Bayesian Solution Uncertainty Quantification for Differential Equations[*]

Oksana A. Chkrebtii[†], David A. Campbell[‡], Ben Calderhead[§], and Mark A. Girolami[¶]

Abstract. We explore probability modelling of discretization uncertainty for system states defined implicitly by ordinary or partial differential equations. Accounting for this uncertainty can avoid posterior under-coverage when likelihoods are constructed from a coarsely discretized approximation to system equations. A formalism is proposed for inferring a fixed but *a priori* unknown model trajectory through Bayesian updating of a prior process conditional on model information. A one-step-ahead sampling scheme for interrogating the model is described, its consistency and first order convergence properties are proved, and its computational complexity is shown to be proportional to that of numerical explicit one-step solvers. Examples illustrate the flexibility of this framework to deal with a wide variety of complex and large-scale systems. Within the calibration problem, discretization uncertainty defines a layer in the Bayesian hierarchy, and a Markov chain Monte Carlo algorithm that targets this posterior distribution is presented. This formalism is used for inference on the JAK-STAT delay differential equation model of protein dynamics from indirectly observed measurements. The discussion outlines implications for the new field of probabilistic numerics.

Keywords: Bayesian numerical analysis, uncertainty quantification, Gaussian processes, differential equation models, uncertainty in computer models.

1 Quantifying uncertainty for differential equation models

Many scientific, economic, and engineering disciplines represent the spatio-temporal evolution of complex systems implicitly as differential equations, using few but readily interpretable parameters. Differential equation models describe the natural dependence between system states and their rates of change on an open spatio-temporal domain, $\mathcal{D} \subset \mathbb{R}^d$. Their relationship, represented by the function F, is fully specified up to some physical constants θ belonging to a parameter space Θ. Mathematically, system states, $u : \mathcal{D} \times \Theta \to \mathbb{R}^p$ satisfy the partial differential equation (PDE),

$$F\left(x, t, \frac{\partial u}{\partial t}, \frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2}, \frac{\partial^2 u}{\partial x_1 \partial x_2}, \ldots, u, \theta\right) = 0, \quad (x, t) \in \mathcal{D} \cup \partial\mathcal{D},$$

[†]Department of Statistics, The Ohio State University, Columbus, OH, USA, oksana@stat.osu.edu
[‡]Department of Statistics & Actuarial Science, Simon Fraser University, Canada, dac5@stat.sfu.ca
[§]Department of Mathematics, Imperial College London, London, UK, b.calderhead@ucl.ac.uk
[¶]Department of Statistics, The University of Warwick, Coventry, UK, M.Girolami@warwick.ac.uk

and constraints at the boundary $\partial\mathcal{D}$ of $\mathcal{D}$. When states evolve with respect to a single variable, such as time, the model simplifies to an ordinary differential equation (ODE). The explicit solution, denoted $u^* : \mathcal{D} \times \Theta \to \mathbb{R}^p$, of a differential equation problem over spatio-temporal locations, $(x, t) \in \mathcal{D}$, is used to explore system dynamics, design experiments, or extrapolate to expensive or dangerous circumstances. Therefore, increasing attention is being paid to the challenges associated with quantifying uncertainty for systems defined by such models, and in particular those based on mathematical descriptions of complex phenomena such as the weather, ocean currents, ice sheet flow, and cellular protein transport (Ghanem and Spanos, 2003; Kaipio et al., 2004; Huttunen and Kaipio, 2007; Marzouk et al., 2007; Marzouk and Najm, 2009; Stuart, 2010).

The main challenge of working with differential equation models, from both mathematical and statistical perspectives, is that solutions are generally not available in closed form. When this occurs, prior exploration of the mathematical model, $u^*(x, t, \theta)$, cannot be performed directly. This issue is dealt with throughout the literature by replacing the exact solution with an N dimensional approximate solution, $\hat{u}^N(x, t, \theta)$, obtained using numerical techniques over a size N grid partitioning of the domain $\mathcal{D}$. Inference and prediction based on $\hat{u}^N(x, t, \theta)$ then inform the reasoning process when, for example, assessing financial risk in deciding on oil field bore configurations, or forming government policy in response to extreme weather events.

Limited computation and coarse mesh size are contributors to numerical error which, if assumed negligible can lead to serious misspecification for these highly nonlinear models. The study of how uncertainty propagates through a mathematical model is known as the *forward problem*. Numerical error analysis provides local and global discretization error bounds that are characterized point-wise and relate to the asymptotic behaviour of the deterministic approximation of the model. However, accounting for this type of *verification error* for the purpose of model inference has proven to be a difficult open problem (Oberkampf and Roy, 2010), causing discretization uncertainty to be often ignored in practice. For complex, large-scale models, maintaining a mesh size that is fine enough to ignore discretization uncertainty demands prohibitive computational cost (Arridge et al., 2006), so that a coarse mesh is the only feasible choice despite associated model uncertainty (Kim, J.-h. et al., 2014). Furthermore, some ODE and PDE models exhibit highly structured but seemingly unpredictable behaviour. Although the long term behaviour of such systems is entrenched in the initial states, the presence of numerical discretization error introduces small perturbations which eventually lead to exponential divergence from the exact solution. Consequently, information about initial states rapidly decays as system solution evolves in time (Berliner, 1991).

The *calibration* or *statistical inverse problem* concerns the uncertainty in unknown model inputs, θ, given measurements of the model states (Bock, 1983; Ionides et al., 2006; Dowd, 2007; Ramsay et al., 2007; Xue et al., 2010; Brunel, 2008; Liang and Wu, 2008; Calderhead and Girolami, 2011; Campbell and Steele, 2011; Gugushvili and Klaassen, 2012; Campbell and Chkrebtii, 2013; Xun et al., 2013). However, inference for differential equation models lacking a closed form solution has mainly proceeded under the unverified assumption that numerical error is negligible over the entire parameter space. From the perspective of Kennedy and O'Hagan (2001), the discrepancy between the numerical approximation, $\hat{u}^N$, and the unobservable mathematical

model, u^*, at the observation locations $(\mathbf{x}, \mathbf{t}) \in \mathcal{D}^T$ would be modelled by the function, $\delta(\mathbf{x}, \mathbf{t}) = u^*(\mathbf{x}, \mathbf{t}, \theta) - \hat{u}^N(\mathbf{x}, \mathbf{t}, \theta)$. Estimation of $\delta(\mathbf{x}, \mathbf{t})$ requires observations $y(\mathbf{x}, \mathbf{t})$ from $u^*(\mathbf{x}, \mathbf{t}, \theta)$. However, magnitude and structure of discretization error change with θ, while $\delta(\mathbf{x}, \mathbf{t})$ is agnostic to the specific form of the underlying mathematical model, treating it as a "black box". Consequently exploring the model uncertainty for given values of θ requires observations for each parameter setting or the assumption that $\delta(\mathbf{x}, \mathbf{t})$ is independent of θ. Alternatively, in the discussion of Kennedy and O'Hagan (2001), H. Wynn argues that the sensitivity equations and other underlying mathematical structures that govern the numerical approximation could be included in the overall uncertainty analysis.

1.1   Contributions and organization of the paper

This paper develops a Bayesian formalism for modelling discretization uncertainty within the forward problem, and explores its contribution to posterior uncertainty in the model parameters. We substantially expand and formalize ideas first described by Skilling (1991) and develop a general class of probability models for solution uncertainty for differential equations in the so-called explicit form,

$$\mathcal{A} = f(x, t, \mathcal{B}, \theta), \quad (x, t) \in \mathcal{D} \cup \partial\mathcal{D}, \tag{1}$$

where $\mathcal{A}$ and $\mathcal{B}$ are subsets of the deterministic solution $u = u(x, t, \theta)$ and its partial derivatives, and the function f is Lipschitz continuous in $\mathcal{B}$. Although (1) is very general, Sections 2 and 3 consider ordinary differential equations for expositional simplicity, where $\mathcal{A} = du(t, \theta)/dt := u_t$ and $\mathcal{B} = u(t, \theta) := u$ for $t \in [0, L], L > 0$, i.e. $u_t = f(t, u, \theta)$, with a given initial condition. Sections 4 and 5 extend this to the generality of (1), showcasing a wide class of ODEs and PDEs, such as mixed boundary value problems with multiple solutions, delay initial function problems, and chaotic differential equations.

Section 2 considers fixed model parameters θ, hyperparameters Ψ, and discretization grid of size N. Prior uncertainty about the solution u and its derivative u_t is encoded in a probability model $[u, u_t \mid \theta, \Psi]$ on a space of functions satisfying the initial condition. Section 3 describes a sequential model interrogation method that consists of (i) drawing a sample u^{n-1} from the marginal $[u \mid f_1, \ldots, f_{n-1}, \theta, \Psi]$, and (ii) using this to compute *model interrogations* $f_n \equiv f(t_n, u^{n-1}, \theta)$ at grid locations $t_n \in [0, L], L > 0$. Conditioning on these incrementally obtained interrogations defines the sequence of models $[u, u_t \mid f_1, \theta, \Psi], \ldots, [u, u_t \mid f_1, \ldots, f_N, \theta, \Psi]$. In Section 3.3 we marginalize over $\mathbf{f} = (f_1, \ldots, f_N)$ to obtain the probability model, or *probabilistic solution*,

$$[u \mid \theta, \Psi, N] = \int \int [u, u_t, \mathbf{f} \mid \theta, \Psi, N] \, du_t \, d\mathbf{f}, \tag{2}$$

which contracts to the true deterministic solution u as $N \to \infty$. Computational complexity is shown in Section 3.2 to be of the same order as comparably sampled explicit numerical solvers, and scalability is illustrated in Section 5 for the chaotic Navier–Stokes system involving the solution of over 16,000 coupled stiff ordinary differential equations. The probabilistic solution defines a trade-off between discretization uncer-
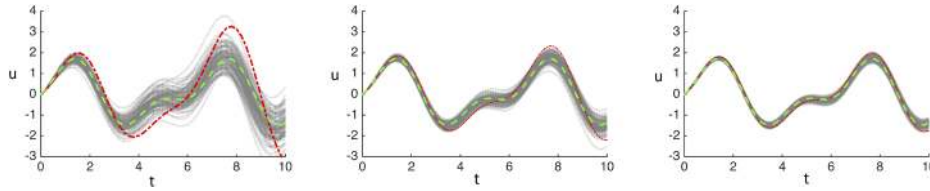
Figure 1: Exact solution (green line) of ODE (6); 100 draws (grey lines) from $p\{u(\mathbf{t}) \mid \Psi, N\}$ over a dense grid $\mathbf{t}$; first order explicit Euler approximation (red line), obtained using equally-spaced domain partitions of sizes $N = 50, 100, 200$ (left to right).

tainty and computational cost. To illustrate this point, Figure 1 compares the exact, but *a priori* unknown solution with sample trajectories from the proposed probability model (2) and a first order numerical scheme obtained over progressively refined discretization grids. The exact solution lies within a high posterior density region of (2), and as the grid is refined yielding more information about the model, knowledge of the solution increases, concentrating to the exact solution. Strong posterior non-stationarity of the error structure (2) is illustrated in Figure 2 for the three states of a system whose trajectory is restricted around a chaotic attractor (details are provided in Section 5.1).

In Section 4 we embed discretization uncertainty in the inferential hierarchy,

$$[\theta, u, \Psi, \Sigma \mid \mathbf{y}, N] \; \propto \; \underbrace{[\mathbf{y} \mid u, \theta, \Sigma]}_{\text{likelihood}} \quad \underbrace{[u \mid \theta, \Psi, N]}_{\text{solution uncertainty}} \quad \underbrace{[\theta, \Psi, \Sigma]}_{\text{prior}}, \tag{3}$$

where Σ includes parameters of the observation process, such as the covariance structure of the error terms. We adopt a forward simulation approach for the middle layer within a Markov chain Monte Carlo (MCMC) algorithm to sample from this posterior distribution. The importance of explicitly modelling discretization uncertainty within the inverse problem is illustrated in Section 4.1 by the nonlinear posterior dependence between solver hyperparameters Ψ and differential equation model parameters θ. This approach provides a useful exploratory tool to diagnose the presence of discretization effects. The problem of posterior under-coverage resulting from ignoring discretization induced solution uncertainty in (3) is illustrated in Section 5.4.

Discussion of various extensions and implications for the emerging field of probabilistic numerics is provided in Section 6. Algorithmic and theoretical details are described in the Supplement (Chkrebtii et al., 2016). MATLAB implementations of all examples are provided at `github.com/ochkrebtii/uqdes`.

1.2 Notation

We use of the mathematical framework for defining numerical problems on function spaces developed through the work of Stuart (2010), Kaipio and Somersalo (2007), O'Hagan (1992), and Skilling (1991). We denote partial derivatives using subscripts to indicate the variable with respect to which a partial derivative is taken. For example
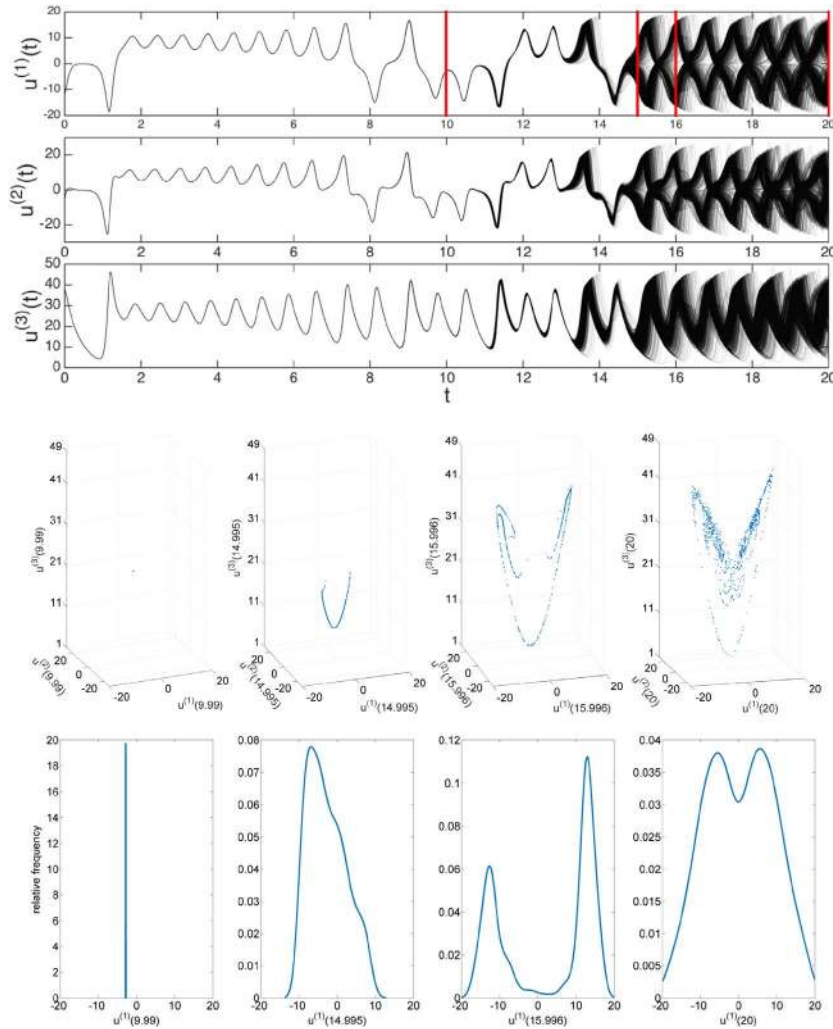
Figure 2: 1000 draws from (2) for the Lorenz63 model with fixed initial states and model parameters in the chaotic regime. Red vertical lines in $u^{(1)}(t)$ correspond to four time points used in the diagrams in the last two rows.

$u_{xx} := \frac{\partial^2}{\partial x^2} u$ refers to the second partial derivative of u with respect to x. Superscripts in parentheses denote elements of a vector, for example, $u^{(k)}$ is the kth element of the P-dimensional vector u. The methodology proposed in this paper is based on a sequentially updated probabilistic solution, consequently we use superscripts without parentheses, e.g., m^k, to refer to k-times updated functions. Superscripts on random variables, e.g., u^k, represent a sample from its k-times updated distribution. For notational simplicity, we omit the dependence of the differential equation solution u, and derivative u_t on θ, Ψ, N, and x and/or t when the context makes this dependence clear.

2 Prior model for a fixed but unknown solution

This section is concerned with probability modelling of initial beliefs about the unknown function, $u^* : \mathcal{D} \times \Theta \to \mathbb{R}^p$, satisfying an ODE initial value problem of the form,

$$u_t = f(t, u, \theta), \quad t \in [0, L], L > 0, \tag{4}$$

with fixed initial value $u(0) = u^*(0)$. We describe how to incorporate boundary and regularity assumptions into a flexible Gaussian process (GP) prior model (Rasmussen and Williams, 2006; Stuart, 2010) for the unknown solution. Prior beliefs are updated in Section 3 by conditioning on information about the differential equation model obtained over a size N partition of the domain $\mathcal{D}$ according to a chosen sampling scheme.

2.1 GP prior on the solution of an ODE initial value problem

GPs are a flexible class of stochastic processes that can be defined jointly with their derivatives in a consistent way through the specification of an appropriate covariance operator (Solak et al., 2003). We model uncertainty jointly for the time derivative in (4) and its solution, (u_t, u), using a GP prior with covariance structure, (C_t^0, C^0), and mean function, (m_t^0, m^0). The marginal means must satisfy the constraint $\int_0^t m_t^0(z)dz = m^0(t)$. The covariance structure is defined as a flexible convolution of kernels (see, e.g., Higdon, 1998, 2002). A covariance operator for the derivative u_t evaluated between locations t_j and t_k is obtained by convolving a square integrable kernel $R_\lambda : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ with itself,

$$C_t^0(t_j, t_k) = \alpha^{-1} \int_{\mathbb{R}} R_\lambda(t_j, z) R_\lambda(t_k, z)dz,$$

parameterized by the length-scale $\lambda > 0$ and prior precision $\alpha > 0$, to be discussed in Section 3.4. An example is the infinitely differentiable kernel, $R_\lambda(t_j, t_k) = \exp\{-(t_j - t_k)^2/2\lambda^2\}$, which decreases exponentially with squared distance between t_j and t_k weighted by the length-scale λ. Its convolution yields the *squared exponential covariance* $C_t^0(t_j, t_k) = \sqrt{\pi}\alpha^{-1}\lambda \exp\{-(t_j - t_k)^2/4\lambda^2\}$. An example of a kernel with bounded support is $R_\lambda(t_j, t_k) = 1_{(t_j - \lambda, t_j + \lambda)}(t_k)$, which yields the piecewise linear *uniform covariance* $C_t^0(t_j, t_k) = \{\min(t_j, t_k) - \max(t_j, t_k) + 2\lambda\} \, 1_{(0,\infty)}\{\min(t_j, t_k) - \max(t_j, t_k) + 2\lambda\}$. Details are provided in Supplement D.4. A marginal covariance operator, C^0, on the state is obtained by integrating C_t^0 with respect to evaluation points t_j and t_k or, equivalently, by convolving the integrated kernel $Q_\lambda(t_j, t_k) = \int_0^{t_j} R_\lambda(z, t_k)dz$ with itself:

$$C^0(t_j, t_k) = \alpha^{-1} \int_{\mathbb{R}} Q_\lambda(t_j, z) Q_\lambda(t_k, z)dz.$$

Defining the covariance over the state by integration ensures that $C^0(0, 0) = 0$, which enforces the boundary condition, $u(0) = m^0(0)$ (see Supplement D.1; an alternative way to enforce the boundary constraint is to obtain the derivative process u_t by differentiation of u and condition the joint prior for (u_t, u) on the exact initial state, $u^*(0)$). This introduces anisotropy over the state that is consistent with exact knowledge of the solution at the initial boundary and increasing uncertainty thereafter. The cross-covariance

terms are defined analogously as $\int_0^{t_j} C_t^0(z, t_k)\mathrm{d}z = \alpha^{-1} \int_{\mathbb{R}} Q_\lambda(t_j, z) R_\lambda(t_k, z)\mathrm{d}z$ and $\int_0^{t_k} C_t^0(t_j, z)\mathrm{d}z = \alpha^{-1} \int_{\mathbb{R}} R_\lambda(t_j, z) Q_\lambda(t_k, z)\mathrm{d}z$, where each is the adjoint of the other.

The initial joint GP prior probability distribution for the solution at a vector of evaluation times $\mathbf{t_k}$ and its time derivative at a possibly different vector of evaluation times $\mathbf{t_j}$, given fixed hyperparameters $\Psi = (m_t^0, m^0, \alpha, \lambda)$ is,

$$\begin{bmatrix} u_t(\mathbf{t_j}) \\ u(\mathbf{t_k}) \end{bmatrix} \sim \mathcal{GP}\left(\begin{bmatrix} m_t^0(\mathbf{t_j}) \\ m^0(\mathbf{t_k}) \end{bmatrix}, \begin{bmatrix} C_t^0(\mathbf{t_j}, \mathbf{t_j}) & \int_0^{\mathbf{t_k}} C_t^0(\mathbf{t_j}, \mathbf{z})\mathrm{d}\mathbf{z} \\ \int_0^{\mathbf{t_k}} C_t^0(\mathbf{z}, \mathbf{t_j})\mathrm{d}\mathbf{z} & C^0(\mathbf{t_k}, \mathbf{t_k}) \end{bmatrix} \right). \tag{5}$$

This prior modelling strategy is straightforwardly generalized to explicit initial value problems of the form (1) by defining the prior jointly on the state and the required higher order derivatives.

As a simple example, consider the second order initial value ODE problem,

$$\begin{cases} u_{tt}(t) = \sin(2t) - u, & t \in [0, 10], \\ u_t(0) = 0, \ u(0) = -1. \end{cases} \tag{6}$$

Its exact solution, $u^*(t) = \{-4\cos(t) + 2\sin(t) - \sin(2t) + \cos(t)\}/(4-1)$, is assumed unknown *a priori*. The first column of Figure 3 shows five draws from the marginal densities of the prior $[u_t(\mathbf{t}), u(\mathbf{t})]$ defined over a fine grid $\mathbf{t} \in [0, 10]^T$ using a squared exponential covariance with $(\lambda, \alpha) = (0.8, 5)$. Section 3 describes a sequential updating procedure of the prior model to obtain the probabilistic solution for this example.

3 Updating prior beliefs for fixed but unknown ODE solution

The prior model (5) on the unknown solution and its derivative will be updated conditionally on model information collected over the ordered partition $\mathbf{s} = (s_1, \ldots, s_N)$ of the domain $[0, L]$. Define a *model interrogation* at time s_n as the output $\mathrm{f}_n \equiv f\{s_n, u^{n-1}(s_n), \theta\}$ from the right hand side of equation (4) evaluated at a sample $u^{n-1}(s_n)$ from the marginal prior $[u(s_n) \mid \mathrm{f}_{n-1}]$. Model interrogations are generated sequentially based on state samples drawn from progressively refined prior models in order to accommodate fast changing dynamics. Because the ODE model (4) defines the solution implicitly, we are unable to measure the exact state or derivative, and therefore both the model interrogations and their error structure will take into account increasing uncertainty with time.

This section begins by expanding on these function estimation steps with full details of the Bayesian updating algorithm. We illustrate the forward solution algorithm on the simple example (6) with known closed form solutions before considering more realistic systems. Section 3.2 examines modelling choices for which this procedure can attain computational complexity of $O(N)$, while Section 3.3 establishes the consistency and first order convergence properties of this approach. Choice of hyperparameters is addressed in Section 3.4. Specific extensions to the full generality of (1) are presented in Sections 4 and 5.
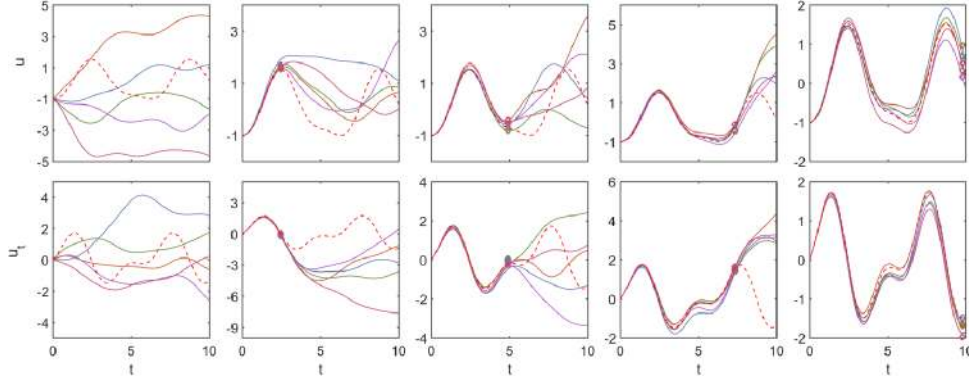
Figure 3: Five draws (solid lines) from the marginal densities of (2) over the state (first row) and its derivative (second row) for ODE (6) after $n = 0, 12, 24, 36, 48$ iterations (from left to right) of Algorithm 1 with discretization grid of size $N = 50$; step-ahead samples u^n and f_n are shown as circles; exact solution is shown as a dotted red line.

3.1   Model updating based on interrogations from sampled states

This section describes a sequential sampling scheme to interrogate the model after each update. Model interrogations occur at pre-determined discretization grid locations $\mathbf{s} = (s_1, \ldots, s_N)$, while the resulting sequence of posterior solution models can be computed at a possibly different vector of time points, $\mathbf{t}$. There are many ways to generalize this basic one-step-ahead sampling scheme, as discussed in Section 6.

First update

The derivative of the exact solution on the boundary, $s_1 = 0$, can be obtained exactly by evaluating the right hand side of (4) at the known initial condition $u^*(s_1)$ as,

$$f_1 \equiv f\{s_1, u^*(s_1), \theta\} = u_t^*(s_1).$$

The first update of the prior probability model (5) is performed by conditioning on the exact derivative f_1. This yields the joint conditional predictive probability distribution at evaluation times $\mathbf{t_j}$ and $\mathbf{t_k}$,

$$\begin{bmatrix} u_t(\mathbf{t_j}) \\ u(\mathbf{t_k}) \end{bmatrix} \Bigg| \ f_1 \sim \mathcal{GP}\left(\begin{bmatrix} m_t^1(\mathbf{t_j}) \\ m^1(\mathbf{t_k}) \end{bmatrix}, \begin{bmatrix} C_t^1(\mathbf{t_j}, \mathbf{t_j}) & \int_0^{\mathbf{t_k}} C_t^1(\mathbf{t_j}, \mathbf{z}) \mathrm{d}\mathbf{z} \\ \int_0^{\mathbf{t_k}} C_t^1(\mathbf{z}, \mathbf{t_j}) \mathrm{d}\mathbf{z} & C^1(\mathbf{t_k}, \mathbf{t_k}) \end{bmatrix} \right), \qquad (7)$$

where marginal means and covariances at the vectors of time $\mathbf{t_j}$ and $\mathbf{t_k}$ are given by:

$$m_t^1(\mathbf{t_j}) = m_t^0(\mathbf{t_j}) + C_t^0(\mathbf{t_j}, s_1) C_t^0(s_1, s_1)^{-1} \left\{ f_1 - m_t^0(s_1) \right\},$$

$$m^1(\mathbf{t_k}) = m^0(\mathbf{t_k}) + C_t^0(s_1, s_1)^{-1} \int\limits_0^{\mathbf{t_k}} C_t^o(\mathbf{z}, s_1) \mathrm{d}\mathbf{z} \left\{ f_1 - m_t^0(s_1) \right\},$$

$$C_t^1(\mathbf{t_j}, \mathbf{t_j}) = C_t^0(\mathbf{t_j}, \mathbf{t_j}) - C_t^0(\mathbf{t_j}, s_1) C_t^0(s_1, s_1)^{-1} C_t^0(s_1, \mathbf{t_j}),$$

$$C^1(\mathbf{t_k}, \mathbf{t_k}) = C^0(\mathbf{t_k}, \mathbf{t_k}) - \left\{ \int_0^{\mathbf{t_k}} C_t^0(\mathbf{z}, s_1) \mathrm{d}\mathbf{z} \right\} C_t^0(s_1, s_1)^{-1} \left\{ \int_0^{\mathbf{t_k}} C_t^0(\mathbf{z}, s_1) \mathrm{d}\mathbf{z} \right\}^\top.$$

Closed forms for means and covariances are obtained by integration (along with cross-covariances, Supplement D.1). We refer to this predictive probability measure as the *first update* of the algorithm and use it as the prior for the next iteration of the algorithm.

Second update

The derivative of the exact solution is unknown beyond $s_1 = 0$. We therefore introduce the random variable f_2, representing a model interrogation at the subsequent time step, $s_2 > 0$. A realization of f_2 is obtained by drawing a sample $u^1(s_2)$ at time s_2 from the marginal predictive distribution of (7),

$$u(s_2) \mid \mathrm{f}_1 \sim \mathcal{N}\left(m^1(s_2), C^1(s_2, s_2)\right),$$

and then applying the transformation from the right hand side of equation (4),

$$\mathrm{f}_2 \equiv f\left\{s_2, u^1(s_2), \theta\right\}.$$

The uncertainty in $u(s_2) \mid \mathrm{f}_1$ means that f_2 is not guaranteed to equal to the exact derivative at time s_2 (the function f is a differential operator only for the input $u = u^*$). Indeed, without the exact solution, the probability distribution of f_2 is unavailable. As with any model for a data generating process, the likelihood (error model) should be elicited on a case-by-case basis. Here we provide general guidance on constructing this likelihood in the absence of additional expert information about the error model for f_2, and all subsequent interrogations f_n. In general, we expect that (i) the discrepancy between $u_t(t)$ and $f\{t, u(t), \theta\}, t \in [0, L]$ is zero when u is the exact solution; and (ii) the function f is smooth in the second argument, therefore discrepancy decreases smoothly with decreasing uncertainty in the derivative of the sampled state realization $u^1(s_2)$. These facts suggest a spherically symmetric error model for f_2 with a single mode at $\mathrm{f}_2 = u_t(s_2)$ and dispersed as an increasing function of distance from $s_1 = 0$. For example, we may consider the Gaussian error model,

$$\mathrm{f}_2 \mid u_t, \mathrm{f}_1 \sim \mathcal{N}\left(u_t(s_2), r_1(s_2)\right), \tag{8}$$

with mean $u_t(s_2)$ and variance $r_1(s_2) = C_t^1(s_2, s_2)$. A degenerate special case of this error model is discussed at the end of this section. We may now update (7) by conditioning on the model interrogation f_2 under the error model (8) to obtain the predictive probability distribution at the vectors of evaluation times $\mathbf{t_j}$ and $\mathbf{t_k}$,

$$\begin{bmatrix} u_t(\mathbf{t_j}) \\ u(\mathbf{t_k}) \end{bmatrix} \middle| \mathrm{f}_2, \mathrm{f}_1 \sim \mathcal{GP}\left(\begin{bmatrix} m_t^2(\mathbf{t_j}) \\ m^2(\mathbf{t_k}) \end{bmatrix}, \begin{bmatrix} C_t^2(\mathbf{t_j}, \mathbf{t_k}) & \int_0^{\mathbf{t_k}} C_t^2(\mathbf{t_j}, \mathbf{z}) \mathrm{d}\mathbf{z} \\ \int_0^{\mathbf{t_j}} C_t^2(\mathbf{z}, \mathbf{t_k}) \mathrm{d}\mathbf{z} & C^2(\mathbf{t_k}, \mathbf{t_k}) \end{bmatrix} \right). \tag{9}$$

Defining $g_2 = C_t^1(s_2, s_2) + r_1(s_2)$, the marginal means and covariances at the vectors of evaluation times $\mathbf{t_j}$ and $\mathbf{t_k}$ are given by,

$$m_t^2(\mathbf{t_j}) = m_t^1(\mathbf{t_j}) + C_t^1(\mathbf{t_j}, s_2) g_2^{-1}\left\{\mathrm{f}_2 - m_t^1(s_2)\right\},$$

$$m^2(\mathbf{t_k}) = m^1(\mathbf{t_k}) + \int\limits_0^{\mathbf{t_k}} C_t^1(\mathbf{z}, s_n)\mathrm{d}\mathbf{z}\, g_2^{-1}\left\{\mathrm{f}_2 - m_t^1(s_2)\right\},$$

$$C_t^2(\mathbf{t_j}, \mathbf{t_j}) = C_t^1(\mathbf{t_j}, \mathbf{t_j}) - C_t^1(\mathbf{t_j}, s_2)\, g_2^{-1} C_t^1(s_2, \mathbf{t_j}),$$

$$C^2(\mathbf{t_k}, \mathbf{t_k}) = C^1(\mathbf{t_k}, \mathbf{t_k}) - \left\{\int\limits_0^{\mathbf{t_k}} C_t^1(\mathbf{z}, s_2)\mathrm{d}\mathbf{z}\right\} g_2^{-1} \left\{\int\limits_0^{\mathbf{t_k}} C_t^1(\mathbf{z}, s_2)\mathrm{d}\mathbf{z}\right\}^{\top},$$

where state mean, covariance, and derivative cross-covariances can be obtained analytically via integration (see Supplement D.1).

Subsequent updates

Subsequent updates, $2 < n \leq N$, begin by drawing a sample $u^{n-1}(s_n)$ at time s_n (illustrated by circles in the top row of Figure 3) from the marginal predictive posterior $[u(s_n) \mid \mathrm{f}_{n-1}, \ldots, \mathrm{f}_1]$, whose mean and covariance are given in Algorithm 1. A realization of the model interrogation f_n is constructed by applying the model transformation (4) to $u^{n-1}(s_n)$. The update is performed relative to the prior $[u_t, u \mid \mathrm{f}_{n-1}, \ldots, \mathrm{f}_1]$, by assimilating f_n under the Gaussian error model,

$$\mathrm{f}_n \mid u_t(s_n), \mathrm{f}_{n-1}, \ldots, \mathrm{f}_1 \sim \mathcal{N}\left\{u_t(s_n), r_{n-1}(s_n)\right\}, \quad 1 \leq n \leq N, \tag{10}$$

where, e.g., $r_{n-1}(s_n) = C_t^{n-1}(s_n, s_n)$, to obtain the predictive probability distribution at the vectors of evaluation times $\mathbf{t_j}$ and $\mathbf{t_k}$,

$$\begin{bmatrix} u_t(\mathbf{t_j}) \\ u(\mathbf{t_k}) \end{bmatrix} \middle| \mathrm{f}_n, \ldots, \mathrm{f}_1 \end{bmatrix} \sim \mathcal{GP}\left(\begin{bmatrix} m_t^n(\mathbf{t_j}) \\ m^n(\mathbf{t_k}) \end{bmatrix}, \begin{bmatrix} C_t^n(\mathbf{t_j}, \mathbf{t_j}) & \int_0^{\mathbf{t_k}} C_t^n(\mathbf{t_j}, \mathbf{z})\mathrm{d}\mathbf{z} \\ \int_0^{\mathbf{t_j}} C_t^n(\mathbf{z}, \mathbf{t_k})\mathrm{d}\mathbf{z} & C^n(\mathbf{t_k}, \mathbf{t_k}) \end{bmatrix}\right). \tag{11}$$

Defining $g_n = C_t^{n-1}(s_n, s_n) + r_{n-1}(s_n)$, the resulting joint predictive probability distribution is Gaussian with marginal means and covariances:

$$m_t^n(\mathbf{t_j}) = m_t^{n-1}(\mathbf{t_j}) + C_t^{n-1}(\mathbf{t_j}, s_n)\, g_n^{-1}\left\{\mathrm{f}_n - m_t^{n-1}(s_n)\right\},$$

$$m^n(\mathbf{t_k}) = m^{n-1}(\mathbf{t_k}) + g_n^{-1}\int\limits_0^{\mathbf{t_k}} C_t^{n-1}(\mathbf{z}, s_n)\mathrm{d}\mathbf{z}\left\{\mathrm{f}_n - m_t^{n-1}(s_n)\right\},$$

$$C_t^n(\mathbf{t_j}, \mathbf{t_k}) = C_t^{n-1}(\mathbf{t_j}, \mathbf{t_k}) - C_t^{n-1}(\mathbf{t_j}, s_n)\, g_n^{-1} C_t^{n-1}(s_n, \mathbf{t_k}),$$

$$C^n(\mathbf{t_j}, \mathbf{t_k}) = C^{n-1}(\mathbf{t_j}, \mathbf{t_k}) - g_n^{-1}\int\limits_0^{\mathbf{t_j}} C_t^{n-1}(\mathbf{z}, s_n)\mathrm{d}\mathbf{z}\left\{\int\limits_0^{\mathbf{t_k}} C_t^{n-1}(\mathbf{z}, s_n)\mathrm{d}\mathbf{z}\right\}^{\top},$$

and with state mean, covariances, and derivative cross-covariances again obtained via integration (see Algorithm 1 and Supplement D.1). Here the state and derivative, u and u_t respectively, may be of arbitrary dimension. Under a bounded support covariance, $C^n(\mathbf{t_j}, \mathbf{t_k})$ and $C_t^n(\mathbf{t_j}, \mathbf{t_k})$ become sparse, band diagonal matrices.

After N updates have been performed, we obtain a joint posterior probability distribution $[u_t, u \mid \mathbf{f}]$ for the unknown solution and its derivative, conditional on realizations of one trajectory of the model interrogation vector $\mathbf{f}$. Draws from the marginal model (2) can be obtained via Monte Carlo, as described in Algorithm 1, yielding realized

trajectories from the forward model (2) at desired time locations **t**. Five draws from the joint distribution of (2) and the interrogations f_n at iterations $n = 12, 24, 36, 48$ are illustrated in the second through fifth columns of Figure 3 for the second-order ODE initial value problem (6). For notational simplicity, we will hereafter assume that the temporal locations of interest, **t**, form a subset of the discretization grid locations, **s**.

Point mass likelihood on model interrogations

A special case of this procedure models each interrogation, f_n, as the exact unknown derivative, $u_t^*(s_j)$, regardless of grid size, N. Thus, model interrogations are interpolated in the derivative space by replacing (8) and (10) with the point mass density:

$$P\left(u_t(s_n) = f_n\right) = 1, \quad 1 \leq n \leq N.$$

Because sampled process realizations can still be inconsistent with the solution, computational implementation of this alternative scheme often requires using an arbitrary covariance nugget increasing with time, or the use of a rougher covariance function than the smoothness implied by the differential equation model. For these reasons we will restrict our attention to explicitly modelling the interrogation error, e.g. (8) and (10).

3.2   Computational complexity

A reasonable modelling assumption for the state is that temporal correlation decays to zero with distance. Indeed, most deterministic numerical methods make the analogous assumption by defining a lag of $M \geq 1$ time steps, beyond which the solution has no direct effect. This Markovian assumption both adds flexibility for modelling very fast-changing dynamics and drastically reduces computational cost. Under a compactly supported covariance structure, Algorithm 1 requires only $O(N)$ operations by allowing truncation of the weight matrices employed in the mean and covariance updates. In contrast, when employing a covariance structure with unbounded support, Algorithm 1 attains computational complexity proportional to the cube of the number of time steps, since the number of operations can be written as a finite sum over $1 \leq n \leq N$ with each iteration requiring an order $O(n^2)$ operations.

A computational comparison of Algorithm 1 with a similarly sampled explicit first order solver is provided in the Supplement C. Although probability models for the solution attain the same computational scaling, the probabilistic method is more expensive. However, the model of solution uncertainty makes up for its added computational cost in practical applications where discretization grids cannot be refined enough to ensure negligible discretization effects. Within the inverse problem, the adaptive estimation of length-scale can substantially increase solver accuracy (see Figure 10 in Supplement C).

3.3   Convergence

We now consider the rate at which the stochastic process with probability distribution (2) generated via Algorithm 1 concentrates around the exact solution of (4). The rate

---

Algorithm 1 For the initial value problem (4), draw one sample from the forward model (2) over $\mathbf{t} = (t_1, \cdots, t_T)$, given Ψ and discretization grid $\mathbf{s} = (s_1, \cdots, s_N)$.

At time $s_1 = 0$ initialize the derivative $\mathrm{f}_1 = f\{s_1, u^*(0), \theta\}$, and define m^0, m_t^0, C^0, C_t^0 as above;

for $n = 1 : N$ **do**

  If $n = 1$, set $g_1 = C_t^0(s_1, s_1)$, otherwise set $g_n = C_t^{n-1}(s_n, s_n) + r_{n-1}(s_n)$;

  Compute for each modelled system component,

$$m^n(\mathbf{s}) = m^{n-1}(\mathbf{s}) + g_n^{-1} \int_0^{\mathbf{s}} C_t^{n-1}(\mathbf{z}, s_n) \mathrm{d}\mathbf{z} \left\{ \mathrm{f}_n - m_t^{n-1}(s_n) \right\},$$

$$m_t^n(\mathbf{s}) = m_t^{n-1}(\mathbf{s}) + g_n^{-1} C_t^{n-1}(\mathbf{s}, s_n) \left\{ \mathrm{f}_n - m_t^{n-1}(s_n) \right\},$$

$$C^n(\mathbf{s}, \mathbf{s}) = C^{n-1}(\mathbf{s}, \mathbf{s}) - g_n^{-1} \int_0^{\mathbf{s}} C_t^{n-1}(\mathbf{z}, s_n) \mathrm{d}\mathbf{z} \left\{ \int_0^{\mathbf{s}} C_t^{n-1}(\mathbf{z}, s_n) \mathrm{d}\mathbf{z} \right\}^\top,$$

$$C_t^n(\mathbf{s}, \mathbf{s}) = C_t^{n-1}(\mathbf{s}, \mathbf{s}) - g_n^{-1} C_t^{n-1}(\mathbf{s}, s_n) C_t^{n-1}(s_n, \mathbf{s}),$$

$$\int_0^{\mathbf{s}} C_t^n(\mathbf{z}, \mathbf{s}) \mathrm{d}\mathbf{z} = \int_0^{\mathbf{s}} C_t^{n-1}(\mathbf{z}, \mathbf{s}) \mathrm{d}\mathbf{z} - g_n^{-1} \int_0^{\mathbf{s}} C_t^{n-1}(\mathbf{z}, s_n) \mathrm{d}\mathbf{z} \, C_t^{n-1}(s_n, \mathbf{s});$$

  if $n < N$ **then**

    Sample one-step-ahead realization $u^n(s_{n+1})$ from the predictive distribution on the state for each modelled system component,

$$u^n(s_{n+1}) \sim p\left\{ u^n(s_{n+1}) \mid \mathrm{f}_n, \ldots, \mathrm{f}_1 \right\} = \mathcal{N}\left\{ u^n(s_{n+1}) \mid m^n(s_{n+1}), C^n(s_{n+1}, s_{n+1}) \right\},$$

    and interrogate the model by computing $\mathrm{f}_{n+1} = f\{s_{n+1}, u^n(s_{n+1}), \theta\}$;

  end if

end for

Sample and return $\mathbf{u} = (u^N(t_1), \cdots, u^N(t_T)) \sim \mathcal{GP}\{m^N(\mathbf{t}), C^N(\mathbf{t}, \mathbf{t})\}$, where $\mathbf{t} \subset \mathbf{s}$.

---

of convergence is as fast as the reduction of local truncation error for the analogously sampled one-step explicit Euler method. Proof is provided in the Supplement D.2.

Theorem 1. *Consider a function $f(t, u(t)) : [0, L] \times \mathbb{R} \to \mathbb{R}$ that is continuous in the first argument and Lipschitz continuous in the second argument. The stochastic process with probability distribution (2) obtained using Algorithm 1 with $r_n(s) = 0, s \in [0, L]$ and a covariance kernel satisfying conditions (D.6) and (D.7) in Supplement D.3, converges in L^1 to the unique solution satisfying (4) as the maximum discretization grid step size, h, the length-scale, λ, and the prior variance, α^{-1}, tend to zero. The rate of convergence is $O(h)$, proportional to the decrease in the maximum step size, h.*

3.4 The choice of hyperparameters

Analogously with choosing a numerical solver, the choice of probabilistic solver hyperparameters must be based on prior knowledge about the solution, such as its smoothness.

Moreover, as with many nonparametric function estimation problems, asymptotic results on convergence may be used to guide our choice of covariance hyperparameters within the forward problem. For example, for the explicit first order sampling scheme, the result of Theorem 1 suggests setting the prior variance, $1/\alpha$, and the length-scale, λ, proportionally to the maximum distance, h, between subsequent discretization grid locations. When experimental data is available, it also becomes possible to obtain marginal posterior densities for the hyperparameters as part of the calibration problem (see Section 4), allowing data driven adaptation over the space of model parameters. In contrast, deterministic numerical differential equation solvers effectively fix the analogues of these hyperparameters, such as the quadrature rule or the step number, both within the forward and inverse problems.

4   Exact Bayesian posterior model inference

The discretization uncertainty model in Section 3 may now be incorporated into the calibration problem. We wish to infer model parameters θ defining states $\mathbf{u} = (u(t_1), \ldots, u(t_T))^\top$ from measurement data $\mathbf{y} = (y(t_1), \ldots, y(t_T))^\top$, via the hierarchical model (3),

$$
\begin{aligned}
{[\mathbf{y} \mid \mathbf{u}, \theta, \Sigma]} &\propto \rho\left\{\mathbf{y} - H\left(\mathbf{u}, \theta\right)\right\}, \\
{[\mathbf{u} \mid \theta, \Psi, N]} &= p\left(\mathbf{u} \mid \theta, \Psi, N\right), \\
{[\theta, \Psi, \Sigma]} &= \pi(\theta, \Psi, \Sigma),
\end{aligned}
$$

where the transformation H maps the differential equation solution to the observation space, ρ is a probability density, π is a prior density, Σ defines the observation process, such as the error structure, and θ may include unknown initial conditions or other model components. We describe a Markov chain Monte Carlo procedure targeting the joint posterior of the state and unknown parameters conditional on a vector of noisy observations from $[\mathbf{y} \mid \mathbf{u}, \theta, \Sigma]$. Section 4.1 describes a case where states are indirectly observed through a nonlinear transformation H.

Algorithm 2 targets the posterior distribution (3) by generating forward model proposals via conditional simulation from Algorithm 1 (Supplement B provides a parallel tempering implementation). Within Algorithm 2, proposed sample paths are conditionally simulated from $p(\mathbf{u} \mid \theta, \Psi) = \int p(\mathbf{u}, \mathbf{f} \mid \theta, \Psi)\mathrm{d}\mathbf{f}$, removing the dependence of the acceptance probability on this intractable marginal density. Such partially likelihood-free Markov chain Monte Carlo methods (Marjoram et al., 2003) are widely used for inference on stochastic differential equations (e.g., Golightly and Wilkinson, 2011).

As with numerical based approaches, computation of the forward model is the rate limiting step in Algorithm 2. Computational savings can be obtained by exploiting the structure of the probabilistic solver, such as by targeting progressively refined posterior models in a Sequential Monte Carlo framework, or using ensemble Markov chain Monte Carlo methods (Neal, 2011) when mixture components of (2) can be quickly computed. We now have the components for accounting for discretization uncertainty in differential equation models of natural and physical systems. The remainder presents a calibration problem illustrating posterior sensitivity to discretization uncertainty.

Algorithm 2 Draw K samples from posterior $p(\theta, \mathbf{u} \mid \mathbf{y}, \alpha, \lambda, \Sigma)$ given observations of the transformed solution of a differential equation with unknown parameters θ.

Initialize $\theta, \alpha, \lambda \sim \pi(\cdot)$ where π is the prior density;
Using a probabilistic solver (e.g., Algorithm 1) simulate a vector of realizations, $\mathbf{u}$ conditional on θ, α, λ, over the size N discretization grid;
for $k = 1 : K$ **do**
    Propose $\theta', \alpha', \lambda' \sim q(\cdot \mid \theta, \alpha, \lambda)$ where q is a proposal density;
    Using a probabilistic solver (e.g., Algorithm 1) simulate a vector of realizations, $\mathbf{u}'$ conditional on $\theta', \alpha', \lambda'$, over the size N discretization grid;
    Compute the rejection ratio,

$$\rho = \frac{p(\mathbf{y} \mid \mathbf{u}', \theta', \Sigma)}{p(\mathbf{y} \mid \mathbf{u}, \theta, \Sigma)} \frac{\pi(\theta', \alpha', \lambda')}{\pi(\theta, \alpha, \lambda)} \frac{q(\theta, \alpha, \lambda \mid \theta', \alpha', \lambda')}{q(\theta', \alpha', \lambda' \mid \theta, \alpha, \lambda)},$$

    where the ratio of marginal densities, $p(\mathbf{u} \mid \mathbf{f}, \theta, \alpha, \lambda)$ to $p(\mathbf{u}' \mid \mathbf{f}', \theta', \alpha', \lambda')$, is omitted because the realizations $\mathbf{u}$ and $\mathbf{u}'$ are simulated directly from these densities;
    Sample $v \sim \mathrm{U}[0, 1]$
    if $v < \min(\rho, 1)$ **then**
        Update $(\theta, \alpha, \lambda) \leftarrow (\theta', \alpha', \lambda')$;
        Update $\mathbf{u} \leftarrow \mathbf{u}'$;
    end if
    Return $(\theta, \alpha, \lambda, \mathbf{u})$.
end for

4.1   Inference on a delay initial function model

The JAK-STAT mechanism describes a series of reversible biochemical reactions of STAT-5 transcription factors initiated by binding of the Erythropoietin (Epo) hormone to cell surface receptors (Pellegrini and Dusanter-Fourt, 1997). After gene activation occurs within the nucleus, the transcription factors revert to their initial state, returning to the cytoplasm to be used in the next activation cycle. This last stage is not well understood and is replaced by the unknown time delay, $\tau > 0$. One model describes changes in 4 reaction states of STAT-5 via the nonlinear delay initial function problem,

$$\begin{cases} u_t^{(1)}(t) &= -\ k_1\, u^{(1)}(t)\, \mathrm{Epo}R_A(t) + \ 2\, k_4\, u^{(4)}(t - \tau), & t \in [0, 60], \\ u_t^{(2)}(t) &= \ k_1\, u^{(1)}(t)\, \mathrm{Epo}R_A(t) - k_2\, u^{(2)^2}(t), & t \in [0, 60], \\ u_t^{(3)}(t) &= -\ k_3\, u^{(3)}(t) + \frac{1}{2}\, k_2\, u^{(2)^2}(t), & t \in [0, 60], \\ u_t^{(4)}(t) &= \ k_3\, u^{(3)}(t) - k_4\, u^{(4)}(t - \tau), & t \in [0, 60], \\ u(t) &= \ \phi(t), & t \in [-\tau, 0]. \end{cases} \qquad (12)$$

The initial function components $\phi^{(2)}(t) = \phi^{(3)}(t) = \phi^{(4)}(t)$ are identically zero, and the constant initial function $\phi^{(1)}(t)$ is unknown. The states for this system cannot be measured directly, but are observed through a nonlinear transformation, $H : \mathbb{R}^3 \times \Theta \to$

Model component	Prior
τ	χ_6^2
$k_i,\ i = 1, \dots, 6$	$\mathrm{Exp}(1)$
$\lambda_i,\ i = 1, \dots, 4$	χ_1^2
$\alpha_i + 100,\ i = 1, \dots, 4$	$\mathrm{Log\text{-}}\mathcal{N}(10, 1)$
$u^{(1)}(0)$	$N(y^{(3)}(0), 40^2)$
$u_t^{(i)},\ i = 1, \dots, 4$	$\mathcal{GP}(0, C_t)$
$u^{(i)},\ i = 1, \dots, 4$	$1\{i = 1\}u^{(1)}(0) + \int_0^t u_t^{(i)}(s)\,\mathrm{d}s$

Table 1: Priors on the states and parameters, assumed to be *a priori* independent.

$\mathbb{R}^4$, defined as,

$$
H\left\{u(t), \theta\right\} = \begin{pmatrix} k_5 \left\{u^{(2)}(t) + 2u^{(3)}(t)\right\} \\ k_6 \left\{u^{(1)}(t) + u^{(2)}(t) + 2u^{(3)}(t)\right\} \\ u^{(1)}(t) \\ u^{(3)}(t) \left\{u^{(2)}(t) + u^{(3)}(t)\right\}^{-1} \end{pmatrix}, \tag{13}
$$

and parameterized by the unknown scaling factors k_5 and k_6. The indirect measurements of the states,

$$
y^{(j)}(t_{j,i}) = H^{(j)}\left\{u(t_{j,i}), \theta\right\} + \varepsilon^{(j)}(t_{j,i}), \quad 1 \le i \le T_j,\ 1 \le j \le 4,
$$

are assumed to be contaminated with additive zero mean Gaussian noise, $\{\varepsilon^{(j)}(t_{j,i})\}_{1 \le i \le T_j,\ 1 \le j \le 4}$, with experimentally determined standard deviations. Analysis is based on experimental data (Swameye et al., 2003; Raue et al., 2009). As per Raue et al. (2009), the forcing function $\mathrm{Epo}R_A$ shares the same smoothness as the solution state (piecewise linear first derivative) as modelled by a GP interpolation.

This data has been used for calibrating the JAK-STAT pathway mechanism by a number of authors (e.g., Schmidl et al., 2013; Swameye et al., 2003; Raue et al., 2009; Campbell and Chkrebtii, 2013) under a variety of modelling assumptions and inferential approaches. The inaccuracy and computational constraints of numerical techniques required to solve the system equations have led some authors to resort to coarse ODE approximations, or to employ semiparametric Generalized Smoothing methods. This motivates analysis of the structure and propagation of discretization error through the inverse problem. A further issue is that the model (12) and its variants may be misspecified, which must be first uncoupled from discretization effects for further study. Indeed, the above cited analyses conclude that the model is not flexible enough to fit the available data, however it is not clear how much of this misfit is due to model discrepancy, and how much may be attributed to discretization error.

The probabilistic approach for forward simulation of delay initial function problems is described in Supplement B, extending the updating strategy to accommodate current and lagged state process realization information. The probabilistic solution is generated using an equally spaced discretization grid of size $N = 500$, and a covariance structure based on the uniform covariance kernel. Prior distributions are listed in Table 1.
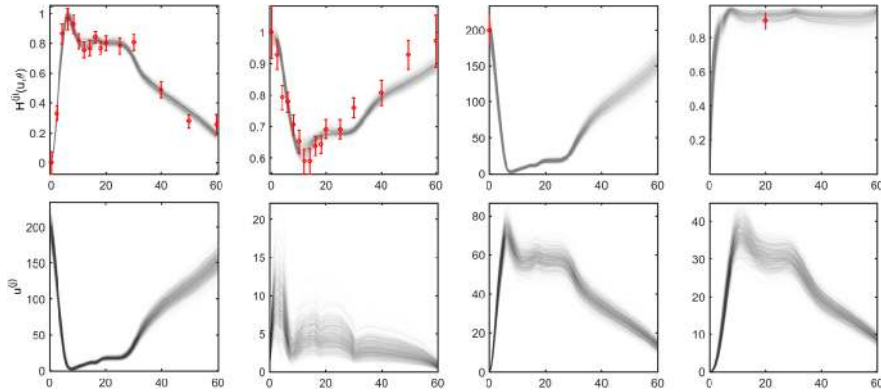
Figure 4: Sample paths (grey lines) of the observation processes (first row) obtained by transforming a sample from the marginal posterior distribution of the states (second row) by the observation function (13); experimental data (red points) and error bars showing two standard deviations of the measurement error.

Markov chain Monte Carlo is used to obtain samples from the posterior distribution (3). A parallel tempering sampler (Geyer, 1991) described in Supplement B efficiently traverses the parameter space of this multimodal posterior distribution.

A sample from the marginal posterior of state trajectories and the corresponding observation process are shown in Figure 4. As expected, the posterior trajectories, which incorporate discretization uncertainty in the forward problem, appear more diffuse than those obtained in e.g., Schmidl et al. (2013), while more flexibly describing the observed data, with the exception of a small number of outliers. Correlation plots and kernel density estimates with priors for the marginal posteriors of rate parameters k_1, k_2, k_3, k_4, observation scaling parameters k_5, k_6, time lag τ, initial function value $u^{(1)}(0)$, and hyperparameters for the solution uncertainty $\Psi = (\alpha, \lambda)$ are shown in Figure 5. All parameters with the exception of the prior precision α are identified by the data, while the rate parameter k_2 appears to be only weakly identified. We observe strong correlations between parameters, consistent with previous studies on this system. For example, there is strong correlation among the scaling parameters k_5, k_6 and the initial first state $u^{(1)}(0)$. Importantly, there is a relationship between the length-scale λ of the probabilistic solver and the first, third and fourth reaction rates. Furthermore, this relationship has a nonlinear structure, where the highest parameter density region seems to change with length scale implying strong sensitivity to the solver specifications. The length-scale is the probabilistic analogue, under a bounded covariance, of the step number in a numerical method. However, in analyses based on numerical integration, the choice of a numerical technique effectively fixes this parameter at an arbitrary value that is chosen a priori. The result here suggests that for this problem, the inferred parameter values are highly and nonlinearly dependent on the choice of the numerical method used. Because we have explicitly quantified discretization uncertainty, the observed lack of fit may suggest additional model refinement may be required.
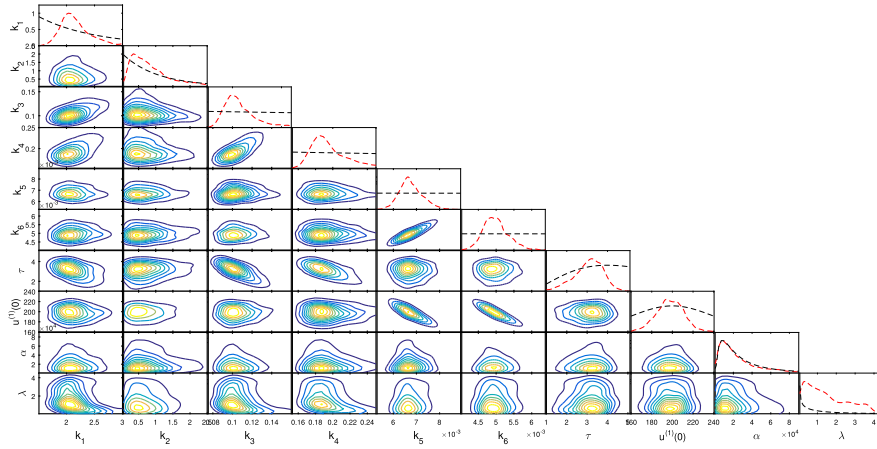
Figure 5: 50,000 posterior samples of the model parameters using a probabilistic solver with a grid of size 500, generated using parallel tempering Algorithm 7 with ten chains. Prior probability densities are shown in black. Note the complex posterior correlation structure between model parameters and solver hyperparameters across the lower rows.

5 Applications

We extend probability modelling of solution uncertainty to a variety of forward problems: chaotic dynamical systems, an ill-conditioned mixed boundary value problem, a stiff, high-dimensional initial value problem, and extensions to PDEs.

5.1 Discretization uncertainty under chaotic dynamics

The classical "Lorenz63" initial value problem (Lorenz, 1963) is a three-state ODE model of convective fluid motion induced by a temperature difference between an upper and lower surface. States $u^{(1)}, u^{(2)}$ and $u^{(3)}$ are proportional to the intensity of the fluid's convective motion, the temperature difference between (hot) rising and (cold) descending currents, and the deviation of the vertical temperature profile from linearity respectively. The model describing the time-evolution of their dynamics,

$$\begin{cases} u_t^{(1)} & = -\sigma(u^{(1)} + u^{(2)}), & t \in [0, 20], \\ u_t^{(2)} & = -ru^{(1)} - u^{(2)} - u^{(1)}u^{(3)}, & t \in [0, 20], \\ u_t^{(3)} & = u^{(1)}u^{(2)} - bu^{(3)}, & t \in [0, 20], \\ u & = u^*, & t = 0, \end{cases} \tag{14}$$

depends on dimensionless parameters σ, r and b. The standard choice of $\theta = (\sigma, r, b) = (10, 8/3, 28)$ with initial state $u^*(0) = (-12, -5, 38)$ falls within the chaotic regime. This system has a unique solution whose trajectory lies on a bounded region of the phase space (e.g., Robinson, 2001, pp. 271–272), and is unknown in closed form.
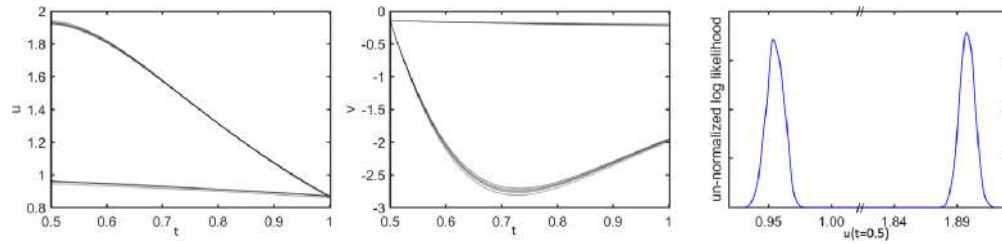
Figure 6: 100 posterior realizations from (17). Left and centre: states, u and v, evaluated over a fine grid, $\mathbf{t}$, over the domain, $\mathcal{D} = [.5, 1]$. Right: marginal un-normalized log posterior density over the unknown initial condition $u(t = .5)$.

A sample of 1000 trajectories generated from the probabilistic solution for this system is shown in Figure 2 given hyperparameters $N = 5001, \alpha = N, \lambda = 2h, h = 20/N$, constant prior mean function and a squared exponential covariance structure. There is a short time window within which there is negligible uncertainty in the solution, but the accumulation of discretization uncertainty results in rapidly diverging, yet highly structured errors. Figure 2 shows posterior samples in the 3 state dimensions (top row) and density estimates of the samples for $u^{(1)}(t)$ (bottom row) at 4 distinct evaluation times. The structured uncertainty in the forward model (2) is not simply defined by widening of the high posterior density region, but instead exhibits a behaviour that is consistent with the dynamics of the underlying stable attractor.

5.2   Solution multiplicity in a mixed boundary value problem

Consider a special case the Lane–Emden ODE mixed boundary problem, which is used to model the density, u, of gaseous spherical objects, such as stars, as a function of radius, t, (Shampine 2003). The second order system can be rewritten as a system of a first order equations on the interval $[.5, 1]$ with mixed boundary values,

$$
\begin{cases}
u_t & = v, \\
v_t & = -2vt^{-1} - u^5, \\
u(t = 1) & = \sqrt{3}/2, \\
v(t = .5) & = -288/2197,
\end{cases}
\tag{15}
$$

where the notation $u(t = 1)$ and $v(t = .5)$ is used to emphasize that boundary conditions are different for each state. The two leftmost panels of Figure 6 illustrate multimodality in the posterior solution, when multiple functions satisfy the mixed boundary value problem. These high posterior density regions concentrate as the number of grid points and the prior precision grow, suggesting the potential for defining intermediate target densities for sequential and particle Markov chain Monte Carlo schemes to sample realizations from the posterior distribution over unknown model parameters within the calibration problem. This example highlights the need for accurately modelling functional rather than point-wise solution uncertainty.

Mixed boundary value problems introduce challenges for many numerical solvers, which employ optimization and the theory of ODE initial value problems to estimate the unknown initial state, here $u(t = .5)$, subject to the state constraints at either boundary of the domain of integration, here $[.5, 1]$. The optimization over $u(t = .5)$ is performed until the corresponding initial value problem solution satisfies the known boundary condition, $u^*(t = 1)$, to within a user specified tolerance. Considering the mixed boundary value problem (15), the probabilistic solution consists of an inference problem over the unspecified initial value, $u(t = .5)$, given its known final value $u^*(t = 1)$. The likelihood naturally defines the mismatch between the boundary value $u(t = 1)$, obtained from the realized probabilistic solution given $u^*(t = 1)$, as follows,

$$u^*(t = 1) \mid u(t = .5), v^*(t = .5) \sim \mathcal{N}\left\{m^{N(u)}(t = 1), C^{N(u)}(t = 1, t = 1)\right\}, \quad (16)$$

where $m^{N(u)}(t = 1)$ and $C^{N(u)}(t = 1, t = 1)$ are the posterior mean and covariance for state u at time $t = 1$, obtained via Algorithm 1. The posterior over the states is,

$$[u, v, u(t = .5) \mid u^*(t = 1), v^*(t = .5)] \quad (17)$$
$$\propto [u, v \mid u(t = .5), u^*(t = 1), v^*(t = 1)] \, [u(t = .5) \mid u^*(t = 1), v^*(t = .5)] \, [u(t = .5)].$$

The possibility of multimodality suggests the use of an appropriate Markov chain Monte Carlo scheme, such as parallel tempering (Geyer, 1991), implemented in Supplement B, which can quickly identify and explore disjoint regions of high posterior probability.

5.3 Discretization uncertainty for a model of fluid dynamics

The following example illustrates that the probabilistic framework developed may be reliably and straightforwardly applied to a high-dimensional dynamical system. The Navier–Stokes system is a fundamental model of fluid dynamics, incorporating laws of conservation of mass, energy and linear momentum for an incompressible fluid over a domain given constraints imposed along the boundaries. It is an important component of complex models in oceanography, weather, atmospheric pollution, and glacier movement. Despite its extensive use, the dynamics of Navier–Stokes models are poorly understood even at small time scales, where they can exhibit turbulence.

The Navier–Stokes PDE model for the time evolution of 2 components of the velocity, $u : \mathcal{D} \to \mathbb{R}^2$, of an incompressible fluid on a torus, $\mathcal{D} = [0, 2\pi) \times [0, 2\pi]$, can be expressed in spherical coordinates as,

$$\begin{cases} u_t - \theta \, \Delta u + (u \cdot \nabla) \, u &= \mathsf{g} - \nabla \mathsf{p}, & (x, t) \in \mathcal{D} \times [a, b], \\ \nabla \cdot u &= 0, & (x, t) \in \mathcal{D} \times [a, b], \\ \int u^{(j)} \, \mathrm{d}x &= 0, & (x, t) \in \mathcal{D} \times [a, b], \ j = 1, 2, \\ u &= u^*, & (x, t) \in \mathcal{D} \times \{0\}, \end{cases} \quad (18)$$

where $\Delta := [\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2}]$ is the Laplacian operator such that $\Delta u = (u_{x_1 x_1}^{(1)} + u_{x_2 x_2}^{(1)}, u_{x_1 x_1}^{(2)} + u_{x_2 x_2}^{(2)})$, and $\nabla := [\frac{\partial}{\partial x_1} + \frac{\partial}{\partial x_2}]$ is the gradient operator such that $\nabla u = (u_{x_1}^{(1)} + u_{x_2}^{(1)}, u_{x_1}^{(2)} + u_{x_2}^{(2)})$. The model is parametrized by the viscosity of the fluid, $\theta \in \mathbb{R}^+$, the
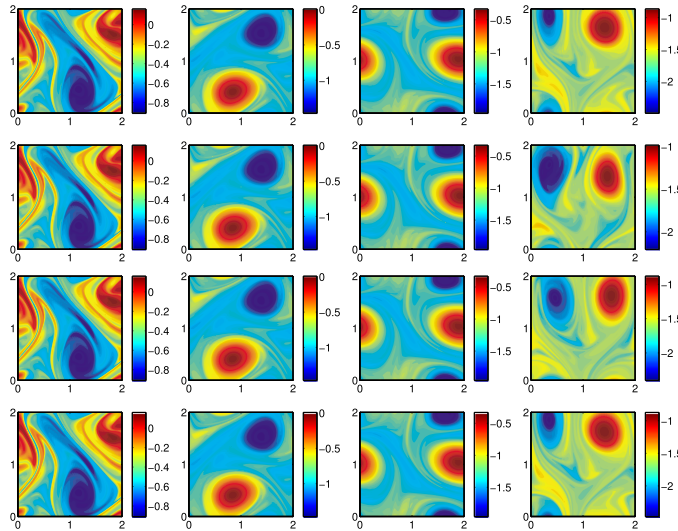
Figure 7: Time evolution of four forward simulated realizations (along rows) of fluid vorticity, governed by the forced Navier–Stokes model (18), over two spatial dimensions: the angle of the inner ring (horizontal axis) and outer ring (vertical axis) of a two dimensional torus. Angles are expressed in radians. Vorticities are evaluated at times $\mathbf{t} = (0.2, 0.4, 0.6, 0.8)$ units (along columns).

pressure function $\mathsf{p} : \mathcal{D} \times [a, b] \to \mathbb{R}$, and the external time-homogeneous forcing function $\mathsf{g} := 5 \times 10^{-3} \cos[(\frac{1}{2}, \frac{1}{2}) \cdot x]$. We further assume periodic boundary conditions, and viscosity $\theta = 10^{-4}$ in the turbulent regime. The exact solution of the Navier–Stokes boundary value problem (18) is not known in closed form.

We will visualize the probabilistic solution of the Navier–Stokes system by reducing the two components of velocity to a one dimensional function: the local spinning motion of the incompressible fluid, called vorticity, which we define as, $\varpi = -\nabla \times u$, where $\nabla \times u$ represents the rotational curl (the cross product of ∇ and u), with positive vorticity corresponding to clockwise rotation. We discretize the Navier–Stokes model (18) over a grid of size 128 in each spatial dimension. A pseudo spectral projection in Fourier space yields 16,384 coupled, stiff ODEs with associated constraints (details provided on the accompanying website). Figure 7 shows four forward simulated vorticity trajectories (along rows), obtained from two components of velocity at four distinct time points (along columns). Differences in the state dynamics arise by the final time shown, where the four trajectories visibly diverge from one another. These differences describe uncertainty resulting from discretizing the exact but unknown infinite dimensional solution.

5.4 Discretization uncertainty for partial differential equations

Similarly to the ODE case, we model the prior belief about the fixed but unknown solution and partial derivatives of a PDE boundary value problem by defining a proba-

bility measure over multivariate trajectories, as well as first or higher order derivatives with respect to spatial inputs. The prior structure will depend on the form of the PDE model under consideration. We illustrate discretization uncertainty modelling for PDEs by considering the parabolic heat equation describing heat diffusion over time along a single spatial dimension implicitly via,

$$
\begin{cases}
u_t(x, t) & = & \kappa u_{xx}(x, t), & t \in [0, 0.25], & x \in [0, 1], \\
u(x, t) & = & \sin{(x\pi)}, & t = 0, & x \in [0, 1], \\
u(x, t) & = & 0, & t \in (0, 0.25], & x \in \{0, 1\},
\end{cases}
\tag{19}
$$

with conductivity parameter of $\kappa = 1$. Note that this PDE has the explicit form (1) with $\mathcal{A} = \{u, u_t\}$ and $\mathcal{B} = u_{xx}$.

One modelling choice for extending the prior covariance to a spatio-temporal domain is to assume separability by defining covariances over time and space independently. Accordingly we choose temporal kernel R_λ as in Section 2.1 and let $R_\nu : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a possibly different spatial kernel function with length-scale ν. We integrate the spatial kernel once to obtain $Q_\nu(x_j, x_k) = \int_0^{x_j} R_\nu(z, x_k)\mathrm{d}z$ and twice to obtain $S_\nu(x_j, x_k) = \int_0^{x_j} Q_\nu(z, x_k)\mathrm{d}z$. The covariance structures between temporal and spatial evaluation points $[x_j, t_j]$ and $[x_k, t_k]$ are obtained from the process convolution formulation with temporal and spatial prior precision parameters α and β respectively, as follows,

$$
\begin{array}{rcl}
C^0_{xx}([x_j, t_j], [x_k, t_k]) & = & \alpha^{-1}\beta^{-1} \int_\mathbb{R} R_\nu(x_j, z)R_\nu(x_k, z)\mathrm{d}z \int_\mathbb{R} Q_\lambda(t_j, w)Q_\lambda(t_k, w)\mathrm{d}w, \\
C^0_t([x_j, t_j], [x_k, t_k]) & = & \alpha^{-1}\beta^{-1} \int_\mathbb{R} S_\nu(x_j, z)S_\nu(x_k, z)\mathrm{d}z \int_\mathbb{R} R_\lambda(t_j, w)R_\lambda(t_k, w)\mathrm{d}w, \\
C^0([x_j, t_j], [x_k, t_k]) & = & \alpha^{-1}\beta^{-1} \int_\mathbb{R} S_\nu(x_j, z)S_\nu(x_k, z)\mathrm{d}z \int_\mathbb{R} Q_\lambda(t_j, w)Q_\lambda(t_k, w)\mathrm{d}w,
\end{array}
$$

with cross-covariances defined analogously. The mean function and covariance specification uniquely determine the joint GP prior on the state and its partial derivatives given hyperparameters Ψ, which include $\alpha, \beta, \lambda, m^0_{xx}, m^0_t, m^0$, at vectors of temporal and spatio-temporal locations $(\mathbf{x_j}, \mathbf{t_j})$, $(\mathbf{x_k}, \mathbf{t_k})$, $(\mathbf{x_l}, \mathbf{t_l})$ as:

$$
\begin{bmatrix} u_{xx}(\mathbf{x_j}, \mathbf{t_j}) \\ u_t(\mathbf{x_k}, \mathbf{t_k}) \\ u(\mathbf{x_l}, \mathbf{t_l}) \end{bmatrix} \sim \mathcal{GP}\left(\begin{bmatrix} m^0_{xx}(\mathbf{x_j}, \mathbf{t_j}) \\ m^0_t(\mathbf{x_k}, \mathbf{t_k}) \\ m^0(\mathbf{x_l}, \mathbf{t_l}) \end{bmatrix}, \begin{bmatrix} C_{1,1} & C_{1,2} & C_{1,3} \\ C_{2,1} & C_{2,2} & C_{2,3} \\ C_{3,1} & C_{3,2} & C_{3,3} \end{bmatrix} \right),
\tag{20}
$$

where,

$$
\begin{array}{ll}
C_{1,1} = C^0_{xx}([\mathbf{x_j}, \mathbf{t_j}], [\mathbf{x_j}, \mathbf{t_j}]), & C_{1,2} = \int_0^{\mathbf{x_k}} \int_0^{\mathbf{z}} \frac{\partial}{\partial \mathbf{t_k}} C^0_{xx}([\mathbf{x_j}, \mathbf{t_j}], [\mathbf{w}, \mathbf{t_k}])\mathrm{d}\mathbf{w}\mathrm{d}\mathbf{z}, \\
C_{1,3} = \int_0^{\mathbf{x_l}} \int_0^{\mathbf{z}} C^0_{xx}([\mathbf{x_j}, \mathbf{t_j}], [\mathbf{w}, \mathbf{t_l}])\mathrm{d}\mathbf{w}\mathrm{d}\mathbf{z}, & C_{2,2} = C^0_t([\mathbf{x_k}, \mathbf{t_k}], [\mathbf{x_k}, \mathbf{t_k}]), \\
C_{2,3} = \int_0^{\mathbf{t_l}} C^0_t([\mathbf{x_k}, \mathbf{t_k}], [\mathbf{x_l}, \mathbf{w}])\mathrm{d}\mathbf{w}, & C_{3,3} = C^0([\mathbf{x_l}, \mathbf{t_l}], [\mathbf{x_l}, \mathbf{t_l}]).
\end{array}
$$

When integrated covariances are available in closed form (e.g., Supplement D.4), the above matrices can be computed analytically.

A generalization of the sequential updating scheme presented in Algorithm 1 is outlined in Supplement B. The *forward in time, continuous in space* (FTCS) sampling scheme takes steps forward in time, while smoothing over spatial model interrogations
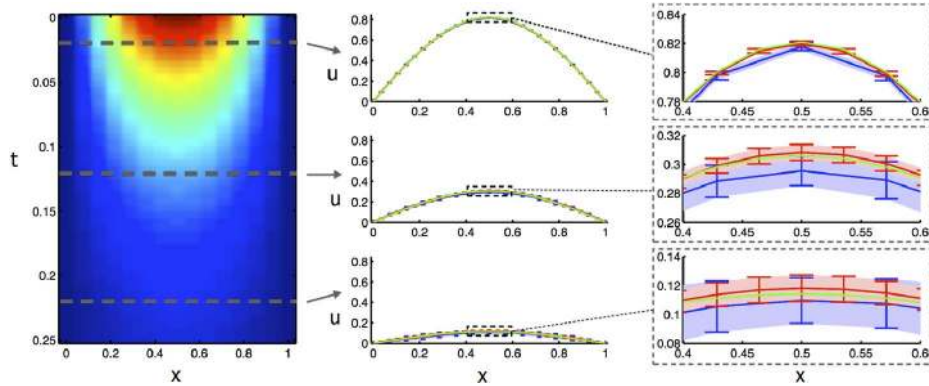
Figure 8: Uncertainty in the solution of the heat equation discretized using two spatio-temporal grids in blue and red respectively: 15×50 and 29×100. Spatial posterior predictions are shown at $t = (0.02, 0.12, 0.22)$. The exact solution is shown in green; error bars show the empirical mean and 2 standard deviations computed from 50 simulations.

at each step. In Figure 8 the exact solution for the PDE initial value problem (19) with $\kappa = 1$ on the domain $[0, 1] \times [0, 0.25]$ is compared with the probability model of uncertainty given two different grid sizes. The simulations illustrate that as the mesh size becomes finer, the uncertainty in the solution decreases.

The characterization of spatial uncertainty is critical for more complex models in cases where there are computational constraints limiting the number of system evaluations that may be performed. We illustrate its effect by performing posterior inference for the parameter κ from data simulated from an exact solution with $\kappa = 1$ and zero mean Gaussian error with standard deviation of 0.005. Figure 9 shows the posterior distribution over κ obtained by using both a deterministic "forward in time, centred in space" (FTCS) finite difference solver and a probabilistic solver under a variety of discretization grids. For the probabilistic solution, a squared exponential covariance structure was chosen over the spatial domain, with length-scale 1.5 times the spatial step size. For computational efficiency a uniform covariance structure was chosen over the temporal domain, with length-scale 2 times the temporal step size. The prior precision was set to 10,000. Note the change in the scale as the mesh is refined and the posterior variance decreases. The use of a deterministic solver illustrates the problem of posterior under-coverage that may occur if discretization uncertainty is not taken into account and too coarse a grid is employed relative to the complexity of model dynamics. If the discretization is not fine enough, the approximate posterior assigns negligible probability mass to the true value of κ, in contrast to the unbiased posterior obtained when employing the exact solution for inference. In this illustrative setting, the use of a probabilistic solver propagates discretization uncertainty in the solution through to the posterior distribution over the parameters. By using a probabilistic solver with the same coarsely discretized grid in this example, the posterior density is more diffuse but approximately centred on the true value of κ.
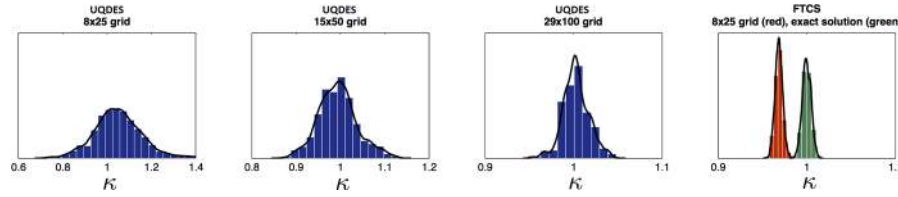
Figure 9: Posterior densities for the conductivity parameter in the heat equation given simulated data over an 8×25 spatio-temporal grid from the exact solution with $\kappa = 1$. Inference is based on the probabilistic differential equation solver using three grid sizes (blue). The posterior density based on a deterministic FTCS numerical scheme (red) and the posterior density based on the exact solution (red) are also provided.

6　Discussion

This paper presents a Bayesian formalism for characterizing the structure of discretization uncertainty in the solution of differential equations. Discretization uncertainty for dynamical systems may be naturally described in terms of degrees of belief, an idea first formulated by Skilling (1991) and O'Hagan (1992). A Bayesian function estimation approach offers a way of defining and updating our beliefs about probable model trajectories conditional on interrogations of the differential equation model. The resulting uncertainty over the solution can then be propagated through the inferential procedure by defining an additional layer in the Bayesian posterior hierarchy. As such, the proposed formulation is a radical departure from the existing practice of constructing an approximate statistical error model for the data using numerical techniques.

The more general goal of studying other sources of model uncertainty requires first understanding how much is known about the dynamics of the proposed model and how that information changes with time, space, and parameter settings. Fundamentally, the probabilistic approach we advocate defines a trade-off between solution accuracy and the size of the inverse problem, by choice of discretization grid size. This relationship has been of deep interest in the uncertainty quantification community (see, e.g., Kaipio et al., 2004; Arridge et al., 2006) and our approach offers a way of resolving the individual contributions to overall uncertainty from discretization, model misspecification (in the sense of Kennedy and O'Hagan, 2001), and Monte Carlo error.

The area of probabilistic numerics is an emerging field at the intersections of Statistics with Applied Mathematics, Numerical Analysis, and Computer Science, and we hope that this paper will provide a starting point for further work in this area. In particular, while the proposed approach is computationally competitive with first order numerical methods, it will be important to implement more efficient sampling algorithms analogous to those of higher order deterministic solvers developed from over a century of research into numerical methods. For example a sampling scheme developed based on multiple model interrogations located between discretization grid points. Because many traditional higher order numerical solvers (e.g. explicit fourth order Runge–Kutta) use such a carefully constructed weighted average of model interrogations, our approach also

provides a way to interpret numerical method in the context of a general probabilistic framework (e.g., Schober et al., 2014). It is not surprising that a self-consistent Bayesian modelling approach has similarities to numerical methods whose form has been chosen for their convergence properties. From the perspective presented in this paper, rather than emulate numerical solvers, quadrature is automatically provided by the Gaussian process integration that occurs at each step of Algorithm 1. If the solution smoothness is correctly specified *a priori* via the covariance structure, then the Gaussian process integration step will naturally yield an appropriate quadrature rule.

Modelling uncertainty about an exact but unknown solution may be viewed as providing both an estimate of the exact solution and its associated functional error analysis. Although proportionally scalable, uncertainty quantification methods are naturally more expensive than deterministic numerical approximations without associated error analyses, making efficient sampling strategies for stochastic model calibration an important goal for future research. As may be expected, low dimensional, stable systems, without any topological restrictions on the solution may not suffer from significant discretization effects at a reasonably fine discretization grid. However, it is often difficult to determine *a priori* whether this stability will persist across the entire parameter space. An example is illustrated by the JAK-STAT case study, where the relationships between model parameters and solver specifications (reflecting changes in the system dynamics) may introduce bias in the estimation of parameters. If model uncertainty is indeed negligible and comparable over all parameter settings, sampled trajectories obtained via the probabilistic method will have little variability around the exact solution. If, however, the chosen discretization grid is too coarse with respect to the rate of change of system dynamics over some subset of the input space, the uncertainty associated with the model error will be accounted and propagated. For this reason, it may be prudent to adopt a probabilistic approach if one cannot be reasonably certain of negligible and comparable discretization error across all parameter settings.

Importantly, the probabilistic framework can be used to optimize computation. For a given discretization grid size, a natural question is how we can arrange the grid on the domain in such a way that the resulting approximation is as close as possible in terms of some specified criterion. Most commercially available numerical differential equation solvers select the step length of the discretization grid sequentially. Controlling the local error by choice of the distance between neighbouring grid locations is called *adaptive step size selection*. For example, in the simplest cases, this is accomplished by evaluating the local truncation error at each step and halving the step size if this exceeds an error tolerance that is pre-specified by the user (the process may be repeated several times per step until an appropriate local truncation error is achieved). In a similar way to local truncation error, a predictive probability model of discretization uncertainty may be used to guide sequential mesh refinement for complex dynamical systems. Optimizing a chosen design criterion would yield the desired discretization grid, as suggested by Skilling (1991). Some initial work in this direction has been conducted in Chkrebtii (2013), where it is shown that the natural Kullback–Leibler divergence criterion may be successfully used to adaptively choose the discretization grid in a probabilistic manner.

Prior specification may be used to rule out physically inconsistent trajectories that may arise from a numerical approximation. In numerical analysis, specialized techniques

are developed and shown to satisfy certain properties or constraints for a given class of differential equations. A probabilistic framework would permit imposing such constraints in a more flexible and general way. Smoothness constraints on the solution may be incorporated through the prior, and anisotropy in the trajectory may be incorporated at each step through the covariance and the error model. Although in such general cases it may no longer be possible to obtain closed form representations of the sequential prior updates, an additional layer of Monte Carlo sampling may be used to marginalize over the realizations required to interrogate the differential equation model.

These developments lie at the frontier of research in uncertainty quantification, dealing with massive nonlinear models, complex or even chaotic dynamics, and strong spatial-geometric effects (e.g. subsurface flow models). Even solving such systems approximately is a problem that is still at the edge of current research in numerical analysis.

Inference and prediction for computer experiments (Sacks et al., 1989) is based on a model of the response surface, or emulator, constructed from numerical solutions of prohibitively expensive large-scale system models over a small subset of parameter regimes. Currently, numerical uncertainty is largely ignored when emulating a computer model, although it is informally incorporated through a covariance nugget (see, e.g., Gramacy and Lee, 2012) whose magnitude is often chosen heuristically. Adopting a probabilistic approach on a large scale will have practical implications in this area by permitting relaxation of the error-free assumption adopted when modelling computer code output, leading to more realistic and flexible emulators. For this reason, Bayesian calibration of stochastic emulators will be an important area of future research.

Numerical approximations of chaotic systems are inherently deterministic, qualitative dynamics (of models such as Lorenz63) are often studied statistically by introducing artificial perturbations on the numerical trajectory. The rate of exponential growth (the *Lyapunov exponent*) between nearby trajectories can be estimated via Monte Carlo by introducing small changes to the numerical trajectory over a grid defined along the domain. Our proposed approach explicitly models discretization uncertainty and therefore offer an alternative to such artificial perturbations, thus opening new avenues for exploring uncertainty in the system's qualitative features.

Supplementary Material

Supplementary Material for "Bayesian Solution Uncertainty Quantification for Differential Equations" (DOI: 10.1214/16-BA1017SUPP; .pdf).

References

Arridge, S. R., Kaipio, J. P., Kolehmainen, V., Schweiger, M., Somersalo, E., Tarvainen, T., and Vauhkonen, M. (2006). "Approximation errors and model reduction with an application in optical diffusion tomography." *Inverse Problems*, 22(1): 175–195. MR2194190. doi: http://dx.doi.org/10.1088/0266-5611/22/1/010. 1240, 1261

Berliner, L. M. (1991). "Likelihood and Bayesian prediction of chaotic systems." *Journal of the American Statistical Association*, 86(416): 938–952. MR1146342. 1240

Bock, H. G. (1983). "Recent advances in parameter identification techniques for ODE." In Deuflhard, P. and Harrier, E. (eds.), *Numerical Treatment of Inverse Problems in Differential and Integral Equations*, 95–121. Birkhäuser. MR0714563. 1240

Brunel, N. J. (2008). "Parameter estimation of ODEs via nonparametric estimators." *Electronic Journal of Statistics*, 2: 1242–1267. MR2471285. doi: http://dx.doi.org/10.1214/07-EJS132. 1240

Calderhead, B. and Girolami, M. (2011). "Statistical analysis of nonlinear dynamical systems using differential geometric sampling methods." *Interface Focus*, 1(6): 821–835. 1240

Campbell, D. and Steele, R. J. (2011). "Smooth functional tempering for nonlinear differential equation models." *Statistics and Computing*, 22: 429–443. MR2865027. doi: http://dx.doi.org/10.1007/s11222-011-9234-3. 1240

Campbell, D. A. and Chkrebtii, O. (2013). "Maximum profile likelihood estimation of differential equation parameters through model based smoothing state estimates." *Mathematical Biosciences*, 246: 283–292. MR3132050. doi: http://dx.doi.org/10.1016/j.mbs.2013.03.011. 1240, 1253

Chkrebtii, O. (2013). "Probabilistic solution of differential equations for Bayesian uncertainty quantification and inference." Ph.D. thesis, Simon Fraser University. 1262

Chkrebtii, O. A., Campbell, D. A., Calderhead, B., and Girolami, M. A. (2016). "Supplementary material for "Bayesian solution uncertainty quantification for differential equations"." *Bayesian Analysis*. doi: http://dx.doi.org/10.1214/16-BA1017SUPP. 1242

Dowd, M. (2007). "Bayesian statistical data assimilation for ecosystem models using Markov chain Monte Carlo." *Journal of Marine Systems*, 68(3–4): 439–456. 1240

Geyer, C. (1991). "Markov chain Monte Carlo maximum likelihood." In *Computing Science and Statistics, Proceedings of the 23rd Symposium on the Interface, 156*. American Statistical Association. 1254, 1257

Ghanem, R. and Spanos, P. (2003). *Stochastic Finite Elements: A Spectral Approach*. Springer-Verlag. MR1083354. doi: http://dx.doi.org/10.1007/978-1-4612-3094-6. 1240

Golightly, A. and Wilkinson, D. J. (2011). "Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo." *Interface Focus*, 1: 807–820. 1251

Gramacy, R. and Lee, H. (2012). "Cases for the nugget in modeling computer experiments." *Statistics and Computing*, 22: 713–722. MR2909617. doi: http://dx.doi.org/10.1007/s11222-010-9224-x. 1263

Gugushvili, S. and Klaassen, C. A. (2012). "Root n-consistent parameter estimation for systems of ordinary differential equations: bypassing numerical integration

via smoothing." *Bernoulli*, 18: 1061–1098. MR2948913. doi: http://dx.doi.org/10.3150/11-BEJ362. 1240

Higdon, D. (1998). "A process-convolution approach to modeling temperatures in the North Atlantic Ocean." *Journal of Environmental and Ecological Statistics*, 5: 173–190. 1244

Higdon, D. (2002). *Space and Space–Time Modeling Using Process Convolutions*, 37–54. London, UK: Springer Verlag. MR2059819. 1244

Huttunen, J. M. J. and Kaipio, J. P. (2007). "Approximation error analysis in nonlinear state estimation with an application to state–space identification." *Inverse Problems*, 23(5): 2141. MR2353332. doi: http://dx.doi.org/10.1088/0266-5611/23/5/019. 1240

Ionides, E., Bretó, C., and King, A. (2006). "Inference for nonlinear dynamical systems." *Proceedings of the National Academy of Sciences of the United States of America*, 103(49): 18438–18443. 1240

Kaipio, J. and Somersalo, E. (2007). "Statistical inverse problems: Discretization, model reduction and inverse crimes." *Journal of Computational and Applied Mathematics*, 198: 493–504. MR2260683. doi: http://dx.doi.org/10.1016/j.cam.2005.09.027. 1242

Kaipio, J. P., Seppanen, A., Somersalo, E., and Haario, H. (2004). "Posterior covariance related optimal current patterns in electrical impedance tomography." *Inverse Problems*, 20(3): 919. MR2067508. doi: http://dx.doi.org/10.1088/0266-5611/20/3/016. 1240, 1261

Kennedy, M. C. and O'Hagan, A. (2001). "Bayesian calibration of computer models." *Journal of the Royal Statistical Society B*, 63(3): 425–464. MR1858398. doi: http://dx.doi.org/10.1111/1467-9868.00294. 1240, 1241, 1261

Kim, J.-h. et al. (2014). "The AGORA high-resolution galaxy simulations comparison project." *The Astrophysical Journal Supplement Series*, 210(1): 14. http://stacks.iop.org/0067-0049/210/i=1/a=14 1240

Liang, H. and Wu, H. (2008). "Parameter estimation for differential equation models using a framework of measurement error in regression models." *Journal of the American Statistical Association*, 103(484): 1570–1583. MR2504205. doi: http://dx.doi.org/10.1198/016214508000000797. 1240

Lorenz, E. N. (1963). "Deterministic nonperiodic flow." *Journal of the Atmospheric Sciences*, 20: 130–141. 1255

Marjoram, P., Molitor, J., Plagnol, V., and Tavaré, S. (2003). "Markov chain Monte Carlo without likelihoods." *Proceedings of the National Academy of Sciences*, 100(26): 15324–15328. 1251

Marzouk, Y. and Najm, H. (2009). "Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems." *Journal of Computational Physics*, 228(6): 1862–1902. MR2500666. doi: http://dx.doi.org/10.1016/j.jcp.2008.11.024. 1240

Marzouk, Y., Najm, H., and Rahn, L. (2007). "Stochastic spectral methods for efficient Bayesian solution of inverse problems." *Journal of Computational Physics*, 224(2): 560–586. MR2330284. doi: http://dx.doi.org/10.1016/j.jcp.2006.10.010. 1240

Neal, R. M. (2011). "MCMC Using Ensembles of States for Problems with Fast and Slow Variables such as Gaussian Process Regression." arXiv:1101.0387. MR2858447. 1251

Oberkampf, W. and Roy, C. (2010). *Verification and Validation in Scientific Computing.* Cambridge University Press. 1240

O'Hagan, A. O. (1992). "Some Bayesian numerical analysis." *Bayesian Statistics*, 4: 345–363. MR1380285. 1242, 1261

Pellegrini, S. and Dusanter-Fourt, I. (1997). "The structure, regulation and function of the Janus Kinases (JAKs) and the signal transducers and activators of transcription (STATs)." *European Journal of Biochemistry*, 248(3): 615–633. 1252

Ramsay, J., Hooker, G., Campbell, D., and Cao, J. (2007). "Parameter estimation for differential equations: a generalized smoothing approach." *Journal of the Royal Statistical Society B*, 69: 741–796. MR2368570. doi: http://dx.doi.org/10.1111/j.1467-9868.2007.00610.x. 1240

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning.* Cambridge, Massachusetts: MIT Press. MR2514435. 1244

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., and Timmer, J. (2009). "Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood." *Bioinformatics*, 25: 1923–1929. 1253

Robinson, J. (2001). *Infinite-Dimensional Dynamical Systems: An Introduction to Dissipative Parabolic PDEs and the Theory of Global Attractors.* Cambridge Texts in Applied Mathematics. Cambridge University Press. MR1881888. doi: http://dx.doi.org/10.1007/978-94-010-0732-0. 1255

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). "Design and analysis of computer experiments." *Statistical Science*, 4(4): 409–423. MR1041765. 1263

Schmidl, D., Czado, C., Hug, S., and Theis, F. J. (2013). "A vine-copula based adaptive MCMC sampler for efficient inference of dynamical systems." *Bayesian Analysis*, 8(1): 1–22. MR3036249. doi: http://dx.doi.org/10.1214/13-BA801. 1253, 1254

Schober, M., Duvenaud, D. K., and Hennig, P. (2014). "Probabilistic ODE solvers with Runge–Kutta means." In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., and Weinberger, K. (eds.), *Advances in Neural Information Processing Systems 27*, 739–747. Curran Associates, Inc. 1262

Shampine, L. (2003). "Singular boundary value problems for ODEs." *Applied Mathematics and Computation*, 138: 99–112. MR1950348. doi: http://dx.doi.org/10.1016/S0096-3003(02)00111-X. 1256

Skilling, J. (1991). *Bayesian Solution of Ordinary Differential Equations*, 23–37. Seattle: Kluwer Academic Publishers.   1241, 1242, 1261, 1262

Solak, E., Murray-Smith, R., Leithead, W. E., Leith, D. J., and Rasmussen, C. E. (2003). "Derivative observations in Gaussian process models of dynamic systems." In Becker, S., Thrun, S., and Obermayer, K. (eds.), *Advances in Neural Information Processing Systems 15*, 1057–1064. MIT Press. http://papers.nips.cc/paper/2287-derivative-observations-in-gaussian-process-models-of-dynamic-systems.pdf   1244

Stuart, A. M. (2010). "Inverse problems: A Bayesian perspective." *Acta Numerica*, 19: 451–559.   MR2652785. doi: http://dx.doi.org/10.1017/S0962492910000061. 1240, 1242, 1244

Swameye, I., Muller, T., Timmer, J., Sandra, O., and Klingmuller, U. (2003). "Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling." *Proceedings of the National Academy of Sciences*, 100: 1028–1033.   1253

Xue, H., Miao, H., and Wu, H. (2010). "Sieve estimation of constant and time-varying coefficients in nonlinear ordinary differential equation models by considering both numerical error and measurement error." *The Annals of Statistics*, 38: 2351–2387. MR2676892. doi: http://dx.doi.org/10.1214/09-AOS784.   1240

Xun, X., Cao, J., Mallick, B., Carroll, R. J., and Maity, A. (2013). "Parameter estimation of partial differential equation models."   MR3174680. doi: http://dx.doi.org/10.1080/01621459.2013.794730.   1240