Image inversion and uncertainty quantification for constitutive laws of pattern formation

Hongbo Zhao [a], Richard D. Braatz [a], Martin Z. Bazant [a,b,*]

[a] *Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, United States of America*
[b] *Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, United States of America*

A B S T R A C T

The forward problems of pattern formation have been greatly empowered by extensive theoretical studies and simulations, however, the inverse problem is less well understood. It remains unclear how accurately one can use images of pattern formation to learn the functional forms of the nonlinear and nonlocal constitutive relations in the governing equation. We use PDE-constrained optimization to infer the governing dynamics and constitutive relations and use Bayesian inference and linearization to quantify their uncertainties in different systems, operating conditions, and imaging conditions. We discuss the conditions to reduce the uncertainty of the inferred functions and the correlation between them, such as state-dependent free energy and reaction kinetics (or diffusivity). We present the inversion algorithm and illustrate its robustness and uncertainties under limited spatiotemporal resolution, unknown boundary conditions, blurry initial conditions, and other non-ideal situations. Under certain situations, prior physical knowledge can be included to constrain the result. Phase-field, reaction-diffusion, and phase-field-crystal models are used as model systems. The approach developed here can find applications in inferring unknown physical properties of complex pattern-forming systems and in guiding their experimental design.

© 2021 Elsevier Inc. All rights reserved.

1. Introduction

Beyond the aesthetic value of pattern formation widely observed in many systems [1], their images can be harnessed to distill useful physical properties and test theoretical models quantitatively. In contrast to natural images and neural network models, images of pattern formation often lie on a manifold that can be described by relatively simple partial differential equations (PDE), despite the high degrees of freedom at the microscopic scale [2] and the richness of the pattern itself [3]. Identifying the PDEs that govern an evolving pattern helps to uncover its mechanism and the constitutive relations with a small number of experiments [4]. Physically, the solution to the inverse problems reveals nonequilibrium behaviors that are difficult to compute from first principles, providing a data-driven approach to modeling complex behavior. However, due to the nonlinearity of the PDEs, it remains unclear to what extent the constitutive laws of pattern formation, especially the uncertainty in the functional form of the nonlinear and nonlocal dependence, can be reliably and robustly identified from images.

---

* Corresponding author.
  E-mail address: bazant@mit.edu (M.Z. Bazant).

Analytical approaches have yielded valuable insights into the scaling laws of pattern formation [5,6] and hence how the dynamics depend on key parameters. For spinodal decompositions, it is well known that the characteristic wavenumber of the initial spinodal pattern is related to the second derivative of the free energy [7], and when the coarsening is governed by diffusion, the domain size grows as $t^{1/3}$ in time [8–10]. The growth law is found to vary with different models of concentration-dependent diffusivity [11]. For nonconserved fields, Ginzburg-Landau theory predicts a well-known $t^{1/2}$ growth law [10,9]. The structure factor of the pattern at different stages of the coarsening in both simulations and experiments has been found to collapse with the appropriate scaling [8,12].

However, scaling analysis does not describe the morphology in real space. The growth law only applies when the domain is sufficiently large and no boundary effects are present [9,10]. In addition, for systems with competing dynamics such as reaction and diffusion, the pattern depends sensitively on both reaction kinetics and diffusivity [13,3], and does not necessarily follow a simple growth law. Hence, in this work, we identify the governing equation from images through PDE-constrained optimization and use uncertainty quantification to assess the accuracy of the constitutive relations.

Previous studies have demonstrated the possibility of inferring constitutive relations or governing equations from images [4,14–21]. Parameter identification (including the nonlocal interaction kernel) and model selection in phase-separating systems has also been studied in detail [22–24,4]. With images, we have the opportunity to refine simplified models and allow unknown closures in the governing equation to be informed by data. However, for pattern-forming systems, the uncertainty in the functional form of the nonlinear or nonlocal constitutive relations remains largely unknown. Similarly, the sensitivity of the pattern morphology to the functional form is unclear, and cannot be easily elucidated by the scaling analysis mentioned above. In addition, while the patterns can be strongly influenced by specific conditions such as the average volume fraction and reaction rate, a systematic analysis of inversion based on images obtained under different conditions is lacking. We discuss solutions to these issues above in this work.

Recently, it is found that the low-dimensional manifold learned from images of spinodal decomposition is well correlated with the free energy barrier and average concentration [25,26]. Here, we take the inverse problem approach; instead of creating a forward mapping from physical properties to patterns, we ask what physical properties can be accurately inferred from full images. In contrast with learning from the low-dimensional representation of the images (which more broadly can be obtained from feature engineering [27] and dimensionality reduction techniques [28]), we utilize the full dataset in defining the objective function and in constructing the likelihood model to maximize the retrieval of useful information.

As with any regression problem, care must be taken with regards to generalization and extrapolation. Regularization is needed for the inversion of functions (infinite-dimensional inverse problems) to ensure the problem is well-posed. For example, neural networks have been used to discover or represent physical laws and constitutive relations by incorporating physical constraints [29–38]. Sparse or symbolic regression is also commonly used to achieve parsimony and better physical interpretability [39–44]. Physical constraints such as force equilibrium have been integrated with material properties to enable data-driven solutions to problems in mechanics either by using data directly without a model or inverting the constitutive model from data [45,46]. Our approach enforces the general form of the governing equation while allowing the unknown dynamics (such as reaction and diffusion) to be identified. We also allow the unknown constitutive relations to be nonlinearly or nonlocally dependent on the order parameters. Additionally, imposing symmetry in the constitutive relations and other prior knowledge such as the miscibility gap can narrow down the uncertainty and prevent overfitting.

Recently, surrogate models such as Gaussian processes [47], polynomial chaos expansion [48,49], deep neural networks [50,51], generative adversarial networks [52], and physics-informed neural networks [53] have been used in inverse uncertainty quantification and they are particularly useful when the model evaluation is expensive. Here, we choose the full Bayesian approach [54,55,17] based on the full PDE model and given snapshots to infer the posterior distribution of the constitutive relations. To address the curse of dimensionality that is associated with the large number of parameters used to represent functions, we use a dimension-independent Markov Chain Monte Carlo [56] to ensure efficient sampling.

Our method offers a top-down approach to the construction of constitutive relations of complex systems directly from macroscopically observed fields. It is complementary to the bottom-up multiscale simulation approaches of scale bridging and learning closure models [57–60]. With advanced imaging capabilities and the availability of more image data, the top-down data-driven approach to modeling pattern formation becomes increasingly relevant and hence calls for a detailed analysis of how well constitutive laws can be extracted from data [61,62].

We formulate the methods for solving PDE-constrained optimization and uncertainty quantification in Section 2 and apply the methods to various examples in Section 3 to study the inversion result, uncertainty and correlation in the inferred quantities, and whether models can be distinguished. In Section 3.1, the convergence of the optimization is demonstrated; the uncertainty quantification is applied to phase field and phase field crystal models to illustrate the quality of the sampling. We study phase-separating systems driven by diffusion and/or reaction as they relax toward equilibrium under different average concentration (Section 3.2), identify the contribution of reaction and/or diffusion (Section 3.3), and when system is chemically driven out of equilibrium (Section 3.4). We also study the effect of imaging conditions including temporal resolution (Section 3.5), spatial resolution (Section 3.6), image domain size (Section 3.7), as well as blurring (Section 3.8) on the inversion results and corresponding uncertainties.

2. Method

We study a class of pattern-forming systems that has a non-convex free energy $F[c]$ as a functional of the order parameter field $c(x)$. Phase separation occurs due to instability, that is, there exists δc such that $\delta^2 F < 0$. When the order parameter is a conserved parameter such as concentration, the dynamics can be driven toward equilibrium by diffusion,

$$\frac{\partial c}{\partial t} = \nabla \cdot (D(c)c\nabla\mu),\tag{1}$$

where the functional derivative $\mu = \delta F/\delta c$ is the chemical potential. In chemically driven systems where the interior of the system is in direct contact with the chemical reservoir, as found in surface adsorption and surface-reaction-limited nanoparticles, and when the driving is small enough to use linear irreversible thermodynamics [63,64], the governing equation is

$$\frac{\partial c}{\partial t} = -R_0(c)\mu.\tag{2}$$

Being in contact with a chemical reservoir, the free energy is now $F[c] - \mu_{\text{res}} \int c dx$ with the addition of a Lagrange multiplier. μ_{res} is the chemical potential of the chemical reservoir.

Thermodynamically, the region of instability $\delta^2 F < 0$ is prohibited unless the system is out of equilibrium. Therefore, the temporal evolution of the patterns gives us access to constitutive relations in the unstable region. In the case of a phase-separating system, the free energy is typically described by a regular-solution (or Ginzburg-Landau) type double-well energy with a gradient penalty term,

$$F[c(\mathbf{r})] = \int \left(g_h(c(\mathbf{r})) + \kappa |\nabla c|^2 \right) d\mathbf{r},\tag{3}$$

where $\mathbf{r}$ is the position in space, $g_h(c)$ is the volumetric free energy when the concentration field $c(\mathbf{r})$ is homogeneous, κ is the coefficient for penalizing concentration gradient, which leads to a diffuse interface when the system phase separates [7]. When coupled with diffusion as in Eq. (1), this expression becomes the well-known Cahn-Hilliard equation used extensively in phase field models [65]. When coupled with Eq. (2), it is known as the Allen-Cahn equation. In this case, we attempt to learn the double-well free energy $g_h(c)$ as well as the dynamics (diffusivity, and reaction kinetics). We define the homogeneous chemical potential to be $\mu_h(c) = dg_h(c)/dc$. Hence the chemical potential is

$$\mu = \mu_h(c) - \kappa \nabla^2 c.\tag{4}$$

Another type of pattern-forming system has a more generic nonlocal form,

$$F[c(\mathbf{r})] = \int g_h(c(\mathbf{r}))d\mathbf{r} + \int c(\mathbf{r}')C_2(|\mathbf{r}' - \mathbf{r}|)c(\mathbf{r})d\mathbf{r}d\mathbf{r}'.\tag{5}$$

The conserved diffusion equation (1) combined with Eq. (5) is also known as the phase field crystal equation (or dynamic density functional theory) [66]. A particular case of a non-conserved Eq. (2) combined with Eq. (5) is known as the Swift-Hohenberg equation [66,67]. Different direct correlation functions C_2 can give rise to spatial patterns such as lamellar and lattice structures. In this case, we are interested in the sensitivity of the pattern with respect to the direct correlation function.

Multiple snapshots taken in time are used as training data, while the first image is used as the initial condition. Suppose the measurement noise of the images is an additive Gaussian white noise, $c_{\text{data}}(t_j, \mathbf{r}') - c(t_i, \mathbf{r}; \mathbf{p}) \sim \mathcal{N}(0, \sigma^2 \delta(t_i - t_j, \mathbf{r} - \mathbf{r}'))$, where c is the model prediction and c_{data} is the observed concentration field, t_i is time, and $\mathbf{p}$ are the parameters for the unknown constitutive relations. The conditional probability of the observed data, aka the likelihood, is

$$P(c_{\text{data}}|\mathbf{p}) \propto \exp -\frac{1}{2\sigma^2} \left[\sum_{i=1}^{M} \int d\mathbf{r}\, (c(t_i, \mathbf{r}; \mathbf{p}) - c_{\text{data}}(t_i, \mathbf{r}))^2 \right],\tag{6}$$

where M is the total number of images in time. Similarly, we can also define the likelihood when the observed data is discrete in space and the noise is spatially and temporally uncorrelated, where the integral becomes a summation over discrete points in space. From a Bayesian perspective, the posterior distribution of the unknown parameters $\mathbf{p}$ satisfies

$$P(\mathbf{p}|c_{\text{data}}) \propto P(c_{\text{data}}|\mathbf{p})P(\mathbf{p}),\tag{7}$$

where $P(\mathbf{p})$ is the prior probability, which depends on the prior knowledge of the constitutive relations discussed in detail below.

For binary mixtures, we express the free energy as a sum of the ideal entropy of mixing (id) to limit the concentration within $[0, 1]$ and a non-ideal excess part (ex). The corresponding chemical potential is

$$\mu_h(c) = \mu_{\mathrm{id}}(c) + \mu_{\mathrm{ex}}(c) = \ln \frac{c}{1-c} + \sum_{n=1}^{N} a_n P_n(c), \tag{8}$$

where the P_n are normalized Legendre polynomials defined on the interval [0,1], and a_n are the coefficients to be determined. Similarly, the diffusivity (or reaction kinetic prefactor) can be parameterized as $\ln D(c) = \sum_n b_n P_n(c)$ to ensure positivity. Assuming the prior for the non-ideal part of the (excess) chemical potential follows a Gaussian distribution $\mu_{\mathrm{ex}}(c) \sim \mathcal{N}(0, \delta(c - c'))$, or similarly if the prior for diffusivity is $\ln D(c) \sim \mathcal{N}(0, \delta(c - c'))$, then due to the orthonormality of P_n,

$$P(\mathbf{p}) \propto \exp -\frac{1}{2} \|\mathbf{p}\|_2^2, \tag{9}$$

where $\mathbf{p}$ are the coefficients a_n. The covariance for the priors may be defined by differential operators to penalize high-frequency components in the constitutive relations. For example, $\ln D(c) \sim \mathcal{N}(0, -\mathcal{L}^{-1})$, where $\mathcal{L}\psi(c) = \frac{d}{dc}\left[c(1-c)\frac{d}{dc}\psi(c)\right]$. In other words, $P(\psi(c)) \propto \exp\left[-0.5 \int_0^1 \psi(c)L\psi(c)dc\right]$. Legendre polynomials (defined on [0,1]) are eigenfunctions of the differential operator, $\mathcal{L}P_n(c) = -n(n+1)P_n(c)$. Hence, in this set of orthonormal functions $\psi(c) = \sum_{i=0}^{\infty} p_i P_i(c)$,

$$P(\psi(c)) \propto \exp -\frac{1}{2} \sum_{i=0}^{\infty} n(n+1)p_i^2. \tag{10}$$

Note that the coefficient for the constant term $P_0(c)$ is 0 (degenerate in p_0).

The direct correlation function in Eq. (5) is represented in Fourier space in the form

$$\hat{C}_2(\mathbf{k}) = \sum_{n=0}^{N} d_n (2^n n! \sqrt{\pi})^{-1/2} e^{-|\mathbf{k}|^2/2} H_n(|\mathbf{k}|), \tag{11}$$

where $\hat{C}_2(\mathbf{k}) = \int C_2(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}} d\mathbf{r}$, H_n is the physicists' Hermite polynomial and the basis functions are orthonormal, and d_n is the corresponding coefficient. Similarly, if we assume that function $r \mapsto \hat{C}_2(r\mathbf{e})$ where $\mathbf{e}$ is a unit vector follows a Gaussian distribution with delta variance, the prior probability distribution can be written as Eq. (9), where $\mathbf{p}$ are the coefficients d_n.

For a general prior $\mathbf{p} \sim \mathcal{N}(0, \mathbf{\Gamma}_{\mathrm{Pr}})$, it is convenient to transform $\mathbf{z} = \mathbf{\Gamma}_{\mathrm{Pr}}^{-1/2}\mathbf{a}$ so that $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$, which is shown below to be useful for the Markov chain Monte Carlo (MCMC) sampling and linear constraints.

In addition to a prior that promotes smoothness and penalizes high-order basis functions, priors can be modified to satisfy certain constraints. For example, in addition to the images that capture the transient behavior, the compositions of equilibrium phases (c_1, c_2) – also known as the miscibility gap – can be easily accessible; hence the chemical potential is subject to the thermodynamic constraint,

$$\int_{c_1}^{c_2} \mu_h(c)dc = \mu_h(c_1)(c_2 - c_1) \tag{12}$$

$$\mu_h(c_1) = \mu_h(c_2). \tag{13}$$

For the Cahn-Hilliard equation, an arbitrary constant term can be added to $\mu_h(c)$, so we can set $\mu_h(c_1) = \mu_h(c_2) = 0$. These three equations are a set of linear constraints on the coefficients $\mathbf{Bz} = \mathbf{d}$. For a general prior on the coefficients, it is convenient to decompose $\mathbf{z}$ into two orthonormal spaces, $\mathbf{A}$ and $\mathbf{A}^\perp$, where $\mathbf{A}$ is the null space of $\mathbf{B}$, $\mathbf{BA} = \mathbf{0}$. Then we have $\mathbf{z} = \mathbf{A}\xi + \mathbf{A}^\perp \xi^\perp$. Thus the constraint is $\mathbf{BA}^\perp \xi^\perp = \mathbf{d}$. It can be shown that the conditional prior is

$$p(\mathbf{z}|\mathbf{Bz} = \mathbf{d}) \propto \exp -\frac{1}{2} \|\xi\|_2^2. \tag{14}$$

According to Bayes theorem, the posterior of the unknown parameters is $P(\mathbf{p}|c_{\mathrm{data}}) \propto P(c_{\mathrm{data}}|\mathbf{p})P(\mathbf{p})$. The maximum *a posteriori* estimate (MAP) is defined by minimizing the objective function

$$S(\mathbf{p}) = \frac{1}{2\sigma^2}\left[\sum_{i=1}^{M} \int d\mathbf{r} \, (c(t_i, \mathbf{r}; \mathbf{p}) - c_{\mathrm{data}}(t_i, \mathbf{r}))^2\right] + \frac{1}{2}\mathbf{p}^*\Gamma_{\mathrm{Pr}}^{-1}\mathbf{p} \tag{15}$$

Without loss of generality, 10 parameters are used for all functions studied here, that is, Legendre polynomials of order 1 to 10 are used for the chemical potential, and of order 0 to 9 are used for the diffusivity and reaction kinetics. When the thermodynamic constraint (Eq. (12)) is applied, Legendre polynomials of order 0 to 12 are used.

The objective function is optimized using a gradient-based optimizer. The gradient of the objective function is

$$\frac{\partial S}{\partial p} = \frac{1}{\sigma^2} \sum_{i=1}^{M} \int d\mathbf{r} \, (c(t_i, \mathbf{r}; \mathbf{p}) - c_{\text{data}}(t_i, \mathbf{r})) \frac{\partial c}{\partial \mathbf{p}} + \mathbf{\Gamma}_{\text{Pr}}^{-1} \mathbf{p}. \tag{16}$$

The model sensitivity $\partial c / \partial \mathbf{p}$ for each parameter can be computed while solving the forward problem. For a general PDE $\frac{\partial c}{\partial t} = g(t, c; \mathbf{p})$

$$\frac{\partial}{\partial t} \left(\frac{\partial c}{\partial \mathbf{p}} \right) = \frac{\partial g}{\partial c} \cdot \frac{\partial c}{\partial \mathbf{p}} + \frac{\partial g}{\partial p}, \tag{17}$$

which is known as forward sensitivity analysis (FSA). From the FSA, we estimate the Hessian of the objective function using Gauss-Newton approximation [68],

$$\mathbf{H}[S] \approx \frac{1}{\sigma^2} \sum_{i=1}^{M} \int d\mathbf{r} \left(\frac{\partial c}{\partial \mathbf{p}} \right)^* \frac{\partial c}{\partial \mathbf{p}} + \mathbf{\Gamma}_{\text{Pr}}^{-1}. \tag{18}$$

The approximation becomes increasingly accurate as the difference between the model and data is decreased. An alternative to FSA is adjoint sensitivity analysis (ASA), which involves solving the adjoint linear sensitivity equation backward in time once to obtain the sensitivity of the objective function with respect to all parameters [69–71] (equivalent to backpropagation for neural networks). When ASA is used, only the gradient is computed and the optimizer can use a gradient descent algorithm. This approach is useful for a large number of parameters, such as a 2D field or weights in neural nets. When the number of parameters is small, such as in the case of parametrizing concentration-dependent functions, the benefit of using FSA to obtain an estimate of the Hessian outweighs its computational cost, as the convergence is much faster. We use the trust-region algorithm [68] for the optimization when the Hessian is available.

Given the posterior distribution $P(\mathbf{p}|c_{\text{data}})$, we quantify the uncertainty by sampling the parameter space using Markov Chain Monte Carlo (MCMC). A popular adaptive Metropolis algorithm updates the covariance of the proposal distribution adaptively from the chain samples [72], that is, the covariance at step n is $C_n = s_d \text{Cov}(X_0, \dots X_{n-1}) + s_d \epsilon I$, where Cov is the sample covariance from the previous n samples and $s_d = 2.4^2/d$, where d is the dimensionality of the parameter space. This method does not compute the gradient or Hessian. In our case, we need an MCMC sampler for functions, which are parameterized in a finite-dimensional space by truncating the polynomials. With an increasingly fine representation of the function (higher dimension), the adaptive MCMC method becomes slow and the mixing quality of the Markov chain deteriorates rapidly with finer representations of the functions [55]. The sensitivity with respect to high-order basis functions leads to poorer sampling, and the sampler stagnates for a prolonged period of time, which is especially problematic for computationally expensive model evaluations of PDEs that we study here. Here we use a dimension-independent and likelihood-informed (DILI) MCMC [56] that takes the local Hessian information and adopts an operator-weighted proposal distribution to achieve better sampling efficiency that is independent of the dimensionality of the parameter space. We start the algorithm from the optimal solution (MAP). The Hessian is computed periodically using Eq. (18) to determine the parameter subspace that are most informed by the model (likelihood-informed subspace). The posterior covariance $\mathbf{\Sigma}$ is more accurately estimated by combining the covariance of projecting the chain samples onto the likelihood-informed subspace and the prior covariance in the complement prior-informed subspace. The proposal distribution comes from the discretization of the Langevin equation,

$$d\mathbf{p} = -\mathbf{\Sigma} \frac{\partial S}{\partial \mathbf{p}} d\tau + \sqrt{2\mathbf{\Sigma}} d\mathbf{W}, \tag{19}$$

whose stationary distribution is the posterior distribution (Eq. (7)), and $\mathbf{W}$ is random noise $\langle \mathbf{W}(\tau)^* \mathbf{W}(\tau') \rangle = \delta(\tau - \tau') \mathbf{I}$. As an alternative, we adopt the proposal distribution,

$$d\mathbf{p} = -\mathbf{\Sigma} \mathbf{\Gamma}_{\text{Pr}}^{-1} \mathbf{p} d\tau + \sqrt{2\mathbf{\Sigma}} d\mathbf{W} \tag{20}$$

based on gradient descent from the prior distribution only to avoid computing the model sensitivity at every step. This expression requires $\mathbf{\Gamma}_{\text{Pr}}$ to be nonsingular. In practice, for a degenerate distribution such as Eq. (10), we set the variance of the degenerate component (the coefficient of the constant basis function $P_0(c)$) to be sufficiently large. We set a maximum number of Hessian evaluations that can be performed. Therefore, the extra computational time needed for this method compared to the adaptive MCMC with the same number of samples N_s becomes vanishingly small as N_s increases.

3. Applications

3.1. Optimization and uncertainty quantification

We generate simulated data for the evolution of the Cahn-Hilliard equation. The boundary conditions are $\mathbf{n} \cdot \nabla c = 0$, which corresponds to no surface wetting, and $\mathbf{n} \cdot \nabla \mu = 0$, which corresponds to no flux [63,73]. Fig. 1 shows the evolution
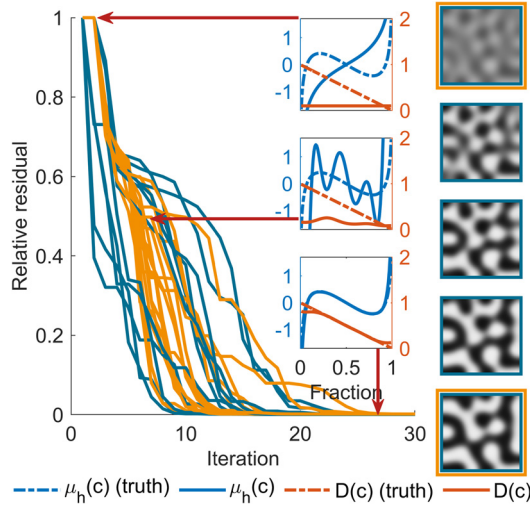
Fig. 1. Training on 11 realizations of spinodal decomposition snapshots. Insets show the initial guess for $\mu_h(c)$ and $D(c)$, a typical state during the training, and its final convergence to the truth. Blue and orange curves in the main plot correspond to taking 5 and 2 snapshots as the training datasets, respectively, as highlighted by the outline of an example set of images on the right. The images are concentration fields $c(x)$ with black and white corresponding to $c = 1$ and $c = 0$, respectively. The same colormap is used throughout this paper. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

of the objective function during the optimization using different realizations of spinodal decomposition snapshots as training data, which are generated with the same image resolution and physical parameters and different random initial conditions. These results demonstrate the robustness of the optimization algorithm. Using FSA, the Gauss-Newton approximation, and the trust-region algorithm, the optimizer converges to the truth within tens of iterations robustly with 5, or as few as 2 snapshots.

10 parameters are used for both μ_{ex} and $\ln D(c)$. The initial guess for μ_{ex} and $\ln D(c)$ is chosen such that it is far from the truth and generates a non-pattern-forming evolution ($\mu_h(c) = \ln \frac{c}{1-c} + (1 - 2c)$ and $D(c)/L^2 = 0.1$, where L is the domain size (see the inset at iteration 1 in Fig. 1; these initial guesses are used throughout the text unless otherwise noted). We recommend choosing a small $D(c)$ as the initial guess to "freeze" the pattern. If the pattern relaxes too fast and becomes uniform beyond the first frame, the sensitivity of the pattern with respect to all model parameters approaches zero, which is known as the vanishing gradient problem in machine learning. The initial guess for the gradient penalty is $\kappa/\kappa_{truth} = 0.05$. The length scale of the diffuse interface is $\sqrt{\kappa_{truth}} = 0.045L$.

To find the MAP, squared L_2 norm is used as the regularizer for $\mu_{ex}(c)$ and $\mathcal{L}$ norm is used for $\ln D(c)$, and regularization parameters for both are 10^{-5} (this set of parameters is used throughout unless otherwise noted). Due to the structure of Eq. (1), $\mu_h(c)$ and $D(c)$ can only be determined up to a constant scale; therefore, throughout the text, we report $\mu_h(c)\kappa_{truth}/\kappa$ and $D(c)\kappa/\kappa_{truth}$. In fact, the optimizer often converges to a different κ, while the scaled quantities above are accurately recovered. Allowing κ to vary in the optimization generally speeds up the convergence.

In the appendix A, we compare the performance of optimization using FSA (trust-region algorithm) versus ASA (gradient descent) and found that FSA leads to a much faster convergence, lower computational cost, and higher success rate of convergence to the global minimum (as opposed to a local minimum).

The optimization allows us to quickly find the MAP, which serves as the starting point for MCMC. For the uncertainty analysis, we assume the measurement is a continuous field. κ is fixed (since $\mu_h(c)$ and $D(c)$ can always be rescaled, as explained above), and $\mu_h(c)$ is subject to the constraint Eq. (12). The prior for $\mu_h(c)$ and $\ln D(c)$ are described by Eqs. (14) and (10), respectively, here and for the rest of the paper. The Markov chain is converted to $\mu_h(c)$ and $D(c)$ and the uncertainty at each c is plotted at 95% confidence level throughout the paper. Fig. 2a shows the uncertainty quantification result and the performance of the DILI MCMC algorithm. The shaded area corresponds to the confidence interval, while the solid line is the marginal mean at each c. The measurement noise σ may be obtained from the knowledge of the instrument or inferred together with other parameters. From here on, unless otherwise noted, we assume $\sigma^2 = 10^{-4}$ to illustrate the sensitivity more clearly. The Markov chain achieves good mixing within less than 100 steps for all parameters (20 in total) as shown by the autocorrelation of the chain in Fig. 2a. Fig. 2 also shows the trajectory of one of the parameters, which is indicative of good sampling. 2×10^4 samples are used unless otherwise noted.

In comparison, the adaptive MCMC (Fig. A.11) has a much longer autocorrelation and hence poorer mixing. The burn-in time is also much longer. Therefore, we use DILI MCMC to perform uncertainty quantification in the following examples.

If the system is non-phase-separating, $\kappa = 0$, Eq. (1) becomes

$$\frac{\partial c}{\partial t} = \nabla \cdot \left(D(c)c \cdot \mu_h'(c)\nabla c \right). \tag{21}$$
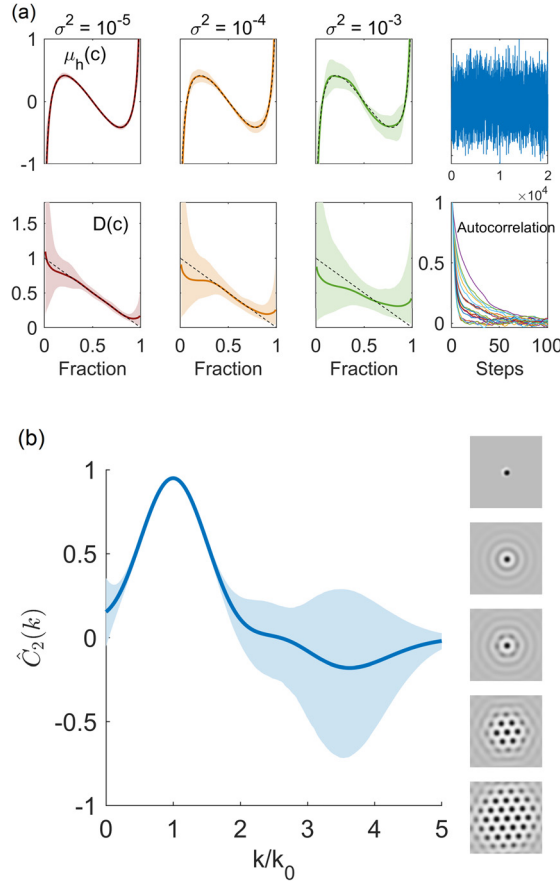
Fig. 2. (a) Uncertainty quantification of the chemical potential and diffusivity from 5 snapshots of spinodal decomposition as shown in Fig. 1. The measurement noise σ^2 is varied. The shaded regions are the 95% confidence interval of the functions at each c. The solid lines are the marginal mean of the functions at each c. The dashed lines are truth. This plotting convention is used throughout the paper unless otherwise noted. The two panels on the right are the Markov chain trajectory of a_1 and the autocorrelation of all 20 parameters, respectively. (b) Uncertainty quantification of the direct correlation from phase-field-crystal model. Images on the right are the training data.

Therefore, only the chemical diffusivity $D_{\text{chem}} = cD\mu_h'(c)$ can be inferred and no thermodynamic information can be obtained from only the concentration field. However, the extraction of $\mu_h(c)$ and $D(c)$ is possible in phase-separating systems, and is largely uncorrelated, as shown by the correlation coefficients between the parameters $a_1, a_2, \ldots, a_{10}$ and $b_0, b_1, \ldots, b_9$ in Fig. B.12. The correlation among odd/even polynomials of low order (of the same function) is strong, while the correlation between $\mu_h(c)$ and $D(c)$ is weak. The correlation between the thermodynamic and transport properties can also be quantified through certain scalars of interest. At $\sigma^2 = 10^{-5}$, $-\mu_h'(c_0)$ and $D(c_0)$ are weakly and negatively correlated (Fig. B.13). As explained in further in detail in Section 3.3, $D(c_0)\left[\mu_h'(c_0)\right]^2$ determines the initial rate of spinodal decomposition, but information from the entire field decouples $\mu_h'(c)$ and $D(c)$.

Another quantity of interest is the interfacial tension, which is defined as the energy of a flat interface at equilibrium $\gamma = \int_{-\infty}^{\infty}\left[g_h(c) + \frac{1}{2}\kappa\left(\frac{\partial c}{\partial x}\right)^2\right]$. Using the equilibrium condition $\mu(x) = \mu_0$ where μ_0 is the chemical potential of the two equilibrium phases, we have

$$\gamma = \int_{c_1}^{c_2} \sqrt{2\kappa\,\Delta g_h(c)}\,dc \tag{22}$$

where $\Delta g_h(c) = g_h(c) - g_h(c_1) - (c - c_1)\mu_0$. At the late stage of coarsening, the domain growth rate is proportional to $\gamma D(c)$ [9]. However, we find that γ and $D(c_0)$ (or $D(c_1)$) are almost uncorrelated (Fig. B.13), due to information at the diffuse interface and early-stage patterns.

For the phase field crystal models (Eq. (5)), previous studies have found correlation functions that generate certain crystal structures and elastic constants [74,75]. In Fig. 2, we use MCMC to determine the uncertainty in the direct correlation function inferred from a set of images of nucleation that forms hexagonal crystal structure ($\sigma^2 = 10^{-4}$). Using the parameterization given by Eq. (11) and the squared L_2 norm, we find the extent that the function is allowed to vary while keeping
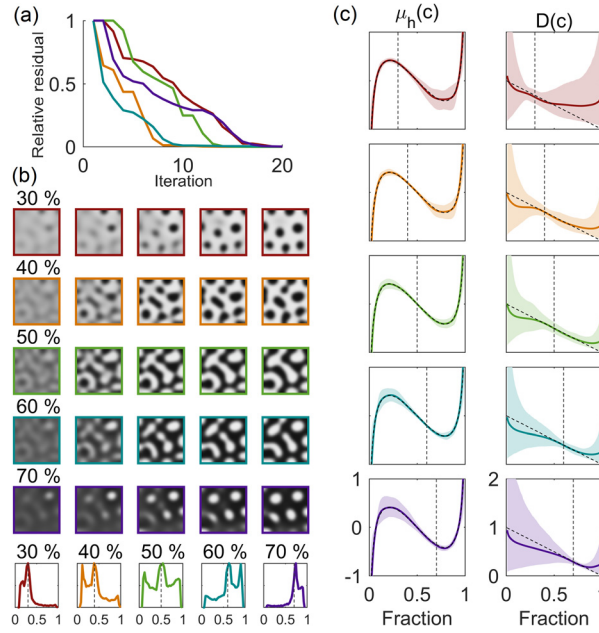
Fig. 3. Training on spinodal decomposition snapshots of mixtures from 30% to 70%. Different composition ratios are colored consistently for all figures. (a) Relative residuals for all ratios. (b) The training datasets and their composition histograms, where the average composition is highlighted by dashed lines. (c) The uncertainty in the learned chemical potential $\mu_h(c)$ and diffusivity $D(c)$.

the same pattern is very small around wavenumbers that correspond to the spatial wavelength k_0 observed in the image. Uncertainty increases for wavenumbers as moving away from k_0 due to the insensitivity of the model to these components and the lack of Fourier components at high wavenumbers in the images provided.

3.2. Starting composition

Images taken at different operating conditions can reveal information with different levels of confidence. Patterns of spinodal decomposition at different average concentration range from an interconnected structure near $c_0 = 0.5$ to dispersed droplets near $c_0 = 0.3$ and $c_0 = 0.7$ (see Fig. 3). First, an optimization is performed to find the MAP, starting from the same initial guesses as Fig. 1, convergence to the truth is achieved within 20 iterations. Then MCMC is performed to study the uncertainty in the $\mu_h(c)$ and $D(c)$. Histograms of the images show a peak near the average concentration (also the initial condition) as strong as the peak at the miscibility gap near 0 and 1. With sufficient pixel values around c_0 and their dynamics driven by $\mu_h'(c_0)$ and $D(c_0)$ at the average concentration ($\mu_h(c_1)$ is fixed to be 0), the uncertainty consistently reaches a minimum at c_0 for both constitutive relations and all average concentrations studied. Overall, images of $c = 0.5$ reduce the maximum uncertainty, while away from $c = 0.5$, multiple experiments may be necessary to accurately determine the functions on a larger range of c.

3.3. Model selection: diffusion and reaction

When the governing dynamics of the system is unknown, or it is unknown whether the order parameter is locally conserved, we consider the possibility of both conserved and nonconserved dynamics, by generalizing Eqs. (1) and (2), and allow the magnitude of both dynamics to be inferred from the patterns,

$$\frac{\partial c}{\partial t} = \nabla \cdot D(c)c\nabla\mu + R_0(c)(\mu_{\text{res}} - \mu).$$

(23)

When solving the inverse problem, the reservoir chemical potential μ_{res} is unknown. We allow it to vary in time such that the average concentration $\int c dV / \int dV$ is equal to that of the images, which is constant in this case. This becomes an algebraic constraint on the model. Reaction-diffusion models have been studied extensively in literature [13,76,77]. Here we focus on a one-component system where the reaction takes place between the system and reservoir. Lithium in lithium iron phosphate (LFP) platelet particles is known to undergo diffusion in the lateral direction via a surface layer while the platelet also exchanges lithium with the electrolyte reservoir [78]. The dynamics of a thin platelet particle may be modeled with Eq. (23) on a 2D plane [63].

The time scale of relaxation can be understood from the dispersion relation of Eq. (23) linearized around a homogeneous state $c(\mathbf{r}) = c_0$, obtained by substituting a perturbation $e^{\omega t + i\mathbf{kr}}$ into the linearized equation [79,63,80], and together with the chemical potential defined by Eq. (3),

$$\omega(k) = -\left(c_0 D(c_0)k^2 + R_0(c_0)\right)\left(\mu_h'(c_0) - \kappa k^2\right), \tag{24}$$

where $k = |\mathbf{k}|$. When diffusion dominates, $-c_0 D(c_0)\mu_h'(c_0) > R_0(c_0)\kappa$, the maximum instability growth rate is

$$\omega_{\max} = \max_k \omega(k) = \frac{\left(c_0 D(c_0)\mu_h'(c_0) - R_0(c_0)\kappa\right)^2}{4c_0 D(c_0)\kappa}, \tag{25}$$

when there is no reaction, $\omega_{\max} = c_0 D(c_0)\left[\mu_h'(c_0)\right]^2/4\kappa$. Otherwise, reaction dominates and

$$\omega_{\max} = -R_0(c_0)\mu_h'(c_0). \tag{26}$$

When diffusion dominates, the maximum growth rate of instability is obtained at a nonzero wavenumber, while when reaction dominates, the maximum growth rate corresponds to $k = 0$. Therefore, the patterns are visually different. However, when both effects are important, the relative strength of reaction and diffusion cannot be easily distinguished, which motivates a systematic approach of identifying the underlying dynamics via model selection.

We generate the images by varying the magnitude of the diffusivity while keeping the characteristic time scale $\omega_{\max}$ constant. The iso-$\omega_{\max}$ curve is shown in Fig. 4b. The images used as the training data for each diffusivity value, ranging from diffusion only to reaction only, are taken at the same time and shown in Fig. 4a. The chemical potential of the external reservoir adjusts passively while the total concentration is conserved. Training data and results are colored consistently based on its diffusivity in Fig. 4. The initial guess for $\mu_h(c)$ and κ are the same as section 3.1. The initial guesses for $R_0(c)$ and $D(c)$ are 0.1 and 0.01 respectively for all cases. 10 parameters are used to represent each function. The residual plot Fig. 4c confirms that the truth model is found for all cases and that the images are sufficient for identifying the underlying dynamics. When the true $D(c)$ or $R_0(c)$ is zero, the solver converges to a vanishingly small value for $D(c)$ and $R_0(c)$.

Fig. 4d shows the uncertainty in the inferred functions. The uncertainty in the chemical potential, or free energy, inferred from reaction-controlled patterns is higher than diffusion-controlled patterns, which is also reflected in the uncertainty of the interfacial tension. For the five cases studied $D(c_0)\left[\mu_h'(c_0)\right]^2/4\kappa = 0, 5, 10, 15, 20$, the mean and 2 standard deviations of $\gamma/\gamma_{\text{truth}}$ is 0.99 ± 0.12, 1.01 ± 0.079, 1.01 ± 0.066, 1.00 ± 0.057, 1.01 ± 0.059, respectively. The interfacial tension is $\gamma_{\text{truth}} = 0.2\sqrt{2\kappa}$. At the late stage of coarsening for Allen-Cahn equation, the interface growth rate is proportional to the local curvature and independent of the interfacial tension (see Section 2.3 in Ref. [9]), which explains the increasing uncertainty in the free energy when reaction dominates. Early-stage snapshots within the miscibility gap are essential for inferring the free energy of an Allen-Cahn system.

When reaction dominates, the diffusivity becomes increasingly uncertain, and when the true diffusivity is zero, the inferred diffusivity can be anything below a threshold. Recall that the constant term in $\ln D(c)$ has a degenerate prior; therefore the upper bound of $D(c)$ is determined by the likelihood only. The same is true for reaction kinetics – when diffusion dominates, the inferred reaction kinetics becomes uncertain. While the patterns are not sensitive to the exact form of $D(c)$ and $R_0(c)$ when either reaction or diffusion dominates, their magnitudes can be identified relatively accurately. Regardless of the magnitude of diffusivity, the uncertainty in $R_0(c)$ becomes increasingly large when c approaches 0 or 1. This effect can be understood from an analysis of the sensitivity of $c(x)$ with respect to $R_0(c)$. In the sensitivity equation (17), when close to the miscibility gap, $\frac{\partial}{\partial c}\left(\frac{\partial c}{\partial t}\right) < 0$ due to thermodynamic stability, the sensitivity of the reaction rate $\delta R_0(c) \cdot (\mu_{\text{res}} - \mu(c))$ and hence the sensitivity of $c(x)$ becomes increasingly small as $\mu \to \mu_{\text{res}}$. Similar to Section 3.1, we observe a weakly negative correlation between $-\mu_h'(c_0)$ and $D(c_0)$ (or $R_0(c_0)$), which can be understood from their negative correlation when $\omega_{\max}$ is constant. The correlation between $D(c_0)$ and $R_0(c_0)$ is weak (see Fig. B.14). The correlation among the three is at its maximum when $-c_0 D(c_0)\mu_h'(c_0) = R_0(c_0)\kappa$, which is the critical point where the system transitions from being reaction-dominated to diffusion-dominated.

3.4. Chemically driven system: the effect of autocatalysis

We study a system that is chemically driven by an external chemical reservoir at a constant average reaction rate, the concentration evolution follows [63]

$$\frac{\partial c}{\partial t} = R_0(c)f(\mu_{\text{res}} - \mu). \tag{27}$$

When driven far from equilibrium, f is no longer necessarily a linear function. Here we use $f(x) = 2\sinh(x/2)$, which is known as symmetric Butler-Volmer kinetics in electrochemistry [81]. The external chemical potential μ_{res} varies in time, subject to the constraint of the average reaction rate R,

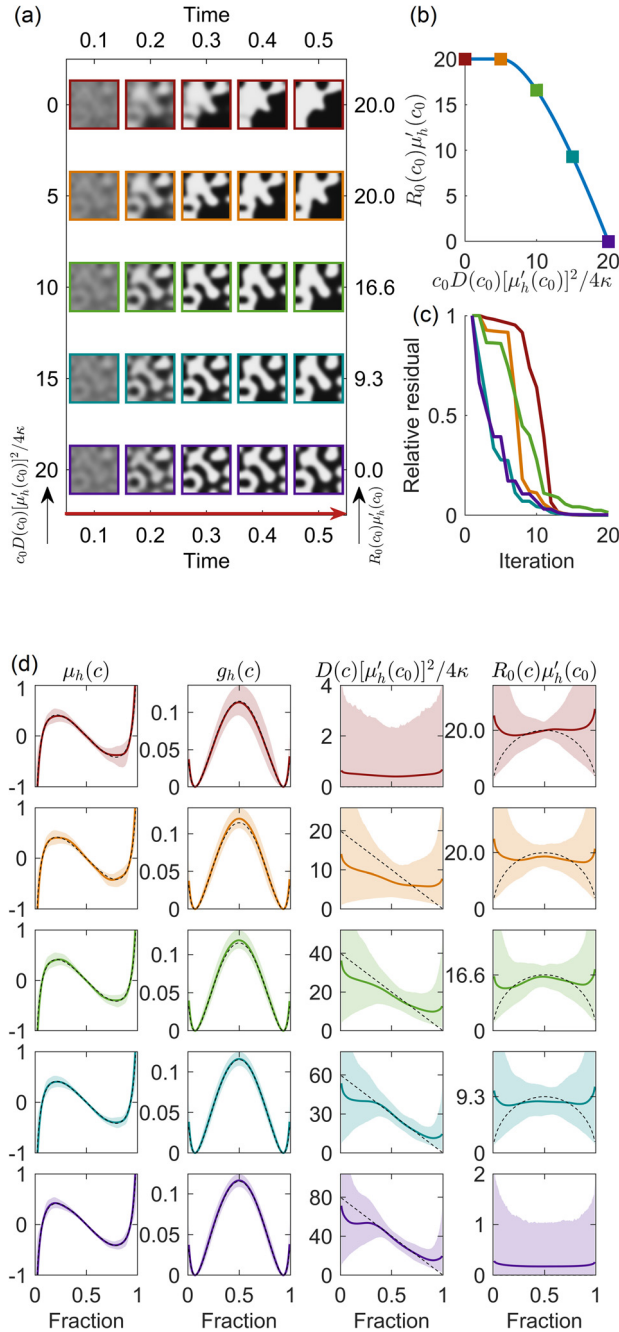$$\int \frac{\partial c}{\partial t}dV = R \int dV. \tag{28}$$

Fig. 4. Dynamics of spinodal decomposition driven by varying degrees of reaction and diffusivity. The five cases studied are colored consistently in all plots. (a) The training datasets and their corresponding physical time, scaled reaction kinetic prefactor and diffusivity. (b) The relationship between scaled reaction kinetic prefactor and diffusivity with constant instability growth rate. (c) The residual plot during the training process for all cases. (d) The uncertainty in chemical potential, free energy, scaled diffusivity and reaction kinetic prefactor for all cases. The scaling constants $\mu'_h(c_0)$ and κ are the known truth.

The linear stability and pattern formation of such a chemically driven system have been studied extensively [63,80,82–84]. The dispersion relation of Eq. (27) is $\omega = R'_0 f - R_0 f' \cdot (\mu'_h - \kappa k^2)$. Depending on the magnitude and direction of the reaction rate R, the state-dependent $R_0(c)$ can alter the linear stability (and hence the pattern) to deviate from its thermodynamic stability as determined by $\mu'_h(c)$.

In Fig. 5, we present an example where $R_0(c)$ is asymmetric about $c = 0.5$. In the range of c where $R'_0(c) < 0$, the pattern is stabilized (destabilized) when $f > 0$ ($f < 0$). With $R_0(c)$ skewed to the left, the pattern under a positive reaction rate R is more homogeneous than a negative one of the same magnitude. We start from a random initial condition and consider two sets of snapshots where $R = 0.08$ and -1, respectively.
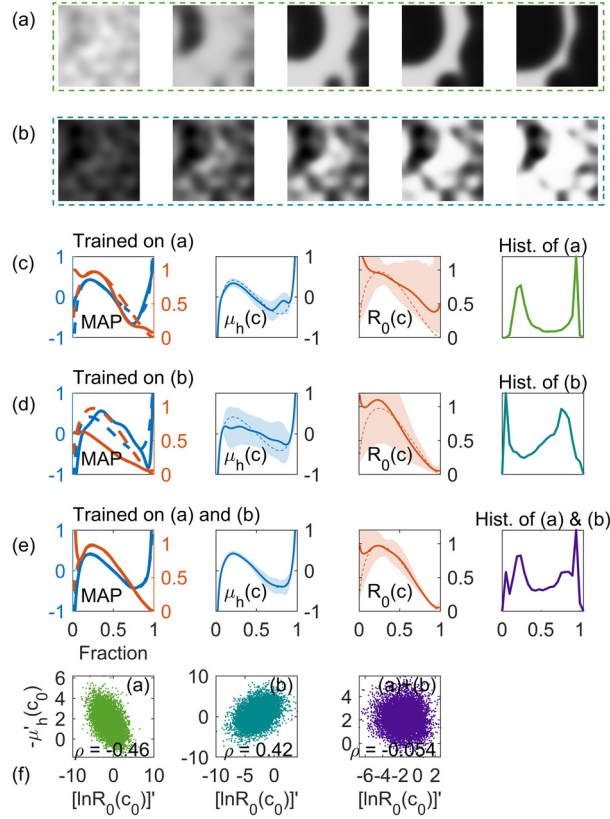
Fig. 5. Learning reaction kinetics and free energy from the spatial patterns of a system chemically driven at a constant average reaction rate. (a,b) Training data where $R = 0.08$ and $R = -1$, respectively. (c,d) Training result based on datasets (a) and (b), respectively. A comparison between the known truth and MAP, the uncertainty of $\mu_h(c)$ and $R_0(c)$, and the histogram of datasets (a) and (b). (e) Training result and uncertainty based on both (a) and (b) and histogram of the both datasets combined. (f) The scatter plots of $[\ln R_0(c_0)]'$ and $-\mu'_h(c_0)$ from the MCMC sample based on datasets (a), (b) and both combined, $c_0 = 0.5$.

In Fig. 5, we compare the inversion results based on image data from a single or both directions. The reservoir chemical potential $\mu_{res}(t)$ is unknown. We find that the systematic error between MAP and the truth is large when only images reacting in a single direction are used as the training data, even though the objective function S is sufficiently small (root-mean-squared error of frames 2–5 is less than 0.5%). MAP becomes almost identical to the truth when both directions are used, and the strongly concentration-dependent reaction kinetics $R_0(c)$ can be captured. This effect is also reflected in the marginal mean of $\mu_h(c)$ and $R_0(c)$ from the MCMC result. The uncertainty of both functions is significantly reduced when both datasets are used, highlighting the necessity of datasets at different operating conditions in order to infer both thermodynamic and kinetic properties for a chemically driven system whose only observed information is the concentration field (assuming μ_{res} is unknown).

Fig. 5f shows the scatter plots of $[\ln R_0(c_0)]'$ and $-\mu'_h(c_0)$ ($c_0 = 0.5$) from the MCMC sample. Using the dataset with $R > 0$ ($R < 0$), the two quantities are negatively (positively) correlated. When both datasets are used, their correlation is reduced. The correlation can be understood from the dispersion relation mentioned above. Linearizing (27) around a uniform field $c(x) = c_0$, we obtain [80,83]

$$
\begin{aligned}
\omega(k) &= \left[R'_0 f - R_0 f' \cdot \left(\mu'_h + \kappa k^2 \right) \right] \Big|_{c=c_0} \\
&= R \left(\frac{d \ln R_0}{dc} - \frac{d \ln f}{dc} \cdot \left(\mu'_h + \kappa k^2 \right) \right) \Big|_{c=c_0}
\end{aligned}
\tag{29}
$$

That is, if $c(x, t = 0) = c_0 + \nu e^{ikx}$, where $\nu \ll 1$ is a small perturbation, then $c(x, t) = c_0 + Rt + \nu e^{\omega(k)t + ikx}$. Note that $R \to R_0(c) f(\mu_{res} - \mu_h(c_0))$ as $\nu \to 0$. Therefore, in the limit of $t \to 0$ and $\nu \to 0$, when multiple or a range of wavenumber **k** exists, only $\omega/R = (\ln R_0)' - (\ln f)' \mu'_h$ and $(\ln f)'$ at c_0 can be inferred from the pattern. Note that ω, also known as the autocatalytic rate [80], is the key to determining the heterogeneity of the pattern: when $\omega > 0$, the pattern becomes linearly unstable and vice versa. Using Butler-Volmer kinetics,
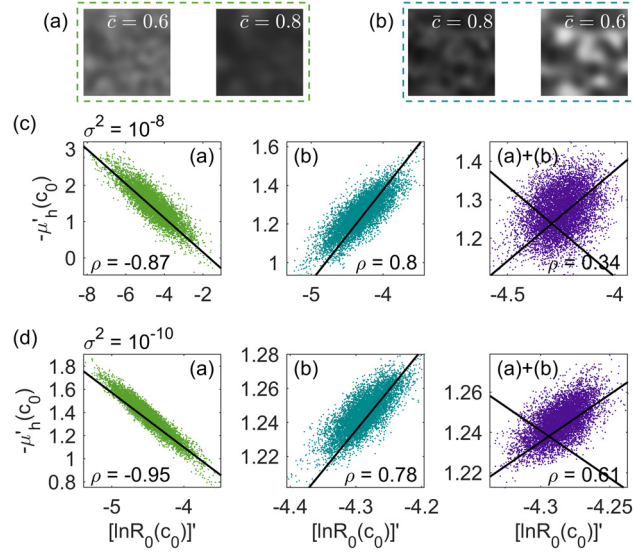
Fig. 6. The scatter plots of $[\ln R_0(c_0)]'$ and $-\mu_h'(c_0)$ from the MCMC sample (c and d, where $\sigma^2 = 10^{-8}$ and 10^{-10}, respectively) based on datasets (a), (b) and both combined, $c_0 = 0.7$. The autocatalytic rate ω is constant on the solid lines (at c_0 and the reaction rate R that corresponds to the dataset used), which analytically predict the correlation observed between $[\ln R_0(c_0)]'$ and $-\mu_h'(c_0)$.

$$\frac{d \ln f}{dc} = \frac{1}{f}\sqrt{\left(\frac{f}{2}\right)^2 + 1}$$
$$= \frac{1}{R}\sqrt{\left(\frac{R}{2}\right)^2 + R_0^2}. \tag{30}$$

Therefore, when $R > 0$ ($R < 0$), $(\ln R_0)'$ and $-\mu_h'$ are negatively (positively) correlated. To further demonstrate the correlation between the two important quantities, we perform MCMC using two datasets, each of which contains two snapshots with an average concentration of 0.6 and 0.8, with $R = 0.2$ and $R = -0.2$. In Fig. 6 with very small observation noise $\sigma^2 = 10^{-8}$ and 10^{-10}, we show that $[\ln R_0(c_0)]'$ and $-\mu_h'(c_0)$ ($c_0 = 0.7$) are strongly correlated with only one dataset used and much weaker correlation is observed when both are used. The solid lines are $(\ln R_0)' - (\ln f)' \mu_h' = $ const, where the constant and $(\ln f)'$ are determined by the known truth. The agreement of the uncertainty quantification with the analytical analysis demonstrates that when only one dataset obtained under a given reaction rate is available, reaction kinetics and thermodynamics are strongly coupled and only autocatalytic rate ω can be determined. To infer both quantities separately (to reduce the correlation of their posterior distribution), datasets under different reaction rates, preferably in different directions, are necessary. Fig. B.15 confirms that R_0 has a small posterior variance and is not correlated with μ_h', since $(\ln f)'$ can be determined independently.

3.5. Temporal resolution

The availability of snapshots over the course of spinodal decomposition and coarsening determines the uncertainty in the inferred parameters. Fig. 7 shows that with increasing number of snapshots (2, 3, and 5), the uncertainty decreases. These snapshots are evenly spaced in terms of the squared L_2 norm of the difference from the first snapshot. In Fig. 7 we define the distance between patterns to be $|\Delta| = \int (c(t, \mathbf{r}) - c(t = 0, \mathbf{r}))^2 \, d\mathbf{r}$.

At the late stage of coarsening, most pixels in the image are found to be near the miscibility gap $c = c_1$ and c_2 (see histograms in Fig. 7), and the coarsening rate is mostly determined by diffusivity near c_1 and c_2. Whenever a late-stage snapshot is provided, we observe a local minimum in the uncertainty of $D(c)$ at c_1 and c_2, as highlighted by the vertical dashed lines. If only the early spinodal stage images are provided, the uncertainty for both $\mu_h(c)$ and $D(c)$ away from the initial concentration $c_0 = 0.5$ increases. Without the early-stage snapshots, the uncertainty of the chemical potential within the miscibility gap is high. While the value of $\mu_h(c)$ at each c may be uncertain, the mean plus and minus 2 standard deviation of $\gamma/\gamma_{\text{truth}}$ is 1.02 ± 0.11, 1.01 ± 0.065, 1.00 ± 0.056, 1.02 ± 0.09, 1.03 ± 0.10, respectively, which indicates that two late-stage snapshots are as useful as two early-stage snapshots in providing information about the interfacial tension.

In fact, it is known in phase field theory that the chemical potential of a sphere of radius R is γ/R (plus some constant), its growth rate is proportional to its difference with the exterior chemical potential and the detailed functional form of $\mu_h(c)$ is not important. Therefore, images of the dynamics within the miscibility gap are critical in measuring the free energy of a phase-separating system.
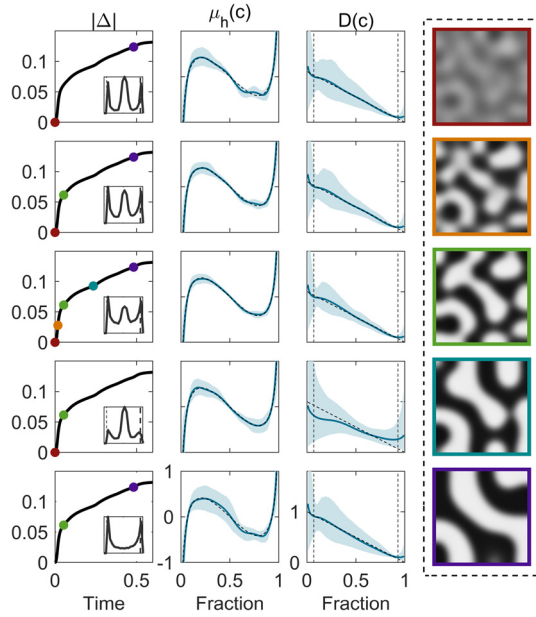
Fig. 7. The uncertainty in the chemical potential and diffusivity given snapshots taken at different time over the course of spinodal decomposition. The left columns show the squared L_2 norm of the difference between a pattern at time t and that at $t = 0$ ($|\Delta|$). The snapshots included in each row are highlighted as dots in the $|\Delta| - t$ plots whose color corresponds to that of the image outline in the far right column. From the top row to the bottom, 2, 3, and 5 snapshots equally spaced in $|\Delta|$, 2 early-stage snapshots, and 2 late-stage snapshots are taken as the training data, respectively. The insets show the histograms of combined training data of each case.

3.6. Spatial resolution

To study the effect of spatial resolution, images are taken from a subset of the field and down-sampled on a rectangular grid. The PDE is numerically solved on a finer grid to resolve the fine details. The first snapshot is linearly interpolated onto the finer simulation grid to be the initial condition. Since the boundary condition for the subset is unknown, all images are also interpolated in space to yield the concentration and its normal gradient at the boundaries, which are then interpolated in time and used as the boundary condition. Assuming that the noise at each pixel and snapshots are independent, the objective function is defined as

$$S(\mathbf{p}) = \frac{1}{2\sigma_p^2} \left[\sum_{i,j} \left(c(t_i, \mathbf{r}_j; \mathbf{p}) - c_{\text{data}}(t_i, \mathbf{r}_j) \right)^2 \right] + \|\mathbf{p}\|_L \tag{31}$$

where the pixel-wise variance is $\sigma_p^2 = 10^{-2}$. The likelihood function is defined similarly. Fig. 8a shows the MAP as well as the uncertainty from MCMC with different resolutions. Note that we place no constraint on $\mu_h(c)$ for finding the MAP but we fix the miscibility gap for MCMC.

Since the initial condition and boundary conditions are inaccurate at low temporal and spatial resolutions, MCMC predicts a systematic error compared to the truth. We also compare the MAP result when we enforce the known boundary condition (zero flux and no surface wetting) in Fig. 8b. In this case, the optimizer fails to get even close to the truth at low resolution. Therefore, when the spatial resolution is low, it is preferable to initialize from the first snapshot and impose boundary conditions from the data itself, despite the fact they are not accurately known. The inaccuracy in the initial condition can be partially compensated by the boundary condition. In fact, comparing the model prediction from the MAP results at different resolutions, the patterns are largely preserved. The optimizer fails to find a reasonable solution at a resolution of 8×8. Therefore, as a rule of thumb, at least 3 pixels per wavelength of spinodal pattern are needed.

Without constraining the miscibility gap, the optimizer converges to a $\mu_h(c)$ with smaller miscibility gap with decreasing spatial resolution. This occurs because the low sampling rate effectively filters out the high-frequency components, blurring the high contrast between the two phases.

3.7. Image domain size

In cases where the field of view is limited to a subset of the entire domain, the boundary condition is unknown. Similar to Section 3.6, the concentration and its normal gradient from each snapshots are interpolated in time and used as the boundary conditions. Fig. 9 shows the training result for images of different domain sizes. The discrepancy between the truth
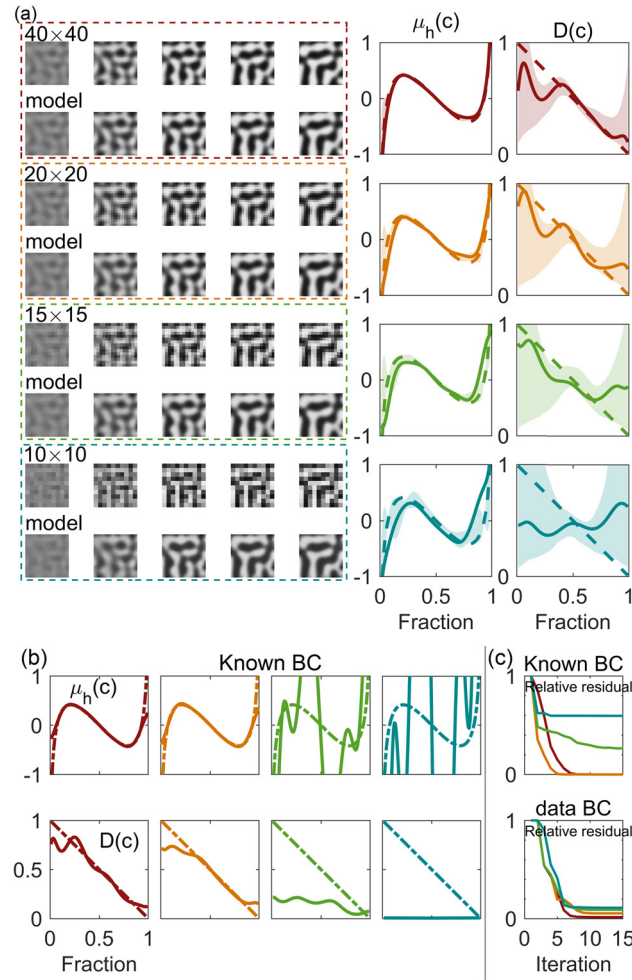
Fig. 8. The effect of image resolution on inversion and uncertainty. (a) Using images of different resolutions shown on the right and their interpolated values as the initial and boundary conditions, inversion is performed using boundary condition from the data. The MAP is shown as solid curves in the corresponding right panels. The uncertainty is shown in the same set of plots as the shaded regions. (b) Results of inversion performed on images of the same resolutions (colored consistently) using the known boundary conditions. (c) The residual during training with known boundary conditions (known BC) and boundary condition from the data (data BC).

and the MAP increases with decreasing domain size, suggesting that with the parameters become less identifiable with less information and local minima of the objective function is likely to be encountered. Only lower order polynomials of $D(c)$ can be inferred when the domain size is too small to contain large concentration variations. However, the computational cost for solving the PDE is greatly reduced if only a smaller subdomain in the training data is used. The regularization parameter is 10^{-5} times the domain size.

3.8. Blurring filter

Imaging systems have a certain point source function (PSF) that may spread over more than a single pixel. The images are a convolution of the object and the PSF. Therefore studying the inversion of these blurred images is important for practical imaging devices. In the example of spinodal decomposition used above, we show that using blurred images directly will result in systematic error in the inferred chemical potential and diffusivity. However, it is possible to invert the characteristic length scale of the PSF and more accurate physical properties simultaneously, effectively leading to a physics-constrained deconvolution.

In Fig. 10, we generate a sequence of images in (a) and convolve them with Gaussian (b and d) and box-averaging PSF (c). The PSF is indicated by the red region in the upper right corner of the last image in the sequence. The Gaussian PSF is common and can be used to approximate an Airy disk, which is the PSF of a circular aperture. Box-averaging PSF is constant within a compact support. We adopt four inversion strategies, denoted in Fig. 10 as A) assume the images are not blurred; B) assume the images are blurred by a Gaussian PSF and invert its standard deviation d together with $\mu_h(c)$ and $D(c)$; C) the same as B, except that $\mu_h(c)$ is constrained to a fixed miscibility gap, which may be measured more accurately after a
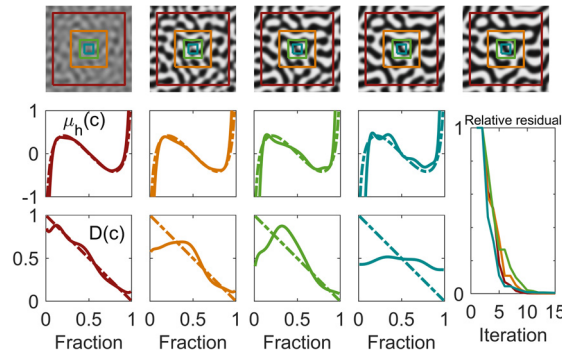
Fig. 9. Inversion performed on images of different physical sizes. Initial and boundary conditions are imposed from data interpolated in time. The chemical potential and diffusivity are the MAP results based on training data outlined by the same color. The residual plot of all cases is shown on the right.
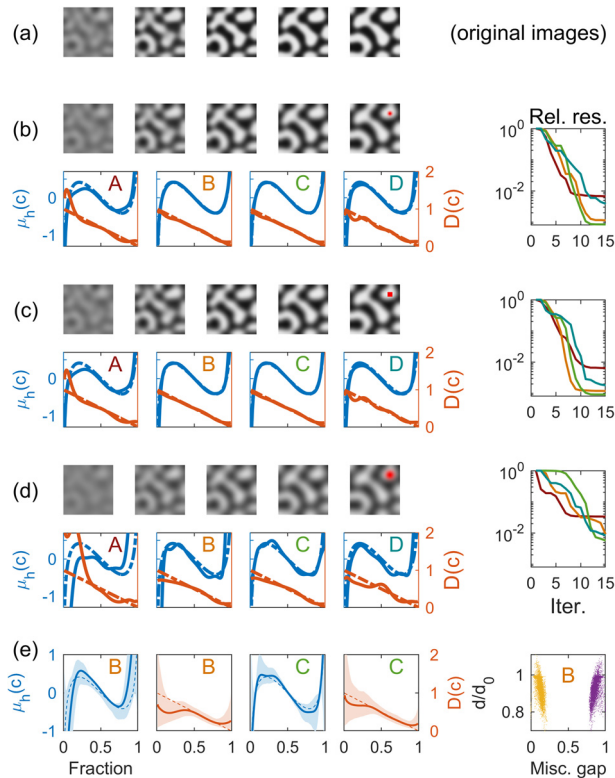

Fig. 10. Inversion performed on blurred images. (a) original images; (b,c,d) Inversion of $\mu_h(c)$ and $D(c)$ based on blurred images whose PSFs are shown in red; (e) Uncertainty quantification based on images in (c). Letters A–D correspond to inversion strategies described in the main text. $\mu_h(c)$ are plotted in blue and $D(c)$ are plotted in orange. Solid lines are the MAP solutions in (b)–(d) and marginal mean in (e). Dashed lines are the known truth. "Rel. res." stands for relative residual. The scatter plot shows the MCMC sample of the Gaussian filter length d and the miscibility gap c_1 and c_2.

long relaxation into two well-separated phases; D) the same as B, except that the first image is deconvolved and then used as the initial condition. In strategies A–C, the first given image (blurred) is used as the initial condition. In all strategies, the regularization parameter is 10^{-2}, the objective function is a discrete summation of the squared error on five 50×50 images. The initial guess is $d/L = 0.1$, $\alpha = 0.1$ (noise-to-signal ratio for deconvolution, see below), $D(c) = 0.1$.

Strategy A (treating blurred images as the truth) underestimates the interfacial tension and the miscibility gap in all cases. The objective function value that the optimizer converged to is also significantly higher than other strategies. This effect becomes more severe when the length scale of the PSF increases and becomes comparable to the spinodal length scale, $l = \sqrt{2\kappa/\mu_h'(c_0)}$. The standard deviation of the Gaussian PSF is $d = 0.03L = 0.67l$ and $d = 0.06L = 1.34l$ respectively for Figs. 10b and d, where L is the image size. For the latter case, $[-3d, 3d]$ extends to about one wavelength of the spinodal pattern $2\pi l$. The side length of the averaging box in Fig. 10c is $0.1L$.

When the PDE solution is convolved with a Gaussian PSF and d is also optimized (strategy B), the inferred $\mu_h(c)$ and $D(c)$ are closer to the truth. For the sets of images blurred by a Gaussian PSF, the correct d is found, while for Fig. 10c the solver converges to $d/L = 0.03$. For images whose unknown PSF is bounded and decays to zero, the Gaussian PSF is often a good estimate. Here, nonzero Fourier components of the images are concentrated in a narrow band around l^{-1}; blurring is only sensitive to the characteristic length scale of the PSF and not its details. Fig. 10c shows that an approximation with a Gaussian PSF is sufficient. In addition, Gaussian PSF is differentiable with respect to its parameter d, hence useful for gradient-based optimizers. Given the physical constraint of miscibility gap, strategy C shows that the objective function is further decreased.

Strategy D deconvolves the first image to use as the initial condition. To reduce the ringing effect of deconvolution due to discontinuity at the boundary, we perform a linear interpolation between the given image and the blurred image convolved with the PSF, weighing the given image more in the interior and the blurred image more at the boundary (to reduce the high frequency components near the boundary). We use $\left(1 - e^{-x^2/4d^2}\right)\left(1 - e^{-y^2/4d^2}\right)$, shifted so that $[x, y] = [0, 0]$ corresponds to the corner. We use Wiener deconvolution which in the Fourier space is $K^*(K^*K + \alpha)^{-1}$, where K is the PSF in Fourier space, α is the noise-to-signal ratio and included as a variable to be optimized. The residual plots and plots of $\mu_h(c)$ and $D(c)$ show that this strategy is close to or sometimes worse than strategies B and C which do not deconvolve the first image, which shows that, while the initial condition can be polluted by noise and blurring, preprocessing may be unnecessary. In fact, the unknown boundary condition may introduce additional error in the process of deconvolution. Inversion of a PDE whose solution diverges when the initial condition is slightly perturbed [85] is beyond the scope of this work.

The uncertainty using strategies B and C based on images in Fig. 10c and $\sigma_p^2 = 0.01$ is summarized in Fig. 10e. The scatter plot shows a strong and negative correlation between the inferred Gaussian filter length d and the width of the miscibility gap when $\mu_h(c)$ does not have any constraint, further demonstrating the importance of imposing the physical constraint. When such a constraint is imposed, the standard deviation of d/d_0 is reduced from 0.07 to 0.035, where d_0 is the truth. The miscibility gap $[c_1, c_2]$ is defined by Eq. (12) together with $\mu_h'(c_1) > 0$, $\mu_h'(c_2) > 0$ and $c_1 \neq c_2$.

4. Conclusion and outlook

In summary, using the approach of PDE-constrained optimization and Bayesian inference, we performed a systematic analysis of the inversion and the uncertainty quantification of the constitutive relations based on images of pattern formation. We showed that the optimization is robust with the use of estimated Hessian, and that the MCMC sampling is efficient in a high dimensional parameter space. For a phase-separating system, the posterior distributions of free energy (or the corresponding interfacial energy) and diffusivity based on the image data are uncorrelated, which is also corroborated by scaling analysis. The uncertainty in the inferred quantities depends strongly on the average composition and the relative contribution of reaction and diffusion. For a chemically reactive system that is driven out of equilibrium, a linear stability analysis confirms the MCMC result that the reaction kinetics and free energy can be inferred separately only when images generated at multiple reaction rates or directions are available. We show that images taken at earlier stages of spinodal decomposition can be used to infer the functional form of nonlinear free energy and diffusivity more accurately than those from the coarsening process, while the latter can also inform the interfacial tension sufficiently accurately. When the spatial resolution of the images is low, the image domain is smaller than the physical domain, or when the boundary condition is unknown, the inversion remains robust by imposing the data at the image boundary as the boundary condition in the simulation, while the uncertainty increases with decreasing spatial resolution and domain size. When the images are blurred, the inversion algorithm can extract the constitutive relation as well as the characteristic size of the blurring kernel using the first frame as the initial condition without deblurring, yet there exists a correlation between the inferred miscibility gap and the kernel size. Including the prior knowledge of the miscibility gap as a constraint further reduces the uncertainty of the inferred quantities. The examples show that when the imaging quality is limited by the instrument, such a PDE-constrained inversion effectively becomes a physics-informed superresolution imaging technique.

The methods and applications discussed here serve as a first step toward quantitative frame-by-frame and pixel-by-pixel matching between experiments and theoretical models, as excellent agreement has already been observed in many complex systems [86–89]. Despite the lack of microscopic information, the macroscopic pattern-formation dynamics can be described parsimoniously by PDEs with relatively few parameters. The inversion of the PDE means that, from the images, these macroscopic physical properties can be measured that would otherwise be inaccessible especially for systems far from equilibrium such as active matter and biology, where free energy is ill-defined, and nonequilibrium thermodynamics is poorly understood. The uncertainty quantification can be applied to optimal design of experiments to carefully probe regions of higher uncertainty as informed by prior experimental data.

The inversion also enables a physically interpretable parametrization of complex systems, which may help in establishing a mapping between the macroscopic and microscopic parameters, and eventually engineering or controlling patterns by tuning physical properties of the constituents, as reported recently in the biological engineering of Turing patterns [90] and the design of PDEs to create desired patterns [91]. Our computational approach can be integrated into the loop to accelerate the search in a high dimensional parameter space by identifying the most important engineering handles.

While phase field (Cahn-Hilliard and Allen-Cahn models) and phase field crystal models were selected as the model systems in this article, the approach can be readily extended to other systems, such as for fluid dynamics and multi-component

Table A.1

A comparison of the performance of FSA and ASA. # of snapshots: number of snapshots used as the training data. Success ratio: for 100 different datasets, the percentage of optimizations that converge to the global minimum. Iterations to success: the number of iterations (10% - median - 90% percentile) taken by the optimizer to converge to the global minimum (cases that converge to a local minima are not considered). Relative time to success: the median computational time relative to the first row for cases that converge to the global minimum.

Method and # of snapshots	Success ratio	Iterations to success	Relative time to success
FSA, 5	75%	21-31-43	1
FSA, 2	81%	19-25-34	0.7
ASA, 5	5%	33-65-93	4.6
ASA, 2	47%	66-80-100	13.0

reaction-diffusion equations, where more complicated patterns may arise [1,92,3], and the sensitivity of patterns and bifurcation dynamics with respect to constitutive relations in a high-dimensional parameter space awaits exploration. For complex systems, further study is needed to quantify the range of phenomenon that a model can describe. Techniques in inverse problems, dynamical systems, and identifiability analysis should be employed when discrepancy between experiments and models arises due to nonidealities such as spatial heterogeneity as well as unknown hidden variables.

CRediT authorship contribution statement

Hongbo Zhao: Conceptualization, Formal analysis, Methodology, Software, Visualization, Writing – original draft. **Richard D. Braatz:** Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing. **Martin Z. Bazant:** Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing.

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.

Acknowledgements

Appendix A. Comparison of optimization and uncertainty quantification algorithms

Similar to the generation of Fig. 1, we generate 100 realizations of spinodal decomposition snapshots and perform the optimization using either all five snapshots or the first and last snapshots, starting from the same initial guess as described in the main text. The termination criterion is 10^{-6} for optimality, function, and step tolerance. The objective function values that the optimizer converges to for different realizations, if at the global minimum, are very close and much smaller than values at local and suboptimal minima, hence can be easily classified. Table A.1 lists the percentage of successful convergence to the global minimum, the number of iterations, and computational time for successful cases using FSA (trust-region algorithm) and ASA (gradient descent with BFGS). We can see the ASA performs poorly on every metric considered, although it improves when initial guess is brought closer to the truth. Note that, using the ASA method, each iteration involves a line search, which usually involves multiple function and gradient evaluation, as shown by its longer average computational time per iteration than that of FSA. In FSA, each iteration requires one function, gradient and Hessian evaluation.

Fig. A.11 shows the results of an adaptive MCMC applied to the same problem in Fig. 2. The covariance of the proposal distribution at step n is $C_n = s_d \text{Cov}(X_0, \ldots X_{n-1}) + s_d \epsilon I$. We choose $\epsilon = 0.001$. Both Fig. A.11 and Fig. 2 generate 20,000 samples. The uncertainty computed in Fig. A.11 discards the first 5000 samples as the burn-in and is sufficiently close to Fig. 2. The autocorrelation of the MCMC chain using this adaptive algorithm is stronger than the DILI algorithm described in the main text.

Appendix B. Correlation between thermodynamic and kinetic properties

Fig. B.12 shows the matrix of correlation coefficients between parameters for $\mu_h(c)$ $(a_1, a_2, \ldots, a_{10})$ and $D(c)$ $(b_0, b_1, \ldots, b_9)$ inferred from spinodal decomposition snapshots (same as used by Fig. 2) and $\sigma^2 = 10^{-4}$. The correlation coefficient between a_i and a_j is $\text{Cov}(a_i, a_j)/\sqrt{\text{Var}(a_i)\text{Var}(a_j)}$, where Cov and Var stand for the covariance and variance, respectively.

Using the same set of images, Fig. B.13 shows the scatter plot of $-\mu_h'(c_0)$ versus $D(c_0)$ from the MCMC chain, as well as $\gamma/\sqrt{2\kappa}$ versus $D(c_0)$, where $c_0 = 0.5$ is the average fraction. ρ is the correlation coefficient between the two parameters
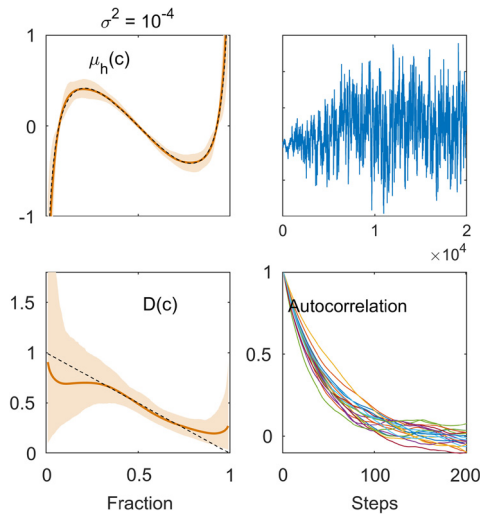
Fig. A.11. Uncertainty quantification of the chemical potential and diffusivity from 5 snapshots of spinodal decomposition as shown in Fig. 1. The measurement noise is $\sigma^2 = 10^{-4}$. The shaded regions are the 95% confidence interval of the functions at each c. The solid lines are the marginal mean of the functions at each c. The dashed lines are truth. The two panels on the right are the Markov chain trajectory of a_1 and the autocorrelation of all 20 parameters, respectively.
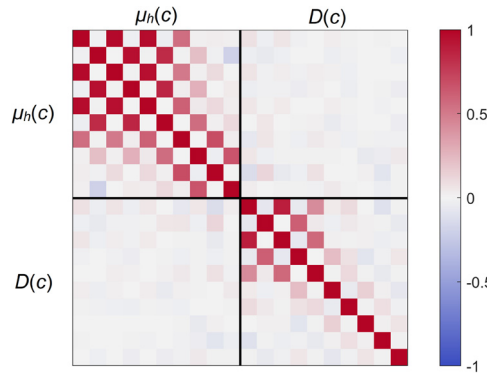

Fig. B.12. Correlation coefficients between parameters for $\mu_h(c)$ and $D(c)$.
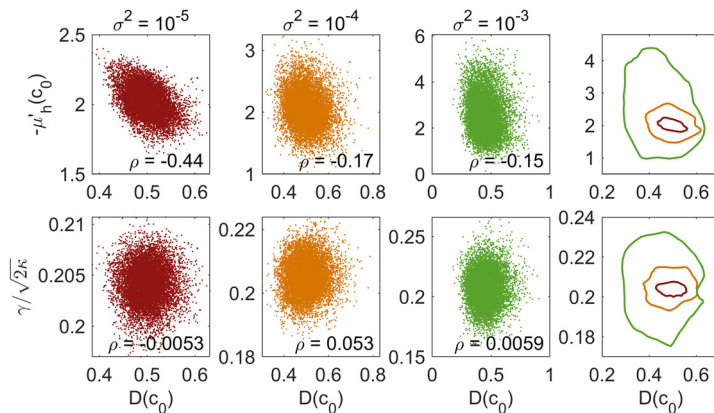
Fig. B.13. Scatter plots and 90% confidence region of $-\mu'_h(c_0)$ versus $D(c_0)$ and $\gamma/\sqrt{2\kappa}$ versus $D(c_0)$ (20,000 samples). The colors of the contours match those of the scatter plots.
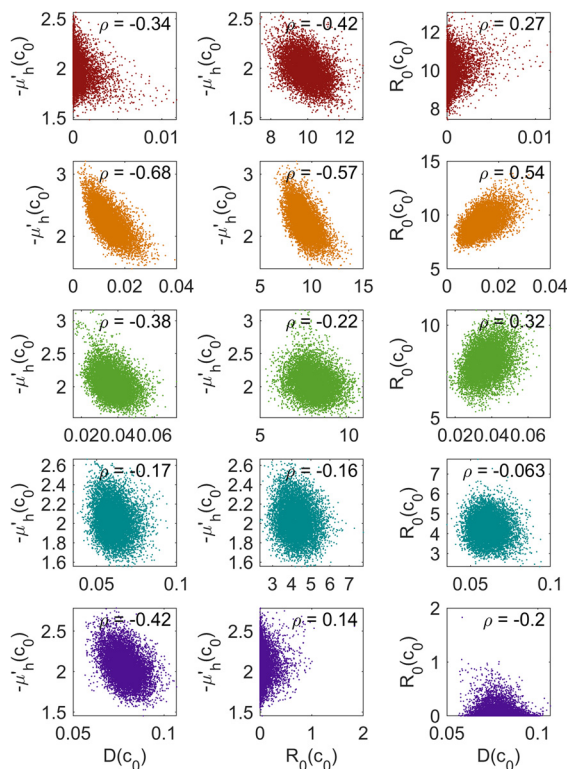
Fig. B.14. Scatter plots of $-\mu'_h(c_0)$, $D(c_0)$, and $R_0(c_0)$ (20,000 samples). Each row has a color that corresponds to the set of images having the same color in Fig. 4a.
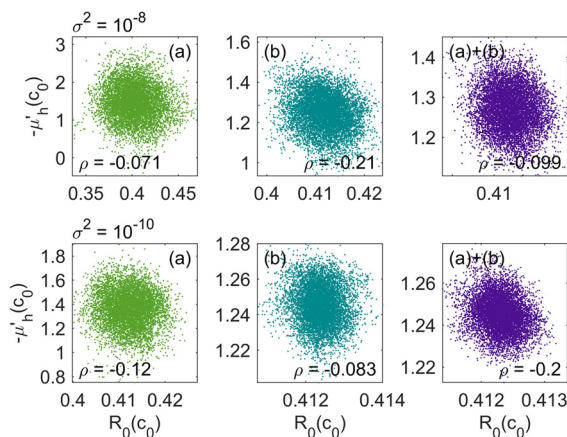

Fig. B.15. Scatter plots of $R_0(c_0)$ versus $\mu'_h(c_0)$ using datasets in Fig. 6 (20,000 samples), $c_0 = 0.7$.

as defined above. The initial rate of decomposition is $c_0 D(c_0) \left[\mu'_h(c_0)\right]^2 / 4\kappa$, which explains the weakly negative correlation between $-\mu'_h(c_0)$ and $D(c_0)$. The contours are the 90% confidence region.

In a system where both reaction and diffusion are present, we study the pairwise correlation among thermodynamics, reaction kinetics, and diffusivity by showing the scatter plots of $-\mu'_h(c_0)$, $D(c_0)$, and $R_0(c_0)$ in Fig. B.14. The samples are drawn based on images in Fig. 4.

Using the chemically driven concentration fields studied in Fig. 6a and b, Fig. B.15 shows the scatter plots of $R_0(c_0)$ and $\mu'_h(c_0)$ where $c_0 = 0.7$ and confirms that they are uncorrelated and $R_0(c_0)$ can be determined with high accuracy.

References

[1] M.C. Cross, P.C. Hohenberg, Pattern formation outside of equilibrium, Rev. Mod. Phys. 65 (3) (1993) 851–1112, https://doi.org/10.1103/RevModPhys.65.851.

[2] A.M. Turing, The chemical basis of morphogenesis, Bull. Math. Biol. 52 (1–2) (1990) 153–197, https://doi.org/10.1007/BF02459572.

[3] J.E. Pearson, Complex patterns in a simple system, Science 261 (5118) (1993) 189–192, https://doi.org/10.1126/science.261.5118.189, arXiv:patt-sol/9304003.

[4] H. Zhao, B.D. Storey, R.D. Braatz, M.Z. Bazant, Learning the physics of pattern formation from images, Phys. Rev. Lett. 124 (6) (2020) 60201, https://doi.org/10.1103/PhysRevLett.124.060201.

[5] W.M. Deen, Analysis of Transport Phenomena, 2nd edition, Oxford University Press, 2011, https://global.oup.com/ushe/product/analysis-of-transport-phenomena-9780199740284?cc=us{&}lang=en{&}.

[6] D. Fraggedakis, N. Nadkarni, T. Gao, T. Zhou, Y. Zhang, Y. Han, R.M. Stephens, Y. Shao-Horn, M.Z. Bazant, A scaling law to determine phase morphologies during ion intercalation, Energy Environ. Sci. (2020) 22–27, https://doi.org/10.1039/d0ee00653j.

[7] J.W. Cahn, J.E. Hilliard, Free energy of a nonuniform system. I. Interfacial free energy, J. Chem. Phys. 28 (2) (1958) 258–267, https://doi.org/10.1063/1.1744102.

[8] H. Furukawa, A dynamic scaling assumption for phase separation, Adv. Phys. 34 (6) (1985) 703–750, https://doi.org/10.1080/00018738500101841.

[9] A.J. Bray, Theory of phase-ordering kinetics, Adv. Phys. 51 (2) (2002) 481–587, https://doi.org/10.1080/00018730110117433.

[10] I. Lifshitz, V. Slyozov, The kinetics of precipitation from supersaturated solid solutions, J. Phys. Chem. Solids 19 (1–2) (1961) 35–50, https://doi.org/10.1016/0022-3697(61)90054-3, http://linkinghub.elsevier.com/retrieve/pii/0022369761900543, arXiv:1011.1669v3.

[11] H. Manzanarez, J.P. Mericq, P. Guenoun, J. Chikina, D. Bouyer, Modeling phase inversion using Cahn-Hilliard equations – influence of the mobility on the pattern formation dynamics, Chem. Eng. Sci. 173 (2017) 411–427, https://doi.org/10.1016/j.ces.2017.08.009.

[12] T. Nishi, T.T. Wang, T.K. Kwei, Thermally induced phase separation behavior of compatible polymer mixtures, Macromolecules 8 (2) (1975) 227–234, https://doi.org/10.1021/ma60044a025.

[13] S. Kondo, T. Miura, Reaction-diffusion model as a framework for understanding biological pattern formation, Science 329 (5999) (2010) 1616–1620, https://doi.org/10.1126/science.1179047.

[14] V. Bukshtynov, O. Volkov, B. Protas, On optimal reconstruction of constitutive relations, Physica D 240 (16) (2011) 1228–1244, https://doi.org/10.1016/j.physd.2011.04.006.

[15] V. Bukshtynov, B. Protas, Optimal reconstruction of material properties in complex multiphysics phenomena, J. Comput. Phys. 242 (2013) 889–914, https://doi.org/10.1016/j.jcp.2013.02.034.

[16] A.K. Sethurajan, S.A. Krachkovskiy, I.C. Halalay, G.R. Goward, B. Protas, Accurate characterization of ion transport properties in binary symmetric electrolytes using in situ NMR imaging and inverse modeling, J. Phys. Chem. B 119 (37) (2015) 12238–12248, https://doi.org/10.1021/acs.jpcb.5b04300.

[17] A. Sethurajan, S. Krachkovskiy, G. Goward, B. Protas, Bayesian uncertainty quantification in inverse modeling of electrochemical systems, J. Comput. Chem. 40 (5) (2019) 740–752, https://doi.org/10.1002/jcc.25759, arXiv:1806.00036.

[18] J. Morales Escalante, W. Ko, J.M. Foster, S. Krachkovskiy, G. Goward, B. Protas, Discerning models of phase transformations in porous graphite electrodes: insights from inverse modelling based on MRI measurements, Electrochim. Acta 349 (2020) 136290, https://doi.org/10.1016/j.electacta.2020.136290.

[19] A.Y. Borisevich, A.N. Morozovska, Y.-M. Kim, D. Leonard, M.P. Oxley, M.D. Biegalski, E.A. Eliseev, S.V. Kalinin, Exploring mesoscopic physics of vacancy-ordered systems through atomic scale observations of topological defects, Phys. Rev. Lett. 109 (6) (2012) 065702, https://doi.org/10.1103/PhysRevLett.109.065702.

[20] X. Li, O. Dyck, R.R. Unocic, A.V. Ievlev, S. Jesse, S.V. Kalinin, Statistical learning of governing equations of dynamics from in-situ electron microscopy imaging data, Mater. Des. 195 (2020) 108973, https://doi.org/10.1016/j.matdes.2020.108973.

[21] H. You, Y. Yu, S. Silling, M. D'Elia, Data-driven learning of nonlocal models: from high-fidelity simulations to constitutive laws, 2020, arXiv:2012.04157.

[22] A. Hawkins-Daarud, S. Prudhomme, K.G. van der Zee, J.T. Oden, Bayesian calibration, validation, and uncertainty quantification of diffuse interface models of tumor growth, J. Math. Biol. 67 (6–7) (2013) 1457–1485, https://doi.org/10.1007/s00285-012-0595-9.

[23] C. Kahle, K.F. Lam, J. Latz, E. Ullmann, Bayesian parameter identification in Cahn–Hilliard models for biological growth, SIAM/ASA J. Uncertain. Quantificat. 7 (2) (2019) 526–552, https://doi.org/10.1137/18M1210034.

[24] E. Rocca, L. Scarpa, A. Signori, Parameter identification for nonlocal phase field models for tumor growth via optimal control and asymptotic analysis, 2020, arXiv:2009.11159, 1–39.

[25] F. Schoeneman, V. Chandola, N. Napp, O. Wodo, J. Zola, Entropy-isomap: manifold learning for high-dimensional dynamic processes, in: Proc. - 2018 IEEE Int. Conf. Big Data, Big Data 2018, 2019, pp. 1655–1660, arXiv:1802.06823.

[26] F. Schoeneman, V. Chandola, N. Napp, O. Wodo, J. Zola, Learning manifolds from dynamic process data, Algorithms 13 (2) (2020) 1–16, https://doi.org/10.3390/a13020030.

[27] K.A. Severson, P.M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M.H. Chen, M. Aykol, P.K. Herring, D. Fraggedakis, M.Z. Bazant, S.J. Harris, W.C. Chueh, R.D. Braatz, Data-driven prediction of battery cycle life before capacity degradation, Nat. Energy 4 (5) (2019) 383–391, https://doi.org/10.1038/s41560-019-0356-8.

[28] G.E. Hinton, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507, https://doi.org/10.1126/science.1127647.

[29] M. Raissi, A. Yazdani, G.E. Karniadakis, Hidden fluid mechanics: learning velocity and pressure fields from flow visualizations, Science 367 (6481) (2020) 1026–1030, https://doi.org/10.1126/science.aaw4741.

[30] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. 378 (2019) 686–707, https://doi.org/10.1016/j.jcp.2018.10.045.

[31] M. Raissi, G.E. Karniadakis, Hidden physics models: machine learning of nonlinear partial differential equations, J. Comput. Phys. 357 (2018) 125–141, https://doi.org/10.1016/j.jcp.2017.11.039.

[32] R. Tipireddy, P. Perdikaris, P. Stinis, A. Tartakovsky, A comparative study of physics-informed neural network models for learning unknown dynamics and constitutive relations, Apr. 2019, arXiv:1904.04058.

[33] A.M. Tartakovsky, C.O. Marrero, P. Perdikaris, G.D. Tartakovsky, D. Barajas-Solano, Learning parameters and constitutive relationships with physics informed deep neural networks, 2018, arXiv:1808.03398.

[34] S. Greydanus, M. Dzamba, J. Yosinski, Hamiltonian neural networks, in: NeurIPS 2019, 2019, pp. 1–11, arXiv:1906.01563.

[35] M. Cranmer, S. Greydanus, S. Hoyer, P. Battaglia, D. Spergel, S. Ho, Lagrangian neural networks, 2020, arXiv:2003.04630, 1–9.

[36] D.Z. Huang, K. Xu, C. Farhat, E. Darve, Learning constitutive relations from indirect observations using deep neural networks, J. Comput. Phys. 416 (2020), https://doi.org/10.1016/j.jcp.2020.109491, arXiv:1905.12530.

[37] K. Xu, D.Z. Huang, E. Darve, Learning constitutive relations using symmetric positive definite neural networks, J. Comput. Phys. 428 (2021) 1–31, https://doi.org/10.1016/j.jcp.2020.110072, arXiv:2004.00265.

[38] K. Linka, M. Hillgärtner, K.P. Abdolazizi, R.C. Aydin, M. Itskov, C.J. Cyron, Constitutive artificial neural networks: a fast and general approach to predictive data-driven constitutive modeling by deep learning, J. Comput. Phys. 429 (2021) 110010, https://doi.org/10.1016/j.jcp.2020.110010, https://linkinghub.elsevier.com/retrieve/pii/S0021999120307841.

[39] S.H. Rudy, S.L. Brunton, J.L. Proctor, J.N. Kutz, Data-driven discovery of partial differential equations, Sci. Adv. 3 (4) (2017) e1602614, https://doi.org/10.1126/sciadv.1602614, arXiv:1609.06401.

[40] Z. Long, Y. Lu, B. Dong, PDE-Net 2.0: learning PDEs from data with a numeric-symbolic hybrid deep network, J. Comput. Phys. 399 (2019) 108925, https://doi.org/10.1016/j.jcp.2019.108925, arXiv:1812.04426.

[41] S.L. Brunton, J.L. Proctor, J.N. Kutz, W. Bialek, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, Proc. Natl. Acad. Sci. USA 113 (15) (2016) 3932–3937, https://doi.org/10.1073/pnas.1517384113.

[42] S. Maddu, B.L. Cheeseman, I.F. Sbalzarini, C.L. Müller, Stability selection enables robust learning of partial differential equations from limited noisy data, 2019, arXiv:1907.07810, 1–19.

[43] S.M. Udrescu, M. Tegmark, AI Feynman: a physics-inspired method for symbolic regression, Sci. Adv. 6 (16) (2020), https://doi.org/10.1126/sciadv.aay2631, arXiv:1905.11481.

[44] H.U. Voss, J. Timmer, J. Kurths, Nonlinear dynamical system identification from uncertain and indirect measurements, Int. J. Bifurc. Chaos 14 (06) (2004) 1905–1933, https://doi.org/10.1142/S0218127404010345.

[45] T. Kirchdoerfer, M. Ortiz, Data-driven computational mechanics, Comput. Methods Appl. Mech. Eng. 304 (2016) 81–101, https://doi.org/10.1016/j.cma.2016.02.001, arXiv:1510.04232.

[46] R. Ibañez, D. Borzacchiello, J.V. Aguado, E. Abisset-Chavanne, E. Cueto, P. Ladeveze, F. Chinesta, Data-driven non-linear elasticity: constitutive manifold construction and problem discretization, Comput. Mech. 60 (5) (2017) 813–826, https://doi.org/10.1007/s00466-017-1440-1.

[47] X. Wu, T. Kozlowski, H. Meidani, K. Shirvan, Inverse uncertainty quantification using the modular Bayesian approach based on Gaussian process, part 1: theory, Nucl. Eng. Des. 335 (May 2018) 339–355, https://doi.org/10.1016/j.nucengdes.2018.06.004.

[48] N.E. Owen, P. Challenor, P.P. Menon, S. Bennani, Comparison of surrogate-based uncertainty quantification methods for computationally expensive simulators, SIAM/ASA J. Uncertain. Quantificat. 5 (1) (2017) 403–435, https://doi.org/10.1137/15m1046812, arXiv:1511.00926.

[49] L. Yan, T. Zhou, Adaptive multi-fidelity polynomial chaos approach to Bayesian inference in inverse problems, J. Comput. Phys. 381 (2019) 110–128, https://doi.org/10.1016/j.jcp.2018.12.025, arXiv:1807.00618.

[50] R.K. Tripathy, I. Bilionis, Deep UQ: learning deep neural network surrogate models for high dimensional uncertainty quantification, J. Comput. Phys. 375 (2018) 565–588, https://doi.org/10.1016/j.jcp.2018.08.036, arXiv:1802.00850.

[51] L. Yan, T. Zhou, An adaptive surrogate modeling based on deep neural networks for large-scale Bayesian inverse problems, Commun. Comput. Phys. 28 (5) (2020) 2180–2205, https://doi.org/10.4208/CICP.OA-2020-0186, arXiv:1911.08926.

[52] Y. Yang, P. Perdikaris, Adversarial uncertainty quantification in physics-informed neural networks, J. Comput. Phys. 394 (2019) 136–152, https://doi.org/10.1016/j.jcp.2019.05.027, arXiv:1811.04026.

[53] D. Zhang, L. Lu, L. Guo, G.E. Karniadakis, Quantifying total uncertainty in physics-informed neural networks for solving forward and inverse stochastic problems, J. Comput. Phys. 397 (2019) 1–19, https://doi.org/10.1016/j.jcp.2019.07.048, arXiv:1809.08327.

[54] A. Tarantola, Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation, SIAM, 2005.

[55] A.M. Stuart, Inverse problems: a Bayesian perspective, Acta Numer. 19 (May 2010) 451–559, https://doi.org/10.1017/S0962492910000061, https://www.cambridge.org/core/product/identifier/S0962492910000061/type/journal_article.

[56] T. Cui, K.J. Law, Y.M. Marzouk, Dimension-independent likelihood-informed MCMC, J. Comput. Phys. 304 (2016) 109–137, https://doi.org/10.1016/j.jcp.2015.10.008, arXiv:1411.3688.

[57] I.G. Kevrekidis, C.W. Gear, J.M. Hyman, P.G. Kevrekidid, O. Runborg, C. Theodoropoulos, Equation-free, coarse-grained multiscale computation: enabling mocroscopic simulators to perform system-level analysis, Commun. Math. Sci. 1 (4) (2003) 715–762, https://projecteuclid.org/euclid.cms/1119655353.

[58] I.G. Kevrekidis, C.W. Gear, G. Hummer, Equation-free: the computer-aided analysis of complex multiscale systems, AIChE J. 50 (7) (2004) 1346–1355, https://doi.org/10.1002/aic.10106.

[59] W. E, B. Engquist, The heterognous multiscale methods, Commun. Math. Sci. 1 (1) (2003) 87–132, https://doi.org/10.4310/cms.2003.v1.n1.a8.

[60] J. Han, C. Ma, Z. Ma, E. Weinan, Uniformly accurate machine learning-based hydrodynamic models for kinetic equations, Proc. Natl. Acad. Sci. USA 116 (44) (2019) 21983–21991, https://doi.org/10.1073/pnas.1909854116.

[61] S.V. Kalinin, B.G. Sumpter, R.K. Archibald, Big-deep-smart data in imaging for guiding materials design, Nat. Mater. 14 (10) (2015) 973–980, https://doi.org/10.1038/nmat4395.

[62] S.V. Kalinin, E. Strelcov, A. Belianinov, S. Somnath, R.K. Vasudevan, E.J. Lingerfelt, R.K. Archibald, C. Chen, R. Proksch, N. Laanait, S. Jesse, Big, deep, and smart data in scanning probe microscopy, ACS Nano 10 (10) (2016) 9068–9086, https://doi.org/10.1021/acsnano.6b04212.

[63] M.Z. Bazant, Theory of chemical kinetics and charge transfer based on nonequilibrium thermodynamics, Acc. Chem. Res. 46 (5) (2013) 1144–1160, https://doi.org/10.1021/ar300145c, arXiv:1208.1587v5.

[64] D. Kondepudi, I. Prigogine, Modern Thermodynamics, John Wiley & Sons, Ltd., Chichester, UK, 2014.

[65] L.-Q. Chen, Phase-field models for microstructure evolution, Annu. Rev. Mater. Res. 32 (1) (2002) 113–140, https://doi.org/10.1146/annurev.matsci.32.112001.132041.

[66] H. Emmerich, H. Löwen, R. Wittkowski, T. Gruhn, G.I. Tóth, G. Tegze, L. Gránásy, Phase-field-crystal models for condensed matter dynamics on atomic length and diffusive time scales: an overview, Adv. Phys. 61 (6) (2012) 665–743, https://doi.org/10.1080/00018732.2012.737555, arXiv:1304.1195.

[67] J. Swift, P.C. Hohenberg, Hydrodynamic fluctuations at the convective instability, Phys. Rev. A 15 (1) (1977) 319–328, https://doi.org/10.1103/PhysRevA.15.319.

[68] R. Fletcher, Practical Methods of Optimization, John Wiley & Sons, Ltd., Chichester, West Sussex England, 2000.

[69] T. Maly, L.R. Petzold, Numerical methods and software for sensitivity analysis of differential-algebraic systems, Appl. Numer. Math. 20 (1–2) (1996) 57–79, https://doi.org/10.1016/0168-9274(95)00117-4, https://linkinghub.elsevier.com/retrieve/pii/0168927495001174.

[70] Y. Cao, S. Li, L. Petzold, Adjoint sensitivity analysis for differential-algebraic equations: algorithms and software, J. Comput. Appl. Math. 149 (1) (2002) 171–191, https://doi.org/10.1016/S0377-0427(02)00528-9, https://linkinghub.elsevier.com/retrieve/pii/S0377042702005289.

[71] Y. Cao, S. Li, L. Petzold, R. Serban, Adjoint sensitivity analysis for differential-algebraic equations: the adjoint DAE system and its numerical solution, SIAM J. Sci. Comput. 24 (3) (2003) 1076–1089, https://doi.org/10.1137/s1064827501380630.

[72] H. Haario, E. Saksman, J. Tamminen, An adaptive Metropolis algorithm, Bernoulli 7 (2) (2001) 223–242, https://doi.org/10.2307/3318737.

[73] J.W. Cahn, On spinodal decomposition, Acta Metall. 9 (9) (1961) 795–801, https://doi.org/10.1016/0001-6160(61)90182-1, https://linkinghub.elsevier.com/retrieve/pii/0001616061901821.

[74] M. Greenwood, N. Provatas, J. Rottler, Free energy functionals for efficient phase field crystal modeling of structural phase transformations, Phys. Rev. Lett. 105 (4) (2010) 045702, https://doi.org/10.1103/PhysRevLett.105.045702.

[75] A. Jaatinen, C.V. Achim, K.R. Elder, T. Ala-Nissila, Thermodynamics of bcc metals in phase-field-crystal models, Phys. Rev. E 80 (3) (2009) 031602, https://doi.org/10.1103/PhysRevE.80.031602.

[76] J.D. Murray, Mathematical Biology: I. An Introduction, Springer Science & Business Media, 2007, https://www.springer.com/gp/book/9780387952239.

[77] Z. Wang, X. Huan, K. Garikipati, Variational system identification of the partial differential equations governing the physics of pattern-formation: inference under varying fidelity and noise, Comput. Methods Appl. Mech. Eng. 356 (2019) 44–74, https://doi.org/10.1016/j.cma.2019.07.007, arXiv:1812.11285.

[78] Y. Li, F. El Gabaly, T.R. Ferguson, R.B. Smith, N.C. Bartelt, J.D. Sugar, K.R. Fenton, D.A. Cogswell, A.L.D. Kilcoyne, T. Tyliszczak, M.Z. Bazant, W.C. Chueh, Current-induced transition from particle-by-particle to concurrent intercalation in phase-separating battery electrodes, Nat. Mater. 13 (12) (2014) 1149–1156, https://doi.org/10.1038/nmat4084.

[79] P. Bai, D.A. Cogswell, M.Z. Bazant, Suppression of phase separation in LiFePO 4 nanoparticles during battery discharge, Nano Lett. 11 (11) (2011) 4890–4896, https://doi.org/10.1021/nl202764f, arXiv:1108.2326.

[80] M.Z. Bazant, Thermodynamic stability of driven open systems and control of phase separation by electro-autocatalysis, Faraday Discuss. 199 (2017) 423–463, https://doi.org/10.1039/C7FD00037E.

[81] J. Newman, K.E. Thomas-Alyea, Electrochemical Systems, John Wiley & Sons, 2012.

[82] D.A. Cogswell, M.Z. Bazant, Coherency strain and the kinetics of phase separation in LiFePO 4 nanoparticles, ACS Nano 6 (3) (2012) 2215–2225, https://doi.org/10.1021/nn204177u, arXiv:1110.3877v2.

[83] H. Zhao, M.Z. Bazant, Population dynamics of driven autocatalytic reactive mixtures, Phys. Rev. E 100 (1) (2019) 012144, https://doi.org/10.1103/PhysRevE.100.012144, arXiv:1901.05575.

[84] D. Fraggedakis, M.Z. Bazant, Tuning the stability of electrochemical interfaces by electron transfer reactions, J. Chem. Phys. 152 (18) (2020) 184703, https://doi.org/10.1063/5.0006833, arXiv:2003.03057.

[85] J. Pathak, B. Hunt, M. Girvan, Z. Lu, E. Ott, Model-free prediction of large spatiotemporally chaotic systems from data: a reservoir computing approach, Phys. Rev. Lett. 120 (2) (2018) 24102, https://doi.org/10.1103/PhysRevLett.120.024102.

[86] J. Berry, S.C. Weber, N. Vaidya, M. Haataja, C.P. Brangwynne, D.A. Weitz, RNA transcription modulates phase transition-driven nuclear body assembly, Proc. Natl. Acad. Sci. USA 112 (38) (2015) E5237–E5245, https://doi.org/10.1073/pnas.1509317112.

[87] H.H. Wensink, J. Dunkel, S. Heidenreich, K. Drescher, R.E. Goldstein, H. Löwen, J.M. Yeomans, Meso-scale turbulence in living fluids, Proc. Natl. Acad. Sci. USA 109 (36) (2012) 14308–14313, https://doi.org/10.1073/pnas.1202032109.

[88] J. Dunkel, S. Heidenreich, K. Drescher, H.H. Wensink, M. Bär, R.E. Goldstein, Fluid dynamics of bacterial turbulence, Phys. Rev. Lett. 110 (22) (2013) 1–5, https://doi.org/10.1103/PhysRevLett.110.228102, arXiv:1302.5277.

[89] G. Liu, A. Patch, F. Bahar, D. Yllanes, R.D. Welch, M.C. Marchetti, S. Thutupalli, J.W. Shaevitz, Self-driven phase transitions drive myxococcus xanthus fruiting body formation, Phys. Rev. Lett. 122 (24) (2019) 248102, https://doi.org/10.1103/PhysRevLett.122.248102.

[90] M. Kryuchkov, O. Bilousov, J. Lehmann, M. Fiebig, V.L. Katanaev, Reverse and forward engineering of Drosophila corneal nanocoatings, Nature 585 (7825) (2020) 383–389, https://doi.org/10.1038/s41586-020-2707-9.

[91] N. Yoshinaga, S. Tokuda, Bayesian modelling of pattern formation from one snapshot of pattern, 2020, arXiv:2006.06125.

[92] S. Chandrasekhar, Hydrodynamic and Hydromagnetic Stability, Courier Corporation, 1981.