for the sample to obtain a sharp crack again. At time step 385, it is hence now visible that the sample changed and resembles the ground truth more closely. The error decreased. At time step 485, a second analysis has been performed. The error is still low.

4.1.1. Reaction forces and extrapolation after the update

An important quantity for the evaluation of the effect of a certain crack path is the remaining strength of a structure. A different path can potentially lead to a different residual load bearing capacity, which could in practice be an important quantity to base decisions on. Naturally, each random sample in the EnKF ensemble shows a different reaction force per time step. Fig. 14 shows the whole time line for the 2D SENS example. Regardless of the given initial damage, all samples behave similarly in the first, elastic only, part of the pseudo time line. Only when the fracture driving energy is large enough to initiate a crack that reaches the initial

Table 5

Point estimates for the data model hyperparameters at each ensemble analysis step in Fig. 12.

	ρ	σ_d	ℓ_d
t_{382}	0.9983	$1.470 e^{-8}$	0.1713
t_{482}	0.9923	$3.516 e^{-8}$	0.1512

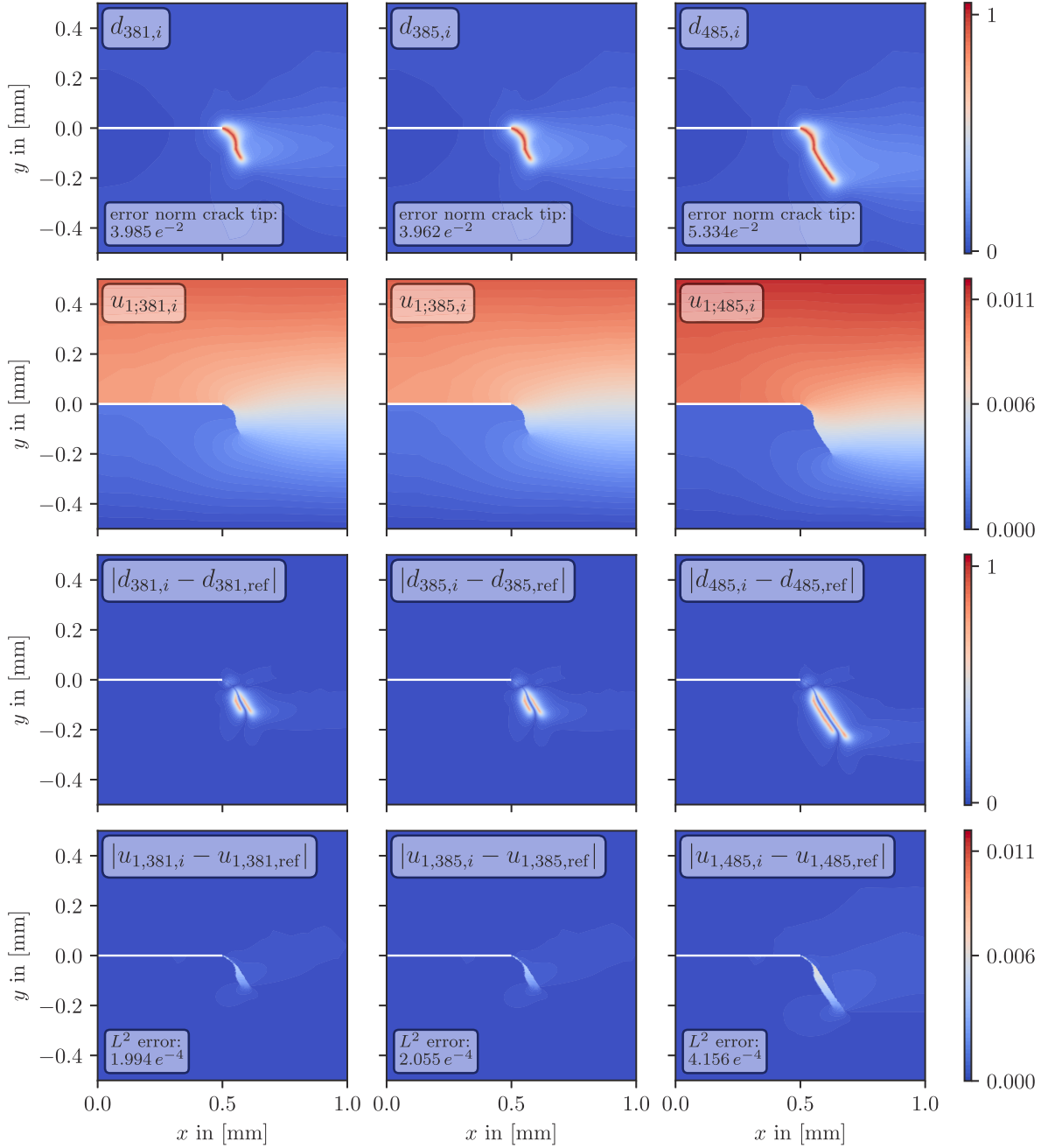


Fig. 13. SENS timeline for a single ensemble member, without data assimilation. It is visible that the error increases with every time step.

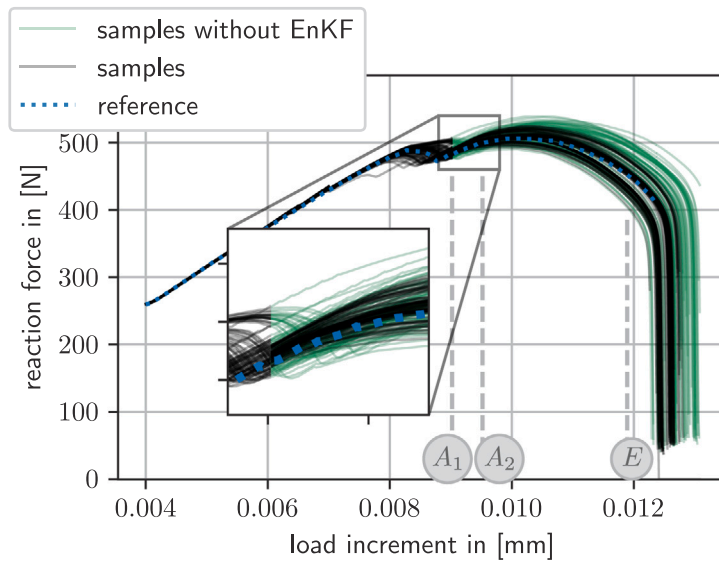


Fig. 14. Reaction forces for the SENS benchmark with a single analysis step at time step A_1 , visible in black. After the Kalman shift, the ensemble members cluster more closely around the reference solution. For comparison, the ensemble without Kalman shift is visible in green.

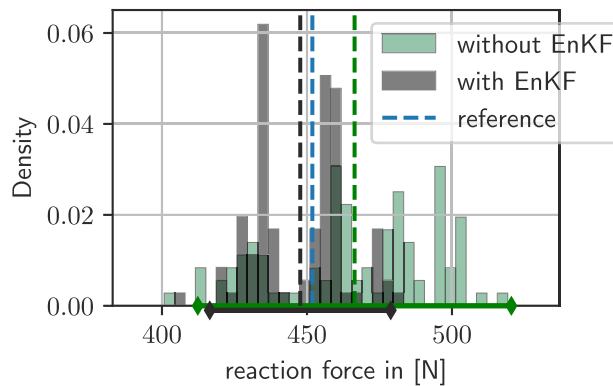


Fig. 15. Histogram of the distribution of reaction forces at extrapolation time step E for the SENS benchmark with a single analysis step at time step A . The spread (2σ) reduced while the reference solution still lies in the range of the posterior covariance.

damage, differences become apparent. The spread of different reaction forces is visible in the plotted samples. At time step A , the EnKF analysis step is performed. A couple of interesting phenomena can be observed at this point:

- Firstly, the spread reduces. While there are some outliers and a noticeable but smaller spread, most ensemble members now show a similar reaction force.
- Further in time, the average reaction force still closely follows the reference, i.e. the solution from which sparse data are drawn. Notice that the spread increases over time also for the updated ensemble. For time step E , Fig. 15 shows a histogram of the updated ensemble compared to the same ensemble simulation without the Kalman analysis step. The standard deviation of the resulting distribution is smaller than the one for the case with no analysis step at A . Also, the reference value is still captured within the PDF.
- Fig. 16 captures the histogram of the peak force of each sample, i.e. the maximum load that the structure can resist. Also here, the variance reduces significantly compared to the prior.

In Fig. 17, two analysis steps are introduced, in alignment with Fig. 12. Notice that, compared to only performing the Kalman shift once, the variance decreases even further. This is also visible in the histograms of the extrapolation time step E and the histogram of the peak force in Figs. 18 and 19, respectively. On an Intel(R) Core(TM) i7-13850HX with 64 GB of RAM, the average runtime for a single forecast time step is ca. 160 s on a single core. Hence, for the whole time line in this study, the computation of the phase field forecast itself took about 26 h. By running the code in parallel, i.e. one ensemble member per core, the runtime

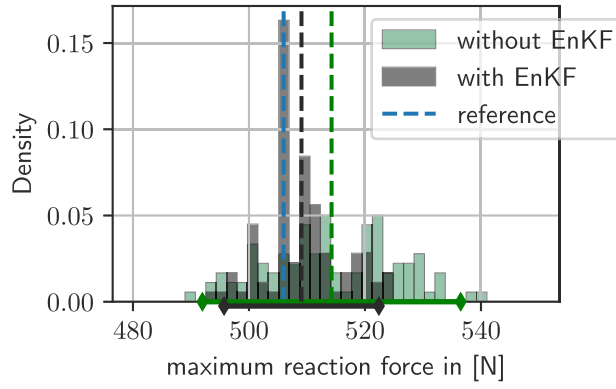


Fig. 16. Histogram of the distribution of the maximum reaction force for the SENS benchmark with a single analysis step at time step A. The spread (2σ) reduced while the reference solution still lies in the range of the posterior covariance.

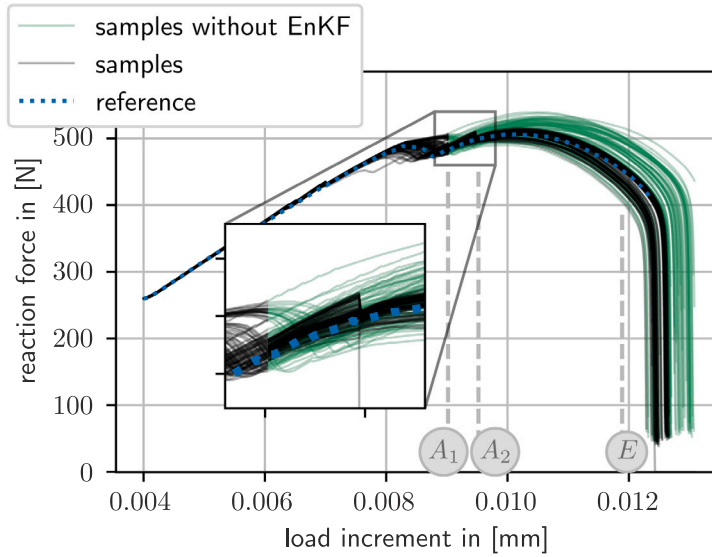


Fig. 17. Reaction forces for the SENS benchmark with two analysis steps at timesteps A_1 and A_2 , visible in black. After the Kalman shift, the ensemble members cluster more closely around the reference solution. For comparison, the ensemble without Kalman shift is visible in green.

can be decreased accordingly. The runtime for a single ensemble analysis, including the regularization step, was measured at ca. 43 min in total. We see great potential for surrogate modeling techniques at this point.

5. Discussion

Recalling the objective function of the ensemble Kalman update

$$J_n(\mathbf{a}_{n,i}^A) = \frac{1}{2} \|\mathbf{y}_n - \mathbf{H} \mathbf{a}_{n,i}^A\|_r^2 + \frac{1}{2} \|\mathbf{a}_{n,i}^A - \mathbf{a}_{n,i}^F\|_{C_n}^2, \quad (35)$$

reveals that the physical model only enters through the simulation of the prior ensemble. Indeed, the ensemble Kalman update is not forced to respect the physical constraints of the phase field model, which is the origin of the stability problems observed in the numerical experiments. The regularization projects the data-driven ensemble update back towards the physically admissible manifold, while also filtering spurious oscillations through an increased length-scale. In this section we will interpret the regularization in terms of proximal correction and formulate an outlook towards strongly constrained variational data assimilation in this context.

To this end, it is instructive to adopt the perspective of energy minimization on the discrete level, see Section 2. From this viewpoint, we can motivate one regularization step of Section 3.3.3 in the following way: given an EnKF-assimilated state $(\mathbf{u}_{n,i}^A, \mathbf{d}_{n,i}^A)$, consider

$$\tilde{\mathbf{d}}_{n,i}^R = \operatorname{argmin}_{\mathbf{d}} \left(\tau \|\mathbf{d} - \mathbf{d}_{n,i}^A\|_{C_d}^2 + E_{h,n}(\mathbf{u}_{n,i}^A, \mathbf{d}; \mathbf{0}, L) \right), \quad (36)$$

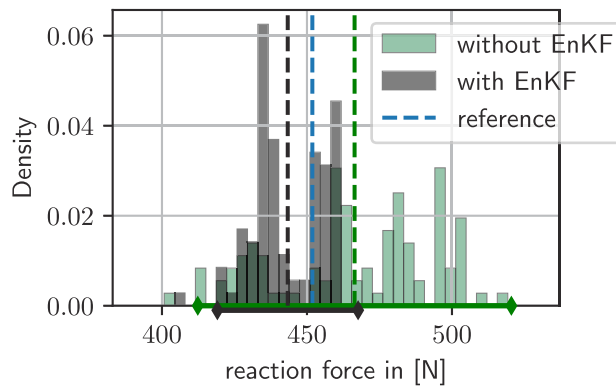


Fig. 18. Histogram of the distribution of reaction forces at extrapolation time step E for the SENS benchmark with two analysis steps at timesteps A_1 and A_2 . The spread (2σ) reduced while the reference solution still lies in the range of the posterior covariance.

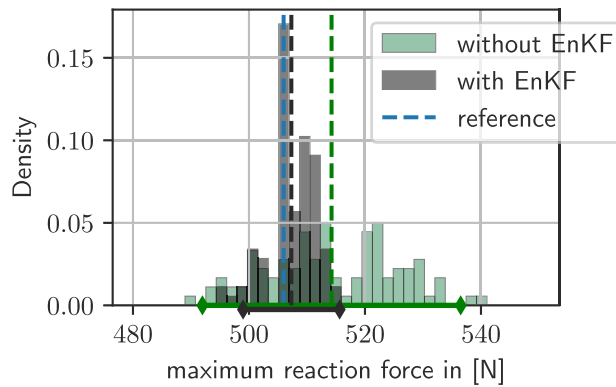


Fig. 19. Histogram of the distribution of the maximum reaction force for the SENS benchmark with two analysis steps at timesteps A_1 and A_2 . The spread (2σ) reduced while the reference solution still lies in the range of the posterior covariance.

where the irreversibility of cracks is lifted by setting $\varphi' = \mathbf{0}$ and the phase fields are smoothed by employing $L > \ell$. Note that we do not solve (36) until convergence, but simply carry out a single approximate step. This approximate proximal step projects the purely data-driven Kalman shift back into the direction of the physical model and is, in practice, realized with a single Newton step with the Kalman analysis as a starting point. In (36), large values of τ force the update to stay close to the starting value. The covariance matrix of the ensemble Kalman update naturally gives high weight to the ensemble parts with low variance. Nevertheless, as storing a complete covariance matrix in memory requires a lot of memory, we replace it with the identity matrix in the examples presented in this paper.

Having highlighted the prominent role of the physical model in phase field data assimilation it appears natural to target strongly constrained variational data assimilation in future work. The 4DVAR approach, for instance, considers smoothing instead of filtering over a time window with $N_T + 1$ steps with the model entering as strong constraint. The respective optimization problem seeks to determine the solution vector at the beginning of that window $\mathbf{a}_0^A$ such that

$$J_0(\mathbf{a}_0^A) = \frac{1}{2} \sum_{n=0}^{N_T-1} \|\mathbf{y}_{n+1} - \mathbf{H} \bar{\mathbf{a}}_{n+1}^A\|_r^2 + \frac{1}{2} \|\mathbf{a}_0^A - \mathbf{m}_0\|_{C_0}^2, \quad (37)$$

subject to the phase field update

$$\bar{\mathbf{a}}_{n+1}^A = \mathcal{M}(\bar{\mathbf{a}}_n^A), \quad (38)$$

is minimized. Here, $\mathbf{m}_0$ is an initial guess, e.g., the forecast $\mathbf{a}_n^F$ at that time step. Part of the optimization problem are all time steps with data inside the assimilation window. This implies that in each iteration of the optimizer, the model has to be solved from time 0 up until time N_T , so that the objective function can be evaluated. Although 4DVAR is formulated for a mean vector instead of an ensemble, one could also replace the mean at time 0 with an ensemble member and the covariance C_0 with the ensemble approximated covariance.

In principle, the assimilation window could be reduced to a single time step, thereby enforcing the model to hold for the analysis vector and thus only permitting physically admissible solutions. A drawback is that no closed form solution exists for the 4DVAR

optimization problem and hence, gradient descent methods or others need to be employed. This in turn leads to a huge computational cost because the model needs to be solved multiple times per time step in the assimilation window and also gradients are typically required.

6. Conclusion and outlook

In this paper, we presented a way to quantify and reduce uncertainty in a phase field brittle fracture problem by means of data assimilation. Random initial conditions lead to randomness in the crack path and hence also the remaining stiffness of the structure under consideration. In practice, knowledge over both these quantities is desired. Therefore, reducing the uncertainty about it by using an Ensemble Kalman Filter, given sparse measurement data, makes it possible to make more precise statements about the load capacity of a given structure compared to a stochastic model or data alone.

The stochastic phase field model is propagated in time up until a time step at which data are available. At that point, the measurement data are assimilated with the model and the result is regularized in order to form a suitable initial condition for the next pseudo time step. The procedure was analyzed in detail along with an illustrative 1D example. A 2D numerical benchmark example served to demonstrate the applicability of the method in more relevant settings. It could be demonstrated that the displacement and phase field states can be updated very accurately. However, a plain Ensemble Kalman Filter does not lead to acceptable results: A subsequent regularization is necessary to provide physically meaningful solutions. The problem of a lack of regularization can potentially be solved by adopting an variational perspective on data assimilation and using methods such as 4DVar, which constrain the analysis to a feasible model output without explicitly projecting the analysis after inference. However, an approach like that requires solving the model again for multiple time steps during inference, which leads to much higher computational cost.

Also, different sources of uncertainty such as randomness in parameters and the mesh should be addressed, as well as an extension of the method towards dynamical phase field methods. These involve derivatives in time for which the EnKF needs to be adapted.

Another research direction that should be explored is the optimal placement of sensors. In order not to miss important information about crack paths, it should be avoided to simply place a large number of sensors throughout the whole domain, which would in practice not be feasible. Possibly the phase field ensemble itself can be used to steer a sensor placement criterion in order to only place sensors in regions which promise a strong uncertainty reduction in the EnKF.

CRediT authorship contribution statement

Lucas Hermann: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ralf Jänicke:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **Knut Andreas Meyer:** Writing – review & editing, Supervision, Software, Methodology, Conceptualization. **Ulrich Römer:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Funding

The research leading to these results received funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Projektnummer 255042459 GRK2075/2.

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.

Appendix. Covariance inflation and localization

An EnKF is usually not used in the large ensemble limit, i.e. with more ensemble members than degrees of freedom. In the present examples, much fewer ensemble members are used compared to the high number of degrees of freedom in the phase field problem, which are necessary to resolve the crack. This can lead to numerical issues such as spurious long-distance correlations and filter collapse due to the rank deficiency of the empirical covariance matrix [38]. Two countermeasures are commonly advised in the EnKF literature: Firstly, covariance inflation is used to ensure that the covariance does not vanish. As put forth in [39], to inflate the covariance, each ensemble member is inflated individually before the analysis step. An inflation factor r , usually slightly greater than 1.0, is used. For an ensemble member $[i]$ at time step n , we obtain

$$\mathbf{a}_{n,i}^{\text{inf}} = r(\mathbf{a}_{n,i} - \bar{\mathbf{a}}_n) + \bar{\mathbf{a}}_n. \quad (39)$$

From the inflated ensemble members, an inflated covariance matrix $\mathbf{C}_n^{\text{inf}}$ is computed.

The second countermeasure is covariance localization which serves to prevent spurious correlations and hence, unphysical effects in the posterior state. According to e.g. [40], covariance localization is accomplished by performing the Schur product of a locally supported correlation matrix ρ^{loc} with the covariance matrix, i.e.

$$\mathbf{C}_n^{\text{loc}} = \rho^{\text{loc}} \circ \mathbf{C}_n^{\text{inf}}. \quad (40)$$

The authors in [40] state that there are many suitable ρ^{loc} , while a commonly used one has been defined by [41] and closely resembles a squared exponential covariance. Hence, in this paper, we use a squared exponential covariance function for the localization which works well for the discussed problems.

Data availability

The source code (written mainly in Julia [42] with Ferrite.jl [43] as a FEM backend) which was developed along this paper is available on GitHub and Zenodo at https://github.com/herluc/EnKF_Phase_Field and <https://doi.org/10.5281/zenodo.20828230>.

References

- [1] G.A. Francfort, J.J. Marigo, Revisiting brittle fracture as an energy minimization problem, *J. Mech. Phys. Solids* 46 (8) (1998) 1319–1342, [http://dx.doi.org/10.1016/S0022-5096\(98\)00034-9](http://dx.doi.org/10.1016/S0022-5096(98)00034-9).
- [2] C. Miehe, F. Welschinger, M. Hofacker, Thermodynamically consistent phase-field models of fracture: Variational principles and multi-field FE implementations, *Internat. J. Numer. Methods Engrg.* 83 (10) (2010) 1273–1311, <http://dx.doi.org/10.1002/nme.2861>.
- [3] M. Ambati, T. Gerasimov, L. De Lorenzis, Phase-field modeling of ductile fracture, *Comput. Mech.* 55 (5) (2015) 1017–1040, <http://dx.doi.org/10.1007/s00466-015-1151-4>.
- [4] M. Kalina, T. Schneider, J. Brummund, M. Kästner, Overview of phase-field models for fatigue fracture in a unified framework, *Eng. Fract. Mech.* 288 (2023) 109318, <http://dx.doi.org/10.1016/j.engfracmech.2023.109318>.
- [5] T. Gerasimov, U. Römer, J. Vondřejc, H.G. Matthies, L. De Lorenzis, Stochastic phase-field modeling of brittle fracture: Computing multiple crack patterns and their probabilities, *Comput. Methods Appl. Mech. Engrg.* 372 (2020) 113353, <http://dx.doi.org/10.1016/j.cma.2020.113353>.
- [6] S. Nagaraja, U. Römer, H.G. Matthies, L. De Lorenzis, Deterministic and stochastic phase-field modeling of anisotropic brittle fracture, *Comput. Methods Appl. Mech. Engrg.* 408 (2023) 115960, <http://dx.doi.org/10.1016/j.cma.2023.115960>.
- [7] T.C. Chu, W.F. Ranson, M.A. Sutton, Applications of digital-image-correlation techniques to experimental mechanics, *Exp. Mech.* 25 (1985) 232–244.
- [8] M. Girolami, E. Febrianto, G. Yin, F. Cirak, The statistical finite element method (statFEM) for coherent synthesis of observation data and model predictions, *Comput. Methods Appl. Mech. Engrg.* 375 (2021) 113533, <http://dx.doi.org/10.1016/j.cma.2020.113533>.
- [9] C. Duffin, E. Cripps, T. Stenler, M. Girolami, Statistical finite elements for misspecified models, *Proc. Natl. Acad. Sci.* 118 (2) (2021) e2015006118, <http://dx.doi.org/10.1073/pnas.2015006118>.
- [10] T. Wu, B. Rosić, L. De Lorenzis, H.G. Matthies, Parameter identification for phase-field modeling of fracture: A Bayesian approach with sampling-free update, *Comput. Mech.* 67 (2) (2021) 435–453, <http://dx.doi.org/10.1007/s00466-020-01942-x>.
- [11] A. Pulikathodi, E. Lacazedieu, L. Chamoine, Real-time inverse crack tracking in uncertain microstructures using PGD-based model reduction and extended Kalman filtering, *Comput. Mech.* 71 (2) (2023) 311–332, <http://dx.doi.org/10.1007/s00466-022-02240-4>.
- [12] F.A. Rochinha, A. Peirce, Monitoring hydraulic fractures: State estimation using an extended Kalman filter, *Inverse Problems* 26 (2) (2010) 025009, <http://dx.doi.org/10.1088/0266-5611/26/2/025009>.
- [13] A. Peirce, F. Rochinha, An integrated extended Kalman filter–implicit level set algorithm for monitoring planar hydraulic fractures, *Inverse Problems* 28 (1) (2011) 015009, <http://dx.doi.org/10.1088/0266-5611/28/1/015009>.
- [14] V. Pandurangan, A. Peirce, Z.R. Chen, R.G. Jeffrey, Tiltmeter mapping of measured nonsymmetric hydraulic-fracture growth in a conglomerate/sandstone formation using the implicit level-set algorithm and the extended Kalman filter, *SPE J.* 23 (01) (2017) 172–185, <http://dx.doi.org/10.2118/186100-pa>.
- [15] E. Kuncham, S. Sen, P. Kumar, H. Pathak, An online model-based fatigue life prediction approach using extended Kalman filter, *Theor. Appl. Fract. Mech.* 117 (2022) 103143, <http://dx.doi.org/10.1016/j.tafmec.2021.103143>.
- [16] J. Ping, D. Zhang, History matching of fracture distributions by ensemble Kalman filter combined with vector based level set parameterization, *J. Pet. Sci. Eng.* 108 (2013) 288–303, <http://dx.doi.org/10.1016/j.petrol.2013.04.018>.
- [17] M. Cervera, G.B. Barbat, M. Chiumenti, J.-Y. Wu, A comparative review of XFEM, mixed FEM and phase-field models for quasi-brittle cracking, *Arch. Comput. Methods Eng.* 29 (2) (2022) 1009–1083, <http://dx.doi.org/10.1007/s11831-021-09604-8>.
- [18] R. Bharali, F. Larsson, R. Jänicke, A micromorphic phase-field model for brittle and quasi-brittle fracture, *Comput. Mech.* 73 (3) (2024) 579–598, <http://dx.doi.org/10.1007/s00466-023-02380-1>.
- [19] L. De Lorenzis, T. Gerasimov, Numerical implementation of phase-field models of brittle fracture, in: L. De Lorenzis, A. Düster (Eds.), *Modeling in Engineering using Innovative Numerical Methods for Solids and Fluids*, Springer International Publishing, Cham, 2020, pp. 75–101.
- [20] D. Khimin, L. Kolditz, V. Kosin, K. Mang, T. Wick, Modeling, Discretization, Optimization, and Simulation of Phase-Field Fracture Problems, 2023, <http://dx.doi.org/10.15488/15172>.
- [21] T. Heister, M.F. Wheeler, T. Wick, A primal-dual active set method and predictor-corrector mesh adaptivity for computing fracture propagation using a phase-field approach, *Comput. Methods Appl. Mech. Engrg.* 290 (2015) 466–495, <http://dx.doi.org/10.1016/j.cma.2015.03.009>.
- [22] P. Kristensen, Phase Field Methods for Fracture - With a Special Focus on Hydrogen Assisted Fracture (Ph.D. thesis), Technical University of Denmark, 2022.
- [23] G. Evensen, *Data Assimilation: The Ensemble Kalman Filter*, Springer, Berlin, Heidelberg, 2009, <http://dx.doi.org/10.1007/978-3-642-03711-5>.
- [24] C.E. Rasmussen, C.K.I. Williams, *Gaussian processes for machine learning*, in: *Adaptive Computation and Machine Learning*, MIT Press, Cambridge, Mass, 2006.
- [25] D. Sanz-Alonso, A.M. Stuart, A. Taeb, Inverse problems and data assimilation, 2023, <http://dx.doi.org/10.48550/arXiv.1810.06191>, arXiv:1810.06191.
- [26] G. Evensen, The Ensemble Kalman Filter: Theoretical formulation and practical implementation, *Ocean. Dyn.* 53 (4) (2003) 343–367, <http://dx.doi.org/10.1007/s10236-003-0036-9>.
- [27] N. Gordon, D. Salmond, A. Smith, Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *IEE Proc. F Radar Signal Process.* 140 (2) (1993) 107, <http://dx.doi.org/10.1049/ip-f-2.1993.0015>.
- [28] M.C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 63 (3) (2001) 425–464, <http://dx.doi.org/10.1111/1467-9868.00294>.
- [29] J.P. Kaipio, E. Somersalo, *Statistical and computational inverse problems*, Applied Mathematical Sciences, Vol. 160, Springer, New York, NY, 2005, <http://dx.doi.org/10.1007/b138659>.
- [30] A. Doucet, S. Godsill, C. Andrieu, On sequential Monte Carlo sampling methods for Bayesian filtering, *Stat. Comput.* 10 (3) (2000) 197–208, <http://dx.doi.org/10.1023/a:1008935410038>.
- [31] D. Giles, M.M. Graham, M. Giordano, T. Koskela, A. Beskos, S. Guillas, ParticleDA.Jl v1.0: A real-time data assimilation software platform, *Geosci. Model. Dev.* 17 (6) (2024) 2427–2445, <http://dx.doi.org/10.5194/gmd-2023-38>.

- [32] T. Bengtsson, P. Bickel, B. Li, Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems, in: *Probability and Statistics: Essays in Honor of David a. Freedman*, Vol. 2, Institute of Mathematical Statistics, 2008, pp. 316–335, <http://dx.doi.org/10.1214/193940307000000518>.
- [33] E. Calvello, J.A. Carrillo, F. Hoffmann, P. Monmarché, A.M. Stuart, U. Vaes, Statistical accuracy of the ensemble Kalman filter in the near-linear setting, 2025, <http://dx.doi.org/10.48550/arXiv.2503.16154>, arXiv:2503.16154.
- [34] E. Calvello, P. Monmarché, A.M. Stuart, U. Vaes, Accuracy of the ensemble Kalman filter in the near-linear setting, *SIAM Journal on Numerical Analysis* 64 (2) (2026) 391–429.
- [35] M. Kirkesæther Brun, T. Wick, I. Berre, J.M. Nordbotten, F.A. Radu, An iterative staggered scheme for phase field brittle fracture propagation with stabilizing parameters, *Comput. Methods Appl. Mech. Engrg.* 361 (2020) 112752, <http://dx.doi.org/10.1016/j.cma.2019.112752>.
- [36] M. Hofacker, C. Miehe, Continuum phase field modeling of dynamic fracture: Variational principles and staggered FE implementation, *Int. J. Fract.* 178 (1–2) (2012) 113–129, <http://dx.doi.org/10.1007/s10704-012-9753-8>.
- [37] J.B. McDonald, Y.J. Xu, A generalization of the beta distribution with applications, *J. Econometrics* 66 (1) (1995) 133–152, [http://dx.doi.org/10.1016/0304-4076\(94\)01612-4](http://dx.doi.org/10.1016/0304-4076(94)01612-4).
- [38] C. Schillings, A.M. Stuart, Analysis of the Ensemble Kalman Filter for inverse problems, *SIAM J. Numer. Anal.* 55 (3) (2017) 1264–1290, <http://dx.doi.org/10.1137/16M105959X>.
- [39] T.M. Hamill, J.S. Whitaker, Accounting for the error due to unresolved scales in ensemble data assimilation: a comparison of different approaches, *Mon. Weather Rev.* 133 (11) (2005) 3132–3147, <http://dx.doi.org/10.1175/MWR3020.1>.
- [40] P.L. Houtekamer, H.L. Mitchell, A sequential ensemble Kalman filter for atmospheric data assimilation, *Mon. Weather Rev.* 129 (1) (2001) 123–137.
- [41] G. Gaspari, S.E. Cohn, Construction of correlation functions in two and three dimensions, *Q. J. R. Meteorol. Soc.* 125 (554) (1999) 723–757, <http://dx.doi.org/10.1002/qj.49712555417>.
- [42] J. Bezanson, A. Edelman, S. Karpinski, V.B. Shah, Julia: A fresh approach to numerical computing, *SIAM Rev.* 59 (1) (2017) 65–98, <http://dx.doi.org/10.1137/141000671>.
- [43] K. Carlsson, F. Ekre, Ferrite. jl contributors, Ferrite.jl, 2026, <http://dx.doi.org/10.5281/ZENODO.13862652>, Zenodo.