ROYAL SOCIETY OPEN SCIENCE

royalsocietypublishing.org/journal/rsos

Research

Cite this article: Hirsh SM, Barajas-Solano DA, Kutz JN . 2022 Sparsifying priors for Bayesian uncertainty quantification in model discovery. *R. Soc. Open Sci.* **9**: 211823.
https://doi.org/10.1098/rsos.211823

Received: 27 November 2021
Accepted: 18 January 2022

Subject Category:
Mathematics

Subject Areas:
applied mathematics/mathematical modelling/artificial intelligence

Keywords:
Bayesian inference, uncertainty quantification, model discovery

Author for correspondence:
J. Nathan Kutz
e-mail: kutz@uw.edu

Sparsifying priors for Bayesian uncertainty quantification in model discovery

Seth M. Hirsh[1], David A. Barajas-Solano[3] and J. Nathan Kutz[2]

[1]Department of Physics, and [2]Department of Applied Mathematics, University of Washington, Seattle, WA, USA
[3]Pacific Northwest National Laboratory, Richland, WA, USA

JNK, 0000-0002-6004-2275

We propose a probabilistic model discovery method for identifying ordinary differential equations governing the dynamics of observed multivariate data. Our method is based on the *sparse identification of nonlinear dynamics* (SINDy) framework, where models are expressed as sparse linear combinations of pre-specified candidate functions. Promoting parsimony through sparsity leads to interpretable models that generalize to unknown data. Instead of targeting point estimates of the SINDy coefficients, we estimate these coefficients via sparse Bayesian inference. The resulting method, *uncertainty quantification SINDy* (UQ-SINDy), quantifies not only the uncertainty in the values of the SINDy coefficients due to observation errors and limited data, but also the probability of inclusion of each candidate function in the linear combination. UQ-SINDy promotes robustness against observation noise and limited data, interpretability (in terms of model selection and inclusion probabilities) and generalization capacity for out-of-sample forecast. Sparse inference for UQ-SINDy employs Markov chain Monte Carlo, and we explore two sparsifying priors: the *spike and slab prior*, and the *regularized horseshoe prior*. UQ-SINDy is shown to discover accurate models in the presence of noise and with orders-of-magnitude less data than current model discovery methods, thus providing a transformative method for real-world applications which have limited data.

1. Introduction

In recent years, there has been a rapid increase in measurements gathered from complex nonlinear dynamics for which their

THE ROYAL SOCIETY PUBLISHING

governing equations are unknown. A key challenge is to discover explicit representations of these equations, which can then be used for system identification, forecasting and control. Measurements are often compromised by noise or may exhibit chaotic behaviour, in which case it is critical to quantify how uncertainty affects the model discovery process. To address this challenge, we introduce the *uncertainty quantification sparse identification of nonlinear dynamics* (UQ-SINDy) framework, which leverages sparsity promotion in a Bayesian probabilistic setting to extract a parsimonious set of governing equations. Our method provides uncertainty estimates of both the parameter values and the inclusion probabilities for different terms in the models.

Discovery of governing equations plays a fundamental role in the development of physical theories. With increasing computing power and data availability in recent years, there have been substantial efforts to identify the governing equations directly from data [1–3]. There has been particular emphasis on parsimonious representations because they have the benefits of promoting interpretability and generalizing well to unknown data [4–11]. The SINDy method was proposed in [5], which leverages dictionary learning and sparse regression to model dynamical systems. This approach has been successful in modelling a diversity of applications, including in chemistry [12], optics [13], engineered systems [14], epidemiology [15] and plasma physics [16]. Furthermore, there has been a variety of modifications, including improved robustness to noise [17–19], generalizations to partial differential equations [20–22], multi-scale physics [23] and libraries of rational functions [24,25].

Although these methods identify the equations, measurements often contain observation errors, which may imperil the predictive capacity of learned models. A common approach to remedy this is to use the Bayesian probability framework where uncertainty is quantified in terms of probability and where priors are employed to encode assumptions and knowledge about model parameters [26,27]. Bayesian methods have been widely used for uncertainty quantification in time series models, with applications to weather forecasting [28–30], disease modelling [31–33], traffic flow [34–36] and finance [37–39], among many others. By leveraging informative priors Bayesian inference can achieve better results for limited data compared to frequentist approaches [40–42]. More recently, these methods have been incorporated into model discovery frameworks, exhibiting state-of-the-art performance for system identification in the presence of noise [3,43,44]. Although these methods provide a range of possible values, realizations of these models are in general not sparse and consequently lack the capability to identify relevant terms in the model.

Sparse regression is a popular tool to identify a small subset of variables that explain the data. However, finding the true minimum is computationally intractable in practice. In the frequentist setting, a popular solution is to use the Lasso, which corresponds to an l_1 penalty term [45]. In the Bayesian setting, sparsity is generated by fundamentally different mechanisms. Most notably, although the corresponding prior (the Laplace prior) shares the same maximum likelihood estimator as the Lasso [46], the distribution has fat tails and thus does not produce sparse realizations [47]. The spike and slab model remedies this by explicitly using Bernoulli variables to determine whether a term is present in the model, and has become the leading method for incorporating sparsity in the Bayesian framework [48–50]. One disadvantage to this prior, however, is its dependence on discrete variables, which makes inference prohibitively expensive for high-dimensional systems. Smooth approximations, such as the horseshoe [51,52], horseshoe+ [53], regularized horseshoe [54], Dirichlet–Laplace [55] and R2-D2 priors [56], have been shown to yield performance comparable to the spike and slab model. For this work, we will primarily focus on the regularized horseshoe prior, also known as the Finnish horseshoe.

In this work, we propose the UQ-SINDy framework, which provides uncertainty estimates of both the parameter value and inclusion probabilities and promotes sparsity in realizations of the model. This model leverages advances sparsity and Bayesian approaches for solving ordinary differential equations (ODEs) to achieve this goal. Importantly, the UQ-SINDy framework is capable of model discovery with limited and noisy data, improving significantly on existing methods by requiring orders-of-magnitude less data to identify a stable, robust and sparse model. Even with as little as 21 time points, the UQ-SINDy framework can identify such a model, making it an ideal tool for application areas like biology where often limited data is available. Indeed, it is the only model discovery method capable of working in this limited and noisy data regime. In §§2.1 and 2.2, we review the SINDy method and Bayesian inference for ordinary differential equations (ODEs), respectively. In §2.3, we review sparsity promoting priors, namely the spike and slab and regularized horseshoe priors, and compare their performance to the Laplace prior. In §3.1, we introduce two sparsity promoting Bayesian methods, spike and slab SINDy and regularized horseshoe SINDy. In §3.2, we illustrate these methods on two synthetic nonlinear datasets, a Lotka–Volterra model and nonlinear oscillator, and one real-world example of lynx and hare

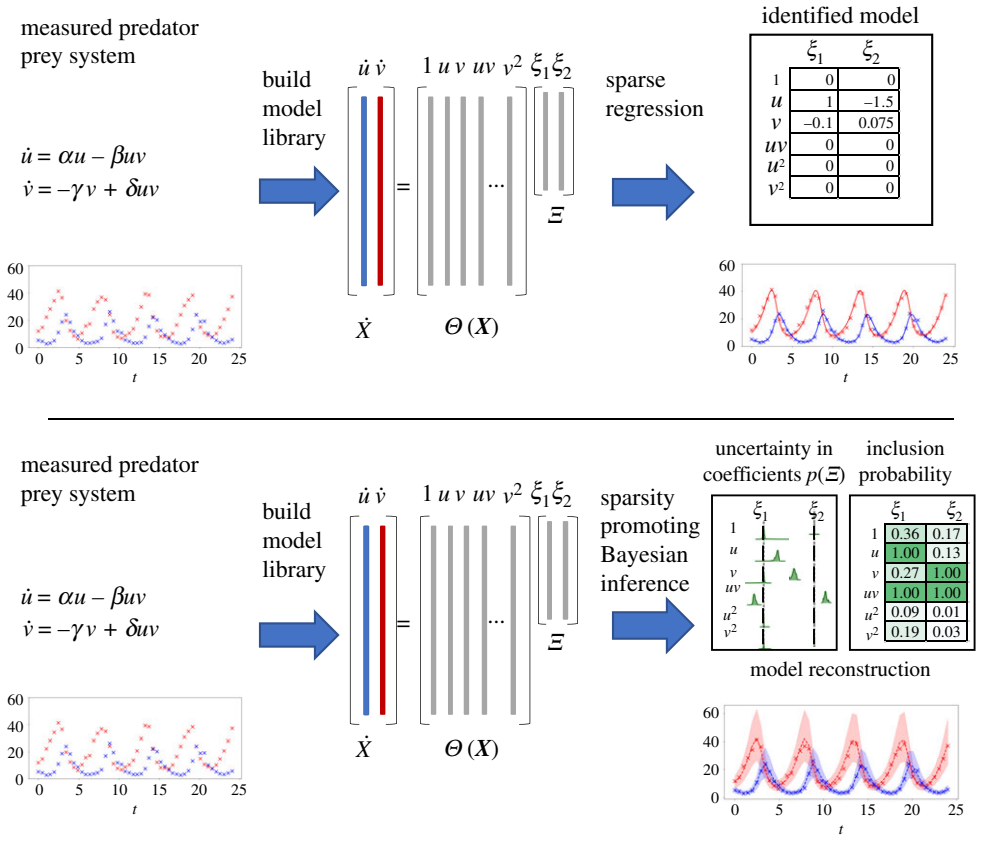

Figure 1. Comparison of SINDy algorithm and UQ-SINDy. (Top) Schematic of SINDy algorithm. A dynamical system governed by unknown governing equations is measured. Next, we compute the derivative of the time series $\dot{\boldsymbol{X}}$ and construct a library $\Theta(\boldsymbol{X})$ of candidate terms. Last, we perform sparse regression to identify the terms in the library that best explain the time series. (Bottom) Schematic of UQ-SINDy algorithm. A dynamical system governed by unknown governing equations is measured. Next, we posit a SINDy library $\Theta(\boldsymbol{X})$ of candidate terms. Last, we perform sparsity promoting Bayesian inference to compute the inclusion probability and the posterior distribution of each term in the SINDy library. An ensemble of reconstructions can then be computed, which quantify the credibility of predictions.

population data. We find that these methods are able to extract accurate and meaningful Bayesian models even in the presence of significant noise and sparse samples. These results are summarized and future improvements are discussed in §4.

2. Background

The UQ-SINDy framework is based on several recent developments in the fields of sparse regression, ODEs and Bayesian inference, and we review these contributions here. In §2.1, we introduce the SINDy algorithm, which employs sparse regression to identify governing equations in the frequentist setting. In §2.2, we review Bayesian inference for ODEs. In §2.3, we review three different priors for sparse inference—the Laplace, spike and slab, and regularized horseshoe priors—and compare their benefits and drawbacks.

2.1. Sparse identification of nonlinear dynamics

The SINDy method is a recently developed technique that leverages sparse regression to identify the governing equations from a given time series (figure 1). We consider a system with state $\boldsymbol{x}(t) = [x_1(t), x_2(t), \ldots, x_d(t)]^\top \in \mathbb{R}^d$ governed by the differential equation

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}),$$

for some unknown function $\boldsymbol{f}: \mathbb{R}^d \to \mathbb{R}^d$. The system's state is observed at the discrete times $t = t_1, \ldots, t_n$. The goal of SINDy is to discover $\boldsymbol{f}$ from these observations.

To do so, we postulate that f can be written as a linear combination of a library of l candidate functions $\theta_i : \mathbb{R}^d \to \mathbb{R}$, $i \in [1, l]$. For example, a commonly used library is the polynomial library

$$\Theta(\boldsymbol{x}) = \begin{bmatrix} x_1(t) & x_2(t) & x_1^2(t) & x_1(t)x_2(t) & \cdots \end{bmatrix} \in \mathbb{R}^l.$$

Next, we define $\boldsymbol{X} = [\boldsymbol{x}(t_1), \boldsymbol{x}(t_2), \ldots, \boldsymbol{x}(t_n)]^\top \in \mathbb{R}^{n \times d}$ as the collection of observed state snapshots, and also define the matrix of library terms evaluated at the observation times,

$$\Theta(\boldsymbol{X}) = \begin{bmatrix} \Theta(\boldsymbol{x}(t_1))^\top & \Theta(\boldsymbol{x}(t_2))^\top & \cdots & \Theta(\boldsymbol{x}(t_n))^\top \end{bmatrix}^\top \in \mathbb{R}^{n \times l}.$$

We then measure or compute the time derivative of the data $\boldsymbol{X}$ and solve the following equation for $\Xi \in \mathbb{R}^{l \times d}$:

$$\dot{\boldsymbol{X}} = \Theta(\boldsymbol{X})\Xi, \tag{2.1}$$

where Ξ denotes the matrix of linear combination coefficients, or *SINDy coefficients*. A key assumption of SINDy is that f may be represented by a small number of library terms, so that the matrix Ξ is sparse. Thus, (2.1) is typically solved through sparse regression, using minimization techniques such as sequential least squares thresholding (STLSQ) [5], Lasso [45], or a relaxed formulation [18]. The SINDy procedure yields the set of identified nonlinear differential equations

$$\dot{\boldsymbol{x}}^\top = \Theta(\boldsymbol{x})\Xi. \tag{2.2}$$

Once identified, this system of differential equations may be used for system identification, prediction and control.

2.2. Bayesian inference for data-driven discovery

Suppose we have the dataset $(\boldsymbol{X}, \boldsymbol{y})$ for which we would like to fit the linear regression model

$$\boldsymbol{y} = \boldsymbol{\beta}^\top \boldsymbol{X} + \boldsymbol{\epsilon}, \tag{2.3}$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I})$ is a vector of independent, identically distributed Gaussian measurement noise with unknown standard deviation σ. In the Bayesian setting, our goal is to determine the posterior distribution of $\boldsymbol{\beta}$ and σ conditioned on the data, i.e. $p(\boldsymbol{\beta}, \sigma | \boldsymbol{X}, \boldsymbol{y})$. To compute this distribution, we leverage Bayes' rule,

$$p(\boldsymbol{\beta}, \sigma | \boldsymbol{X}, \boldsymbol{y}) \propto p(\boldsymbol{y} | \boldsymbol{\beta}, \boldsymbol{X}) p(\sigma) p(\boldsymbol{\beta}),$$

where $p(\boldsymbol{y} | \boldsymbol{\beta}, \boldsymbol{X})$ denotes the data likelihood, and $p(\sigma)$ and $p(\boldsymbol{\beta})$ denote the prior distribution of the noise standard deviation and the regression coefficients. These prior distributions incorporate any available domain knowledge about the distribution of the noise standard deviation and the β_js.

In this work, we are interested in identifying ODE models from noisy data. In particular, given noisy time series and a SINDy model of the form $\dot{\boldsymbol{x}}^\top = \Theta(\boldsymbol{x})\Xi$, our goal is to compute the posterior distribution of the initial conditions $\boldsymbol{x}_0$ and SINDy coefficients Ξ. We assume that the dataset $\boldsymbol{X}$ consists of n noisy snapshots of the observed dynamics, that is $\boldsymbol{X} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n]^\top \in \mathbb{R}^{n \times d}$, where $\boldsymbol{y}_i \in \mathbb{R}^d$ is the noisy snapshot of the system state at time $t = t_i$. For a given probabilistic model of the observation noise, the data is modelled as deviations from the SINDy predictions; for example, for additive noise models,

$$\boldsymbol{y}_i^\top = \boldsymbol{x}_0^\top + \int_0^{t_i} \Theta(\boldsymbol{x}(t'))\Xi \, \mathrm{d}t' + \boldsymbol{\epsilon}_i, \tag{2.4}$$

where $\boldsymbol{\epsilon}_i$ denotes the additive noise for the ith snapshot. Bayes' rule then takes the form

$$p(\Xi, \boldsymbol{x}_0, \boldsymbol{\phi} | \boldsymbol{X}) \propto p(\boldsymbol{X} | \Xi, \boldsymbol{x}_0, \boldsymbol{\phi}) p(\boldsymbol{\phi}) p(\Xi) p(\boldsymbol{x}_0), \tag{2.5}$$

where $\boldsymbol{\phi}$ denotes auxiliary variables of the probabilistic model such as the noise standard deviation. The data likelihood $p(\boldsymbol{X} | \Xi, \boldsymbol{x}_0, \boldsymbol{\phi})$ is given by the chosen observation model (e.g. by (2.4) and the distribution of the noise for additive observation noise).

Computing the posterior distribution (2.5) is in general not analytically tractable, in which case sampling-based methods such as Markov chain Monte Carlo (MCMC) may be used. Once the posterior distribution has been approximated, we may then compute state reconstructions and forecasts conditioned on the observed data [57,58]. Specifically, to estimate the distribution of predicted values of $\boldsymbol{x}$ at an arbitrary time t, we marginalize the data likelihood times the posterior

distribution over Ξ, $\boldsymbol{x}_0$ and $\boldsymbol{\phi}$, that is,

$$p(\boldsymbol{x}(t)|\boldsymbol{X}) = \int p(\boldsymbol{x}(t)|\Xi, \boldsymbol{x}_0, \boldsymbol{\phi}) p(\Xi, \boldsymbol{x}_0, \boldsymbol{\phi}|\boldsymbol{X})\, \mathrm{d}\Xi\, \mathrm{d}\boldsymbol{x}_0\, \mathrm{d}\boldsymbol{\phi}. \tag{2.6}$$

The distribution $p(\boldsymbol{x}(t)|\boldsymbol{X})$ is referred to as the *posterior predictive distribution (PPD)*. The integral in (2.6) can be approximated via sampling by taking the expectation of the data likelihood over posterior samples drawn via MCMC.

2.3. Sparsity promoting priors

Consider the regression problem in (2.3). In many cases, we assume only a few components of $\boldsymbol{x}_i$ are relevant for predicting y_i, in which case we expect $\boldsymbol{\beta}$ to be sparse. In the Bayesian setting, multiple sparsity-inducing priors have been proposed. We describe a few of these approaches below, namely the Laplace, spike and slab, and regularized horseshoe priors.

2.3.1. Laplace prior

Originally proposed by Laplace [59], the Laplace distribution, also known as the double exponential distribution [26], corresponds to the probability distribution function (PDF) $f(x\,|\,\mu, b)$ given by

$$f(x|\mu, b) = \frac{1}{2b} \exp\left\{-\left|\frac{x-\mu}{b}\right|\right\}.$$

Most notably, maximum *a posteriori* (MAP) estimation for this prior corresponds to regression with ℓ_1 regularization, that is [46],

$$\hat{\boldsymbol{\beta}}^{\text{Laplace}} = \arg\max_{\boldsymbol{\beta}} p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{X}) p(\boldsymbol{\beta}) = \arg\min_{\boldsymbol{\beta}} ||\boldsymbol{y} - \boldsymbol{\beta}^\top \boldsymbol{X}|_2^2 + \lambda \|\boldsymbol{\beta}\|_1.$$

In the frequentist setting, solving this regression problem, known as the Lasso problem, has been shown to yield sparse solutions for $\boldsymbol{\beta}$ [45]. This sparsifying behaviour of the Laplace distribution is attributed to the fact that for values of x smaller than b, the distribution is sharply peaked, thus pushing many terms toward 0. Additionally, for values of x greater than b, the distribution has longer tails than the Gaussian distribution, allowing elements to escape significant shrinkage.

Although l_1 regularization induces sparsity in the frequentist case, in the Bayesian setting realizations of the corresponding posterior distributions are not sparse [47]. In particular, in the Bayesian setting we must consider the whole distribution simultaneously. With the Laplace prior, every β_j has probability mass simultaneously pushed toward and away from the origin, forcing relevant β_js to shrink toward the origin and irrelevant terms to have significant probability mass far away from the origin.

To illustrate this, we generate 400 data samples $(\boldsymbol{x}_i, y_i)$ that satisfy (2.3), where $\boldsymbol{x}_i \sim \mathcal{N}(0, 1) \in \mathbb{R}^{10}$, $\boldsymbol{\epsilon}_i \sim \mathcal{N}(0, 0.5^2)$, and $\boldsymbol{\beta} \in \mathbb{R}^{10}$ is chosen to be the sparse vector

$$\boldsymbol{\beta} = [0.3,\ \ 0.2,\ \ -0.3,\ \ 0,\ \ 0,\ \ 0,\ \ 0,\ \ 0,\ \ 0,\ \ 0]^\top.$$

We perform Bayesian inference to estimate $\boldsymbol{\beta}$ using a Laplace prior, and we plot the resulting posterior distribution in figure 2. We note that when using the Laplace prior, the posterior distributions are centred about the true value $\boldsymbol{\beta}$. However, many distributions are peaked at non-zero values, making it difficult to differentiate between relevant and irrelevant variables. Further, due to the wide widths of all the distributions, samples from this posterior distribution will not be sparse. To better enforce sparsity in a Bayesian setting and induce sparse realizations, the distribution of each β_j must either be fully shrunk towards the origin or pushed away from the origin. In §§2.3.2 and 2.3.3 we discuss two priors that satisfy these properties.

2.3.2. Spike and slab prior

The spike and slab prior is one of the most popular sparsifying priors and is typically referred to as the 'gold standard' [48–50] sparsity-inducing prior in the Bayesian setting. For this prior, each β_j is generated using the hierarchical model

$$\beta_j|\lambda_j \sim \mathcal{N}(0, c^2)\lambda_j$$

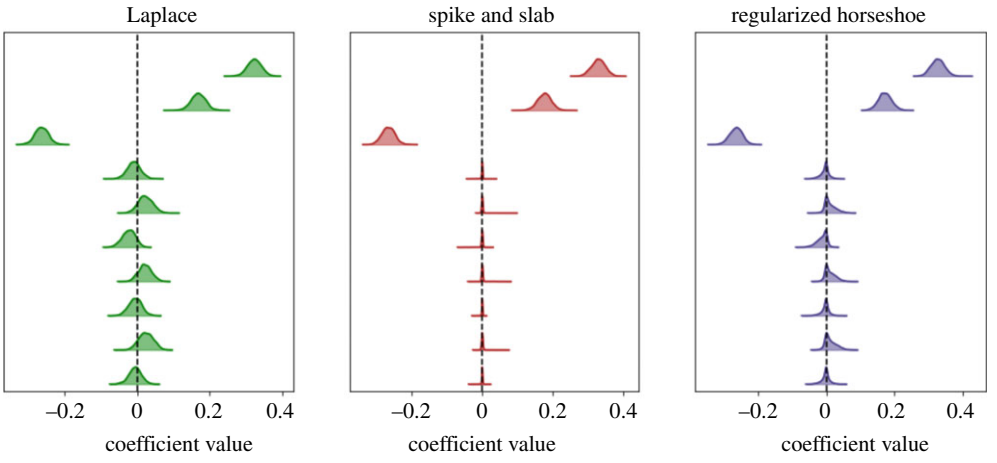

Figure 2. Comparison of posterior distributions for Laplace, spike and slab, and regularized horseshoe priors for a linear regression problem. Both the spike and slab and regularized horseshoe priors promote sparsity in the posterior distributions, while the Laplace prior does not.

and

$$\lambda_j \sim \mathrm{Ber}(\pi),$$

where $\mathrm{Ber}(\pi)$ denotes the Bernoulli distribution with probability of success π. Here, π is the prior probability that λ_j is 1. Otherwise λ_j is 0. From this it can be seen that if λ_j is 1, then the jth term belongs to the model and β_j follows the 'slab' distribution, a normal distribution with variance c^2. If λ_j is 0, then the jth term is not in the model and β_j follows the 'spike' distribution, a Dirac delta distribution centred at zero.

The distribution may be relaxed to

$$\beta_j|\lambda_j \sim \lambda_j\mathcal{N}(0, c^2) + (1 - \lambda_j)\mathcal{N}(0, \epsilon^2)$$

and

$$\lambda_j \sim \mathrm{Ber}(\pi),$$

where $\varepsilon \ll c$. This is similar to before, except when $\lambda_j = 0$, β_j follows a narrow normal distribution with variance ε^2.

The spike and slab prior for $\boldsymbol{\beta}$ is very intuitive and has shown robust performance in practical applications. In figure 2, we plot the resulting posterior distribution for the example in §2.3.2. Most notably, we see that similar to the Laplace prior, the spike and slab prior extracts out wide distributions for the three non-zero coefficients. For the seven zero coefficients, on the other hand, the distribution is sharply spiked at the origin. Consequently, any samples drawn from this posterior distribution will be truly sparse. Compared to the Laplace distribution, the non-zero terms are much more easily identifiable. Furthermore, the mean of λ_j corresponds to the estimate of the 'inclusion probability', that is the likelihood that a particular β_j is relevant to the model.

Although the spike and slab prior has many beneficial properties, one downside is that because of its discrete nature, inference with this prior requires exploring the combinatorial space of possible models. To address this challenge, many smooth approximations to the spike and slab prior distribution have been proposed. We discuss one recent approach in §2.3.3.

2.3.3. Regularized horseshoe prior

The horseshoe prior and the recently developed regularized horseshoe prior are smooth priors that have shown comparable performance to the spike and slab model. The horseshoe is defined as the hierarchical prior

$$\beta_i|\lambda_i, \tau \sim \mathcal{N}(0, \lambda_i^2\tau^2),$$
$$\lambda_i \sim \mathrm{C}^+(0, 1)$$

and

$$\tau \sim \mathrm{C}^+(0, \tau_0),$$

where $C^+(\cdot, \cdot)$ denotes the half-Cauchy distribution [51,52,60]. The key intuition behind this prior is that τ promotes global sparsity, shrinking the posterior distributions of all β_is. The λ_is, known as the local shrinkage parameters, also have half-Cauchy priors, allowing some of the β_is to escape significant shrinkage. Many analyses have focused on choosing an optimal value for τ_0, and in Piironen *et al.* values are recommended for sparse linear regression [54]. In this work, we employ $\tau_0 = 0.1$ unless specified otherwise. We note that decreasing the value of τ_0 increases the sparsity of $\boldsymbol{\beta}$ estimates.

One downside of the horseshoe is that relevant terms that 'escape' shrinkage are not regularized, and thus elements of the posterior distribution may become arbitrarily large. In [54] it was proposed to include a small amount of regularization on each λ_i, resulting in the *regularized horseshoe* prior

$$\begin{aligned}
\beta_i|\tilde{\lambda}_i, \tau, c &\sim \mathcal{N}(0, \tilde{\lambda}_i^2\tau^2),\\
\tilde{\lambda}_i &= \frac{c\lambda_i}{\sqrt{c^2 + \tau^2\lambda_i^2}},\\
\lambda_i &\sim C^+(0, 1),\\
c^2 &\sim \text{Inv-Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}s^2\right)
\end{aligned}$$

and

$$\tau \sim C^+(0, \tau_0),$$

where Inv-Gamma$(\cdot, \cdot)$ denotes the inverse Gamma distribution, and ν and s are parameters that control the shape of the slab. For small values of λ_i, $\lambda_i\tau \ll c$ and $\tilde{\lambda}_i \to \lambda_i$, thus approximating the original horseshoe prior. However, for large values of λ_i, $\lambda_i\tau \gg c$ and $\tilde{\lambda}_i \to c/\tau$, leading to β_i being normally distributed with variance c^2. This regularizes β_i, constraining it to be on the order of c. In this work, we employ the values $\nu = 4$ and $s = 2$.

We illustrate the performance of this prior in figure 2 for the example in §2.3.2. Similar to the spike and slab model, the non-zero coefficients have wide distributions. The zero coefficients, on the other hand are more spiked than those of the Laplace prior, thus resulting in sparser posterior realizations.

Unlike for the spike and slab prior, there is no explicit estimate for the inclusion probabilities. A popular alternative for identifying the relevant terms is to compute the shrinkage factor of the coefficients. Specifically, we compute the MAP estimate $\hat{\beta}_i^{\text{Flat}}$ with a flat prior (i.e. no prior) and compare it to the MAP estimate with the regularized horseshoe prior, $\hat{\beta}_i^{\text{RH}}$. The ratio of these two values is called the shrinkage factor

$$\kappa_i = \frac{\hat{\beta}_i^{\text{RH}}}{\hat{\beta}_i^{\text{Flat}}}. \tag{2.7}$$

The shrinkage factor of the coefficients has been used to define inclusion 'pseudo-probabilities' for sparsity-promoting models [52,54,60]. We employ this approach in this work. In general, these ratios may not lie between 0 and 1.

For our work, we have observed that computing $\hat{\beta}_i^{\text{Flat}}$ with flat priors is challenging. To remedy this, we use normal priors $\beta_i \sim \mathcal{N}(0, 1)$ instead of flat priors. Further, we note that (2.7) can be computed directly from MAP estimates, without having to sample the full posterior distributions. Thus, the shrinkage factors can be estimated using optimization techniques instead of full Bayesian inference. However, in practice the associated optimization problems may be non-convex and highly sensitive to the initial guess. Consequently, for this work we use full Bayesian inference to estimate shrinkage factors.

3. UQ-SINDy

In this section, we combine advances in model discovery for dynamical systems and sparsity promoting Bayesian inference to propose the UQ-SINDy framework, which aims to quantify the uncertainty of estimated SINDy coefficients due to measurement, and to estimate the inclusion probabilities for each term in the SINDy library. In particular, within this framework we introduce two methods: spike and slab SINDy (ss-SINDy) and regularized horseshoe SINDy (rh-SINDy). The ss-SINDy method provides state-of-the-art performance for estimating uncertainty of coefficients and inclusion probability, while the rh-SINDy is a smooth approximation that shows comparable performance. We outline this framework below.

3.1. Method

We start with a set of time series measurements $\boldsymbol{X} \in \mathbb{R}^{n\times d}$ contaminated by measurement noise. We assume that our data are governed by the SINDy model

$$\dot{\boldsymbol{x}}^\top = \boldsymbol{\Theta}(\boldsymbol{x})\boldsymbol{\Xi} \quad \text{and} \quad \boldsymbol{x}(0) = \boldsymbol{x}_0, \tag{3.1}$$

for some sparse matrix of SINDy coefficients $\boldsymbol{\Xi}$ and initial condition $\boldsymbol{x}_0$. Our goal is to determine the posterior distribution $p(\boldsymbol{\Xi}, \boldsymbol{x}_0, \boldsymbol{\phi}|\boldsymbol{X})$.

Step 1: **Construct library.** We posit a library $\boldsymbol{\Theta}: \mathbb{R}^d \to \mathbb{R}^l$ of candidate functions. We emphasize here that $\boldsymbol{\Theta}$ is a symbolic vector function of the system's state $\boldsymbol{x}$. This is in contrast to the original SINDy algorithm, in which $\boldsymbol{\Theta}(\boldsymbol{X})$ is a fixed matrix computed from the time-series data.

Depending on the library, solving the ODE in (3.1) for certain values of initial conditions and parameters may be unstable. Practically, this leads to exploding gradients with respect to SINDy coefficients and initial conditions, and integration steps taken by the ODE solver becoming negligibly small. To remedy this, we add a higher-order polynomial term with a small negative coefficient to the ODE model. For example, for a library of terms up to quadratic order, we add a cubic term, leading to the ODE model

$$\dot{x}_j = \sum_i \theta_i(\boldsymbol{x})\xi_{i,j} - \varepsilon x_j^3, \tag{3.2}$$

where $\xi_{i,j}$ is the i, jth element of $\boldsymbol{\Xi}$. The parameter ε is chosen to be sufficiently small so that the ODE is not affected for values of the system's state that lie within the range of the data, but sufficiently large so that $\dot{\boldsymbol{x}}$ does not grow too large. In general, if the library $\boldsymbol{\Theta}$ includes polynomial terms up to order n, we add a term $-\varepsilon x_i^{n+1}$ if n is even, or $-\varepsilon x_i^{n+2}$ if n is odd. This guarantees that the values $\dot{\boldsymbol{x}}$ remain finite for both positive and negative values of $\boldsymbol{x}$.

Step 2: **Construct model priors and model likelihood.** Let $\hat{\boldsymbol{x}}(t; \boldsymbol{\Xi}, \boldsymbol{x}_0)$ denote the SINDy prediction at time t for given values of $\boldsymbol{\Xi}$ and $\boldsymbol{x}_0$, given by

$$\hat{\boldsymbol{x}}^\top(t; \boldsymbol{\Xi}, \boldsymbol{x}_0) = \boldsymbol{x}_0^\top + \int_{t_0}^{t} \boldsymbol{\Theta}(\boldsymbol{x}(t'))\boldsymbol{\Xi}\,\mathrm{d}t'.$$

For normally distributed measurement noise, the data likelihood takes the form

$$p(\boldsymbol{X}|\boldsymbol{\Xi}, \boldsymbol{x}_0, \boldsymbol{\phi}) = \prod_{i=1}^{n}\prod_{j=1}^{d} \frac{1}{\sigma_j\sqrt{2\pi}} \exp\left[\frac{1}{2\sigma_j^2}|y_{i,j} - \hat{x}_j(t_i; \boldsymbol{\Xi}, \boldsymbol{x}_0)|^2\right]. \tag{3.3}$$

For some cases, the values of $\boldsymbol{X}$ takes non-negative values, such as for populations, in which case we may choose to use a lognormal likelihood instead,

$$p(\boldsymbol{X}|\boldsymbol{\Xi}, \boldsymbol{x}_0, \boldsymbol{\phi}) = \prod_{i=1}^{n}\prod_{j=1}^{d} \frac{1}{y_{i,j}\sigma_j\sqrt{2\pi}} \exp\left[\frac{1}{2\sigma_j^2}|\log y_{i,j} - \log \hat{x}_j(t_i; \boldsymbol{\Xi}, \boldsymbol{x}_0)|^2\right]. \tag{3.4}$$

We must choose priors for the noise level parameters σ_j and the initial conditions $\boldsymbol{x}_0$. These priors are chosen using knowledge about the type of parameter (i.e. whether the parameter is non-negative) and the scales of the data.

In this work, we assume that the noise is uncorrelated in time and among the different state variables. In the case where these correlations exist, the model likelihood and prior can be adjusted. In particular, in the case where the state variables are correlated, [61,62] recommend replacing σ_i with a matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$ and using an LKJ prior. If there are correlations in time, for example, in the case of coloured noise, a Whittle likelihood, with an inverse χ^2 prior for the noise may be used [63].

Step 3: **Choose a sparsity promoting prior for the SINDy coefficients.** Following §2.3, for *spike and slab SINDy (ss-SINDy)* we use the hierarchical prior

$$\xi_{i,j}|\lambda_j \sim \mathcal{N}(0, 1)\lambda_{i,j}\alpha_{i,j}$$

and

$$\lambda_{i,j} \sim \mathrm{Ber}(\pi).$$

For *regularized horseshoe SINDy (rh-SINDy)*, we use the hierarchical prior

$$\xi_{i,j}|\tilde{\lambda}_{i,j}, \tau, c \sim \mathcal{N}(0, 1)\tilde{\lambda}_{i,j}\tau\alpha_{i,j},$$
$$\tilde{\lambda}_{i,j} = \frac{c\lambda_{i,j}}{\sqrt{c^2 + \tau^2\lambda_{i,j}^2}},$$
$$\lambda_{i,j} \sim \mathrm{C}^+(0, 1),$$
$$c^2 \sim \text{Inv-Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}s^2\right)$$

and

$$\tau \sim \mathrm{C}^+(0, \tau_0).$$

For ss-SINDy, we have that $\boldsymbol{\phi}$ consists of the noise-level parameter σ and the local shrinkage paramaters $\lambda_{i,j}$. For rh-SINDy, $\boldsymbol{\phi}$ consists of σ, the $\lambda_{i,j}$s, c and τ. The coefficients $\alpha_{i,j}$, which we choose as constants for this analysis, allow us to incorporate any knowledge about the scales of different parameters. For this work, we choose $\alpha_i = 1$ unless stated otherwise.

Step 4: **Bayesian Inference.** Once the priors and the data likelihood are specified, we employ MCMC to draw samples from the posterior distribution $p(\Xi, \boldsymbol{x}_0, \boldsymbol{\phi}|\boldsymbol{X})$. Furthermore, we estimate the PPD (2.6) for the reconstruction and forecasting tasks of interest. We employ MCMC algorithms as implemented in the Python library PyMC3 [64]; specifically, for rh-SINDy we use the No-U-Turn Sampler (NUTS) [65], and for ss-SINDy we use the compound step sampler implemented in PyMC3.

In the UQ-SINDy framework, NUTS leverages the gradients of the SINDy model prediction $\hat{\boldsymbol{x}}(t; \Xi, \boldsymbol{x}_0)$ with respect to Ξ and $\boldsymbol{x}_0$. These gradients are computed using Sunode [66], a Python wrapper for the CVODES library [67] for solving forward and adjoint ODE problems.

3.2. Examples and applications

In this section, we apply the spike and slab and regularized horseshoe priors in the UQ-SINDy framework and illustrate their performance on three examples: two synthetic datasets and one real-world dataset of lynx and hare populations. For each example, we quantify the likelihood of each term of the SINDy library belonging to the underlying dynamical equations, providing both an estimate of the inclusion probability and a distribution of likely values for each SINDy coefficient. We compare these results to the original SINDy algorithm and show that UQ-SINDy significantly outperforms SINDy in identifying the underlying dynamics for noisy observations.

3.2.1. Lotka–Volterra model

We first study data from the Lotka–Volterra model, also commonly referred to as the predator–prey model, which is a popular system used to model the interaction between two competing groups [68,69]. Originally developed by Lotka to model chemical reactions [70], the system has also been studied as a model in economics [71] and for biological systems [72–74]. We explore one real-world example in §3.2.3.

The Lotka–Volterra model is given by the two nonlinear differential equations

$$\begin{aligned} \dot{u} &= \alpha u - \beta uv \\ \text{and} \quad \dot{v} &= -\gamma v + \delta uv. \end{aligned} \tag{3.5}$$

For this example, we simulate the system with the initial condition $[u_0, v_0] = [10, 5]$ and parameters $\alpha = 1$, $\beta = 0.1$, $\gamma = 1.5$ and $\delta = 0.075$, as in [75], which results in a periodic trajectory. We sample 50 snapshots over a time interval of $t \in [0, 24]$. Additionally, we contaminate this trajectory with lognormal multiplicative noise with distribution lognormal(0, 0.1), which corresponds for this dataset to approximately 10% additive noise. The lognormal distribution is non-negative and is commonly used to model observation errors for state variables restricted to non-negative values. The resulting time series is shown in figure 3, from which we see that the trajectory covers approximately four periods of oscillation.

For this example, we normalize the data as a preprocessing step by dividing each time series (of x and y) by the standard deviation of the data. The normalized data is governed by a differential equation of the same form as the unnormalized data, but with modified parameters $\tilde{\alpha} = 1$, $\tilde{\beta} = -0.68$, $\tilde{\gamma} = -1.5$ and

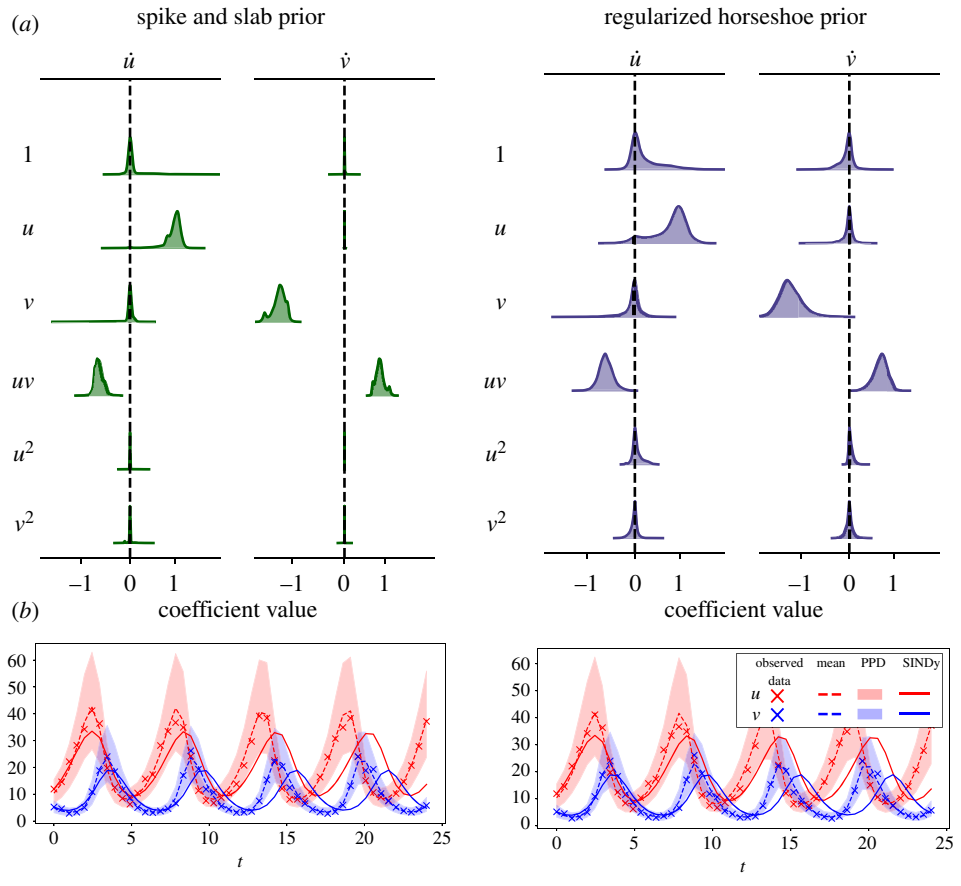

Figure 3. UQ-SINDy applied to a synthetic Lotka–Volterra system with lognormal noise. (*a*) Marginal ss-SINDy and rh-SINDy posterior distributions. (*b*) Observed (crosses) and predicted time series together with the corresponding PPD means (dashed lines) and 90% credibility intervals (shaded areas). SINDy predictions presented as continuous lines.

$\tilde{\delta} = 0.82$. This preprocessing step can be beneficial for systems in which the parameters are of different orders of magnitude.

We apply UQ-SINDy for both the spike and slab prior (ss-SINDy) and regularized horseshoe prior (rh-SINDy). We use a library of polynomial terms $\Theta(u, v) = [1,\ u,\ v,\ u^2,\ v^2,\ uv]$, resulting in a 6×2 matrix of SINDy coefficients Ξ. The SINDy model then reads

$$[\,\dot{u}\quad \dot{v}\,] = \begin{bmatrix} 1 & u & v & u^2 & v^2 & uv \end{bmatrix} \Xi, \quad u(0) = u_0, \quad v(0) = v_0.$$

For the noise level and initial condition, we employ the priors σ_u, $\sigma_v \sim \text{Lognormal}(\mu = -1, \sigma = 0.1)$ and u_0, $v_0 \sim \text{Lognormal}(\mu = 0, \sigma = 1)$, respectively.

In table 1, we present the inclusion probability (for ss-SINDy) and pseudo-probability (for rh-SINDy) of each term in the library. We see significantly higher probabilities for the four true non-zero terms compared to all other terms, indicating that both ss-SINDy and rh-SINDy correctly identify the structure of the governing equation. We note that although the inclusion pseudo-probabilities are not constrained between zero and one, the relevant terms are easily identified with values near to or greater than 1.

In figure 3, we present the marginal posterior distributions of the SINDy coefficient. From this, we immediately see that for both priors, the parameters that belong to the model have broad distributions centred about the true means, while the other eight terms have narrow peaks centred about 0. In table 1, we compare the posterior modes of the SINDy coefficients against the true values of the model parameters. We additionally apply the original SINDy algorithm to the data. We see that SINDy is unable to identify the correct dynamics due to the presence of observation noise. Furthermore, we note that due to their sparsifying behaviours, the posterior mode of the SINDy coefficients for both the spike and slab and regularized horseshoe priors are close to the true values.

In figure 3, we present the mean and 90% credibility interval of the PPDs of the UQ-SINDy reconstructions of the system's states. Furthermore, we also present the prediction using SINDy (solid

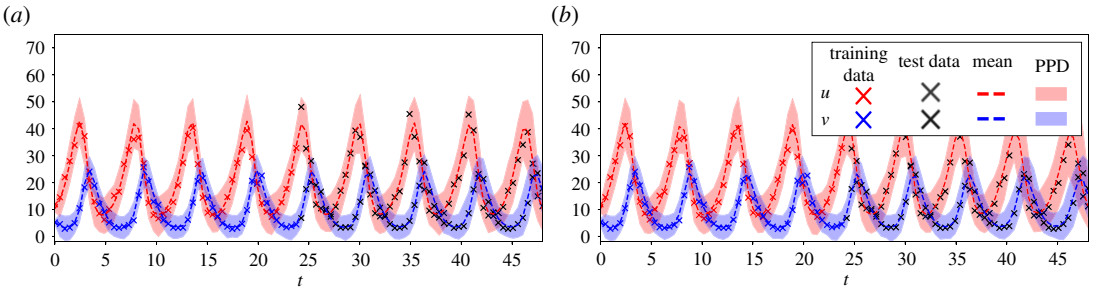

Figure 4. Forecasting using ss-SINDy (*a*) and rh-SINDy (*b*) for the Lotka–Volterra model. We train using samples the time interval [0, 24] (red and blue crosses) and test on samples over the time interval (24, 48] (black crosses). The mean (dashed lines) and 90% credibility intervals (dashed areas) of the PPDs are plotted for the entire time interval.

Table 1. (Above) Posterior modes of SINDy coefficients for the Lotka–Volterra model. (Below) Corresponding inclusion probabilities and pseudo-probabilities. The true non-zero terms in the Lotka–Volterra model are shaded.

	TRUE	SINDy	ss-SINDy	rh-SINDy
$\dot{u}:1$	0	0.62	0.00	0.02
$\dot{v}:1$	0	0	0.00	−0.01
$\dot{u}:u$	1	0.54	1.06	0.98
$\dot{v}:u$	0	0	0.00	0.00
$\dot{u}:v$	0	−0.49	0.00	−0.01
$\dot{v}:v$	−1.5	−1.32	−1.44	−1.39
$\dot{u}:uv$	−0.68	−0.321	−0.73	−0.67
$\dot{v}:uv$	0.82	0.71	0.78	0.73
$\dot{u}:u^2$	0	0	0.00	0.00
$\dot{v}:u^2$	0	0	0.00	0.00
$\dot{u}:v^2$	0	0	0.00	0.00
$\dot{v}:v^2$	0	0	0.00	0.00
		ss-SINDy		rh-SINDy
$\dot{u}:1$		0.36		0.02
$\dot{v}:1$		0.17		−0.05
$\dot{u}:u$		1.00		3.38
$\dot{v}:u$		0.13		0.00
$\dot{u}:v$		0.27		0.03
$\dot{v}:v$		1.00		1.03
$\dot{u}:uv$		1.00		1.17
$\dot{v}:uv$		1.00		1.17
$\dot{u}:u^2$		0.09		0.02
$\dot{v}:u^2$		0.01		0.04
$\dot{u}:v^2$		0.19		0.04
$\dot{v}:v^2$		0.03		−0.02

lines) and the observed values (crosses). The means of the PPDs for each of the model states are close in value to the true data and provide an accurate continuous reconstruction of the data. In addition, both the regularized horseshoe and spike and slab priors result in similar credibility intervals that bound the true samples. The SINDy reconstruction on the other hand degrades for samples at later times.

Finally, we demonstrate how the UQ-SINDy framework can be used for forecasting. To do this, we first simulate noisy data over the time interval (24, 48] (black crosses) and use this as our test set

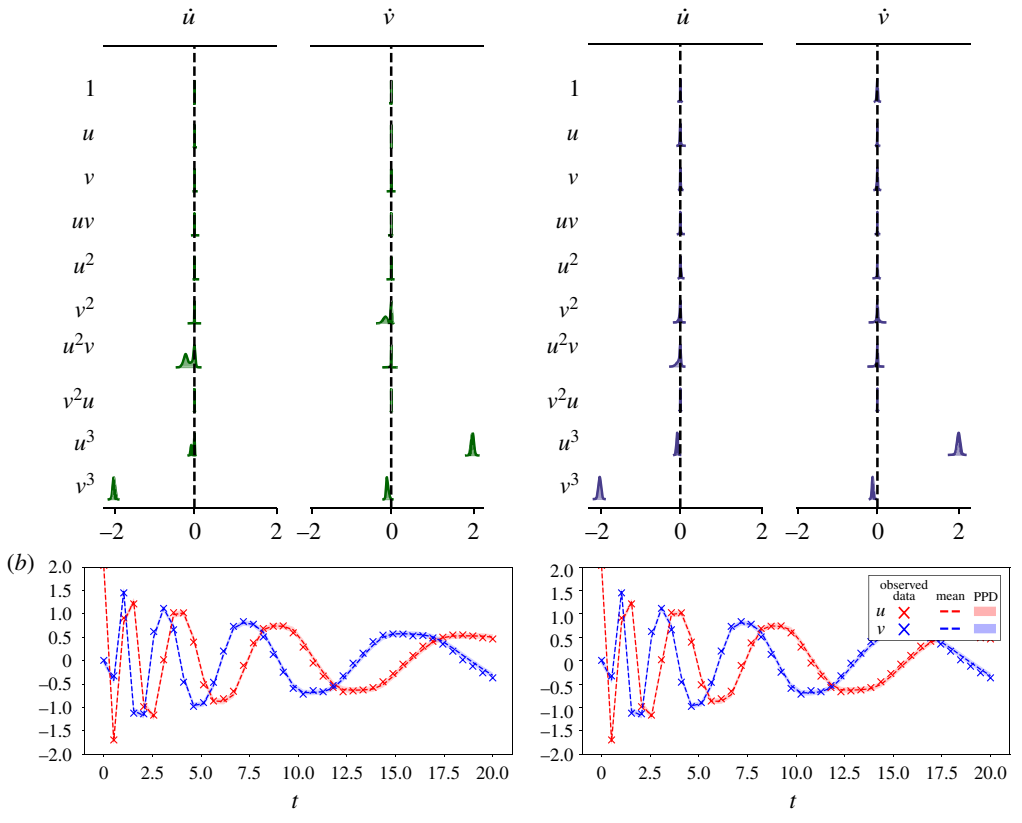

Figure 5. UQ-SINDy applied to a synthetic nonlinear oscillator system with normal noise. (*a*) Marginal ss-SINDy and rh-SINDy posterior distributions. (*b*) Observed (crosses) and predicted time series together with the corresponding PPD means (dashed lines) and 90% credibility intervals (shaded areas).

(figure 4). We then compute the PPD over the entire time interval [0, 48] by sampling from (2.6), and plot the mean and 90% credibility interval of this distribution. We find that the mean of the PPD is very close in value to the true values in the test set. Further, we note that some samples in the test set lie near the bounds of the credibility intervals. This shows that our credibility bounds are tight and accurately capture the uncertainty due to measurement noise.

3.2.2. Nonlinear oscillator and model indeterminacy

As a second example, we consider the damped nonlinear oscillator model of the form

$$\dot{u} = \alpha u^3 + \beta v^3$$

and

$$\dot{v} = \gamma v^3 + \delta u^3.$$

Following [20], we use the values $\alpha = -0.1$, $\beta = -2$, $\gamma = 2$, $\delta = -0.1$ and the initial conditions $[u_0, v_0] = [2, 0]$. Data are generated by sampling this model over the interval $t \in [0, 20]$ with a sampling period of $\Delta t = 0.2$, and adding normally distributed observation noise with distribution $\mathcal{N}(0, 0.02^2)$. The observed trajectory is shown in figure 5. We use a library of polynomial terms $\Theta(u, v) = [1, u, v, u^2, v^2, uv, u^3, v_3, u^2v, v^2u]$, resulting in a 10×2 matrix of SINDy coefficients Ξ. Since the observation noise is normally distributed noise we employ the data likelihood in (3.3). For the noise level and initial condition, we employ the priors $\sigma_u, \sigma_v \sim \text{Gamma}(\alpha = 1, \beta = 0.1)$ and $u_0, v_0 \sim \text{Laplace}(\mu = 0, b = 1)$, respectively.

First, we apply SINDy to the data, resulting in the estimated SINDy coefficients presented in table 2. It can be seen that SINDy does not identify the relevant terms in the model or correctly estimate the values of the model parameters. In fact, none of the terms in the SINDy model are zero. This example is particularly challenging for SINDy because of the sparse data sampling, the size of the library and

Table 2. (Above) Posterior modes of SINDy coefficients for the nonlinear oscillator model. (Below) Corresponding inclusion probabilities and pseudo-probabilities. The true non-zero terms in the nonlinear oscillator model are shaded.

	TRUE	SINDy	ss-SINDy	rh-SINDy
$\dot{u}:1$	0	0.46	0.00	0.00
$\dot{v}:1$	0	0.06	0.00	0.00
$\dot{u}:u$	0	0.54	0.00	0.00
$\dot{v}:u$	0	−0.62	0.00	0.00
$\dot{u}:v$	0	0.81	0.00	0.00
$\dot{v}:v$	0	−0.09	0.00	0.00
$\dot{u}:uv$	0	−0.45	0.00	0.00
$\dot{v}:uv$	0	−0.14	0.00	0.00
$\dot{u}:u^2$	0	−1.82	0.00	0.00
$\dot{v}:u^2$	0	−0.38	0.00	0.00
$\dot{u}:v^2$	0	0.43	0.00	0.00
$\dot{v}:v^2$	0	0.34	0.00	0.00
$\dot{u}:u^2v$	0	0.39	0.00	0.00
$\dot{v}:u^2v$	0	0.15	0.00	0.00
$\dot{u}:v^2u$	0	1.37	0.00	0.00
$\dot{v}:v^2u$	0	−0.22	0.00	0.00
$\dot{u}:u^3$	−0.1	−1.41	0.00	−0.08
$\dot{v}:u^3$	2	−0.53	2.04	2.02
$\dot{u}:v^3$	−2	0.02	−1.96	−1.96
$\dot{v}:v^3$	−0.1	−0.15	−0.11	−0.12

	ss-SINDy	rh-SINDy
$\dot{u}:1$	0.00	−0.04
$\dot{v}:1$	0.01	−0.19
$\dot{u}:u$	0.10	−0.02
$\dot{v}:u$	0.00	−0.07
$\dot{u}:v$	0.05	0.03
$\dot{v}:v$	0.08	−0.03
$\dot{u}:uv$	0.08	−0.02
$\dot{v}:uv$	0.06	0.00
$\dot{u}:u^2$	0.07	0.01
$\dot{v}:u^2$	0.07	0.00
$\dot{u}:v^2$	0.14	0.25
$\dot{v}:v^2$	0.5	−0.01
$\dot{u}:u^2v$	0.70	0.47
$\dot{v}:u^2v$	0.17	0.01
$\dot{u}:v^2u$	0.01	0.38
$\dot{v}:v^2u$	0.02	0.14
$\dot{u}:u^3$	0.50	1.24
$\dot{v}:u^3$	1.00	1.05
$\dot{u}:v^3$	1.00	0.99
$\dot{v}:v^3$	1.00	0.82

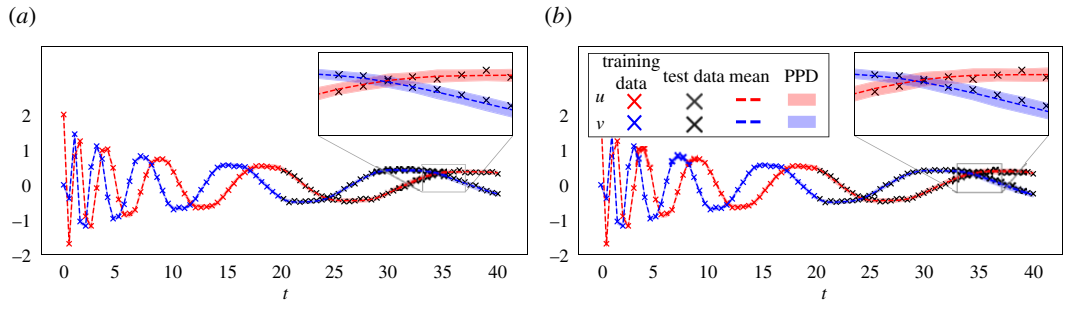

Figure 6. Forecasting using ss-SINDy (*a*) and rh-SINDy (*b*) for the nonlinear oscillator model. We train using samples from the nonlinear oscillator model over the time interval [0, 20] (red and blue crosses) and test on samples over the time interval (20, 40] (black crosses). The mean (dashed lines) and 90% credibility intervals (dashed areas) of the PPDs are plotted for the entire time interval.

the large range of magnitudes of the non-zero coefficients (specifically, note that $|\alpha|$ and $|\delta|$ are much smaller than $|\beta|$ and $|\gamma|$). Next, we apply ss-SINDy and rh-SINDy to these data. The posterior modes are shown in table 2.

We present the marginal posterior distributions of the SINDy coefficients in figure 5. It can be seen that rh-SINDy correctly identifies the governing equation; specifically, we see that the marginal posterior distribution of the SINDy coefficients for the terms in the equation are centred away from zero, while the distributions of all other terms are sharply centred at zero. On the other hand, ss-SINDy identifies the four terms in the governing equation, while also identifying an additional mode corresponding to a model without the $\dot{u}:u^3$ term but with the $\dot{u}:u^2v$ and $\dot{v}:v^2$ terms. These results are reflected in table 2, for which we show the posterior modes of the SINDy coefficients and the corresponding inclusion probabilities and pseudo-probabilities. For rh-SINDy, the four non-zero terms are clearly identified with modes close to the true values and inclusion pseudo-probabilities for the four terms close to one. For ss-SINDy, three of the terms are clearly identified with an inclusion probability close to one, while the terms $\dot{u}:u^3$, $\dot{u}:u^2v$ and $\dot{v}:v^2$ have inclusion probabilities of 0.5, 0.7 and 0.5, respectively.

In figure 5, we present the mean and 90% credibility intervals of the PPDs of the reconstruction of the system's states, together with the training data. Similarly, in figure 6 we present the 90% credibility intervals of the PPDs of future state forecasting for testing data over the time interval (20, 40]. Both rh-SINDy and ss-SINDy lead to similar credibility intervals for both reconstruction and forecasting. We note that the range of predicted model states is much narrower than for the Lotka–Volterra model, which is expected due to the lower noise level present in these measurements. We also emphasize that these PPDs are much tighter than those presented in [3] for this test case, even though we train rh-SINDy and ss-SINDy with substantially less data than in that work. Furthermore, it can be seen that the test data lie within the 90% credibility intervals of the PPDs of each state. Although some of the draws from the ss-SINDy PPD contain terms not in the model, the credibility intervals for both ss-SINDy and rh-SINDy are similar. This suggests that the ambiguity identified by the spike and slab prior is due to model indeterminacy inherent in these dataset.

This indeterminacy can be attributed to the range of values spanned by the coefficients in the governing equation. In particular, the coefficients of $\dot{u}:u^3$ and $\dot{v}:v^3$ are an order of magnitude smaller than the coefficients of $\dot{u}:v^3$ and $\dot{u}:u^3$. To further investigate this indeterminacy, we re-applied ss-SINDy and rh-SINDy with $\alpha_{i,j}=0.1$ for the terms $\dot{u}:u^3$ and $\dot{v}:v^3$. This scaling of the prior incorporates the knowledge that these two terms have coefficients of magnitude $O(0.1)$. The resulting marginal posterior distributions, presented in figure 7, show that this scaling of the prior removes this ambiguity. The corresponding posterior modes and inclusion probabilities and pseudo-probabilities are presented in table 3.

3.2.3. Lynx-hare population model

As a final example, we apply ss-SINDy and rh-SINDy as described in §3.2.1 to model the population dynamics of two species in Canada. In particular, we consider data consisting of measurements by the Hudson Bay Company of lynx and hare pelts between 1900 and 1920 [75,76] (figure 8). The number of pelts for these two species is thought to be proportional to the true populations. Hares are a herbivorous relative of the rabbit, while the lynx is a type of wildcat whose diet depends heavily on hares. This predator–prey interdependence between the two species has been shown to be well

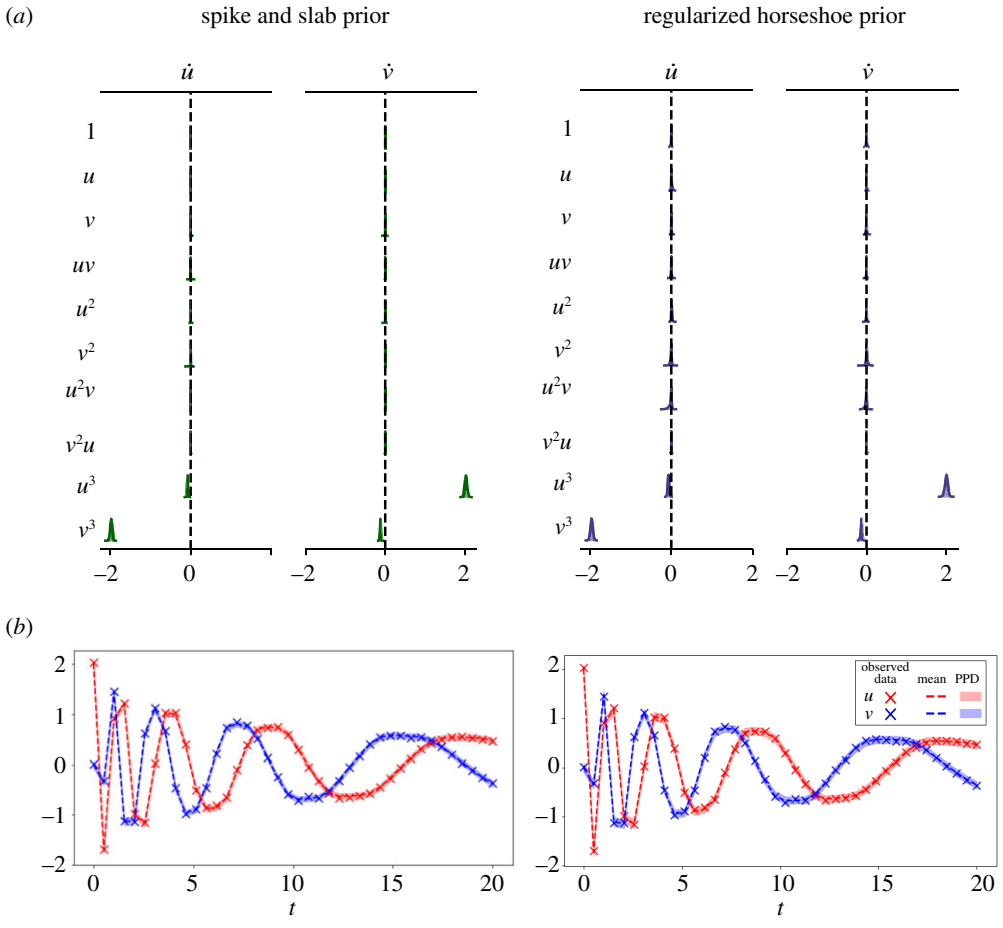

Figure 7. UQ-SINDy, with scaled priors for the terms $\dot{u} : u^3$ and $\dot{v} : v^3$, applied to a synthetic nonlinear oscillator system with normal noise. (*a*) Marginal ss-SINDy and rh-SINDy posterior distributions. (*b*) Observed (crosses) and predicted time series together with the corresponding PPD means (dashaded lines) and 90% credibility intervals (shaded areas).

characterized to first-order by the Lotka–Volterra model (3.5), where u and v correspond to the populations of hares and lynx, respectively. Empirically, the data have been found to be very noisy, with a noise level of approximately 25% [75].

Figure 8 presents the number of pelts recorded yearly for these two species over 21 years. Modelling these data with SINDy is particularly challenging because we have relatively few samples that cover only two cycles. In addition, factors such as the weather and the consistency of trapping between years adds uncertainty to the measurements. Here we compare the performance of ss-SINDY, and rh-SINDY for model discovery under uncertainty. The SINDy library, as in the Lotka–Volterra example, contains all constant, linear and quadratic terms. In addition, as a preprocessing step we normalize the data as described in §3.2.1

The marginal posterior distributions computed using ss-SINDy and rh-SINDy are presented in figure 8. The posterior modes and inclusion probabilities and pseudo-probabilities are presented in table 4, together with maximum-likelihood estimates of the coefficients of the Lotka–Volterra model for the lynx-hare data [75], and estimates computed using the original SINDy algorithm. It can be seen that for ss-SINDy the distinct non-zero peaks correspond to the terms in (3.5). The likelihood of these four terms belonging to the model are very high. We additionally see a small peak near zero for $\dot{u} : u$. This term is highly correlated with a non-zero constant term. We see a similar but more pronounced peak for rh-SINDy. Table 4 shows that ss-SINDy correctly identifies the Lotka–Volterra model and assigns high inclusion probabilities to the four terms in such a model. On the other hand, rh-SINDy identifies three of the four terms correctly. Furthermore, it can be seen that SINDy fails to identify the Lotka–Volterra model, and that the posterior modes for ss-SINDy and rh-SINDy are closer to the maximum-likelihood estimates than the SINDy estimates.

Last, in figure 8, we present the mean and 90% credibility intervals of the PPDs of the time series reconstruction. We note that all data lie within these credibility bounds. The SINDy reconstructions, on the other hand, appear to deviate from the time series for later times.

Table 3. (Above) Posterior modes of SINDy coefficients, with scaled priors for the terms $\dot{u} : u^3$ and $\dot{v} : v^3$, for the nonlinear oscillator model. (Below) Corresponding inclusion probabilities and pseudo-probabilities. The true non-zero terms in the nonlinear oscillator model are shaded.

	TRUE	ss-SINDy	rh-SINDy
$\dot{u} : 1$	0	0.00	0.00
$\dot{v} : 1$	0	0.00	0.00
$\dot{u} : u$	0	0.00	0.00
$\dot{v} : u$	0	0.00	0.00
$\dot{u} : v$	0	0.00	0.00
$\dot{v} : v$	0	0.00	0.00
$\dot{u} : uv$	0	0.00	0.00
$\dot{v} : uv$	0	0.00	0.00
$\dot{u} : u^2$	0	0.00	0.00
$\dot{v} : u^2$	0	0.00	0.00
$\dot{u} : v^2$	0	0.00	0.00
$\dot{v} : v^2$	0	0.00	0.00
$\dot{u} : u^2v$	0	0.00	0.00
$\dot{v} : u^2v$	0	0.00	0.00
$\dot{u} : v^2u$	0	0.00	0.00
$\dot{v} : v^2u$	0	0.00	0.00
$\dot{u} : u^3$	−0.1	−0.08	−0.07
$\dot{v} : u^3$	2	2.03	2.01
$\dot{u} : v^3$	−2	−1.97	−1.97
$\dot{v} : v^3$	−0.1	−0.12	−0.12

	ss-SINDy	rh-SINDy
$\dot{u} : 1$	0.00	0.01
$\dot{v} : 1$	0.00	−0.01
$\dot{u} : u$	0.00	0.02
$\dot{v} : u$	0.00	0.03
$\dot{u} : v$	0.03	0.01
$\dot{v} : v$	0.11	−0.01
$\dot{u} : uv$	0.10	−0.03
$\dot{v} : uv$	0.10	0.01
$\dot{u} : u^2$	0.08	0.08
$\dot{v} : u^2$	0.05	0.00
$\dot{u} : v^2$	0.17	0.11
$\dot{v} : v^2$	0.00	0.014
$\dot{u} : u^2v$	0.00	0.01
$\dot{v} : u^2v$	0.00	0.03
$\dot{u} : v^2u$	0.03	−0.56
$\dot{v} : v^2u$	0.01	−2.53
$\dot{u} : u^3$	1.00	1.23
$\dot{v} : u^3$	1.00	1.05
$\dot{u} : v^3$	1.00	1.00
$\dot{v} : v^3$	1.00	0.93

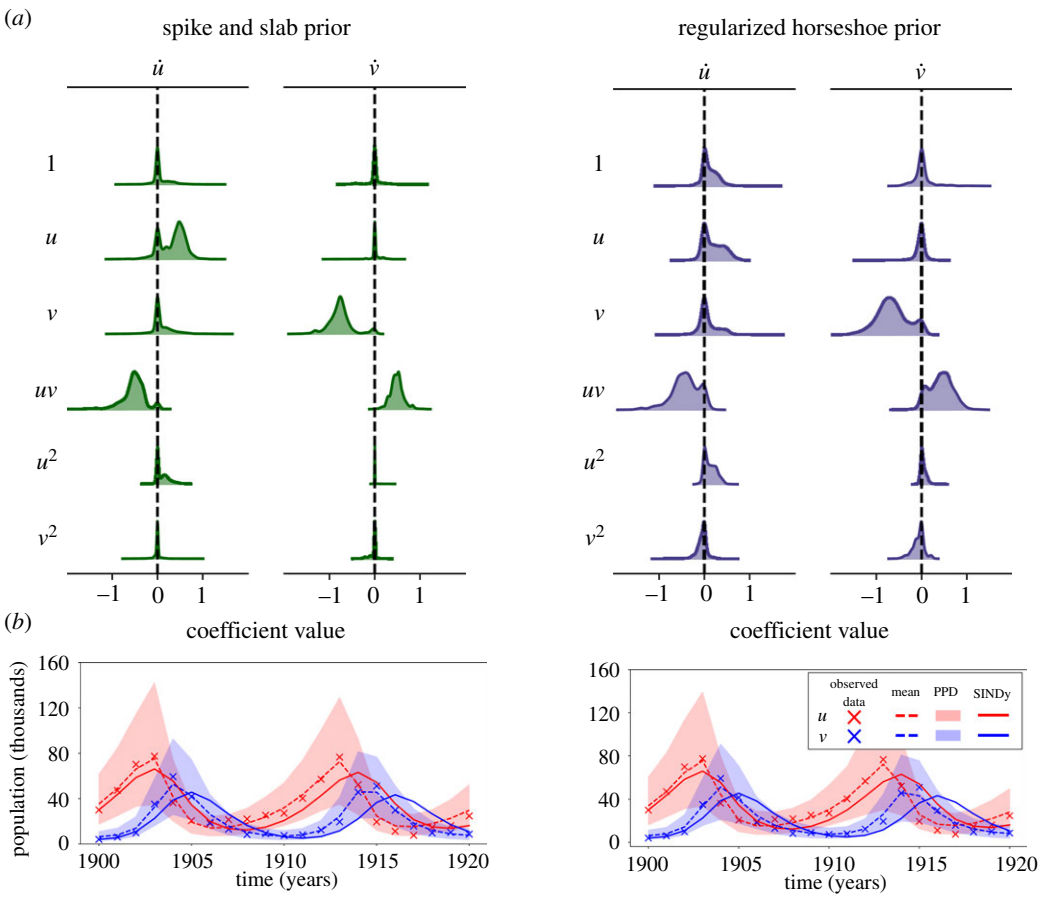

Figure 8. UQ-SINDy applied to the lynx-hare population data. (*a*) Marginal ss-SINDy and rh-SINDy posterior distributions. (*b*) Observed (crosses) and predicted time series together with the corresponding PPD means (dashed lines) and 90% credibility intervals (shaded areas). SINDy predictions presented as continuous lines.

4. Conclusion and future work

In this work, we proposed UQ-SINDy, a new uncertainty quantification framework for identifying governing ODEs directly from noisy and sparse time series data. We leverage advances in model discovery for dynamical systems and sparsity promoting Bayesian inference to identify a sparse set of SINDy library functions that best explain the observed data and quantify the uncertainty in the SINDy coefficients due to measurement noise and the probability of inclusion of each term in the final model. We have applied UQ-SINDy to two synthetic examples and one real-world example of lynx-hare population data. By using the spike-and-slab and regularized horseshoe priors, UQ-SINDy yields posterior distributions of SINDy coefficients with truly sparse draws, and thus results in truly sparse probabilistic model discovery; in contrast, the use of the Laplace prior does not lead to sparse model discovery. We observe that the proposed approach is robust against observation noise and can accommodate sparse samples and small datasets.

Going forward, one of the primary limitations of this method is its scalability to very large SINDy libraries, such as libraries of rational functions, which are common in many biological and physical systems [24,25]. This is primarily due to the computational cost of sampling high-dimensional posterior distributions using MCMC. One remedy for this is to use variational inference, which matches classes of distributions to the posterior distribution by maximizing a lower bound on the marginal likelihood of the data. This method has been particularly effective for high-dimensional models, most notably neural networks, with comparable accuracy to sampling-based methods. Furthermore, in this work we are primarily focused on situations in which the coordinates that induce a sparse representation are known. However, in general this 'effective' set of coordinates may be unknown. Recent work merges SINDy together with neural network architectures in order to simultaneously learn parsimonious governing equations and the associated sparsity-inducing coordinate transformation [17]. Extending UQ-SINDy to this coordinate discovery framework could

Table 4. (Above) Posterior modes of SINDy coefficients for the lynx-hare data. (Below) Corresponding inclusion probabilities and pseudo-probabilities. The non-zero terms in the Lotka–Volterra model are shaded.

	param. est.	SINDy	ss-SINDy	rh-SINDy
$\dot{u} : 1$	0	0	0.00	0.01
$\dot{v} : 1$	0	0	0.00	0.00
$\dot{u} : u$	0.55	0.48	0.47	0.00
$\dot{v} : u$	0	0	0.00	−0.01
$\dot{u} : v$	0	−0.143	0.00	0.00
$\dot{v} : v$	−0.84	−0.71	−0.76	−0.7
$\dot{u} : uv$	−0.455	−0.36	−0.51	−0.42
$\dot{v} : uv$	0.5433	0.42	0.52	0.52
$\dot{u} : u^2$	0	0	0.00	0.01
$\dot{v} : u^2$	0	0	0.00	0.00
$\dot{u} : v^2$	0	0	0.00	0.00
$\dot{v} : v^2$	0	0	0.00	−0.01

	ss-SINDy	rh-SINDy
$\dot{u} : 1$	0.47	0.04
$\dot{v} : 1$	0.35	0.00
$\dot{u} : u$	0.85	0.01
$\dot{v} : u$	0.34	0.02
$\dot{u} : v$	0.54	0.00
$\dot{v} : v$	0.99	0.73
$\dot{u} : uv$	0.96	0.78
$\dot{v} : uv$	1	2.01
$\dot{u} : u^2$	0.581	0.03
$\dot{v} : u^2$	0.08	0.02
$\dot{u} : v^2$	0.31	−0.50
$\dot{v} : v^2$	0.35	−0.06

greatly improve the robustness of the learning process under uncertainty and the quality of the resulting forecasts.

Data accessibility. Data and relevant code for this research work are stored in GitHub: https://github.com/sethhirsh/BayesianSindy and have been archived within the Zenodo repository: https://doi.org/10.5281/zenodo.5893738.

Authors' contributions. S.M.H.: conceptualization, data curation, investigation, methodology, software, writing—original draft, writing—review and editing; D.A.B.-S.: conceptualization, funding acquisition, investigation, methodology, writing—original draft, writing—review and editing; J.N.K.: conceptualization, funding acquisition, investigation, methodology, project administration, supervision, writing–original draft, writing–review and editing. All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Competing interests. We declare we have no competing interests.

Funding. This research was supported by Laboratory Directed Research and Development Program and Mathematics for Artificial Reasoning for Scientific Discovery investment at the Pacific Northwest National Laboratory, a multiprogram national laboratory operated by Battelle for the US Department of Energy under Contract DE-AC05- 76RLO1830.

References

1. Bongard J, Lipson H. 2007 Automated reverse engineering of nonlinear dynamical systems. *Proc. Natl Acad. Sci. USA* **104**, 9943–9948. (doi:10.1073/pnas.0609476104)
2. Schmidt M, Lipson H. 2009 Distilling free-form natural laws from experimental data. *Science* **324**, 81–85. (doi:10.1126/science.1165893)
3. Yang Y, Aziz Bhouri M, Perdikaris P. 2020 Bayesian differential programming for robust systems identification under uncertainty. Preprint. (https://arxiv.org/abs/2004.06843)

4. Bai Z, Wimalajeewa T, Berger Z, Wang G, Glauser M, Varshney PK. 2015 Low-dimensional approach for reconstruction of airfoil data via compressive sensing. *AIAA J.* **53**, 920–933. (doi:10.2514/1.J053287)
5. Brunton SL, Proctor JL, Kutz JN. 2016 Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl Acad. Sci. USA* **113**, 3932–3937. (doi:10.1073/pnas.1517384113)
6. Brunton SL, Tu JH, Bright I, Kutz JN. 2014 Compressive sensing and low-rank libraries for classification of bifurcation regimes in nonlinear dynamical systems. *SIAM J. Appl. Dyn. Syst.* **13**, 1716–1732. (doi:10.1137/130949282)
7. Mackey A, Schaeffer H, Osher S. 2014 On the compressive spectral method. *Multiscale Model. Simul.* **12**, 1800–1827. (doi:10.1137/140965909)
8. Ozoliņš V, Lai R, Caflisch R, Osher S. 2013 Compressed modes for variational problems in mathematics and physics. *Proc. Natl Acad. Sci. USA* **110**, 18 368–18 373. (doi:10.1073/pnas.1318679110)
9. Proctor JL, Brunton SL, Brunton BW, Kutz JN. 2014 Exploiting sparsity and equation-free architectures in complex systems. *Eur. Phys. J. Spec. Top.* **223**, 2665–2684. (doi:10.1140/epjst/e2014-02285-8)
10. Tran G, Ward R. 2017 Exact recovery of chaotic systems from highly corrupted data. *Multiscale Model. Simul.* **15**, 1108–1129. (doi:10.1137/16M1086637)
11. Wang W-X, Yang R, Lai Y-C, Kovanis V, Grebogi C. 2011 Predicting catastrophes in nonlinear dynamical systems by compressive sensing. *Phys. Rev. Lett.* **106**, 154101. (doi:10.1103/PhysRevLett.106.154101)
12. Hoffmann M, Fröhner C, Noé F. 2019 Reactive SINDy: discovering governing reactions from concentration data. *J. Chem. Phys.* **150**, 025101. (doi:10.1063/1.5066099)
13. Sorokina M, Sygletos S, Turitsyn S. 2016 Sparse identification for nonlinear optical communication systems: SINO method. *Opt. Express* **24**, 30 433–30 443. (doi:10.1364/OE.24.030433)
14. Li S, Kaiser E, Laima S, Li H, Brunton SL, Kutz JN. 2019 Discovering time-varying aerodynamics of a prototype bridge by sparse identification of nonlinear dynamical systems. *Phys. Rev. E* **100**, 022220. (doi:10.1103/PhysRevE.100.022220)
15. Horrocks J, Bauch CT. 2020 Algorithmic discovery of dynamic models from infectious disease data. *Sci. Rep.* **10**, 1–18. (doi:10.1038/s41598-020-63877-w)
16. Dam M, Brøns M, Juul Rasmussen J, Naulin V, Hesthaven JS. 2017 Sparse identification of a predator-prey system from simulation data of a convection model. *Phys. Plasmas* **24**, 022310. (doi:10.1063/1.4977057)
17. Champion K, Lusch B, Kutz JN, Brunton SL. 2019 Data-driven discovery of coordinates and governing equations. *Proc. Natl Acad. Sci. USA* **116**, 22 445–22 451. (doi:10.1073/pnas.1906995116)
18. Champion K, Zheng P, Aravkin AY, Brunton SL, Kutz JN. 2020 A unified sparse optimization framework to learn parsimonious physics-informed models from data. *IEEE Access* **8**, 169 259–169 271. (doi:10.1109/ACCESS.2020.3023625)
19. Kaheman K, Brunton SL, Kutz JN. 2020 Automatic differentiation to simultaneously identify nonlinear dynamics and extract noise probability distributions from data. Preprint. (https://arxiv.org/abs/2009.08810)
20. Raissi M, Karniadakis GE. 2018 Hidden physics models: machine learning of nonlinear partial differential equations. *J. Comput. Phys.* **357**, 125–141. (doi:10.1016/j.jcp.2017.11.039)
21. Rudy S, Alla A, Brunton SL, Kutz JN. 2019 Data-driven identification of parametric partial differential equations. *SIAM J. Appl. Dyn. Syst.* **18**, 643–660. (doi:10.1137/18M1191944)
22. Rudy SH, Brunton SL, Proctor JL, Kutz JN. 2017 Data-driven discovery of partial differential equations. *Sci. Adv.* **3**, e1602614. (doi:10.1126/sciadv.1602614)
23. Champion KP, Brunton SL, Kutz JN. 2019 Discovery of nonlinear multiscale systems: sampling strategies and embeddings. *SIAM J. Appl. Dyn. Syst.* **18**, 312–333. (doi:10.1137/18M1188227)
24. Kaheman K, Kutz JN, Brunton SL. 2020 SINDy-PI: a robust algorithm for parallel implicit sparse identification of nonlinear dynamics. Preprint. (https://arxiv.org/abs/2004.02322)
25. Mangan NM, Brunton SL, Proctor JL, Kutz JN. 2016 Inferring biological networks by sparse identification of nonlinear dynamics. *IEEE Trans. Mol. Biol. Multi-Scale Commun.* **2**, 52–63. (doi:10.1109/TMBMC.2016.2633265)
26. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. 2013 *Bayesian data analysis*. Boca Raton, FL: CRC Press.
27. West M, Harrison J. 2006 *Bayesian forecasting and dynamic models*. Berlin, Germany: Springer Science & Business Media.
28. Abramson B, Brown J, Edwards W, Murphy A, Winkler RL. 1996 Hailfinder: a Bayesian system for forecasting severe weather. *Int. J. Forecast.* **12**, 57–71. (doi:10.1016/0169-2070(95)00664-8)
29. Elsner JB, Jagger TH. 2004 A hierarchical Bayesian approach to seasonal hurricane modeling. *J. Clim.* **17**, 2813–2827. (doi:10.1175/1520-0442(2004)017<2813:AHBATS>2.0.CO;2)
30. Yu R, Abdel-Aty M, Ahmed M. 2013 Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors. *Accid. Anal. Prev.* **50**, 371–376. (doi:10.1016/j.aap.2012.05.011)
31. Best N, Richardson S, Thomson A. 2005 A comparison of Bayesian spatial models for disease mapping. *Stat. Methods Med. Res.* **14**, 35–59. (doi:10.1191/0962280205sm388oa)
32. Lawson AB. 2013 *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. Boca Raton, FL: CRC Press.
33. Yuen JE, Hughes G. 2002 Bayesian analysis of plant disease prediction. *Plant Pathol.* **51**, 407–412. (doi:10.1046/j.0032-0862.2002.00741.x)
34. Castillo E, Menéndez JM, Sánchez-Cambronero S. 2008 Predicting traffic flow using Bayesian networks. *Transp. Res. Part B: Methodol.* **42**, 482–509. (doi:10.1016/j.trb.2007.10.003)
35. Sun S, Zhang C, Yu G. 2006 A Bayesian network approach to traffic flow forecasting. *IEEE Trans. Intell. Transp. Syst.* **7**, 124–132. (doi:10.1109/TITS.2006.869623)
36. Zheng W, Lee D-H, Shi Q. 2006 Short-term freeway traffic flow prediction: Bayesian combined neural network approach. *J. Transp. Eng.* **132**, 114–121. (doi:10.1061/(ASCE)0733-947X(2006)132:2(114))
37. Gerlach RH, Chen CWS, Chan NYC. 2011 Bayesian time-varying quantile forecasting for value-at-risk in financial markets. *J. Bus. Econ. Stat.* **29**, 481–492. (doi:10.1198/jbes.2010.08203)
38. Ticknor JL. 2013 A Bayesian regularized artificial neural network for stock market forecasting. *Expert Syst. Appl.* **40**, 5501–5506. (doi:10.1016/j.eswa.2013.04.013)
39. Wright JH. 2008 Bayesian model averaging and exchange rate forecasts. *J. Econom.* **146**, 329–341. (doi:10.1016/j.jeconom.2008.08.012)
40. Lee S-Y, Song X-Y. 2004 Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behav. Res.* **39**, 653–686. (doi:10.1207/s15327906mbr3904_4)
41. Price LR. 2012 Small sample properties of Bayesian multivariate autoregressive time series models. *Struct. Equ. Model.: Multidiscip. J.* **19**, 51–64. (doi:10.1080/10705511.2012.634712)
42. Zitzmann S, Lüdtke O, Robitzsch A, Hecht M. 2021 On the performance of Bayesian approaches in small samples: a comment on Smid, McNeish, Miocevic, and van de Schoot (2020). *Struct. Equ. Model.: Multidiscip. J.* **28**, 40–50. (doi:10.1080/10705511.2020.1752216)
43. Galioto N, Gorodetsky A. 2020 Bayesian system ID: optimal management of parameter, model, and measurement uncertainty. Preprint. (https://arxiv.org/abs/2003.02359)
44. Niven RK, Mohammad-Djafari A, Cordier L, Abel M, Quade M. 2020 Bayesian identification of dynamical systems. *Multidiscip. Digit. Publish. Inst. Proc.* **33**, 33. (doi:10.3390/proceedings2019033033)
45. Tibshirani R. 1996 Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Series B (Methodol.)* **58**, 267–288. (doi:10.1111/j.2517-6161.1996.tb02080.x)
46. Park T, Casella G. 2008 The Bayesian lasso. *J. Am. Stat. Assoc.* **103**, 681–686. (doi:10.1198/016214508000000337)
47. Castillo I, Schmidt-Hieber J, Van der Vaart A. 2015 Bayesian linear regression with sparse priors. *Ann. Stat.* **43**, 1986–2018. (doi:10.1214/15-AOS1334)
48. Ishwaran H, Rao JS. 2005 Spike and slab variable selection: frequentist and Bayesian strategies. *Ann. Stat.* **33**, 730–773. (doi:10.1214/009053604000001147)
49. Madigan D, Raftery AE. 1994 Model selection and accounting for model uncertainty in graphical models using Occam's window. *J. Am. Stat. Assoc.* **89**, 1535–1546. (doi:10.1080/01621459.1994.10476894)
50. Mitchell TJ, Beauchamp JJ. 1988 Bayesian variable selection in linear regression. *J. Am. Stat. Assoc.* **83**, 1023–1032. (doi:10.1080/01621459.1988.10478694)

51. Carvalho CM, Polson NG, Scott JG. 2009 Handling sparsity via the horseshoe. In *Proc. of the 12th Int. Conf. on Artificial Intelligence and Statistics, Clearwater Beach, FL, 16–18 April*, pp. 73–80. Proceedings of Machine Learning Research (PMLR).
52. Carvalho CM, Polson NG, Scott JG. 2010 The horseshoe estimator for sparse signals. *Biometrika* **97**, 465–480. (doi:10.1093/biomet/asq017)
53. Bhadra A, Datta J, Polson NG, Willard B. 2017 The horseshoe+ estimator of ultra-sparse signals. *Bayesian Anal.* **12**, 1105–1131. (doi:10.1214/16-BA1028)
54. Piironen J, Vehtari A. 2017 Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electron. J. Stat.* **11**, 5018–5051. (doi:10.1214/17-EJS1337SI)
55. Bhattacharya A, Pati D, Pillai NS, Dunson DB. 2015 Dirichlet–Laplace priors for optimal shrinkage. *J. Am. Stat. Assoc.* **110**, 1479–1490. (doi:10.1080/01621459.2014.960967)
56. Zhang Y, Reich BJ, Bondell HD. 2016 High dimensional linear regression via the R2-D2 shrinkage prior. Preprint. (https://arxiv.org/abs/1609.00046)
57. Gelman A, Meng X-L, Stern H. 1996 Posterior predictive assessment of model fitness via realized discrepancies. *Stat. Sin.* **6**, 733–760.
58. Tran D, Kucukelbir A, Dieng AB, Rudolph M, Liang D, Blei DM. 2016 Edward: a library for probabilistic modeling, inference, and criticism. Preprint. (https://arxiv.org/abs/1610.09787)
59. De Laplace PS. 1774 Mémoire sur la probabilité des causes par les événements. *Mém. Math. Phys. Présentés l'Acad. R. Sci.* **6**, 621–656.
60. Bhadra A, Datta J, Polson NG, Willard B. 2019 Lasso meets horseshoe: a survey. *Stat. Sci.* **34**, 405–427. (doi:10.1214/19-STS700)
61. Gelman A, Hill J. 2006 *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: Cambridge University Press.
62. Stan Development Team. 2016 Stan modeling language users guide and reference manual. Technical report.
63. Röver C, Meyer R, Christensen N. 2010 Modelling coloured residual noise in gravitational-wave signal processing. *Classical Quantum Gravity* **28**, 015010. (doi:10.1088/0264-9381/28/1/015010)
64. Salvatier J, Wiecki TV, Fonnesbeck C. 2016 Probabilistic programming in Python using PyMC3. *PeerJ Comput. Sci.* **2**, e55. (doi:10.7717/peerj-cs.55)
65. Hoffman MD, Gelman A. 2014 The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15**, 1593–1623.
66. Seyboldt A, Osthege M, Störiko A, Widmer L. 2021 Aseyboldt/sunode: version 0.2.1, April.
67. Serban R, Hindmarsh AC. 2003 CVODES: an ODE solver with sensitivity analysis capabilities. Technical Report UCRL-JP-200039, Lawrence Livermore National Laboratory.
68. Goel NS, Maitra SC, Montroll EW. 1971 On the Volterra and other nonlinear models of interacting populations. *Rev. Mod. Phys.* **43**, 231. (doi:10.1103/RevModPhys.43.231)
69. Volterra V. 1927 Variazioni e fluttuazioni del numero d'individui in specie animali conviventi. *Memoria della Reale Accademia Nazionale dei Lincei* **2**, 31–113.
70. Lotka AJ. 2002 Contribution to the theory of periodic reactions. *J. Phys. Chem.* **14**, 271–274. (doi:10.1021/j150111a004)
71. Goodwin RM. 1982 A growth cycle. In *Essays in economic dynamics* (ed. RM Goodwin), pp. 165–170. Berlin, Germany: Springer.
72. Kingsland SE, Kingsland SE. 1995 *Modeling nature*. Chicago, IL: University of Chicago Press.
73. Thingstad TF. 2000 Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnol. Oceanogr.* **45**, 1320–1328. (doi:10.4319/lo.2000.45.6.1320)
74. Varon M, Zeigler BP. 1978 Bacterial predator-prey interaction at low prey density. *Appl. Environ. Microbiol.* **36**, 11–17. (doi:10.1128/aem.36.1.11-17.1978)
75. Carpenter B. 2018 Predator-prey population dynamics: the Lotka-Volterra model in Stan. See https://mc-stan.org/users/documentation/case-studies/lotka-volterra-predator-prey.html (accessed 28 August 2019).
76. Hewitt CG. 1921 *The conservation of the wild life of Canada*. New York, NY: C. Scribner.