Statistica Sinica Preprint No: SS-2024-0213	
Title	Inference for Delay Differential Equations Using Manifold-Constrained Gaussian Processes
Manuscript ID	SS-2024-0213
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0213
Complete List of Authors	Yuxuan Zhao and Samuel W.K. Wong
Corresponding Authors	Samuel W.K. Wong
E-mails	samuel.wong@uwaterloo.ca

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

Inference for Delay Differential Equations Using Manifold-Constrained Gaussian Processes

Yuxuan Zhao and Samuel W.K. Wong *

University of Waterloo

Abstract: Dynamic systems described by differential equations often involve feedback among system components. When there are time delays for components to sense and respond to feedback, delay differential equation (DDE) models are commonly used. This paper considers the problem of inferring unknown system parameters, including the time delays, from noisy and sparse experimental data observed from the system. We propose an extension of manifold-constrained Gaussian processes to conduct parameter inference for DDEs, whereas the time delay parameters have posed a challenge for existing methods that bypass numerical solvers. Our method uses a Bayesian framework to impose a Gaussian process model over the system trajectory, conditioned on the manifold constraint that satisfies the DDEs. For efficient computation, a linear interpolation scheme is developed to approximate the values of the time-delayed system outputs, along with corresponding theoretical error bounds on the approximated derivatives. Two simulation examples, based on Hutchinson's equation and the *lac* operon system, together with a real-world application using Ontario COVID-19 data,

* Author for correspondence: samuel.wong@uwaterloo.ca

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

are used to illustrate the efficacy of our method.

Key words and phrases: Bayesian inference, dynamic systems, Hutchinson's equation, *lac* operon, parameter estimation.

1. Introduction

Delay differential equations (DDEs) are commonly used to model dynamic processes with time delays in ecology (Hutchinson, 1948), biology (Yildirim and Mackey, 2003), and epidemiology (Ma et al., 2004). In this setting, components of the system require time to respond to feedback, thereby introducing time delay parameters into the model. For example, when modeling disease transmission mechanisms, the system of DDEs could incorporate constant time delays to accommodate the incubation period of the infectious disease (Ma et al., 2004). In models for gene regulation, the time required for transcription and translation activities could be encoded via time delay parameters (Yildirim and Mackey, 2003).

Our focus lies in models described by a set of DDEs,

$$dx_i(t)/dt = f_i\left\{\boldsymbol{x}(t), \boldsymbol{x}(t-\boldsymbol{\tau}), \boldsymbol{\theta}, t\right\}, \quad t \in [0, T], i \in \{1, \cdots, m\}, \qquad (1.1)$$

where m is the number of system components and the vector $\boldsymbol{x}(t) = (x_1(t), x_2(t), \cdots, x_m(t))$ denotes the system output at time t. Here, we let $\boldsymbol{\tau} = (\tau_1, \tau_2, \cdots, \tau_m)$ denote the vector of time delay parameters for the

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

system components (with $\tau_i = 0$ if the i-th component does not involve a time delay) and $\boldsymbol{\theta}$ denote the l-dimensional vector of other unknown model parameters. The term $\boldsymbol{x}(t-\boldsymbol{\tau}) = (x_1(t-\tau_1), x_2(t-\tau_2), \cdots, x_m(t-\tau_m))$ refers to the time-delayed system output at time $t-\boldsymbol{\tau}$, or *historical output* for short. In many applications, the m DDEs share the same vector of time delay parameters (Glass et al., 2021), which we write as $\boldsymbol{\tau}$ for simplicity of notation throughout this paper; if there are equation-specific time delays, we may instead write $\boldsymbol{\tau}_i = (\tau_{i,1}, \tau_{i,2}, \ldots, \tau_{i,m})$ to denote the vector of time delay parameters in the i-th equation.

In contrast to ordinary differential equations (ODEs), DDEs require information about the history of the system dynamics: we let $\mathcal{H}_{\tau_i} = \{x_i(t), t \in [-\tau_i, 0]\}$ denote the values of the i-th system component prior to $t = 0$. The widely-used history function $\mathcal{H}_{\tau_i} = \{x_i(t) = x_i(0), t \in [-\tau_i, 0]\}$ is assumed throughout this paper, which sets $x_i(t-\tau_i) = x_i(0)$ for all $t \leq \tau_i$ (Yildirim and Mackey, 2003; Ma et al., 2004; Bihorel, 2011; Wang and Cao, 2012; Wang et al., 2022). The functions $f_i(\cdot) : \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^l \times \mathbb{R} \to \mathbb{R}$ are treated as known from the scientific context. Due to experimental limitations, data observed from the system are usually noisy and collected at a discrete set of time points $\boldsymbol{\gamma}$. Letting $\boldsymbol{y}(\boldsymbol{\gamma})$ denote the observations, we assume $\boldsymbol{y}(\boldsymbol{\gamma}) = \boldsymbol{x}(\boldsymbol{\gamma}) + \boldsymbol{\epsilon}(\boldsymbol{\gamma})$, where $\boldsymbol{\epsilon}(\boldsymbol{\gamma})$ is Gaussian noise. The goal is to

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

infer the unknown parameters in $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$ given the observations $\boldsymbol{y}(\boldsymbol{\gamma})$.

Inferring parameters in differential equations (DEs) can be a difficult task, as nonlinear systems tend to lack analytic solutions. Thus, the first and oldest general category of methods for the parameter inference problem uses (deterministic) numerical solvers. Horbelt et al. (2002) minimized the weighted sum of squares between the observed data and the model trajectories with respect to the unknown parameters; this nonlinear least squares (NLS) approach does not require distributional assumptions. A likelihood-based approach, which estimates the parameters by maximizing the likelihood function of the observed data given the numerical solution, can be used (Bihorel, 2011). A corresponding Bayesian approach imposes priors on the parameters and can use Markov chain Monte Carlo (MCMC) samplers for inference (Boersch-Supan et al., 2017). While these numerical solver-based methods are better known in the context of ODEs, they are also generally applicable to DDE parameter inference with the help of a numerical solver that supports DDEs (Bellen and Zennaro, 2013). Even in the ODE setting, the repeated use of numerical solvers required for parameter estimation incurs a high computational cost. Numerically solving DDE systems further requires an appropriate interpolation scheme for the historical output; moreover, the historical output can reduce the stability of

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

the system and introduce additional oscillations, so that the solver requires smaller discretization time steps and higher computational cost. Even a system that is stable when modeled with ODEs can become unstable with the addition of a delay (Erneux, 2009). A simple DDE example is presented in Section S1 and Figure 1 of the Supplementary Material, which illustrates the trajectory's sensitivity to small variations in the delay parameter and the increased cost of numerical solvers in the DDE setting. As noted by Liang and Wu (2008), optimization-based algorithms using a numerical solver (e.g., NLS) may only converge to local optima due to the sensitivity of the numerical solution to the parameters and initial conditions. Therefore, it is advantageous to consider methods that approximate the DE solution and conduct parameter inference with the help of a statistical model.

Methods based on collocation and Gaussian processes encompass the two main types of statistical approaches for DE inference without the use of numerical solvers, which we subsequently review. Other related works include the recent development of Physics-Informed Neural Networks (PINNs) in deep learning, which integrate physical laws into neural networks by applying penalty terms to enforce differential equation constraints (Karniadakis et al., 2021). However, unlike statistical approaches, PINNs fail to properly quantify uncertainty in the estimated parameters and trajec-

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

tories; while Bayesian PINNs have been proposed to partially address this limitation (Yang et al., 2021), they nonetheless tend to be computationally inefficient for inverse problems due to the complexity of training.

To bypass numerical solvers, the class of collocation methods approximates the DE solution via basis function expansion (Varah, 1982; Ramsay et al., 2007; Wang and Cao, 2012; Wang et al., 2022). Introducing this idea for ODEs, Varah (1982) used splines to first smooth the noisy data and then in a second step applied least squares on the difference between the spline derivative and the ODEs; doing so requires a relatively accurate estimate of $\boldsymbol{x}(t)$ in the first step. Ellner et al. (1997) applied similar strategies for the case of DDEs. Ramsay et al. (2007) pioneered a generalized profiling procedure for ODEs, wherein the model parameters and spline coefficients are optimized together using a penalized likelihood, to achieve a balance between data fitting and fidelity to the ODE. Extending this approach to a semiparametric method for inference of DDEs, Wang and Cao (2012) used cross-validation to choose the smoothing parameter in the penalized likelihood and noted that a denser set of knots may be needed to handle sharp changes and oscillations in the DDE solution. Subsequently, Wang et al. (2022) proposed a more comprehensive Bayesian approach to collocation for ODE and DDE models, and developed an annealed sequential Monte Carlo

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

(SMC) algorithm to draw posterior samples. By treating the smoothing parameter as part of the posterior distribution, the method of Wang et al. (2022) avoids the potential drawbacks of over-fitting and expensive computation associated with cross-validation, and might be regarded as the state-of-the-art collocation-based inference method for DDEs.

Within a Bayesian framework, Gaussian processes (GPs) can serve as an alternative to collocation methods. For ODE inference, GP-based methods begin by placing a GP prior with hyper-parameters ϕ on $\boldsymbol{x}(t)$, which gives a closed form of the conditional distribution of $\boldsymbol{x}'(t)$ given $\boldsymbol{x}(t)$ to help bypass numerical solvers (Calderhead et al., 2008; Dondelinger et al., 2013; Wenk et al., 2019; Yang et al., 2021). The earliest GP-based method used gradient matching together with heuristics to combine the GP and ODE specifications (Calderhead et al., 2008); an application to DDEs was also considered, but in practice would require numerical solvers for the delay parameters. Subsequent refinements to GP gradient matching (e.g., Dondelinger et al., 2013; Wenk et al., 2019) nonetheless did not resolve the incompatible specification of $\boldsymbol{x}'(t)$ between the GP (via $p\{\boldsymbol{x}'(t)|\boldsymbol{x}(t), \phi\}$) and the ODEs (via the functions f_i). To address this incompatibility, Yang et al. (2021) proposed the manifold-constrained Gaussian process inference (MAGI) method for ODEs. In contrast to gradient matching, MAGI explicitly conditions

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

$\boldsymbol{x}(t)$ on the manifold that satisfies the ODEs, thereby providing a coherent Bayesian framework for inference. MAGI demonstrated accurate parameter estimation and fast computation speed on sparse and noisy data; Wong et al. (2023) further showcased the promising performance of MAGI on a larger ODE model with 10 components and 16 unknown parameters. It is however challenging to incorporate historical outputs for DDEs into GP-based methods, since system outputs at a non-fixed set of time points are required (as $\boldsymbol{\tau}$ is random), which is computationally demanding due to the covariance structure of GPs. Historical outputs pose a similar challenge for probabilistic ODE solvers that employ fixed time steps (Tronarp et al., 2022; Wu and Lysy, 2024). To the best of our knowledge, the use of GPs to facilitate parameter inference for DDEs, without any numerical solvers, remains to be explored.

Therefore, as the main contribution of this paper, we develop methodology that extends the MAGI GP-based framework to DDEs, given its principled Bayesian construction and favorable performance on ODEs. The key challenge is to incorporate historical outputs of DDEs into the GP manifold constraints, while achieving fast computation speed and estimation accuracy. We tackle this challenge by employing a linear interpolation scheme that only involves sparse matrix computations and simplifies the formula-

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

tion of the manifold constraints. Theoretical error bounds for the resulting $\boldsymbol{x}'(t)$ along with numerical validations are provided to justify the approximation. Our proposed method, MAnifold-constrained Gaussian process inference for delay differential equations (MAGIDDE), is applied to simulated and real data examples. We obtain fast and accurate inference for the parameters and system trajectories, compared to other representative methods for DDE inference. The R package and code and that provide our method implementation for the examples in this paper are available at `https://github.com/YuxuanZhao1/magidde`.

2. Review of the MAGI method for ODE inference

In this section, we review the MAGI method (Yang et al., 2021) as applied to parameter inference for ODEs. In the ODE setting, (1.1) reduces to $dx_i(t)/dt = f_i\{\boldsymbol{x}(t), \boldsymbol{\theta}, t\}$, $i = 1, \ldots, m$, or $d\boldsymbol{x}(t)/dt = \mathbf{f}\{\boldsymbol{x}(t), \boldsymbol{\theta}, t\}$ in vector form. MAGI imposes an independent GP prior on each component $x_i(t)$ such that

$$x_i(t) \sim \mathcal{GP}(\mu_i, \mathcal{K}_i), \quad t \in [0, T], \tag{2.1}$$

where $\mathcal{K}_i : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ is a positive definite covariance kernel and $\mu_i : \mathbb{R} \to \mathbb{R}$ is the mean function, usually taken to be $\mu_i(t) \equiv 0$. Let $\pi(\cdot)$ generically denote the prior on the model parameters $\boldsymbol{\theta}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

The data consist of the noisy observations $\boldsymbol{y}(\boldsymbol{\gamma}) = (\boldsymbol{y}_1(\boldsymbol{\gamma}_1), \cdots, \boldsymbol{y}_m(\boldsymbol{\gamma}_m))$, where $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_m)$ denotes the collection of observation time points for each component, i.e., component i is observed at the N_i time points $\boldsymbol{\gamma}_i = (\gamma_{i,1}, \cdots \gamma_{i,N_i})$. Assume that the observed data $\boldsymbol{y}_i(\boldsymbol{\gamma}_i)$ are subject to additive Gaussian noise, i.e.,

$$\boldsymbol{y}_i(\boldsymbol{\gamma}_i) = \boldsymbol{x}_i(\boldsymbol{\gamma}_i) + \boldsymbol{\epsilon}_i(\boldsymbol{\gamma}_i), \quad \boldsymbol{\epsilon}_i(\boldsymbol{\gamma}_i) \sim N(\mathbf{0}, \sigma_i^2 \mathbf{I}_{N_i}), \tag{2.2}$$

where $\mathbf{I}_{N_i}$ is an $N_i \times N_i$ dimensional identity matrix. (The notation t is used to represent time generically, while $\boldsymbol{\gamma}$ denotes the observation time points.)

With a suitably chosen $\mathcal{K}_i$, the conditional distribution of $\boldsymbol{x}'(t)$ given $\boldsymbol{x}(t)$ is also a GP with a fully specified mean function and covariance kernel. To link this GP-specified distribution of $\boldsymbol{x}'(t)$ with the ODE model structure, define W as a random variable measuring the uniform deviation between the stochastic process and ODE, i.e., $W = \sup_{t \in [0,T], i \in \{1, \cdots, m\}} \left| x_i'(t) - f_i \left\{ \boldsymbol{x}(t), \boldsymbol{\theta}, t \right\} \right|$. Setting $W = 0$ thus constrains the GP to lie on the manifold that satisfies the ODEs. In practice, W needs to be approximated by taking the uniform deviation over a finite set of n discretization points $\boldsymbol{I} = \{t_1, \cdots, t_n\}$, so we define $W_{\boldsymbol{I}} = \max_{t \in \boldsymbol{I}, i \in \{1, \cdots, m\}} \left| x_i'(t) - f_i \left\{ \boldsymbol{x}(t), \boldsymbol{\theta}, t \right\} \right|$, with $\boldsymbol{\gamma} \subset \boldsymbol{I} \subset [0, T]$. The key idea is to then condition the GP on the manifold constraint $W = 0$, as approximated by setting $W_{\boldsymbol{I}} = 0$. Following a Bayesian framework, the computable joint posterior of $\boldsymbol{\theta}$ and $\boldsymbol{x}(\boldsymbol{I})$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

conditional on $W_{\boldsymbol{I}} = 0$ and the noisy measurements $\boldsymbol{y}(\boldsymbol{\gamma})$ is given by

$$\begin{aligned} & p\left\{\boldsymbol{\theta}, \boldsymbol{x}(\boldsymbol{I}) | W_{\boldsymbol{I}} = 0, \boldsymbol{y}(\boldsymbol{\gamma})\right\} \propto p\left\{\boldsymbol{\theta}, \boldsymbol{x}(\boldsymbol{I}), W_{\boldsymbol{I}} = 0, \boldsymbol{y}(\boldsymbol{\gamma})\right\} \\ =& \pi(\boldsymbol{\theta}) \times p\left\{\boldsymbol{x}(\boldsymbol{I})\right\} \times p\left\{\boldsymbol{y}(\boldsymbol{\gamma}) | \boldsymbol{x}(\boldsymbol{I})\right\} \times p\left[\boldsymbol{x}'(\boldsymbol{I}) = \mathbf{f}\left\{\boldsymbol{x}(\boldsymbol{I}), \boldsymbol{\theta}, \boldsymbol{I}\right\} | \boldsymbol{x}(\boldsymbol{I})\right]. \end{aligned} \tag{2.3}$$

A brief description of these four terms follows, with details of the closed forms for each term in (2.3) provided in Section S2 of the Supplementary Material. The prior density of the model parameters is $\pi(\boldsymbol{\theta})$. As any finite collection of random variables from the GP $\boldsymbol{x}(t)$ defined in (2.1) follows a multivariate normal distribution, $p\left\{\boldsymbol{x}(\boldsymbol{I})\right\}$ is multivariate normal. According to (2.2), $p\left\{\boldsymbol{y}(\boldsymbol{\gamma}) | \boldsymbol{x}(\boldsymbol{I})\right\}$ is the normal likelihood of the noisy observations. The last term evaluates the GP $\boldsymbol{x}'(\boldsymbol{I})$ at $\boldsymbol{x}'(\boldsymbol{I}) = \mathbf{f}\left\{\boldsymbol{x}(\boldsymbol{I}), \boldsymbol{\theta}, \boldsymbol{I}\right\}$ to satisfy $W_{\boldsymbol{I}} = 0$, and hence is also multivariate normal.

Hamiltonian Monte Carlo (HMC, Neal, 2011) is used to draw samples of $\boldsymbol{\theta}$ and $\boldsymbol{x}(\boldsymbol{I})$ from (2.3). Denser discretization sets $\boldsymbol{I}$ may provide more accurate inference as the manifold constraint is better approximated, at the cost of computation time. Yang et al. (2021) notes that as the cardinality $|\boldsymbol{I}|$ increases, the terms involving the GP prior in (2.3), namely $p\left\{\boldsymbol{x}(\boldsymbol{I})\right\}$ and $p\left[\boldsymbol{x}'(\boldsymbol{I}) = \mathbf{f}\left\{\boldsymbol{x}(\boldsymbol{I}), \boldsymbol{\theta}, \boldsymbol{I}\right\} | \boldsymbol{x}(\boldsymbol{I})\right]$, will become more dominant, while the likelihood remains unchanged, i.e., only the points in the observation set $\boldsymbol{\gamma}$ contribute to the likelihood term $p\left\{\boldsymbol{y}(\boldsymbol{\gamma}) | \boldsymbol{x}(\boldsymbol{I})\right\}$. To achieve a balance

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

between the likelihood and GP prior, a tempering hyper-parameter $\beta = m|\boldsymbol{I}|/\sum_{i=1}^{m}|\boldsymbol{\gamma}_i|$ is introduced, i.e., the ratio between the total number of discretization points and the total number of observations. The GP-related terms are then tempered as $(p\{\boldsymbol{x}(\boldsymbol{I})\} \times p[\boldsymbol{x}'(\boldsymbol{I}) = \mathbf{f}\{\boldsymbol{x}(\boldsymbol{I}), \boldsymbol{\theta}, \boldsymbol{I}\}|\boldsymbol{x}(\boldsymbol{I})])^{1/\beta}$.

3. Methodology

Our main contribution is the MAGIDDE method which provides fast and accurate inference for DDEs, built upon the MAGI inference framework. To fix ideas, we begin by considering a fully Bayesian construction that incorporates the time-delay parameters $\boldsymbol{\tau}$ and historical outputs $\boldsymbol{x}(t-\boldsymbol{\tau})$ into the posterior distribution. A key challenge that becomes apparent is that the historical outputs are not directly available and require a computationally intensive sampling step. To address this challenge, we consider two approximation schemes and derive their corresponding theoretical error bounds. Our practical implementation of MAGIDDE uses a linear interpolation scheme for the historical outputs, which results in a computationally efficient method that maintains estimation accuracy.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

3.1 Fully Bayesian Scheme

We first consider a fully Bayesian scheme that extends MAGI to the DDE setting. To incorporate the time-delay parameters, we place a general prior $\pi(\cdot)$ on $\boldsymbol{\tau}$ and define $W_{\boldsymbol{I}} = \max_{t\in\boldsymbol{I},i\in\{1,\cdots,m\}} |x_i'(t) - f_i\{\boldsymbol{x}(t), \boldsymbol{x}(t-\boldsymbol{\tau}), \boldsymbol{\theta}, t\}|$. The full posterior distribution then consists of $\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{x}(\boldsymbol{I}), \boldsymbol{x}(\boldsymbol{I}-\boldsymbol{\tau})$ and can be written as

$$p\{\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{x}(\boldsymbol{I}), \boldsymbol{x}(\boldsymbol{I}-\boldsymbol{\tau}) | W_{\boldsymbol{I}} = 0, \boldsymbol{y}(\boldsymbol{\gamma})\} \propto p\{\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{x}(\boldsymbol{I}), \boldsymbol{x}(\boldsymbol{I}-\boldsymbol{\tau}), W_{\boldsymbol{I}} = 0, \boldsymbol{y}(\boldsymbol{\gamma})\}, \tag{3.1}$$

where $\boldsymbol{x}(\boldsymbol{I}-\boldsymbol{\tau}) = (\boldsymbol{x}_1(\boldsymbol{I}-\tau_1), \cdots, \boldsymbol{x}_m(\boldsymbol{I}-\tau_m))$, and each $\boldsymbol{x}_i(\boldsymbol{I}-\tau_i) = (x_i(t_1-\tau_i), \cdots, x_i(t_n-\tau_i))^\top$.

Factorizing (3.1) yields

$$p\{\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{x}(\boldsymbol{I}), \boldsymbol{x}(\boldsymbol{I}-\boldsymbol{\tau}), W_{\boldsymbol{I}} = 0, \boldsymbol{y}(\boldsymbol{\gamma})\} = \underbrace{\pi(\boldsymbol{\theta}) \times \pi(\boldsymbol{\tau})}_{(1)} \times \underbrace{p\{\boldsymbol{x}(\boldsymbol{I}-\boldsymbol{\tau}), \boldsymbol{x}(\boldsymbol{I}) | \boldsymbol{\theta}, \boldsymbol{\tau}\}}_{(2)}$$
$$\times \underbrace{p\{\boldsymbol{y}(\boldsymbol{\gamma}) | \boldsymbol{x}(\boldsymbol{I}-\boldsymbol{\tau}), \boldsymbol{x}(\boldsymbol{I}), \boldsymbol{\theta}, \boldsymbol{\tau}\}}_{(3)} \times \underbrace{p\{W_{\boldsymbol{I}} = 0 | \boldsymbol{y}(\boldsymbol{\gamma}), \boldsymbol{x}(\boldsymbol{I}-\boldsymbol{\tau}), \boldsymbol{x}(\boldsymbol{I}), \boldsymbol{\theta}, \boldsymbol{\tau}\}}_{(4)}.$$

Due to the prior independence between the GP and parameters in $\boldsymbol{\theta}$, the second term simplifies as $p\{\boldsymbol{x}(\boldsymbol{I}), \boldsymbol{x}(\boldsymbol{I}-\boldsymbol{\tau}) | \boldsymbol{\tau}\}$. Likewise, the likelihood of the noisy observations at time points $\boldsymbol{\gamma}$ does not depend on $\boldsymbol{\theta}$, $\boldsymbol{\tau}$, and $\boldsymbol{x}(\boldsymbol{I}-\boldsymbol{\tau})$, so the third term simplifies to $p\{\boldsymbol{y}(\boldsymbol{\gamma}) | \boldsymbol{x}(\boldsymbol{I})\}$. For the fourth term, after substituting the definition of $W_{\boldsymbol{I}} = 0$, the resulting density of

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

$\boldsymbol{x}'(\boldsymbol{I})$ depends on $\boldsymbol{x}(\boldsymbol{I})$, $\boldsymbol{x}(\boldsymbol{I}-\boldsymbol{\tau})$, and $\boldsymbol{\tau}$. Hence, we obtain

$$p\left\{\boldsymbol{\theta},\boldsymbol{\tau},\boldsymbol{x}(\boldsymbol{I}),\boldsymbol{x}(\boldsymbol{I}-\boldsymbol{\tau}),W_{\boldsymbol{I}}=0,\boldsymbol{y}(\boldsymbol{\gamma})\right\}=\underbrace{\pi(\boldsymbol{\theta})\times\pi(\boldsymbol{\tau})}_{(1)}\times\underbrace{p\left\{\boldsymbol{x}(\boldsymbol{I}-\boldsymbol{\tau}),\boldsymbol{x}(\boldsymbol{I})|\boldsymbol{\tau}\right\}}_{(2)}\times\underbrace{p\left\{\boldsymbol{y}(\boldsymbol{\gamma})|\boldsymbol{x}(\boldsymbol{I})\right\}}_{(3)}$$

$$\times\underbrace{p\left[\boldsymbol{x}'(\boldsymbol{I})=\mathbf{f}\left\{\boldsymbol{x}(\boldsymbol{I}),\boldsymbol{x}(\boldsymbol{I}-\boldsymbol{\tau}),\boldsymbol{\theta},\boldsymbol{I}\right\}|\boldsymbol{x}(\boldsymbol{I}-\boldsymbol{\tau}),\boldsymbol{x}(\boldsymbol{I}),\boldsymbol{\tau}\right]}_{(4)}.$$

The first term is the prior density of the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$. For the second term, the joint distribution of $\boldsymbol{x}_i(\boldsymbol{I}-\tau_i)$ and $\boldsymbol{x}_i(\boldsymbol{I})$ given τ_i for the $i-$th component is multivariate normal from the GP prior for $x_i(t)$, i.e.,

$$\boldsymbol{x}_i(\boldsymbol{I}-\tau_i),\boldsymbol{x}_i(\boldsymbol{I})|\tau_i\sim N\left\{\begin{pmatrix}\boldsymbol{\mu}_i(\boldsymbol{I}-\tau_i)\\ \boldsymbol{\mu}_i(\boldsymbol{I})\end{pmatrix},\begin{pmatrix}\mathcal{K}_i(\boldsymbol{I}-\tau_i,\boldsymbol{I}-\tau_i) & \mathcal{K}_i(\boldsymbol{I}-\tau_i,\boldsymbol{I})\\ \mathcal{K}_i(\boldsymbol{I},\boldsymbol{I}-\tau_i) & \mathcal{K}_i(\boldsymbol{I},\boldsymbol{I})\end{pmatrix}\right\},\tag{3.2}$$

where $\boldsymbol{\mu}(\boldsymbol{I}-\tau_i)$, $\mathcal{K}_i(\boldsymbol{I}-\tau_i,\boldsymbol{I}-\tau_i)$ and $\mathcal{K}_i(\boldsymbol{I},\boldsymbol{I}-\tau_i)$ respectively represent the mean function at $\boldsymbol{I}-\tau_i$, covariance matrix of $\boldsymbol{x}_i(\boldsymbol{I}-\tau_i)$ and cross-covariance matrix between $\boldsymbol{x}_i(\boldsymbol{I})$ and $\boldsymbol{x}_i(\boldsymbol{I}-\tau_i)$. The third term is the normal likelihood of the noisy observations. The fourth term evaluates the density of $\boldsymbol{x}'(\boldsymbol{I})$ at $\mathbf{f}\left\{\boldsymbol{x}(\boldsymbol{I}),\boldsymbol{x}(\boldsymbol{I}-\boldsymbol{\tau}),\boldsymbol{\theta},\boldsymbol{I}\right\}$, and by the property of GPs $\boldsymbol{x}'(\boldsymbol{I})|\boldsymbol{x}(\boldsymbol{I}-\boldsymbol{\tau}),\boldsymbol{x}(\boldsymbol{I}),\boldsymbol{\tau}$ has a multivariate normal distribution, provided the covariance kernel $\mathcal{K}$ is associated with twice-differentiable curves, i.e.,

$$\boldsymbol{x}_i'(\boldsymbol{I})|\boldsymbol{x}_i(\boldsymbol{I}-\tau_i),\boldsymbol{x}_i(\boldsymbol{I}),\tau_i\sim N\left[\boldsymbol{\mu}_i'(\boldsymbol{I})+\boldsymbol{m}_i\left\{\begin{pmatrix}\boldsymbol{x}_i(\boldsymbol{I}-\tau_i)\\ \boldsymbol{x}_i(\boldsymbol{I})\end{pmatrix}-\begin{pmatrix}\boldsymbol{\mu}_i(\boldsymbol{I}-\tau_i)\\ \boldsymbol{\mu}_i(\boldsymbol{I})\end{pmatrix}\right\},\boldsymbol{\zeta}_i\right],\tag{3.3}$$

where $\boldsymbol{m}_i=\begin{pmatrix}'\mathcal{K}_i(\boldsymbol{I},\boldsymbol{I}-\tau_i) & '\mathcal{K}_i(\boldsymbol{I},\boldsymbol{I})\end{pmatrix}\begin{pmatrix}\mathcal{K}_i(\boldsymbol{I}-\tau_i,\boldsymbol{I}-\tau_i) & \mathcal{K}_i(\boldsymbol{I}-\tau_i,\boldsymbol{I})\\ \mathcal{K}_i(\boldsymbol{I},\boldsymbol{I}-\tau_i) & \mathcal{K}_i(\boldsymbol{I},\boldsymbol{I})\end{pmatrix}^{-1}$, and $\boldsymbol{\zeta}_i=\mathcal{K}_i''(\boldsymbol{I},\boldsymbol{I})-\boldsymbol{m}_i\begin{pmatrix}\mathcal{K}_i'(\boldsymbol{I}-\tau_i,\boldsymbol{I})\\ \mathcal{K}_i'(\boldsymbol{I},\boldsymbol{I})\end{pmatrix}$ with $'\mathcal{K}_i=\frac{\partial}{\partial s}\mathcal{K}_i(s,t)$, $\mathcal{K}_i'=\frac{\partial}{\partial t}\mathcal{K}_i(s,t)$, and $\mathcal{K}_i''=\frac{\partial^2}{\partial s\partial t}\mathcal{K}_i(s,t)$ for any two time points s and t.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

In practice, we suggest using the Matern class where the covariance of the i-th component between time points s and t is given by $\mathcal{K}_i(s,t) = \phi_{i,1}\frac{2^{1-\nu}}{\Gamma(\nu)}\left(\sqrt{2\nu}\frac{d}{\phi_{i,2}}\right)^{\nu} B_\nu\left(\sqrt{2\nu}\frac{d}{\phi_{i,2}}\right)$, where $d = |s-t|$, Γ is the Gamma function, B_ν is the modified Bessel function of the second kind, and ν is the degree of freedom. Reasonable values are $\nu = 2.01$ or 2.5 to satisfy the requirement of twice-differentiable curves; $\nu = 2.01$ is a good default choice that is suitable for rougher curves, while $\nu = 2.5$ is adequate for smoother curves and has faster computation speed (Wong et al., 2024). $\mathcal{K}_i(s,t)$ has two hyper-parameters, $\phi_{i,1}$ and $\phi_{i,2}$, that respectively control the overall variance and bandwidth of the i-th component.

However, sampling the historical outputs from the conditional distribution in (3.2) (and subsequently $\boldsymbol{x}_i'(\boldsymbol{I})$ from (3.3)) is a computational bottleneck, as recalculation of cross-covariances and the conditional covariance matrix $\boldsymbol{\zeta}_i$ is required each time the value of τ_i is updated. Thus in what follows, we describe two schemes for approximating $\boldsymbol{x}(\boldsymbol{I}-\boldsymbol{\tau})$: conditional expectation and linear interpolation. For both schemes, we construct the approximation and derive theoretical error bounds for the derivatives of the system outputs.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

3.2 Approximation of Historical Outputs

To reduce the computational cost of the fully Bayesian scheme, we consider deterministic approximations for $\boldsymbol{x}(\boldsymbol{I}-\boldsymbol{\tau})$ in terms of $\boldsymbol{x}(\boldsymbol{I})$ and $\boldsymbol{\tau}$. Let $\hat{\boldsymbol{x}}(\boldsymbol{I}-\boldsymbol{\tau})$ denote the approximated values of the historical outputs. By dropping the explicit dependence on $\boldsymbol{x}(\boldsymbol{I}-\boldsymbol{\tau})$ in (3.1) and treating it as known, the factorization of the joint posterior simplifies to

$$\begin{aligned} p\left\{\boldsymbol{\theta},\boldsymbol{\tau},\boldsymbol{x}(\boldsymbol{I})|W_{\boldsymbol{I}}=0,\boldsymbol{y}(\boldsymbol{\gamma})\right\} \propto & p\left\{\boldsymbol{\theta},\boldsymbol{\tau},\boldsymbol{x}(\boldsymbol{I}),W_{\boldsymbol{I}}=0,\boldsymbol{y}(\boldsymbol{\gamma})\right\} \\ = & \pi(\boldsymbol{\theta})\times\pi(\boldsymbol{\tau})\times p\left\{\boldsymbol{x}(\boldsymbol{I})|\boldsymbol{\theta},\boldsymbol{\tau}\right\}\times p\left\{\boldsymbol{y}(\boldsymbol{\gamma})|\boldsymbol{x}(\boldsymbol{I}),\boldsymbol{\theta},\boldsymbol{\tau}\right\} \\ & \times p\left[\boldsymbol{x}'(\boldsymbol{I})=\mathrm{f}\left\{\boldsymbol{x}(\boldsymbol{I}),\boldsymbol{x}(\boldsymbol{I}-\boldsymbol{\tau}),\boldsymbol{\theta},\boldsymbol{I}\right\}|\boldsymbol{y}(\boldsymbol{\gamma}),\boldsymbol{x}(\boldsymbol{I}),\boldsymbol{\theta},\boldsymbol{\tau}\right]. \end{aligned}$$

Then, using the previous conditional independence properties and substituting the approximation $\hat{\boldsymbol{x}}(\boldsymbol{I}-\boldsymbol{\tau})$ for $\boldsymbol{x}(\boldsymbol{I}-\boldsymbol{\tau})$ yields

$$\begin{aligned} & p\left\{\boldsymbol{\theta},\boldsymbol{\tau},\boldsymbol{x}(\boldsymbol{I})|W_{\boldsymbol{I}}=0,\boldsymbol{y}(\boldsymbol{\gamma})\right\} \propto \pi(\boldsymbol{\theta})\times\pi(\boldsymbol{\tau})\times p\left\{\boldsymbol{x}(\boldsymbol{I})\right\}\times p\left\{\boldsymbol{y}(\boldsymbol{\gamma})|\boldsymbol{x}(\boldsymbol{I})\right\} \\ & \times p\left[\boldsymbol{x}'(\boldsymbol{I})=\mathrm{f}\left\{\boldsymbol{x}(\boldsymbol{I}),\hat{\boldsymbol{x}}(\boldsymbol{I}-\boldsymbol{\tau}),\boldsymbol{\theta},\boldsymbol{I}\right\}|\boldsymbol{x}(\boldsymbol{I}),\boldsymbol{\tau}\right]. \end{aligned} \tag{3.4}$$

A first approximation scheme for $\boldsymbol{x}(\boldsymbol{I}-\boldsymbol{\tau})$ is via conditional expectation. Specifically, $E\left\{\boldsymbol{x}_i(\boldsymbol{I}-\tau_i)|\boldsymbol{x}_i(\boldsymbol{I})\right\}$ stands as a natural candidate to approximate $\boldsymbol{x}_i(\boldsymbol{I}-\tau_i)$, since it is the best linear predictor and has a closed-form expression (Wang et al., 2020). Thus, we have the approximation

$$\hat{\boldsymbol{x}}_i(\boldsymbol{I}-\tau_i)=E\left\{\boldsymbol{x}_i(\boldsymbol{I}-\tau_i)|\boldsymbol{x}_i(\boldsymbol{I})\right\}=\boldsymbol{\mu}_i(\boldsymbol{I}-\tau_i)+\mathcal{K}_i(\boldsymbol{I},\boldsymbol{I}-\tau_i)\mathcal{K}_i(\boldsymbol{I},\boldsymbol{I})^{-1}\left\{\boldsymbol{x}_i(\boldsymbol{I})-\boldsymbol{\mu}_i(\boldsymbol{I})\right\}. \tag{3.5}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

A second approximation scheme for $\boldsymbol{x}(\boldsymbol{I}-\boldsymbol{\tau})$ uses linear interpolation. Define the interpolated weight $w(t)=\sum_{j=1}^{n-1}\mathbb{1}(t\in[t_j,t_{j+1}])\times\frac{t-t_j}{t_{j+1}-t_j}$, where $\mathbb{1}(\cdot)$ is the indicator function. Then a linearly interpolated value of the i-th component at any $t\in\{\boldsymbol{I}-\tau_i\}$ is given by

$$\hat{x}_i(t)=\sum_{j=1}^{n-1}\mathbb{1}(t\in[t_j,t_{j+1}])\times\{(1-w(t))\times x_i(t_j)+w(t)\times x_i(t_{j+1})\}\,. \quad (3.6)$$

(3.6) can be written in matrix form. Define a $n\times n$ scalar matrix $\boldsymbol{S}_i$ with its (j,q)-th entry given by

$$S_i(j,q)=\begin{cases}1-w(t_j-\tau_i) & \text{for } j\in\{1,\cdots,n\}, q=k(t_j-\tau_i)\\ w(t_j-\tau_i) & \text{for } j\in\{1,\cdots,n\}, q=k(t_j-\tau_i)+1\\ 0 & \text{otherwise}\end{cases}$$

where $k(t_j-\tau_i)$ denotes the integer $c\in\{1,\cdots,n\}$ satisfying $t_j-\tau_i\in[t_c,t_{c+1}]$. Then (3.6) can be rewritten as $\hat{\boldsymbol{x}}_i(\boldsymbol{I}-\tau_i)=\boldsymbol{S}_i\cdot\boldsymbol{x}_i(\boldsymbol{I})$.

Note that both approximation schemes can be viewed as a deterministic linear transformation of $\boldsymbol{x}(\boldsymbol{I})$. Given the history function $\mathcal{H}_{\tau_i}$, the approximation schemes are only applied to the historical output $x_i(t-\tau_i)$ for $t\in\boldsymbol{I}$ where $t>\tau_i$, as we have $x_i(t-\tau_i)=x_i(0)$ for $t\leq\tau_i$. Conditional expectation eliminates the need to work with the full covariance matrix of $(\boldsymbol{x}_i(\boldsymbol{I}-\tau_i),\boldsymbol{x}_i(\boldsymbol{I}))^\top$ and draw samples of $\boldsymbol{x}_i(\boldsymbol{I}-\tau_i)$ from the larger joint posterior in (3.1). However, it still requires computation of the cross-covariance $\mathcal{K}_i(\boldsymbol{I},\boldsymbol{I}-\tau_i)$ and dense matrix multiplication. The computation cost can be

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

substantively reduced further using linear interpolation, as then only sparse matrix ($\boldsymbol{S}_i$) operations are needed.

We next investigate the approximation error of $\mathbf{f}\{\boldsymbol{x}(\boldsymbol{I}), \hat{\boldsymbol{x}}(\boldsymbol{I}-\boldsymbol{\tau}), \boldsymbol{\theta}, \boldsymbol{I}\}$ under the two proposed schemes. It suffices to find the stochastic error bound for any $i \in \{1, \cdots, m\}$, of $\max_{t\in\boldsymbol{I}} \left| f_i\{\boldsymbol{x}(t), \boldsymbol{x}(t-\boldsymbol{\tau}), \boldsymbol{\theta}, t\} - f_i\{\boldsymbol{x}(t), \hat{\boldsymbol{x}}(t-\boldsymbol{\tau}), \boldsymbol{\theta}, t\} \right|$. Under some mild differentiability conditions (listed in Section S3 of the Supplementary Material), we have Theorems 1 and 2, with proofs provided in S4 and S5 of the Supplementary Material. These results guarantee that the approximation $f_i\{\boldsymbol{x}(t), \hat{\boldsymbol{x}}(t-\boldsymbol{\tau}), \boldsymbol{\theta}, t\}$ will be reasonable under either scheme, given that $|\boldsymbol{I}|$ is sufficiently large.

Theorem 1. *Suppose $x_i(t)$ is a GP with mean function $\mu_i(t)$ and Matern covariance kernel $\mathcal{K}_i$ with $\nu \in \{2.01, 2.5\}$, for $i = 1, \cdots, m$, and that the n points in the discretization set $\boldsymbol{I}$ are equally-spaced over $[0, T]$. If the values of the historical outputs are approximated by conditional expectation as described in (3.5), then for any $i \in \{1, \cdots, m\}$,*

$$\max_{t\in\boldsymbol{I}} \left| f_i\{\boldsymbol{x}(t), \boldsymbol{x}(t-\boldsymbol{\tau}), \boldsymbol{\theta}, t\} - f_i\{\boldsymbol{x}(t), \hat{\boldsymbol{x}}(t-\boldsymbol{\tau}), \boldsymbol{\theta}, t\} \right| = O_P\left[\{\log(n)\}^{\frac{1}{2}} n^{-\nu}\right].$$

Theorem 2. *Consider the same setup as in Theorem 1. If the values of the historical outputs are approximated by linear interpolation as described*

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

in (3.6)*, then for any* $i \in \{1, \cdots, m\}$*,*

$$\max_{t \in \boldsymbol{I}} \left| f_i \left\{ \boldsymbol{x}(t), \boldsymbol{x}(t-\boldsymbol{\tau}), \boldsymbol{\theta}, t \right\} - f_i \left\{ \boldsymbol{x}(t), \hat{\boldsymbol{x}}(t-\boldsymbol{\tau}), \boldsymbol{\theta}, t \right\} \right| = O_P \left[\{\log(n)\}^{\frac{1}{2}} n^{-1.5} \right].$$

Remark 1. Theorems 1 and 2 assume an equally-spaced discretization set for $\boldsymbol{I}$. While the MAGIDDE method does not require an equally-spaced $\boldsymbol{I}$, these results suggest that a good practical guideline is to choose an equally-spaced (or approximately equally-spaced) set that includes the observation times.

Remark 2. ν controls the level of smoothness of the Matern covariance; larger values of ν provide smoother covariance functions. If the degree of freedom in the Matern covariance kernel is set as $\nu = 1.5$, the rate of convergence of $f_i \left\{ \boldsymbol{x}(t), \hat{\boldsymbol{x}}(t-\boldsymbol{\tau}), \boldsymbol{\theta}, t \right\}$ will be similar whether we approximate the historical outputs by linear interpolation or conditional expectation. Noting that MAGIDDE chooses $\nu \in \{2.01, 2.5\}$, Theorems 1 and 2 indicate that the conditional expectation scheme has a faster convergence rate than linear interpolation.

Remark 3. If $\boldsymbol{\tau}$ in (1.1) is replaced by an equation-specific vector of time-delay parameters $\boldsymbol{\tau}_i = (\tau_{i,1}, \tau_{i,2}, \ldots, \tau_{i,m})$ for the i-th DDE, the results of Theorems 1 and 2 on $\max_{t \in \boldsymbol{I}} \left| f_i \left\{ \boldsymbol{x}(t), \boldsymbol{x}(t-\boldsymbol{\tau}_i), \boldsymbol{\theta}, t \right\} - f_i \left\{ \boldsymbol{x}(t), \hat{\boldsymbol{x}}(t-\boldsymbol{\tau}_i), \boldsymbol{\theta}, t \right\} \right|$ still hold under the same differentiability conditions, by simply substituting

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

$\boldsymbol{\tau}$ with $\boldsymbol{\tau}_i$.

Remark 2 suggests a trade-off between the faster theoretical convergence rate of conditional expectation and faster computation speed of linear interpolation. We numerically validate and compare the performance of the three schemes on a benchmark system in Section S6 of the Supplementary Material, using $|\boldsymbol{I}| = \{16, 31, 61, 121\}$ to illustrate. The results (Table 1, Figures 2 and 3 in the Supplementary Material) indicate that linear interpolation has a significant speed advantage over the other schemes and yields comparable estimates of parameters and trajectories when $|\boldsymbol{I}|$ is large enough, hence is our preferred choice in practice.

3.3 Practical Implementation

The practical steps to implement the MAGIDDE method with the linear approximation scheme are as follows. We begin by initializing the required parameter and hyper-parameter values needed for HMC sampling. First, we fit a GP to the noisy observations $\boldsymbol{y}(\boldsymbol{\gamma}_i)$ for each component i. If the noise level σ_i^2 is unknown, we obtain values of $\phi_{i,1}$, $\phi_{i,2}$ and σ_i^2 by maximizing the marginal likelihood $p\left\{\phi_{i,1}, \phi_{i,2}, \sigma_i^2 | \boldsymbol{y}(\boldsymbol{\gamma}_i)\right\}$; if σ_i^2 is known, we maximize the marginal likelihood $p\left\{\phi_{i,1}, \phi_{i,2} | \boldsymbol{y}(\boldsymbol{\gamma}_i)\right\}$ instead. The obtained values of the covariance hyper-parameters $\phi_{i,1}$ and $\phi_{i,2}$ are held fixed during subsequent

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

HMC sampling, as in Yang et al. (2021). Second, we use the observations $\boldsymbol{y}(\boldsymbol{\gamma})$ as the starting value of $\boldsymbol{x}(\boldsymbol{\gamma})$ for HMC sampling, and $\boldsymbol{x}(\boldsymbol{I})$ for time points in $\boldsymbol{I}\backslash\boldsymbol{\gamma}$ is initialized at the mean of the GP fit to the noisy observations. Third, we optimize the posterior in (3.4) with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$ (holding other quantities fixed at their initialized values) to provide their corresponding starting values for HMC sampling. Recall that under the approximation scheme, it is not necessary to explicitly initialize $\hat{\boldsymbol{x}}(\boldsymbol{I}-\boldsymbol{\tau})$, since it can be viewed as a function of $\boldsymbol{\tau}$ and $\boldsymbol{x}(\boldsymbol{I})$.

We then proceed to jointly sample $\boldsymbol{x}(\boldsymbol{I})$, together with unknown parameters in $\boldsymbol{\theta}$, $\boldsymbol{\tau}$, and $\boldsymbol{\sigma^2}$ from their posterior distribution (3.4) using HMC (Neal, 2011). Recall that we fix the GP hyper-parameters and use the tempering scheme described in Section 2 to balance the contributions of the GP prior and likelihood during HMC sampling. Our implementation of HMC adjusts the step sizes of the leapfrog integrator automatically during the burn-in phase to ensure that the acceptance rate falls within the range of 60% to 90%. After discarding the samples from the burn-in period, the posterior means of $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$ are treated as the parameter estimates. We call the posterior mean of $\boldsymbol{x}(\boldsymbol{I})$ the *inferred trajectory*, which represents our estimate of the system trajectory. The uncertainties of $\boldsymbol{\theta}$, $\boldsymbol{\tau}$, and $\boldsymbol{x}(\boldsymbol{I})$ are quantified via 95% pointwise credible intervals.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

An important practical consideration is the choice of discretization set $\boldsymbol{I}$ for computation. As a qualitative rule-of-thumb, $\boldsymbol{I}$ should be dense enough to infer a smooth trajectory for the system. In cases where this is not immediately apparent, we can follow the guideline suggested by Wong et al. (2024): start with the smallest equally-spaced set $\boldsymbol{I}_0$ that contains the observation time points $\boldsymbol{\gamma}$; then, construct $\boldsymbol{I}_j \supset \boldsymbol{I}_{j-1}$, $j \geq 1$ by inserting one equally-spaced point between each adjacent pair of points in $\boldsymbol{I}_j$ and re-run MAGIDDE, and stop when stable estimates are obtained (i.e., the credible intervals based on $\boldsymbol{I}_j$ and $\boldsymbol{I}_{j-1}$ largely overlap). Moreover, MAGIDDE provides a natural way to generate future predictions: by extending $\boldsymbol{I}$ to include time points beyond the last observation in $\boldsymbol{\gamma}$, the samples of $\boldsymbol{x}(\boldsymbol{I})$ for $\{t \in \boldsymbol{I} \mid t > \max(\boldsymbol{\gamma})\}$ constitute the model predictions for the future time points. Examples that illustrate these aspects of choosing $\boldsymbol{I}$ are provided in Sections S7 and S11.4 of the Supplementary Material.

4. Simulation Studies

We assess the performance of MAGIDDE and compare with other representative methods for DDE inference, using Hutchinson's equation (May, 1976) as a benchmark system in Section 4.1. To further demonstrate the capability of MAGIDDE to handle more complex models, we consider the

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

lac operon system of Yildirim and Mackey (2003) in Section 4.2.

4.1 Benchmark System

As a benchmark system, we focus on parameter inference for Hutchinson's equation to evaluate the estimation accuracy of different methods under varying observation sample sizes. Hutchinson's equation was proposed for modelling single-species population dynamics, and May (1976) introduced the following DDE for the population size of blowflies (Nicholson, 1954), $dP(t)/dt = rP(t) \times \{1 - P(t-\tau)/(1000 \times K)\}$, where $P(t)$ is the population size at time t, r is the reproduction rate, τ is a time delay, and $1000 \times K$ is the maximum population sustainable by the limited food supply. A noisy observation taken at time t is assumed to follow a lognormal distribution with mean $P(t)$ and variance σ^2 (Wang et al., 2022). Since blowfly counts are strictly positive, we consider a logarithmic transformation by defining $P(t) = \exp\{N(t)\}$. Then, the log-transformed DDE is

$$dN(t)/dt = r \times [1 - \exp\{N(t-\tau)\}/(1000 \times K)], \tag{4.1}$$

and we can equivalently estimate r, K and τ in (4.1) based on the noisy observations $y(t) \sim N\{N(t), \sigma^2\}$ for $t \in \boldsymbol{\gamma}$. Following Wang et al. (2022), we set the true values of the parameters $(\boldsymbol{\theta}, \boldsymbol{\tau}) = (r, K, \tau) = (0.8, 2, 3)$, the initial condition on the log-scale as $N(0) = \log(3500)$, the noise level $\sigma =$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

0.1, and the history function $\mathcal{H}_\tau = \{N(t) = N(0), t \in [-\tau, 0]\}$. We take the time interval of interest as $t \in [0, 30]$ (Wang and Cao, 2012), and consider scenarios with $|\boldsymbol{\gamma}| = 16$, 31, 61, and 121 equally-spaced observations. To create the simulation data, the R package 'deSolve' (Soetaert et al., 2010) is used to numerically solve the DDE trajectories as defined by (4.1).

We note that even for this simple DDE system, the NLS approach with a numerical solver often yields incorrect parameter estimates, because of convergence to local optima (as noted in the Introduction). The presence of local optima is associated with the numerical solution's sensitivity to the parameter values (especially the time delay τ, see Figure 1 in the Supplementary Material) and initial condition. Running multiple tries of NLS from different starting parameter values can only partially mitigate the convergence issues, and becomes computationally inefficient compared to model-based inference methods; see Supplementary Material S8.3 for a detailed analysis of this system using NLS. This suggests the NLS approach may not be recommended even for inferring simple DDE systems.

We compare MAGIDDE with two other representative methods for DDE inference: the 'deBInfer' R package and the semiparametric Bayesian collocation method (SMCDE, Wang et al., 2022). We describe these methods and how they are run to conduct parameter inference for each simulated

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

dataset in Supplementary Material S8.1. We use two metrics to assess the performance of a given method. First, we calculate the root mean square error (RMSE) of the estimated parameters relative to the truth. Second, following Yang et al. (2021), we also calculate the *trajectory RMSE* to assess the accuracy of the system trajectory implied by the parameter estimates. The trajectory RMSE is computed by the following steps: first, we use a numerical solver to construct the true trajectory based on the true parameter values and history function; then, we use the numerical solver to reconstruct the trajectory implied by the estimates of the parameters and history function. Last, we calculate the RMSE between the true trajectory and the reconstructed trajectory at the observation time points on the original scale of the measurements.

For each of the three methods, we compute the RMSEs of the parameters and reconstructed trajectories, as summarized in Table 1 based on 300 simulated datasets. Further discussion of these results, along with boxplots of the parameter estimates and trajectory RMSEs for each individual dataset, are provided in Supplementary Material S8.2. MAGIDDE consistently outperforms the other two methods in terms of recovering the trajectory and parameters with lower error. Sparse observations do not inherently pose a problem for MAGIDDE, since the accuracy of the linear

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

approximation scheme depends on the number of discretization points, and not the number of observations. To visualize the quality of the inferred trajectories, Figure 1 shows that MAGIDDE well-recovers the true underlying trajectory in the sparse 16 observation scenario; the 95% pointwise credible interval becomes narrower for scenarios with denser observations, e.g., Figure 9 in Supplementary Material S9 shows the corresponding plot for 61 observations. MAGIDDE is also the fastest of the three methods; while SMCDE is reasonably fast, deBInfer is an order of magnitude slower due to its dependence on the numerical DDE solver.

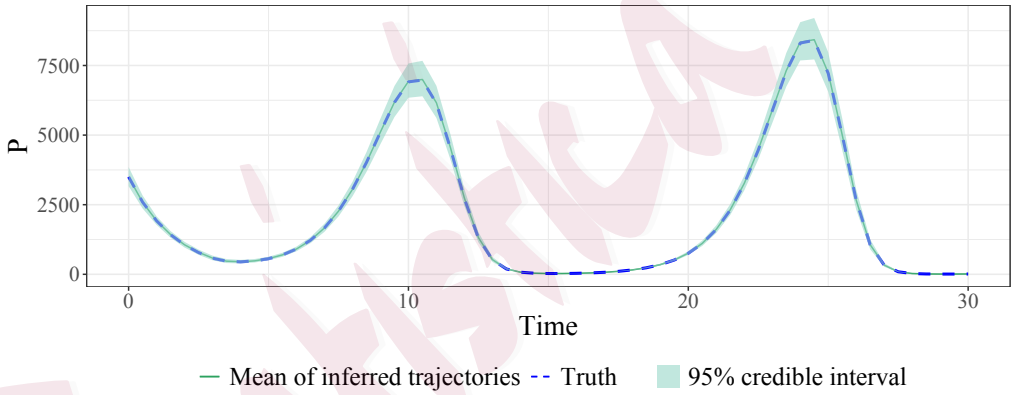

Figure 1: Inferred trajectory of Hutchinson's equation from 16 observations using MAGIDDE. The solid line represents the mean inferred trajectory over 300 simulated datasets, and the dashed line is the truth. The shaded area is the 95% pointwise credible interval, constructed by taking the average 0.025 and 0.975 quantiles of the inferred trajectories across the simulated datasets.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

Table 1: Average parameter estimates (with RMSEs in parentheses) for the log-transformed Hutchinson's equation and average trajectory RMSEs over the 300 simulated datasets. The last column gives the average runtime (in minutes, on a single CPU core).

$\|\gamma\|$	Method	r	K	τ	$N(0)$	σ	Trajectory	Runtime (min.)
16	MAGIDDE	0.80(< **0.01**)	2.00(**0.07**)	3.00(**0.01**)	8.16(**0.04**)	0.11(**0.02**)	**154.33**	**0.43**
	deBInfer	0.74(0.26)	1.79(0.62)	2.72(0.81)	8.04(0.33)	0.13(0.05)	618.48	14.4
	SMCDE	0.70(0.10)	2.01(1.08)	3.14(0.15)	8.14(0.10)	0.16(0.06)	1376.95	1.55
31	MAGIDDE	0.80(< **0.01**)	2.00(**0.05**)	3.00(**0.01**)	8.16(**0.03**)	0.10(**0.01**)	**126.11**	**0.43**
	deBInfer	0.75(0.25)	1.81(0.60)	2.75(0.79)	8.01(0.45)	0.13(0.05)	591.18	19.1
	SMCDE	0.76(0.05)	2.01(0.14)	2.96(0.07)	8.15(0.10)	0.16(0.06)	1233.87	1.99
61	MAGIDDE	0.80(< **0.01**)	2.00(**0.04**)	3.00(**0.01**)	8.16(**0.02**)	0.10(**0.01**)	**101.62**	**0.44**
	deBInfer	0.76(0.23)	1.83(0.55)	2.78(0.77)	7.99(0.42)	0.13(0.05)	598.99	27.0
	SMCDE	0.76(0.04)	2.00(0.06)	2.94(0.07)	8.15(0.09)	0.16(0.06)	1363.14	2.76
121	MAGIDDE	0.80(< **0.01**)	2.00(**0.02**)	3.00(**0.01**)	8.16(**0.02**)	0.10(**0.01**)	**61.34**	**1.16**
	deBInfer	0.74(0.25)	1.79(0.61)	2.72(0.89)	7.96(0.54)	0.13(0.05)	662.11	38.9
	SMCDE	0.77(0.04)	2.00(0.03)	2.94(0.07)	8.14(0.08)	0.16(0.06)	1325.22	3.91

4.2 *Lac* Operon System

Among previous methods for DDE inference, performance was typically assessed using simple DDE models with only one or two components with four or fewer parameters to be estimated (Wang and Cao, 2012; Dondelinger et al., 2013; Wang et al., 2022), e.g., Hutchinson's equation. In this section, we showcase MAGIDDE's capability to perform inference for a more complex DDE model with a larger number of parameters. The setup is sufficiently challenging such that the NLS approach struggles to converge

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

to reasonable parameter values and the associated computational burden of using the numerical solver becomes prohibitive; see Supplementary Material S10.3 for a detailed investigation of this model using NLS.

The *lac* operon is a well-studied gene expression system that enables bacteria to use lactose as an energy source when glucose is scarce (Pardee et al., 1959). Yildirim and Mackey (2003) modelled the regulation of induction in the *lac* operon in *Escherichia coli* via the following system of DDEs to describe the dynamics among five components, where β-galactosidase (B) converts intracellular lactose (L) into allolactose (A), which allows DNA transcription and mRNA (M) translation to increase the levels of β-galactosidase (B) and permease (P):

$$\begin{cases} M'(t) &= \alpha_M \times \left[\dfrac{1 + K_1 \left\{e^{-\mu\tau_M} A(t-\tau_M)\right\}^n}{K + K_1 \left\{e^{-\mu\tau_M} A(t-\tau_M)\right\}^n}\right] + \Gamma_0 - (\gamma_M + \mu) M(t) \\ B'(t) &= \alpha_B e^{-\mu\tau_B} M(t-\tau_B) - (\gamma_B + \mu) B(t) \\ A'(t) &= \alpha_A B(t) \dfrac{L(t)}{K_L + L(t)} - \beta_A B(t) \dfrac{A(t)}{K_A + A(t)} - (\gamma_A + \mu) A(t) \\ L'(t) &= \alpha_L P(t) \dfrac{L_e}{K_{L_e} + L_e} - \beta_{L_1} P(t) \dfrac{L(t)}{K_{L_1} + L(t)} - \beta_{L_2} B(t) \dfrac{L(t)}{K_L + L(t)} - (\gamma_L + \mu) L(t) \\ P'(t) &= \alpha_P e^{-\mu(\tau_B + \tau_P)} M\left\{t - (\tau_B + \tau_P)\right\} - (\gamma_P + \mu) P(t) \end{cases}.$$

Our goal is to estimate the time-delay parameters τ_B, τ_M, τ_P (which represent time required for transcription and translation), the model parameters $\gamma_A, \alpha_M, \alpha_B, \alpha_P, \mu$, and the initial conditions; other parameters are treated

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

as known. We simulate observations from the system at the collection of 23 unevenly-spaced time points $\boldsymbol{\gamma} = \{0, 0.25, 0.5, \cdots, 2, 3, 4, \cdots, 10, 12, 14, \cdots, 20, 25\}$. To mimic noisy measurements, we add Gaussian noise with standard deviation $\sigma_M = 3 \times 10^{-5}, \sigma_B = 1 \times 10^{-5}, \sigma_A = 0.02, \sigma_L = 0.01, \sigma_P = 5 \times 10^{-4}$ to the model trajectories of each component, which we assume to be known from repeated measures. Further scientific background of the system, choice of parameters to estimate and observation times, and a list of all the parameters and their corresponding values from Yildirim and Mackey (2003) are provided in Supplementary Material S10.1.

We generate 100 simulated datasets based on the above setup for the *lac* operon system. We provide the implementation details of MAGIDDE in Section S10.2 of the Supplementary Material. Table 2 summarizes the average parameter estimates and standard deviations across the 100 simulated datasets. MAGIDDE provides reasonable estimates of $\tau_B, \tau_P, \gamma_A, \alpha_M, \alpha_B, \alpha_P$, in terms of relatively low standard deviation and bias. However, τ_M tends to be overestimated, while μ tends to be underestimated, and both have a high standard deviation. A possible reason is that the observation set is not dense enough to capture the short time-delay of 0.1. Moreover, μ is closely related to τ_M as the product $\mu \times \tau_M$ appears in the exponential in the first equation of the system. Indeed, the product of the estimated μ and τ_M is

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

0.0032, which is relatively close to the true value of 0.0043.

Table 2: Average parameter estimates using MAGIDDE (with standard deviation after ± sign) for the *lac* operon model across 100 simulated datasets.

Parameter	Truth	Estimate	Parameter	Truth	Estimate	Parameter	Truth	Estimate
τ_B	2	2.0024 ± 0.3074	α_B	0.0166	0.0143 ± 0.0011	$A(0)$	0.038	0.0319 ± 0.0124
τ_M	0.1	0.2990 ± 0.1275	α_P	10	9.6792 ± 0.3904	$L(0)$	0.372	0.4018 ± 0.0131
τ_P	0.83	0.8283 ± 0.3063	μ	0.0226	0.0144 ± 0.0071	$P(0)$	0.0149	0.0143 ± 0.0004
γ_A	0.52	0.4916 ± 0.0255	$10^2 \times M(0)$	0.0626	0.0630 ± 0.0014			
$10^2 \times \alpha_M$	0.0997	0.0985 ± 0.0032	$10^3 \times B(0)$	0	0.0018 ± 0.0014			

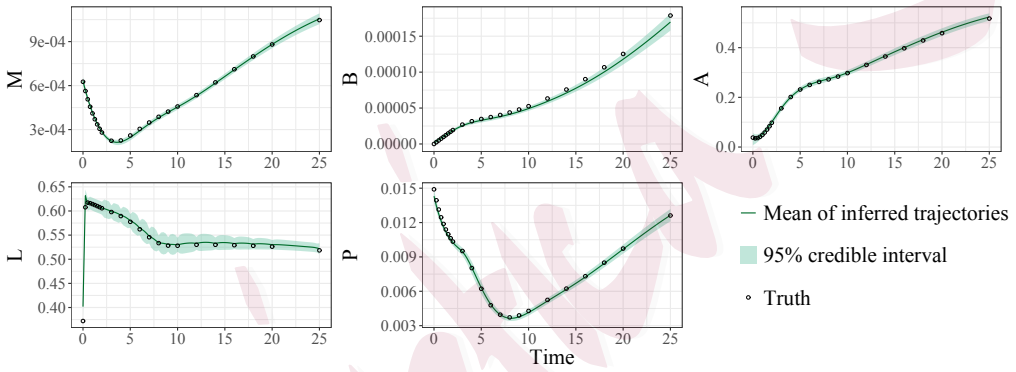

Figure 2: Inferred trajectories obtained by MAGIDDE for each component of the *lac* operon system over 100 simulated datasets. The solid line is the mean of the inferred trajectories. The shaded area represents the corresponding 95% pointwise credible interval, which is constructed by averaging the 0.025 and 0.975 quantiles of the inferred trajectories across the simulated datasets. The dots indicate the true model trajectory at the observation time points, which are unevenly spaced and relatively sparse for the last 15 minutes.

Figure 2 shows the mean inferred trajectories and 95% pointwise credible intervals obtained by MAGIDDE across the 100 simulated datasets,

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

which indicate that our method can reliably recover the trajectories for the five components. Most of the true model outputs are well-covered by the narrow interval, except for a few data points from the L and P components during the first half minute. The 95% credible intervals for the L component look different from the other four components, in terms of having a more "bumpy" shape; this is due to a small estimated GP bandwidth hyper-parameter (ϕ_2) to accommodate the sharp increase from the first observation point. The small GP bandwidth in the L component in turn leads to greater uncertainty in its estimated trajectory at discretization time points that are farther away from observations.

5. Application

This section applies MAGIDDE to estimate the parameters of a time-delayed compartmental model using COVID-19 data from Ontario, Canada. Omicron was the dominant variant in Ontario as of January 2022, and our focus is to infer the parameters for the Omicron variant over the 30-day observation period from January 24 to February 22, 2022, during the peak of the Omicron wave. To account for the disease incubation period, a time-delay parameter can be introduced to the basic SIR model (Ma et al., 2004). Inspired by Ma et al. (2004), we adapt their delayed SIR model

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

to accommodate the Ontario COVID-19 data of interest. First, we add a new compartment D to directly model the death counts associated with COVID-19. Second, we represent the population in each compartment as proportions. Third, we ignore natural birth and death rates since only a short time period is considered. This leads to our DDE system of interest as follows:

$$\begin{cases} S'(t) &= -\tilde{\beta} S(t) I(t-h) \\ I'(t) &= \tilde{\beta} S(t) I(t-h) - \mu_d I(t) - \lambda I(t) \\ R'(t) &= \lambda I(t) \\ D'(t) &= \mu_d I(t) \end{cases}, \tag{5.1}$$

where $\tilde{\beta}$ is the normalized disease transmission rate that governs the flow from the susceptible (S) to the infected (I) compartment, μ_d represents the death rate of infected individuals, and h is the time-delay accounting for the incubation period of the disease (in days), λ is the recovery rate that governs the flow from I to the recovered (R) compartment. Since $1 = S(t)+I(t)+R(t)+D(t)$, we drop the $S'(t)$ equation as it suffices to model (I, R, D) and recover S using this relationship. In Sections S11.1–S11.3 of the Supplementary Material, we describe the data processing steps to obtain noisy observations of the daily population size in the compartments along with other implementation details.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

Table 3: Estimated parameters obtained by MAGIDDE for the time-delayed SIRD model, with 95% credible intervals, based on Ontario COVID-19 data from January 24 to February 22, 2022.

Parameter	Estimate	95% CI	Parameter	Estimate	95% CI	Parameter	Estimate	95% CI
$\tilde{\beta}$	0.0254	(0.0227, 0.0285)	$I(0)$	0.0145	(0.0142, 0.0148)	$10^4 \times \sigma_2$	3.4756	(2.6522, 4.6196)
h	3.0360	(1.2356, 4.7994)	$R(0)$	0.0053	(0.0050, 0.0056)	$10^7 \times \sigma_3$	2.1557	(0.9873, 3.5252)
$10^3 \times \mu_d$	0.3327	(0.3258, 0.3397)	$10^5 \times D(0)$	7.4850	(7.4471, 7.5170)			
λ	0.0751	(0.0730, 0.0774)	$10^4 \times \sigma_1$	3.4730	(2.6768, 4.5814)			

Table 3 summarizes the parameter estimates with 95% credible intervals. The model fitting results indicate that the estimated incubation period (3.03 days) is slightly shorter than the prior mean of 3.5 days, and the estimated recovery period ($1/0.0751 \approx 13$ days) exceeds the typical recovery period (6.87 days) from the existing literature (Wise, 2022). Noting that we used the number of patients admitted in hospital as a proxy to estimate the current infected population, such patients may have needed more time to recover. Shao et al. (2022) pointed out that the median recovery period for patients with moderate or acute symptoms was 13 days, which aligns with our model fitting result.

The trajectory of the S compartment is estimated by subtracting the pointwise estimate of the I, R, D compartments from the total population; the same technique is used to construct its 95% pointwise credible interval. Figure 3 visualizes the estimated trajectories of all four compartments (in

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

terms of population size) using MAGIDDE, which appear reasonable even though the credible intervals do not cover all the observations. Notably, we observe a weekend effect in the hospitalization data, evidenced by a 3-day flatter "ledge" after every four or five observations in the I and R components. This contributes to the inherent noise within the data, and thus the fitted trajectories appear smoother than the actual observations.

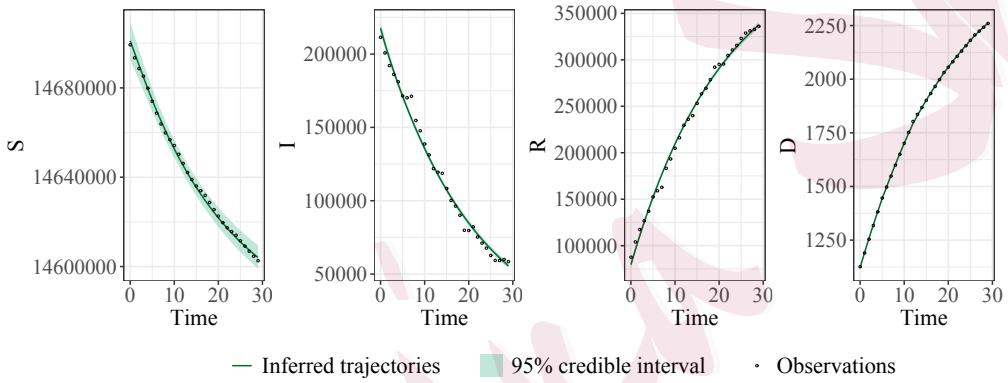

Figure 3: Estimated trajectories of the time-delayed SIRD model based on Ontario COVID-19 data from January 24 to February 22, 2022. The solid line represents the inferred trajectory and the shaded area represents the 95% pointwise credible interval. The dots are the observed data.

Finally, we utilize this DDE system to showcase the predictive performance of MAGIDDE, where we use the first half of the observations for model fitting and the remainder for assessing model predictions. These results are presented in the Supplementary Material S11.4 and illustrate the capability of MAGIDDE to predict "future" observations.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

6. Discussion

In this paper, we presented the MAGIDDE method for inference of DDE models. The method provides fast and accurate inference with the help of GPs and a linear interpolation scheme for handling historical outputs. Simulation results indicate that MAGIDDE well-recovers the parameters and trajectory in a simple benchmark system compared to other representative methods, and a larger *lac* operon system demonstrates the capability of MAGIDDE to recover more complex dynamic systems with multiple time delays. Finally, we fit a time-delayed compartmental model to Ontario COVID-19 data as a practical application.

We outline some directions for future work. First, we focused on DDEs with the history function $\mathcal{H}_{\tau_i} = \{x_i(t) = x_i(0), t \in [-\tau_i, 0]\}$, primarily due to its prevalence in both theoretical and practical applications (Kuang, 1993). However, more complex history functions also exist and could be worthy of investigation, as they can significantly alter the dynamics and stability of DDEs. Second, while DDEs with time-constant parameters cover a broad range of examples, an extension of the method to DDEs with time-varying parameters could also be considered. For example, Liu and Wang (2020) studied the dynamics of genetic regulatory networks with time-varying delays required for transcription and translation processes. In

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

epidemiology, Pei and Liu (2023) introduced the delayed SIR model with a time-varying disease transmission rate and removal rate.

Acknowledgments

This work was partially supported by Discovery Grant RGPIN-2019-04771 from the Natural Sciences and Engineering Research Council of Canada.

Supplementary Material

Includes the proofs, Tables, and Figures referenced in Sections 1,3,4,5.

References

Bellen, A. and M. Zennaro (2013). *Numerical methods for delay differential equations*. Numerical Mathematics and Scientific Computation.

Bihorel, S. (2011). The scaRabee Package: An R-based Tool for Model Simulation and Optimization in Pharmacometrics. `https://cran.r-project.org/package=scaRabee`.

Boersch-Supan, P. H., S. J. Ryan, and L. R. Johnson (2017). deBInfer: Bayesian inference for dynamical models of biological systems in R. *Methods in Ecology and Evolution 8*(4), 511–518.

Calderhead, B., M. Girolami, and N. Lawrence (2008). Accelerating Bayesian inference over

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

nonlinear differential equations with Gaussian processes. *Advances in neural information processing systems 21*, 217–224.

Dondelinger, F., D. Husmeier, S. Rogers, and M. Filippone (2013). ODE parameter inference using adaptive gradient matching with Gaussian processes. In *Artificial intelligence and statistics*, pp. 216–228. PMLR.

Ellner, S. P., B. E. Kendall, S. N. Wood, E. McCauley, and C. J. Briggs (1997). Inferring mechanism from time-series data: delay-differential equations. *Physica D: Nonlinear Phenomena 110*(3-4), 182–194.

Erneux, T. (2009). *Applied delay differential equations*. Springer.

Glass, D. S., X. Jin, and I. H. Riedel-Kruse (2021). Nonlinear delay differential equations and their application to modeling biological network motifs. *Nat. Commun. 12*(1), article 1788.

Horbelt, W., J. Timmer, and H. U. Voss (2002). Parameter estimation in nonlinear delayed feedback systems from noisy data. *Physics Letters A 299*(5-6), 513–521.

Hutchinson, G. E. (1948). Circular causal systems in ecology. *Annals of the New York Academy of Sciences 50*(4), 221–246.

Karniadakis, G. E., I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang (2021). Physics-informed machine learning. *Nature Reviews Physics 3*(6), 422–440.

Kuang, Y. (1993). *Delay differential equations: with applications in population dynamics*. Academic press.

Liang, H. and H. Wu (2008). Parameter estimation for differential equation models using a

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

framework of measurement error in regression models. *Journal of the American Statistical Association 103*(484), 1570–1583.

Liu, C. and F. Wang (2020). Parameter identification of genetic regulatory network with time-varying delays via adaptive synchronization method. *Advances in Difference Equations 2020*, article 127.

Ma, W., M. Song, and Y. Takeuchi (2004). Global stability of an SIR epidemic model with time delay. *Applied Mathematics Letters 17*(10), 1141–1145.

May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature 261*, 459–467.

Neal, R. M. (2011). MCMC Using Hamiltonian Dynamics. In *Handbook of Markov Chain Monte Carlo*, pp. 113–162. Chapman and Hall/CRC.

Nicholson, A. J. (1954). Compensatory reactions of populations to stresses, and their evolutionary significance. *Australian Journal of Zoology 2*(1), 1–8.

Pardee, A. B., F. Jacob, and J. Monod (1959). The genetic control and cytoplasmic expression of "inducibility" in the synthesis of β-galactosidase by E. coli. *Journal of Molecular Biology 1*(2), 165–178.

Pei, L. and D. Liu (2023). Long-term prediction of the COVID-19 epidemics induced by Omicron-virus in China based on a novel non-autonomous delayed SIR model. *Journal of Physics A: Mathematical and Theoretical 56*(23), 235601.

Ramsay, J. O., G. Hooker, D. Campbell, and J. Cao (2007). Parameter estimation for differential

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

equations: a generalized smoothing approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology 69*(5), 741–796.

Shao, J., R. Fan, J. Hu, T. Zhang, C. Lee, X. Huang, F. Wang, H. Liang, Y. Jin, Y. Jiang, et al. (2022). Clinical progression and outcome of hospitalized patients infected with SARS-CoV-2 omicron variant in Shanghai, China. *Vaccines 10*(9), article 1409.

Soetaert, K., T. Petzoldt, and R. W. Setzer (2010). Solving differential equations in R: package deSolve. *Journal of statistical software 33*, 1–25.

Tronarp, F., N. Bosch, and P. Hennig (2022). Fenrir: Physics-enhanced regression for initial value problems. In *International Conference on Machine Learning*, pp. 21776–21794. PMLR.

Varah, J. M. (1982). A spline least squares method for numerical parameter estimation in differential equations. *SIAM Journal on Scientific and Statistical Computing 3*(1), 28–46.

Wang, L. and J. Cao (2012). Estimating parameters in delay differential equation models. *Journal of agricultural, biological, and environmental statistics 17*, 68–83.

Wang, S., S. Ge, R. Doig, and L. Wang (2022). Adaptive semiparametric Bayesian differential equations via sequential Monte Carlo. *Journal of Computational and Graphical Statistics 31*(2), 600–613.

Wang, W., R. Tuo, and C. Jeff Wu (2020). On prediction properties of kriging: Uniform error bounds and robustness. *Journal of the American Statistical Association 115*(530), 920–930.

Wenk, P., A. Gotovos, S. Bauer, N. S. Gorbach, A. Krause, and J. M. Buhmann (2019). Fast

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

Gaussian process based gradient matching for parameter identification in systems of non-linear ODEs. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 1351–1360. PMLR.

Wise, J. (2022). Covid-19: Symptomatic infection with omicron variant is milder and shorter than with delta, study reports. *BMJ 377*, o922.

Wong, S. W., S. Yang, and S. Kou (2023). Estimating and Assessing Differential Equation Models with Time-Course Data. *The Journal of Physical Chemistry B 127*(11), 2362–2374.

Wong, S. W. K., S. Yang, and S. C. Kou (2024). magi: A package for inference of dynamic systems from noisy and sparse data via manifold-constrained Gaussian processes. *Journal of Statistical Software 109*(4), 1–47.

Wu, M. and M. Lysy (2024). Data-Adaptive Probabilistic Likelihood Approximation for Ordinary Differential Equations. In *International Conference on Artificial Intelligence and Statistics*, pp. 1018–1026. PMLR.

Yang, L., X. Meng, and G. E. Karniadakis (2021). B-PINNs: Bayesian physics-informed neural networks for forward and inverse PDE problems with noisy data. *Journal of Computational Physics 425*, 109913.

Yang, S., S. W. Wong, and S. Kou (2021). Inference of dynamic systems from noisy and sparse data via manifold-constrained Gaussian processes. *Proceedings of the National Academy of Sciences 118*(15), e2020397118.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0213

REFERENCES

Yildirim, N. and M. C. Mackey (2003). Feedback regulation in the lactose operon: a mathematical modeling study and comparison with experimental data. *Biophysical journal 84*(5), 2841–2851.

Department of Statistics and Actuarial Science, University of Waterloo

E-mail: yuxuan.zhao@uwaterloo.ca

Department of Statistics and Actuarial Science, University of Waterloo

E-mail: samuel.wong@uwaterloo.ca