

A Data-Assimilation-Augmented Optimization Framework for Parameter Estimation in Dynamical Systems

Muhammad Jalil Ahmad^{*1}, Animikh Biswas¹, and Kathleen Hoffman¹

¹Department of Mathematics and Statistics, University of Maryland, Baltimore County, USA

Abstract

Parameter estimation in nonlinear dynamical systems from observational data is a fundamental inverse problem with applications in many disciplines such as epidemiology, systems biology, climate science, and related fields. In practice, this is further complicated by the fact that observational data are often noisy, sparse, and available only for a subset of the state variables. Furthermore, the initial condition may be unknown or inaccurate, causing further complications for chaotic systems with sensitive dependence on initial conditions. In this work, we develop a data-assimilation-augmented optimization framework for parameter estimation in ordinary differential equation systems using partial state observations. The method introduces a nudged system driven by the available observed component and estimates the unknown parameters by minimizing a cost functional, defined as a time-delayed mismatch between the observations and the corresponding observed component of the nudged solution over the admissible parameter space. Since the nudged system can be arbitrarily initialized, this approach eliminates the dependence on accurate initial-state information.

The Lorenz system arises as the truncation of the Rayleigh–Bénard convection equation. Using the Lorenz–63 system as a test case, while noting that the proposed framework is applicable to a wide variety of dissipative systems admitting *determining functionals*, we establish theoretical results showing synchronization of the nudged solution under parameter agreement, stability under parameter mismatch, and well-posedness of the data-to-parameter inverse map under suitable nondegeneracy conditions. Structural identifiability, practical identifiability, and Sobol sensitivity analyses are incorporated to assess which parameters can be reliably estimated from the available observations. Numerical experiments in both chaotic and non-chaotic regimes show that the proposed framework accurately recovers parameters from noisy partial observations. Comparisons with an existing on-the-fly parameter learning method and with Bayesian MCMC estimation demonstrate that the proposed method remains accurate under partial observations and higher noise levels while requiring substantially lower computational cost.

Keywords: data assimilation; parameter estimation; dynamical systems; nudging; inverse problems; parameter sensitivity; parameter identifiability

MSC 2010 Classifications: 34D06, 34A55, 34H10

1 Introduction

Differential equations (DEs) are the mathematical backbone of models used to describe physical, chemical, biological, and even financial systems. The effectiveness of these models is determined by their ability to replicate real-world behavior. A primary challenge when using such models is the need for accurate initial conditions and parameter estimates. However, in many applications, these quantities are unknown and must be inferred from observational data. This inference problem is often challenging because real-world data may be sparse, noisy, or available only for a subset of the state variables. These limitations can lead to inverse problems that are ill-posed, nonconvex, and computationally demanding, particularly for nonlinear dynamical systems.

These challenges involve several interconnected issues: whether the available data contain enough information to determine the unknown parameters, how strongly different parameters influence the observed dynamics, and how observational data can be incorporated into the model in a stable and computationally efficient way. We therefore organize the related work around parameter identifiability, parameter sensitivity, parameter estimation methods for dynamical systems, limitations of these methods and motivation for data-assimilation.

^{*}Corresponding author: Muhammad Jalil Ahmad, Department of Mathematics and Statistics, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA. Email: LS47576@umbc.edu.

1.1 Related Work and Motivation

Parameter Identifiability: The question of whether it is possible to uniquely estimate parameters for a given model and data set is addressed through the concept of *parameter identifiability* [1–4]. This notion is typically examined in two distinct forms. *Structural identifiability* concerns the uniqueness of parameter estimates under ideal conditions—that is, assuming noise-free, continuous, and complete measurements [1, 3–5]. In this sense, structural identifiability characterizes the maximum amount of parameter information that can be extracted from a specified observation scheme and serves as a necessary prerequisite for reliable estimation from real-world data.

Practical identifiability, on the other hand, accounts for realistic limitations such as measurement noise, finite sampling, and imperfect observability, and investigates whether these factors hinder the ability to reliably estimate parameters in practice [1, 6–12]. As discussed in [6], such limitations may render a structurally identifiable model practically non-identifiable. Conversely, in the idealized limit of noise-free and sufficiently informative data, structural identifiability represents the best-case scenario for practical parameter recovery [13]. A more detailed discussion of parameter identifiability is provided in Sections 2.1 and 2.3.

Parameter Sensitivity: A significant challenge that can arise in the parameter estimation process is an imbalance between the number of unknown parameters and the number of observable states [11, 12, 14]. However, it is crucial to recognize that not all parameters significantly influence the observables in a dynamical system. This issue is addressed through the concept of *parameter sensitivity*, which refers to how changes in a model's parameters influence its outputs [15, 16]. Here, the term “output” may refer to individual state variables, linear combinations thereof, or other derived quantities, including those that are not themselves state variables. For instance, in epidemiological modeling, sensitivity analysis is frequently conducted with respect to the basic reproduction number, R_0 , which represents the average number of secondary cases generated by a single infectious individual. Numerous methods are available in the literature to assess which parameters are sensitive in a model (see, e.g., [17–22]). A more detailed discussion on parameter sensitivity is provided in Section 2.4. Since non-sensitive parameters do not significantly impact the dynamics of the system's observed output, it may not be necessary to estimate them with high precision, and they may therefore be excluded from the parameter estimation process for that particular observation setting.

Parameter Estimation: The literature presents a variety of methods for parameter estimation in dynamical systems. Some approaches leverage neural networks [23–27], which often require large datasets—an assumption that is impractical in many real-world scenarios. Classical approaches often formulate parameter estimation as a nonlinear least-squares or inverse problem, in which the parameters are chosen to minimize the discrepancy between observed data and model predictions [28–31]. These methods are widely used and mathematically well established, but they may become computationally expensive for nonlinear dynamical systems because each evaluation of the cost function typically requires solving the forward model.

In the Bayesian framework, the mismatch between model predictions and observed data is modeled probabilistically through a likelihood function, which facilitates parameter estimation through either maximum likelihood estimation or fully Bayesian inference. In the latter case, prior distributions are introduced to incorporate prior information and regularize the inverse problem. In practice, Bayesian inference is often implemented using Markov chain Monte Carlo (MCMC) methods [32–35], which enable sampling from the posterior distribution of the parameters. However, these methods can be computationally expensive, as they require numerous forward simulations of the model.

Motivation for Data Assimilation in Parameter Estimation: Despite the variety of available parameter estimation methods, several challenges remain. Neural-network-based approaches may require large training data sets, while classical least-squares, Bayesian, and MCMC-based methods can become computationally expensive because they often require many forward simulations of the model. In addition, many parameter estimation methods are sensitive to the choice of initial conditions, which is especially problematic in chaotic systems where small initial errors can lead to large deviations in the trajectory. Another major difficulty is that full-state observations are rarely available in practice; data are often sparse, noisy, indirect, or restricted to only a subset of the state variables.

Data assimilation (DA) provides a natural framework for addressing these issues because it combines observational data with the governing dynamics to stabilize state reconstruction and prediction [36–39]. In this work, we focus on nudging, a continuous data assimilation method originally used in weather prediction and geophysical models [40–42], and later developed into a rigorous mathematical framework by Azouani, Olson, and Titi [43] for geophysical systems. Nudging introduces a feedback control term into the model equations, driving the assimilated solution toward the observed components of the reference solution.

The use of nudging for parameter estimation has been studied in several recent works. Carlson et al. [44] proposed an on-the-fly parameter learning algorithm in which unknown parameters are adjusted using the persistent synchronization error between the true and nudged systems. Martinez et al. [45] developed relaxation-based schemes,

including relaxation least-squares and relaxation Newton iteration methods, for simultaneous state and parameter recovery in dissipative systems. More recently, Newey et al. [46] interpreted related CDA parameter-learning algorithms through finite-dimensional root-finding and optimization methods, including Newton, Gauss–Newton and Levenberg–Marquardt-type approaches.

The work of Newey et al. [46] is particularly relevant to our work because it also considers an observable-error cost functional arising from a nudged system. However, their approach involves finding the zero of the gradient of this cost functional with respect to the parameter pointwise in time and uses successive temporal updates of the parameter estimates thus obtained. This formulation is primarily designed for on-the-fly parameter updates based on a single large-time observable error, which necessitates observing the variables the equations of which contain the parameter. By contrast, the present work formulates parameter recovery as a direct minimization problem for a time-averaged cost functional, where this restriction is not present. Furthermore, our simulation study indicates that the synchronization error may oscillate in time, especially away from the true parameter value, and particularly in the presence of noise in the observed data. Additionally, our computations indicate that the cost functional is nonconvex; thus solving for the gradient may converge to a local minimum, particularly in case of observational error in the data.

While the comparison with Newey et al. [46] clarifies the distinction between our approach and theirs, the inverse-problem viewpoint adopted here is closely related to the determining-map framework for parameter recovery developed in the PDE setting by Biswas and Hudson [47]. In the context of the two-dimensional incompressible Navier–Stokes equations, they formulated viscosity recovery from finitely many modal observations as an inverse problem based on the determining map. Their work used a nudging-generated cost functional and established rigorous conditions for uniqueness, well-posedness, control of the parameter error by the cost functional, and convergence of an associated algorithm. The present work is inspired by this viewpoint, but extends it from a single-parameter PDE setting to a multi-parameter ODE setting with partial state observations.

Using the Lorenz–63 system as a canonical nonlinear test problem with both chaotic and non-chaotic regimes, rather than as a model-specific setting, we establish theoretical results showing synchronization of the nudged solution under parameter agreement, stability under parameter mismatch, and well-posedness of the data-to-parameter inverse map under suitable nondegeneracy conditions. Although the analysis is carried out for Lorenz–63, the proposed nudging-based optimization formulation is not restricted to this system and can be applied to more general dissipative systems including partial differential equations which admit *determining functionals*. The framework also allows for models in which unknown parameters occur in nonlinear terms. Structural identifiability, practical identifiability, and Sobol sensitivity analyses are incorporated to assess which parameters can be reliably estimated from the available observations. Numerical experiments in both chaotic and non-chaotic regimes show that the proposed framework accurately recovers parameters from noisy partial observations. Comparisons with an existing on-the-fly parameter learning method and with Bayesian MCMC estimation demonstrate that the proposed method remains accurate under partial observations and higher noise levels while requiring substantially lower computational cost.

1.2 Contributions

The main contribution of this work is a data-assimilation-augmented optimization framework for estimating unknown parameters in nonlinear ODE systems from partial state observations. The approach uses a nudged system driven by the available observed component and estimates the unknown parameters by minimizing a time-delayed cost functional, defined by the mismatch between the observations and the corresponding observed component of the nudged solution.

The proposed framework is supported by theoretical analysis, including well-posedness and stability estimates. For the Lorenz–63 system, we study synchronization, stability under parameter mismatch, and the well-posedness of the inverse map from observations to parameters. We also identify dynamical situations in which parameter recovery may fail, thereby clarifying when the proposed inverse problem is well-posed and when it is degenerate.

The main contributions of this work are as follows:

- We formulate a nudging-based optimization framework for parameter estimation in finite-dimensional nonlinear ODE systems from partial state observations.
- We extend the determining-map-based philosophy of Biswas and Hudson [47], developed for viscosity recovery in the Navier–Stokes equations, from a single-parameter PDE setting to a multi-parameter ODE setting.
- We introduce a time-delayed, time-averaged cost functional based on the mismatch between the observed data and the corresponding observed component of the nudged solution.
- We prove synchronization results for the Lorenz–63 nudged system under correct parameter values, showing convergence of the nudged trajectory to the true trajectory from arbitrary initial conditions.

- We establish stability estimates under parameter mismatch, showing that the nudged trajectory converges to a neighborhood of the true trajectory whose size is controlled by the parameter error.
- We characterize the well-posedness of the data-to-parameter inverse map in terms of the underlying Lorenz dynamics and identify degenerate trajectories and equilibrium-related cases where parameter recovery may fail or become unstable.
- We derive local error estimates showing that, under appropriate nondegeneracy assumptions, small values of the time-delayed cost functional imply small error in the recovered parameter vector.
- We extend the analysis from full observation to partial observation by showing that, near the true parameter, the nudged trajectory remains close enough to preserve the relevant well-posedness and stability mechanisms.
- We incorporate structural identifiability, practical identifiability, and Sobol sensitivity analysis into the parameter estimation workflow to assess which parameters can be reliably estimated from the available observations.
- We validate the proposed framework on the Lorenz-63 system in both chaotic and non-chaotic regimes, using noisy partial observations, and compare the resulting estimates with existing on-the-fly CDA parameter-learning methods and Bayesian MCMC estimation.

The remainder of this paper is organized as follows. Section 2 reviews key aspects of parameter estimation, including structural identifiability, practical identifiability, well-posedness of the data-to-parameter inverse map, and parameter sensitivity. Section 3 introduces the data-assimilation framework and formulates the nudging-augmented parameter estimation approach. Section 4 presents the Lorenz-63 system and analyzes the corresponding nudged system, including synchronization, stability under parameter mismatch, and well-posedness properties under full and partial observations. Section 5 presents the numerical results for the Lorenz-63 system, including forward prediction, structural and practical identifiability, Sobol sensitivity analysis, and parameter estimation in both chaotic and non-chaotic regimes with noisy observations. Finally, Section 6 summarizes the main findings, compares the proposed approach with an existing data-assimilation-based parameter estimation method and Bayesian MCMC estimation, and outlines directions for future work.

2 Aspects of Parameter Estimation

Parameter estimation for dynamical systems involves several related questions. Model calibration refers to the process of selecting parameter values so that model trajectories agree with observed data [48]. However, before carrying out calibration, one must understand whether the available observations contain sufficient information to determine the unknown parameters. This leads to the notions of structural identifiability, and practical identifiability. In addition, parameter sensitivity analysis provides information about how strongly each parameter influences the model output.

A related but conceptually distinct issue is the stability of the inverse map from observed data to parameters [49]. In other words, even when identifiability guarantees uniqueness, one must still ask whether small changes in the observed data lead only to controlled changes in the recovered parameters. We refer to this stability requirement as the well-posedness of the data-to-parameter inverse map. Although this viewpoint is closely connected to practical identifiability and inverse-problem stability, it is often not stated explicitly in the dynamical-systems parameter-estimation literature. We include it here to clarify the role it plays in the proposed framework and to set the stage for the numerical and theoretical analyses below. This section reviews these concepts and explains how they are used in the numerical investigations below.

2.1 Structural Parameter Identifiability

Structural identifiability analysis investigates whether the parameters of a model can be uniquely determined from the observed output, based only on the model structure and the specified observation scheme, without accounting for limitations in data quantity or measurement noise. More precisely, it asks whether two parameter values that generate the same observed output must necessarily be identical. Thus, structural identifiability is a property of the parameter-to-observation map under idealized conditions, namely continuous and noise-free observations, and provides a theoretical prerequisite for reliable parameter estimation from data.

Structural identifiability is further classified into global and local structural identifiability [2–4]. Consider a general model

$$\begin{cases} \dot{\mathbf{x}}_\theta(t) = f(\mathbf{x}_\theta(t), \theta), \\ \mathbf{x}_\theta(0) = \mathbf{x}_0, \\ \mathcal{O}_\theta(t) = g(\mathbf{x}_\theta(t), \theta), \end{cases} \quad (1)$$

where the state vector $\mathbf{x}_\theta(t) \in \mathbb{R}^n$ represents the state of the system at time t , $\theta \in \Theta \subseteq \mathbb{R}^p$ is the vector of parameters, $\mathbf{x}_0 \in \mathbb{R}^n$ is the initial condition, and $\mathcal{O}_\theta(t) \in \mathbb{R}^q$ denotes the observable outputs, with $q \leq n$. The functions f and g are taken to be defined on an open set and are locally Lipschitz. In particular, this implies that $\mathcal{O}$ is a continuous function of the state $\mathbf{x}$. The initial condition is assumed to be fixed throughout the identifiability analysis. When the initial condition is known, it is suppressed from the notation. When it is unknown and must be estimated, it is treated as part of an extended parameter vector $[\theta^\top, \mathbf{x}(0)^\top]^\top$. When the initial condition is included, the observation is denoted by $\mathcal{O}_\theta(t; \mathbf{x}(0))$. For simplicity, we omit this dependence unless it is explicitly needed.

Definitions 1–3 are stated over a time interval $I \subset [0, +\infty)$ [3, 4]. For instance, one may take $I = [0, +\infty)$ or a finite interval $I = [a, b] \subset [0, +\infty)$. Definitions 1 and 2 characterize the structural identifiability of a single parameter component $\theta^{(i)}$, for $i \in \{1, \dots, p\}$.

Definition 1. Following [3], the parameter $\theta^{(i)}$ is said to be *structurally locally identifiable* if, for almost all $\theta_1 \in \Theta$, there exists a neighborhood $B(\theta_1)$ such that

$$\theta_2 \in B(\theta_1), \quad \mathcal{O}_{\theta_1}(t) = \mathcal{O}_{\theta_2}(t) \text{ for all } t \in I \Rightarrow \theta_2^{(i)} = \theta_1^{(i)}.$$

The model (1) is said to be *structurally locally identifiable* if all parameters $\theta^{(i)}$, $i = 1, \dots, p$, are structurally locally identifiable.

Definition 2. As formulated in [3], the parameter $\theta^{(i)}$ is said to be *structurally globally identifiable* if, for almost all $\theta_1, \theta_2 \in \Theta$,

$$\mathcal{O}_{\theta_1}(t) = \mathcal{O}_{\theta_2}(t) \text{ for all } t \in I \Rightarrow \theta_2^{(i)} = \theta_1^{(i)}.$$

The model (1) is said to be *structurally globally identifiable* if all parameters $\theta^{(i)}$, $i = 1, \dots, p$, are structurally globally identifiable.

It follows immediately that structural global identifiability implies structural local identifiability.

Definition 3. The model (1) is said to be *structurally non-identifiable* if there exists at least one parameter $\theta^{(i)}$ that is not structurally identifiable [3].

We note that in the original formulation [3], the time domain over which the equality $\mathcal{O}_{\theta_1}(t) = \mathcal{O}_{\theta_2}(t)$ is required is not explicitly specified. In the context of dynamical systems, this equality is naturally interpreted over a time interval. To avoid any potential ambiguity and to emphasize that identifiability is not a pointwise notion in time, we explicitly require the equality $\mathcal{O}_{\theta_1}(t) = \mathcal{O}_{\theta_2}(t)$ to hold for all $t \in I$, rather than at a single time instance. We also refer the reader to the more recent review [4], which provides a comprehensive overview of various notions of structural identifiability and discusses the relationships among these notions.

The differential algebra approach to structural identifiability and the associated computational algorithms have been developed in [2, 50–52]; see also the review [1]. The main idea is to eliminate the unobserved state variables and obtain one or more input–output differential equations involving only the measured output, its derivatives, known inputs when present, and the unknown parameters. In the notation of Eq. (1), suppose that such an input–output relation has the form

$$h(\mathcal{O}_\theta(t), \mathcal{O}_\theta^{(1)}(t), \dots, \mathcal{O}_\theta^{(\ell)}(t), \theta) = 0,$$

where $\mathcal{O}_\theta^{(j)}$ denotes the j -th derivative of the observed output and h is an algebraic, or polynomial, expression in the observed output, its derivatives, and the unknown parameters. The existence of such an input–output equation is an important first step, since it shows that the unknown parameters enter an algebraic-differential relation determined by the observed output. However, the existence of this equation alone does not automatically imply structural identifiability. One must still verify that the resulting algebraic equations determine the parameter vector uniquely.

In many models, a single input–output equation does not provide enough independent equations to determine all unknown parameters. One possible approach is to differentiate the input–output equation further in time until a sufficient number of equations is obtained. This is the basis of the implicit-function-theorem approach of Xia and Moog [53], in which an identification function is constructed from the input–output relation and its derivatives. If the Jacobian of this identification function with respect to the unknown parameters has full rank at the true parameter value, then local structural identifiability follows from the implicit function theorem. However, this procedure may require high-order derivatives of the measured output. Such derivatives may be difficult to justify analytically unless sufficient smoothness is known, and they are also difficult to evaluate reliably from finite or noisy data.

An alternative, proposed by Wu et al. [54], is to evaluate the same input–output relation at several distinct observation times rather than repeatedly differentiating it. To describe this idea, let θ^* denote the true parameter vector and define

$$Y_j^* = (\mathcal{O}_{\theta^*}(t_j), \mathcal{O}_{\theta^*}^{(1)}(t_j), \dots, \mathcal{O}_{\theta^*}^{(\ell)}(t_j)), \quad j = 1, \dots, J,$$

where $\{t_j\}_{j=1}^J$ is a finite collection of observation times. Evaluating the input–output equation at these times gives

$$h(Y_j^*, \theta) = 0, \quad j = 1, \dots, J.$$

Equivalently, define the identification map

$$\Phi(\theta; Y^*) = \begin{pmatrix} h(Y_1^*, \theta) \\ \vdots \\ h(Y_J^*, \theta) \end{pmatrix}.$$

Then local structural identifiability can be studied through the rank of the Jacobian of this map with respect to the parameter vector. In particular, if

$$\text{rank } D_\theta \Phi(\theta^*; Y^*) = p,$$

where p is the dimension of the parameter vector, then the identification equations have full rank with respect to the unknown parameters at θ^* . In the square case, this condition reduces to

$$\det D_\theta \Phi(\theta^*; Y^*) \neq 0.$$

By the implicit function theorem, this full-rank condition implies local uniqueness of the parameter vector. More precisely, there exists a neighborhood U of θ^* such that

$$\Phi(\theta; Y^*) = \Phi(\theta^*; Y^*), \quad \theta \in U,$$

implies

$$\theta = \theta^*.$$

Thus, after deriving an input–output equation, one must still verify an appropriate nondegeneracy condition. The multiple-time-point formulation is useful because it can reduce the need for repeatedly differentiating the output equation, while still producing enough algebraic equations to test local structural identifiability. This point is especially relevant when only limited smoothness of the observed trajectory is available, or when higher-order derivatives of noisy data would be numerically unstable.

Symbolic tools such as *DAISY* and *SIAN* provide algorithmic implementations of differential-algebraic methods for structural identifiability analysis. In particular, they can be used to eliminate unobserved state variables and derive input–output relations involving only the observed variables, their derivatives, known inputs when present, and the unknown parameters. *DAISY* was among the first software tools to employ differential algebra techniques for structural identifiability analysis [55]. *SIAN* is another symbolic tool for structural identifiability analysis that combines differential algebra ideas with Taylor-series-based computations [56]. These tools are useful because they automate part of the symbolic analysis that would otherwise be difficult to carry out by hand. At the same time, the outcome of any such analysis should be interpreted in terms of the underlying identifiability question: whether the available output relations determine the unknown parameters uniquely, either globally or locally. Thus, input–output equations, differential-algebraic reductions, and software-generated identifiability conclusions should be viewed as part of the same structural-identifiability framework, with the final conclusion depending on the corresponding uniqueness or rank conditions for the model under consideration.

2.2 Well-posedness of the Data-to-Parameter Inverse Map

While identifiability addresses the issue of injectivity of the map from parameter to data, another important question in parameter estimation is the well-posedness of the inverse map from the observed data to the parameters. In other words, how reliably parameters of a dynamical system can be recovered from observed data can be quantified by an estimate of the form

$$\|\theta - \theta'\| \leq \omega(\|\mathcal{O}_\theta(\cdot) - \mathcal{O}_{\theta'}(\cdot)\|_{\mathcal{B}}),$$

where ω is a suitable modulus of continuity and $\mathcal{B}$ denotes a suitable Banach space containing the observations $\mathcal{O}_\alpha(\cdot)$ for all parameters $\alpha \in \Theta$. This concept is crucial for the stability of any parameter estimation algorithm. While identifiability has been widely addressed, the issue of well-posedness of the inverse map has received much less attention. Practical parameter identifiability analysis, described in Section 2.3, can be regarded as a tool for addressing the issue of well-posedness of the data-to-parameter inverse map numerically [57].

2.3 Practical Parameter Identifiability

While structural identifiability characterizes the theoretical ability to recover parameters from ideal data, practical identifiability assesses whether parameters can be reliably estimated from real, noisy, and finite data. To analyze practical identifiability, we employ a Monte Carlo (MC) approach [1, 7]. The data for this analysis are synthetically generated by numerically solving the system given in Eq. (1), using true parameter values θ^* . For this analysis, the initial condition x_0 is assumed to be known. The Monte Carlo (MC) procedure used for practical identifiability analysis consists of the following steps:

1. Integrate the system in Eq. (1) using the true parameter vector θ^* and some fixed initial data to generate the reference output $g(\mathbf{x}(t), \theta^*)$ at discrete observation times t_i , for $i = 1, \dots, n$.
2. Construct $M = 1,000$ synthetic data sets by perturbing the reference output with measurement noise. Assuming Gaussian noise with zero mean and standard deviation ϵ , the noisy observations are generated as

$$\mathcal{O}_{i,j} = g(\mathbf{x}(t_i), \theta^*)(1 + \xi_{i,j}), \quad \xi_{i,j} \sim \mathcal{N}(0, \epsilon). \quad (2)$$

The standard deviation ϵ determines the noise magnitude; for example, $\epsilon = 0.02$ corresponds to 2% noise. Depending on the application, observational noise may be modeled additively or multiplicatively. In this work, a multiplicative noise model is adopted throughout for consistency.

3. For each synthetic data set, estimate a parameter vector θ_j by fitting the model to the perturbed observations through least-squares minimization:

$$\theta_j \approx \arg \min_{\theta} \sum_{i=1}^n (\mathcal{O}_{i,j} - g(\mathbf{x}(t_i), \theta))^2. \quad (3)$$

The resulting optimization problem is solved using a numerical optimization algorithm, namely the Nelder–Mead algorithm [58].

4. Quantify estimation accuracy for each parameter using the average relative error (ARE), defined by

$$\text{ARE}(\theta^{(k)}) = \frac{100\%}{M} \sum_{j=1}^M \frac{|\theta^{*(k)} - \theta_j^{(k)}|}{|\theta^{*(k)}|}, \quad (4)$$

where $\theta^{(k)}$ represents the k -th parameter in the set θ , $\theta^{*(k)}$ is the k -th parameter in the true parameter vector θ^* , and $\theta_j^{(k)}$ denotes the k -th element of θ_j .

This process is repeated across varying noise levels. Following [6, 7], a parameter is considered practically identifiable if its Average Relative Error (ARE) is less than or equal to the measurement error, ϵ .

2.4 Parameter Sensitivity

Parameter sensitivity analysis is used to quantify how strongly variations in the model parameters affect a chosen model output [15, 16]. In the present work, the output may be an observed state component, a collection of observed components, or another quantity derived from the state variables. Sensitivity information is useful for parameter estimation because parameters that have little influence on the available output are typically more difficult to estimate reliably from that observation scheme. Thus, sensitivity analysis provides a complementary diagnostic to identifiability analysis: it helps indicate which parameters are expected to be influential in the data and which parameters may be weakly informed by the observations.

A widely used tool for determining parameter sensitivity is the Sobol sensitivity analysis. Briefly, the Sobol methodology decomposes the total variance of the model output into contributions associated with individual parameters and combinations of parameters [18]. The resulting *first-order sensitivity index* measures the contribution of a single parameter acting alone, while the *total-order sensitivity index* accounts for both the direct effect of that parameter and all interaction effects involving it. Parameters with small sensitivity indices have a limited influence on the output, whereas parameters with large indices play a dominant role in shaping the observed dynamics. We refer the reader to [16, 18] for a complete mathematical formulation of the variance decomposition and index definitions.

3 Parameter Estimation Augmented by Data Assimilation

This section introduces the data assimilation framework used for parameter estimation in this work. The main idea is that full-state observation represents the ideal setting for trajectory-based parameter estimation, since all components of the model state are directly available for comparison with the model output. In many applications, however, only partial observations are available. Data assimilation provides a mechanism for using these partial observations, together with the model dynamics, to reconstruct the unobserved components of the state. Thus, if the available observations are sufficiently informative, a data-assimilation-based estimator may be expected to approach the performance of a full-observation estimator.

We first review nudging as a continuous data assimilation method for reconstructing the state of a dynamical system from partial observations. We then describe how this framework is incorporated into a parameter estimation procedure by defining a nudging-based cost functional and minimizing it over the admissible parameter set. This approach formulates parameter recovery as an optimization problem while retaining the stabilizing effect of data assimilation.

3.1 Data Assimilation

Using the nudging approach, one introduces a replica of the original system, called the *nudged system*, and adds a feedback term to the equations corresponding to the observed components. This feedback term penalizes the mismatch between the observed data and the corresponding components of the nudged solution.

We now describe the nudging formulation used in this work. Consider the nonlinear dynamical system given in Eq. (1). Assume that only partial observations of the state are available:

$$\mathcal{O}(t) = H\mathbf{x}(t), \quad (5)$$

where $\mathcal{O}(t) \in \mathbb{R}^q$ denotes the observed data and $H : \mathbb{R}^n \rightarrow \mathbb{R}^q$ is a linear observation operator, with $q \leq n$. A typical example of an observation operator is an orthogonal projection onto a subspace of the phase space. In this case, denoting by $P_{\text{Ran } H}$ the orthogonal projection onto the range of H , we have

$$H = H^\top = H^\top H = P_{\text{Ran } H} \quad \text{and} \quad \mathcal{O}(t) = P_{\text{Ran } H} \mathbf{x}(t).$$

To incorporate the observations into the model dynamics, we introduce a replica of Eq. (1), referred to as the nudged system:

$$\dot{\tilde{\mathbf{x}}}(t) = F(\tilde{\mathbf{x}}(t), \tilde{\theta}) + \mathcal{B}(\mathcal{O}(t) - H\tilde{\mathbf{x}}(t)), \quad \tilde{\mathbf{x}}(0) = \tilde{\mathbf{x}}_0. \quad (6)$$

Here, $\mathcal{B} : \mathbb{R}^q \rightarrow \mathbb{R}^n$ is a chosen linear feedback operator that incorporates the observation residual into the state equations, $\tilde{\mathbf{x}}(t)$ denotes the nudged state, and $\tilde{\theta}$ denotes the parameter values used in the nudged model. The nudging term $\mathcal{B}(\mathcal{O}(t) - H\tilde{\mathbf{x}}(t))$ acts as a feedback correction that penalizes the discrepancy between the observed data and the corresponding observations of the nudged solution. In many applications, the true initial condition $\mathbf{x}_0$ is unknown. The nudged system can therefore be initialized independently of the true system; for instance, one may choose $\tilde{\mathbf{x}}(0) = 0$.

The goal is to choose, when possible, a feedback operator $\mathcal{B}$ such that

$$\lim_{t \rightarrow \infty} \|\mathbf{x}(t) - \tilde{\mathbf{x}}(t)\| = 0,$$

regardless of how the nudged system is initialized. In control theory terminology, Eq. (6) is a type of Luenberger observer [59].

When H is an orthogonal projection, a suitable choice often turns out to be $\mathcal{B} = \mu H^\top$, where μ is the nudging coefficient. Therefore, in this case, we have

$$\mathcal{O}(t) = P_{\text{Ran } H} \mathbf{x}(t) = H^\top H \mathbf{x}(t) \quad \text{and} \quad \mathcal{B}(\mathcal{O}(t) - H\tilde{\mathbf{x}}(t)) = \mu P_{\text{Ran } H} (\mathbf{x}(t) - \tilde{\mathbf{x}}(t)).$$

Thus, the nudging term acts on the observed components of the state and drives the nudged solution toward the reference trajectory through the available observations. As we shall see later, in many situations μ can be chosen sufficiently large to ensure

$$\lim_{t \rightarrow \infty} \|\mathbf{x}(t) - \tilde{\mathbf{x}}(t)\| = 0.$$

The effectiveness of this approach depends on whether the observation operator H captures enough information about the underlying dynamics. This is closely related to the theory of determining functionals in geophysical fluid dynamics [60–63]. More precisely, if the observation operator has the *determining property* given in Definition 4, then partial observations can, in principle, determine the full asymptotic dynamics. This is precisely the mechanism that motivates the use of data assimilation for parameter estimation from partial observations.

Definition 4. Following [60, 64], let $\mathcal{H} = \mathbb{R}^n$ be the state space of the dynamical system and let $\{S(t)\}_{t \geq 0}$ denote the associated flow. The observation operator H is called determining if, for any two trajectories $u(t) = S(t)u_0$ and $v(t) = S(t)v_0$, the condition

$$\lim_{t \rightarrow \infty} \|H(u(t) - v(t))\| = 0 \quad \text{implies} \quad \lim_{t \rightarrow \infty} \|u(t) - v(t)\| = 0.$$

The concept of determining operators is related to detectability in linear control theory [65]. In the parameter estimation setting considered here, this property indicates that partial observations may contain enough information to reconstruct the relevant state dynamics. Therefore, data assimilation provides a way to reduce the gap between partial-observation parameter estimation and the ideal full-observation case.

3.2 Nudging-Augmented Parameter Estimation

This subsection describes how the nudging framework is used to construct a parameter estimation method from partial observations. The main motivation is the following. If all state variables were observed, then parameter estimation could be carried out by directly comparing the full model trajectory with the full observed trajectory. This represents the ideal observation setting and provides the best possible information available to any trajectory-based estimation framework. In most applications, however, only a subset of the state variables is observed. The purpose of data assimilation is to use these partial observations, together with the model dynamics, to reconstruct the missing components of the state. Thus, if the nudged system synchronizes with the true system when the correct parameter is used, then the parameter estimation problem based on partial observations may be expected to approach the corresponding full-observation problem.

This observation provides the rationale for the proposed framework. For a trial parameter value $\tilde{\theta}$, the nudged system uses the available data to correct the observed components of the model trajectory. When $\tilde{\theta} = \theta^*$, the feedback term drives the nudged trajectory toward the true trajectory, even when the nudged system is initialized arbitrarily. In this case, the observed trajectory mismatch should decay after a transient time. On the other hand, when $\tilde{\theta}$ is far from θ^* , the model dynamics are inconsistent with the observed trajectory, and the nudging term cannot fully remove the resulting discrepancy. Therefore, the time-delayed cost functional provides a natural objective for parameter recovery.

A closely related approach was developed by Biswas and Hudson [47] in the PDE setting for the two-dimensional incompressible Navier–Stokes equations. In that work, the authors recovered the viscosity from observations of finitely many Fourier modes by using the determining map associated with a nudging algorithm to define a cost functional. The minimization of this cost functional was used to solve the inverse problem of identifying the true viscosity. Their analysis established conditions for uniqueness and well-posedness of the inverse problem, showed that smallness of the cost functional implies proximity to the true viscosity, and provided an algorithm with convergence guarantees.

The present work extends this determining-map-based philosophy to a different setting. Instead of recovering a single viscosity parameter in an infinite-dimensional PDE from finitely many modal observations, we consider finite-dimensional ODE systems with possibly several unknown parameters and only partial state observations. Thus, the framework developed here can be viewed as an ODE and multi-parameter analogue of the nudging-based inverse-problem approach in [47]. The central question is whether data assimilation can reduce the gap between partial-observation parameter estimation and the ideal full-state observation problem.

Suppose that the observation data $\mathcal{O}(t) = H\mathbf{x}(t)$ are available on the time interval $[\tau, \tau + T]$, and that either the full parameter vector θ or a subset of its components is unknown. For each trial parameter value $\tilde{\theta} \in \Theta$, we solve the nudged system

$$\dot{\tilde{\mathbf{x}}}(t) = F(\tilde{\mathbf{x}}(t), \tilde{\theta}) + \mathcal{B}(\mathcal{O}(t) - H\tilde{\mathbf{x}}(t)), \quad \tilde{\mathbf{x}}(0) = \tilde{\mathbf{x}}_0.$$

Here, $\tilde{\mathbf{x}}_0$ can be chosen independently of the true initial condition. The estimated parameter is then obtained by minimizing the time-delayed cost functional, which is defined as a time-delayed mismatch between the observed data and the corresponding observed component of the nudged solution:

$$C(\tilde{\theta}) = \frac{1}{T} \int_{\tau}^{\tau+T} \left\| \mathcal{O}(t) - H\tilde{\mathbf{x}}(t; \tilde{\theta}) \right\|_{\mathbb{R}^q}^2 dt, \quad \tau \geq 0, \quad T > 0. \quad (7)$$

The use of the time-delay τ in the cost functional is important in practice because the nudged system may be initialized far from the true state, so the initial part of the trajectory may contain synchronization error caused by the arbitrary initial condition rather than by the parameter mismatch. Discarding this initial period makes the cost functional more closely reflect the compatibility between the trial parameter and the observed dynamics. Moreover, averaging the mismatch over a time interval reduces sensitivity to instantaneous oscillations in the synchronization error, as well as to error in the observed data, and provides a more robust cost functional than a mismatch evaluated at a single time.

This formulation differs significantly from the standard least-squares trajectory fitting. In standard trajectory fitting with partial observations, each trial parameter generates an unconstrained model trajectory from a prescribed initial condition, and the resulting mismatch may reflect both parameter error and error in the initial condition. In the present framework, the trajectory is continuously corrected by the data through the nudging term. Consequently, the cost functional measures whether the trial parameter can produce a dynamically consistent assimilated trajectory, rather than whether an uncorrected trajectory happens to remain close to the observations over a finite time interval. This is particularly useful in chaotic systems, where small errors in the initial condition can rapidly dominate the trajectory mismatch.

The same viewpoint also clarifies the relationship between the present framework and the continuous data assimilation parameter-learning method of Newey et al. [46]. Their work considers a related observable-error cost functional and interprets CDA-based parameter learning through finite-dimensional root-finding and optimization methods, including Newton, Gauss–Newton, and Levenberg–Marquardt-type updates. A key difference is that their update formulas are based on the observable error at a sufficiently large time and rely on assumptions about the long-time behavior and differentiability of this error. In contrast, the present work directly formulates and studies a time-delayed, time-averaged cost-functional minimization problem.

This distinction is not merely algorithmic. A single-time mismatch may fluctuate in time, especially away from the true parameter or in the presence of noise. The time-averaged cost functional in Eq. (7) is designed to reduce the effect of such fluctuations. In addition, root-finding or local derivative-based updates may identify a local stationary point of the parameter landscape, while the present formulation treats parameter recovery as an inverse problem based on minimization of the nudging-induced cost functional. This allows us to connect small values of the cost functional to small parameter error under suitable nondegeneracy assumptions.

Another difference concerns the observation requirements. In the Lorenz–63 example considered in [46], the authors note that their parameter update appears to require the number of estimated parameters to be no larger than the rank of the observation operator; in practice, estimating all three Lorenz parameters leads them to nudge the full state. In contrast, the framework studied here is designed specifically to use partial state observations. For Lorenz–63, we focus on the case where only one component, such as $x(t)$, is observed, and we show theoretically and numerically how nudging can use this partial observation to recover the missing state information relevant for parameter estimation.

Algorithm 1 should therefore be interpreted not merely as a numerical recipe, but as a way of approximating the ideal full-observation parameter estimation problem using only partial observations: the nudging step reconstructs dynamically consistent state information from the available data, while the optimization step selects the parameter value for which this assimilated trajectory best agrees with the observations.

Algorithm 1: Data-assimilation-augmented parameter estimation

Input: Observed data $\mathcal{O}(t) = H\mathbf{x}(t)$ on $[\tau, \tau + T]$; initial condition $\tilde{\mathbf{x}}_0$ for the nudged system; initial parameter guess $\tilde{\theta}^{(0)} \in \Theta$; feedback operator $\mathcal{B}$; stopping criterion η

Output: Estimated parameter vector θ_e

1 Define the cost functional

$$C(\tilde{\theta}) = \frac{1}{T} \int_{\tau}^{\tau+T} \left\| \mathcal{O}(t) - H\tilde{\mathbf{x}}(t; \tilde{\theta}) \right\|_{\mathbb{R}^q}^2 dt,$$

where $\tilde{\mathbf{x}}(t; \tilde{\theta})$ solves the nudged system (6) with parameter value $\tilde{\theta}$;

2 Initialize

$$\tilde{\theta} \leftarrow \tilde{\theta}^{(0)}.$$

3 **while** the stopping criterion η is not satisfied **do**

4 Solve the nudged system (6) on $[\tau, \tau + T]$ using the current parameter value $\tilde{\theta}$;

5 Evaluate the cost functional

$$C(\tilde{\theta}) = \frac{1}{T} \int_{\tau}^{\tau+T} \left\| \mathcal{O}(t) - H\tilde{\mathbf{x}}(t; \tilde{\theta}) \right\|_{\mathbb{R}^q}^2 dt.$$

6 Update $\tilde{\theta}$ using a numerical optimization algorithm;

7 **end**

8 Set

$$\theta_e \leftarrow \tilde{\theta}.$$

9 **return** θ_e

4 The Lorenz'63 System

In 1963, Edward Lorenz introduced a mathematical model to describe atmospheric convection [66]. This model, widely recognized as the Lorenz system, is characterized by three ordinary differential equations (ODEs):

$$\begin{cases} \dot{x} = \sigma(y - x), \\ \dot{y} = x(\rho - z) - y, \\ \dot{z} = xy - \beta z, \end{cases} \quad (8)$$

where $\theta = (\sigma, \rho, \beta)$ is the parameter vector for the system. These equations describe the dynamics of a two-dimensional fluid layer that is uniformly heated from below and cooled from above. They specifically capture the temporal rate of change of three critical variables: x , which is proportional to the rate of convection; y , which denotes the horizontal temperature variation; and z , which reflects the vertical temperature variation [67]. The parameter σ represents the Prandtl number, while ρ corresponds to the Rayleigh number and β represents certain physical dimensions of the layer itself. From a technical perspective, the Lorenz system is classified as a nonlinear, aperiodic, three-dimensional, deterministic ODE.

The dynamics of the Lorenz system is highly sensitive to the values of its parameters, particularly the Rayleigh number ρ . While all three parameters (σ, β, ρ) influence the dynamics, we focus here on the role of the bifurcation parameter ρ in determining the structure and stability of equilibrium points. For $\rho > 1$, the system admits three equilibrium points: the origin $(0, 0, 0)$, and two symmetric and nontrivial equilibrium points given by

$$\left(\pm \sqrt{\beta(\rho - 1)}, \pm \sqrt{\beta(\rho - 1)}, \rho - 1 \right). \quad (9)$$

In contrast, when $\rho < 1$, the origin is the unique equilibrium and is globally asymptotically stable; all trajectories converge to the origin as $t \rightarrow \infty$. As ρ increases beyond unity, the origin loses stability, and the nontrivial equilibrium points emerge through a pitchfork bifurcation. These remain locally asymptotically stable for values of ρ up to the critical threshold

$$\rho_H := \sigma \cdot \frac{\sigma + \beta + 3}{\sigma - \beta - 1},$$

provided $\sigma > \beta + 1$. For $\rho > \rho_H$, nontrivial equilibrium points lose stability via a Hopf bifurcation, and the system transitions to a regime characterized by chaotic dynamics, marked by the emergence of the Lorenz attractor—a *strange* attractor with sensitive dependence on initial conditions [68]. The bifurcation structure and stability properties of these equilibria are central to understanding the rich dynamical behavior exhibited by the Lorenz system. It is also well known that the Lorenz system is dissipative and possesses an absorbing ball in the phase space, which contains the attractor. In particular, one can show that [69],

$$\limsup_{t \rightarrow \infty} \sqrt{x^2 + y^2 + z^2} \leq R \text{ where } R = \frac{3(\rho + \sigma)}{\sqrt{\min\{\sigma, \beta, 1\}}}. \quad (10)$$

Moreover, the ball $B(0; R)$ is absorbing and invariant for the Lorenz semigroup and contains the Lorenz attractor. Subsequently, we will assume that the parameter vector $\theta = (\sigma, \rho, \beta)$ belongs to a compact set. Due to (10), we will henceforth assume that all trajectories $\{X_\theta(t) : t \in \mathcal{I}\}$ satisfy the bound

$$\sup_{t \in \mathcal{I}} \|X_\theta(t)\| \leq \mathfrak{K}, \quad (11)$$

where $\mathfrak{K} > 0$ is a constant.

4.1 Nudging and Forward Prediction for Lorenz'63

We first illustrate the synchronization properties that allow us to establish the well-posedness of the data-to-parameter inverse map based on partial state observation. It also motivates the parameter estimation framework outlined in Section 5.3. In the Lorenz system, observing either the x or the y component is sufficient to synchronize the system and recover the full state, while observing only the z component is insufficient for this purpose.

The *determining functional* property enables recovery of the full system dynamics, and consequently the unknown parameters, using only partial observations. In what follows, we assume that the x -component is continuously observed. The nudged system corresponding to the Lorenz system (8) is given by

$$\begin{cases} \dot{\tilde{x}} &= \tilde{\sigma}(\tilde{y} - \tilde{x}) + \mu(x - \tilde{x}), \\ \dot{\tilde{y}} &= \tilde{x}(\tilde{\rho} - \tilde{z}) - \tilde{y}, \\ \dot{\tilde{z}} &= \tilde{x}\tilde{y} - \tilde{\beta}\tilde{z}, \end{cases} \quad (12)$$

where μ denotes the nudging coefficient and x corresponds to the data obtained by numerically solving the Lorenz system in Eq. (8). Next, we analyze the behavior of the nudged system when the parameter values used in the nudged model do not coincide with the true parameters. This situation naturally arises in parameter estimation: the true parameter vector is unknown, and the nudged system must be solved using an initial guess. Thus, the parameter mismatch represents the difference between the true parameter vector and the current trial parameter vector used in the nudged dynamics. Theorem 1 quantifies how the nudged solution remains exponentially stable and converges to a neighborhood of the true trajectory determined by the magnitude of the parameter mismatch.

The result in Theorem 1 does not rely on chaotic behavior of the Lorenz system. Rather, it uses only the uniform boundedness of the reference trajectory, as given in Eq. (11), and a sufficiently large nudging coefficient. Consequently, the estimate applies in both chaotic and non-chaotic parameter regimes, provided the stated hypotheses are satisfied.

Theorem 1 (Exponential synchronization and the determining map). *Let $X_\theta(t) = (x(t), y(t), z(t))$ be a solution of the Lorenz-63 system (8) corresponding to the parameter vector $\theta = (\sigma, \rho, \beta)$. Assume that X_θ satisfies the uniform bound (11).*

- (i) *Let $(\tilde{x}(t), \tilde{y}(t), \tilde{z}(t))$ be a solution of the nudged system (12) initialized from an arbitrary initial condition $(\tilde{x}(0), \tilde{y}(0), \tilde{z}(0)) \in \mathbb{R}^3$, with mismatched parameters $(\tilde{\sigma}, \tilde{\rho}, \tilde{\beta}) \neq (\sigma, \rho, \beta)$. Assume that the x -component of the reference solution is observed. Here $\tilde{\sigma}, \tilde{\beta} > 0$, $\tilde{\rho} \in \mathbb{R}$, and $\mu > 0$. Define the errors*

$$\bar{x} = x - \tilde{x}, \quad \bar{y} = y - \tilde{y}, \quad \bar{z} = z - \tilde{z},$$

and the parameter mismatches

$$\delta_\sigma := \sigma - \tilde{\sigma}, \quad \delta_\rho := \rho - \tilde{\rho}, \quad \delta_\beta := \beta - \tilde{\beta}.$$

If μ satisfies

$$\mu > (|\tilde{\sigma} + \tilde{\rho}| + \mathfrak{K})^2 + \frac{\mathfrak{K}^2}{\tilde{\beta}} + \frac{1}{4} - \tilde{\sigma}, \quad (13)$$

then there exists $\gamma > 0$ such that, for all $t \geq t_0$,

$$\bar{x}^2(t) + \bar{y}^2(t) + \bar{z}^2(t) \leq (\bar{x}^2(t_0) + \bar{y}^2(t_0) + \bar{z}^2(t_0))e^{-2\gamma(t-t_0)} + \frac{C_{\mathfrak{K}}}{\gamma}(\delta_\sigma^2 + \delta_\rho^2 + \delta_\beta^2), \quad (14)$$

where $C_{\mathfrak{K}} = \mathfrak{K}^2 \max \left\{ 4, 1, \frac{1}{\tilde{\beta}} \right\}$. In particular,

$$\limsup_{t \rightarrow \infty} (\bar{x}^2(t) + \bar{y}^2(t) + \bar{z}^2(t)) \leq \frac{C_{\mathfrak{K}}}{\gamma}(\delta_\sigma^2 + \delta_\rho^2 + \delta_\beta^2).$$

So the nudged solution converges exponentially fast to a $O(|\delta|)$ -neighborhood of the true solution.

- (ii) *If $\tilde{\sigma} = \sigma^*, \tilde{\rho} = \rho^*, \tilde{\beta} = \beta^*$ and the trajectory $\{\mathbf{x}(t), t \in \mathbb{R}\}$ is on the global attractor $\mathcal{A}$ of the Lorenz system, then*

$$\lim_{t \rightarrow \infty} \|\tilde{\mathbf{x}}(t) - \mathbf{x}(t)\| = 0.$$

- (iii) *Assume that $\tilde{X}_i(t) = (\tilde{x}_i(t), \tilde{y}_i(t), \tilde{z}_i(t))$, $i = 1, 2$, be two solutions of (12) with $x(t)$ replaced by $x_i(t)$, $i = 1, 2$ respectively, where $x_i(\cdot) \in C_{bd}(\mathcal{I}; \mathbb{R})$. Here $C_{bd}(\mathcal{I}; \mathbb{R})$ denotes the space of bounded continuous real-valued functions on the time interval $\mathcal{I}$, and $\mathcal{I}$ contains an interval of the form $[t_0, \infty)$. Then there exists a constant $\tilde{C}_{\mathfrak{K}}$ and μ sufficiently large depending on $\mathfrak{K}$ and $(\tilde{\sigma}, \tilde{\rho}, \tilde{\beta})$ such that*

$$\|\tilde{X}_1(\cdot) - \tilde{X}_2(\cdot)\|_\infty \leq e^{-\gamma(t-t_0)} \|\tilde{X}_1(t_0) - \tilde{X}_2(t_0)\| + \tilde{C}_{\mathfrak{K}} \|x_1(\cdot) - x_2(\cdot)\|_\infty.$$

In particular, if either $\tilde{X}_1(t_0) = \tilde{X}_2(t_0)$, or the trajectories $\tilde{X}_i$, $i = 1, 2$ are on the attractor with $\mathcal{I} = \mathbb{R}$, then the following estimate holds:

$$\|\tilde{X}_1(\cdot) - \tilde{X}_2(\cdot)\|_\infty \leq \tilde{C}_{\mathfrak{K}} \|x_1(\cdot) - x_2(\cdot)\|_\infty. \quad (15)$$

Proof. We first prove (i). Without loss of generality, assume $t_0 = 0$. Subtract the nudged system from the true system and set $\bar{x} = x - \tilde{x}$, $\bar{y} = y - \tilde{y}$, $\bar{z} = z - \tilde{z}$. A direct computation gives:

$$\begin{aligned} \dot{\bar{x}} &= \sigma(y - x) - \tilde{\sigma}(\tilde{y} - \tilde{x}) - \mu(x - \tilde{x}) = -(\tilde{\sigma} + \mu)\bar{x} + \tilde{\sigma}\bar{y} + \delta_\sigma(y - x), \\ \dot{\bar{y}} &= x(\rho - z) - y - (\tilde{x}(\tilde{\rho} - \tilde{z}) - \tilde{y}) = \tilde{\rho}\bar{x} - \bar{y} + (\tilde{x}\tilde{z} - xz) + \delta_\rho x, \\ \dot{\bar{z}} &= xy - \beta z - (\tilde{x}\tilde{y} - \tilde{\beta}\tilde{z}) = (xy - \tilde{x}\tilde{y}) - \tilde{\beta}\bar{z} - \delta_\beta z. \end{aligned} \quad (16)$$

Multiply the three equations in Eq. (16) by $\bar{x}$, $\bar{y}$, $\bar{z}$, respectively, and add. This yields:

$$\frac{1}{2} \frac{d}{dt} (\bar{x}^2 + \bar{y}^2 + \bar{z}^2) + (\tilde{\sigma} + \mu) \bar{x}^2 + \bar{y}^2 + \tilde{\beta} \bar{z}^2 = (\tilde{\sigma} + \tilde{\rho}) \bar{x} \bar{y} + (\tilde{x} \tilde{z} - xz) \bar{y} + (xy - \tilde{x} \tilde{y}) \bar{z} + \delta_\sigma (y - x) \bar{x} + \delta_\rho x \bar{y} - \delta_\beta z \bar{z}. \quad (17)$$

Expand the nonlinear terms:

$$(\tilde{x} \tilde{z} - xz) \bar{y} = ((x - \bar{x})(z - \bar{z}) - xz) \bar{y} = (-x\bar{z} - z\bar{x} + \bar{x}\bar{z}) \bar{y},$$

$$(xy - \tilde{x} \tilde{y}) \bar{z} = (xy - (x - \bar{x})(y - \bar{y})) \bar{z} = (x\bar{y} + y\bar{x} - \bar{x}\bar{y}) \bar{z}.$$

Adding these two identities cancels the terms $x\bar{y}\bar{z}$ and $\bar{x}\bar{y}\bar{z}$, leaving

$$(\tilde{x} \tilde{z} - xz) \bar{y} + (xy - \tilde{x} \tilde{y}) \bar{z} = y\bar{x}\bar{z} - z\bar{x}\bar{y}. \quad (18)$$

Substituting Eq. (18) into Eq. (17) gives

$$\frac{1}{2} \frac{d}{dt} V + (\tilde{\sigma} + \mu) \bar{x}^2 + \bar{y}^2 + \tilde{\beta} \bar{z}^2 = (\tilde{\sigma} + \tilde{\rho}) \bar{x} \bar{y} + y\bar{x}\bar{z} - z\bar{x}\bar{y} + \delta_\sigma (y - x) \bar{x} + \delta_\rho x \bar{y} - \delta_\beta z \bar{z}, \quad (19)$$

where $V(t) := \bar{x}^2 + \bar{y}^2 + \bar{z}^2$. By the uniform boundedness of Lorenz (see [66]), there exists $\mathfrak{K} > 0$ such that $|x(t)|, |y(t)|, |z(t)| \leq \mathfrak{K}$ for all $t \geq 0$, hence

$$(\tilde{\sigma} + \tilde{\rho}) \bar{x} \bar{y} - z\bar{x}\bar{y} \leq (|\tilde{\sigma} + \tilde{\rho}| + \mathfrak{K}) |\bar{x} \bar{y}|, \quad y\bar{x}\bar{z} \leq \mathfrak{K} |\bar{x} \bar{z}|,$$

and also $|y - x| \leq |y| + |x| \leq 2\mathfrak{K}$. Therefore, from Eq. (19),

$$\frac{1}{2} \dot{V} + (\tilde{\sigma} + \mu) \bar{x}^2 + \bar{y}^2 + \tilde{\beta} \bar{z}^2 \leq (|\tilde{\sigma} + \tilde{\rho}| + \mathfrak{K}) |\bar{x} \bar{y}| + \mathfrak{K} |\bar{x} \bar{z}| + 2\mathfrak{K} |\delta_\sigma| |\bar{x}| + \mathfrak{K} |\delta_\rho| |\bar{y}| + \mathfrak{K} |\delta_\beta| |\bar{z}|. \quad (20)$$

By Young's inequality, we have:

$$(|\tilde{\sigma} + \tilde{\rho}| + \mathfrak{K}) |\bar{x} \bar{y}| \leq (|\tilde{\sigma} + \tilde{\rho}| + \mathfrak{K})^2 \bar{x}^2 + \frac{1}{4} \bar{y}^2,$$

$$\mathfrak{K} |\bar{x} \bar{z}| \leq \frac{\mathfrak{K}^2}{\tilde{\beta}} \bar{x}^2 + \frac{\tilde{\beta}}{4} \bar{z}^2,$$

$$2\mathfrak{K} |\delta_\sigma| |\bar{x}| \leq 4\mathfrak{K}^2 \delta_\sigma^2 + \frac{1}{4} \bar{x}^2, \quad \mathfrak{K} |\delta_\rho| |\bar{y}| \leq \mathfrak{K}^2 \delta_\rho^2 + \frac{1}{4} \bar{y}^2, \quad \mathfrak{K} |\delta_\beta| |\bar{z}| \leq \frac{\mathfrak{K}^2}{\tilde{\beta}} \delta_\beta^2 + \frac{\tilde{\beta}}{4} \bar{z}^2.$$

Substituting these bounds into Eq. (20) yields

$$\begin{aligned} \frac{1}{2} \dot{V} + \left(\tilde{\sigma} + \mu - (|\tilde{\sigma} + \tilde{\rho}| + \mathfrak{K})^2 - \frac{\mathfrak{K}^2}{\tilde{\beta}} - \frac{1}{4} \right) \bar{x}^2 + \frac{1}{2} \bar{y}^2 + \frac{\tilde{\beta}}{2} \bar{z}^2 \\ \leq 4\mathfrak{K}^2 \delta_\sigma^2 + \mathfrak{K}^2 \delta_\rho^2 + \frac{\mathfrak{K}^2}{\tilde{\beta}} \delta_\beta^2 \leq C_{\mathfrak{K}} (\delta_\sigma^2 + \delta_\rho^2 + \delta_\beta^2). \end{aligned} \quad (21)$$

Assumption in Eq. (13) ensures that the coefficient of $\bar{x}^2$ in Eq. (21) is positive. Define

$$\gamma := \min \left\{ \tilde{\sigma} + \mu - (|\tilde{\sigma} + \tilde{\rho}| + \mathfrak{K})^2 - \frac{\mathfrak{K}^2}{\tilde{\beta}} - \frac{1}{4}, \frac{1}{2}, \frac{\tilde{\beta}}{2} \right\} > 0.$$

Since $V = \bar{x}^2 + \bar{y}^2 + \bar{z}^2$, inequality (21) implies

$$\frac{1}{2} \dot{V} + \gamma V \leq C_{\mathfrak{K}} (\delta_\sigma^2 + \delta_\rho^2 + \delta_\beta^2).$$

By Grönwall's inequality,

$$V(t) \leq V(0) e^{-2\gamma t} + \frac{C_{\mathfrak{K}}}{\gamma} (\delta_\sigma^2 + \delta_\rho^2 + \delta_\beta^2),$$

which is exactly Eq. (14). The limsup bound follows by letting $t \rightarrow \infty$.

Proof of (ii) follows from (14) with $\delta_\sigma = \delta_\rho = \delta_\beta = 0$, recalling that all trajectories on the attractor satisfy the uniform bound (11) and letting $t \rightarrow \infty$. The proof of (iii) is similar to the proof of (i) and is omitted. $\square$

Remark 1. Similar convergence results can be obtained when only the y -component is observed.

4.2 Well-posedness of the Data-to-Parameter Map for Lorenz '63

We start by providing a comprehensive analysis of this issue when *all* state variables are observed. Clearly, if the parameter determination problem is ill-posed when the full state is observed, it will be so for partial state observations as well. In Theorem 2, we provide a complete analysis of the well-posedness of the inverse map in this case.

Theorem 2 (Well-posedness when the full state is observed). *Let $\mathbf{x}(t) = (x(t), y(t), z(t))$ be a C^1 trajectory of the Lorenz system (8), where $\sigma, \rho, \beta \in \mathbb{R}$ are unknown parameters.*

- (i) *Fix a time t_0 , and assume that we observe the values $(x(t_0), y(t_0), z(t_0), \dot{x}(t_0), \dot{y}(t_0), \dot{z}(t_0))$. If $y(t_0) \neq x(t_0)$, $x(t_0) \neq 0$, and $z(t_0) \neq 0$, then (σ, ρ, β) is uniquely determined by*

$$\sigma = \frac{\dot{x}}{y - x}, \quad \rho = \frac{\dot{y} + xz + y}{x}, \quad \beta = \frac{xy - \dot{z}}{z}.$$

- (ii) *Let $\mathcal{S} = \{(x, y, z) : y = x \text{ or } x = 0 \text{ or } z = 0\}$ and for $\delta > 0$, let $\mathcal{S}_\delta = \{\mathbf{x} \in \mathbb{R}^3 : d(\mathbf{x}, \mathcal{S}) \geq \delta\}$. Let $\theta_i = (\sigma_i, \rho_i, \beta_i)$, $i = 1, 2$ be two parameter vectors for the Lorenz system such that the corresponding trajectories $\mathbf{x}_i$ belong to $\mathcal{S}_\delta$ for all $t \in \mathcal{I}$. Then, we have the (well-posedness) estimate*

$$\|\theta_1 - \theta_2\|_{\mathbb{R}^3} \leq C_{\delta, M} \|\mathbf{x}_1 - \mathbf{x}_2\|_{C_{bd}^1(\mathcal{I}; \mathbb{R}^3)}, \quad (22)$$

where $M = \max_{i=1,2} \|\mathbf{x}_i\|_{C_{bd}^1(\mathcal{I}; \mathbb{R}^3)}$, $C_{bd}^1(\mathcal{I}; \mathbb{R}^3) = \left\{ \mathbf{x} \in C^1(\mathcal{I}; \mathbb{R}^3) : \|\mathbf{x}\|_{C_{bd}^1(\mathcal{I}; \mathbb{R}^3)} < \infty \right\}$, and $\|\mathbf{x}\|_{C_{bd}^1(\mathcal{I}; \mathbb{R}^3)} = \sup_{t \in \mathcal{I}} \|\mathbf{x}(t)\|_{\mathbb{R}^3} + \sup_{t \in \mathcal{I}} \|\dot{\mathbf{x}}(t)\|_{\mathbb{R}^3}$.

- (iii) *If the trajectory is the equilibrium solution*

$$(x(t), y(t), z(t)) = (0, 0, 0), \quad \text{for all } t,$$

then none of the parameters σ, ρ, β is identifiable. On the other hand, if the trajectory is one of the two non-trivial equilibrium points given in (9), then β and ρ are uniquely identifiable while σ is not identifiable.

- (iv) *For the trajectory $x(t) = y(t) = 0, z(t) = e^{-\beta t} z(0)$, where $z(0) \neq 0$, only the parameter β is identifiable.*

- (v) *Assume that $\sigma \neq 0$ and $\rho \neq 0$. If $\mathbf{x}(\cdot)$ is any trajectory of the Lorenz system, then unless it is an equilibrium point, or the trajectory given in (iv), there exists an interval $\mathcal{I}$ and a constant $\delta > 0$ such that $\{\mathbf{x}(t) : t \in \mathcal{I}\} \subset \mathcal{S}_\delta$ in which case the well-posedness estimate (22) holds.*

Proof. (i) Fix t_0 and denote

$$x = x(t_0), \quad y = y(t_0), \quad z = z(t_0), \quad \dot{x} = \dot{x}(t_0), \quad \dot{y} = \dot{y}(t_0), \quad \dot{z} = \dot{z}(t_0).$$

The Lorenz equations may be written as a linear algebraic system for the parameters:

$$\begin{bmatrix} y - x & 0 & 0 \\ 0 & x & 0 \\ 0 & 0 & -z \end{bmatrix} \begin{bmatrix} \sigma \\ \rho \\ \beta \end{bmatrix} = \begin{bmatrix} \dot{x} \\ \dot{y} + xz + y \\ \dot{z} - xy \end{bmatrix}.$$

If $y \neq x$, $x \neq 0$, and $z \neq 0$, then the coefficient matrix is invertible. Hence

$$\sigma = \frac{\dot{x}}{y - x}, \quad \rho = \frac{\dot{y} + xz + y}{x}, \quad \beta = \frac{xy - \dot{z}}{z}.$$

- (ii) Note first that if one observes the trajectory in a time interval, then one can determine its derivative as well. In view of part (i), define the recovery map

$$\Phi(x, y, z, \dot{x}, \dot{y}, \dot{z}) = \left(\frac{\dot{x}}{y - x}, \frac{\dot{y} + xz + y}{x}, \frac{xy - \dot{z}}{z} \right).$$

By part (i), for each trajectory $\mathbf{x}_i = (x_i, y_i, z_i)$, $i = 1, 2$, the corresponding parameter vector satisfies

$$\theta_i = \Phi(x_i, y_i, z_i, \dot{x}_i, \dot{y}_i, \dot{z}_i).$$

Since $\mathbf{x}_i(t) \in \mathcal{S}_\delta$ for all $t \in \mathcal{I}$, the denominators $y_i(t) - x_i(t)$, $x_i(t)$, and $z_i(t)$ are uniformly bounded away from zero. Moreover, the quantities $(x_i, y_i, z_i, \dot{x}_i, \dot{y}_i, \dot{z}_i)$ are bounded in terms of

$$M = \max_{i=1,2} \|\mathbf{x}_i\|_{C_{bd}^1(\mathcal{I}; \mathbb{R}^3)}.$$

Hence Φ is smooth, and therefore Lipschitz, on the corresponding bounded subset of $\mathbb{R}^6 \setminus \{y = x \text{ or } x = 0 \text{ or } z = 0\}$. Consequently, there exists a constant $C > 0$ such that

$$\begin{aligned} \|\theta_1 - \theta_2\|_{\mathbb{R}^3} &\leq \frac{CM}{\delta} \|(x_1, y_1, z_1, \dot{x}_1, \dot{y}_1, \dot{z}_1) - (x_2, y_2, z_2, \dot{x}_2, \dot{y}_2, \dot{z}_2)\|_{L^\infty(\mathcal{I})} \\ &\leq \frac{CM}{\delta} \|\mathbf{x}_1 - \mathbf{x}_2\|_{C_{bd}^1(\mathcal{I}; \mathbb{R}^3)}. \end{aligned}$$

This proves the well-posedness estimate.

- (iii) At the equilibrium point $(0, 0, 0)$, we have $\dot{x} = \dot{y} = \dot{z} = 0$. Substitution into the Lorenz equations gives only the identities

$$0 = 0, \quad 0 = 0, \quad 0 = 0.$$

Thus no condition is imposed on σ , ρ , or β , and none of the parameters is identifiable from this equilibrium trajectory.

Now consider a non-trivial equilibrium point given in (9). At such a point,

$$\dot{x} = \dot{y} = \dot{z} = 0 \quad \text{and} \quad y = x.$$

The first Lorenz equation becomes $0 = \sigma(y - x) = 0$, and therefore imposes no condition on σ . The remaining two equations give

$$0 = x(\rho - z) - x, \quad 0 = x^2 - \beta z.$$

Since the non-trivial equilibrium satisfies $x \neq 0$ and $z \neq 0$, these equations imply

$$\rho = z + 1, \quad \beta = \frac{x^2}{z}.$$

Thus ρ and β are uniquely determined by the nonzero equilibrium state, while σ remains unidentifiable.

- (iv) Suppose $x(t) = y(t) = 0, z(t) = e^{-\beta t} z(0)$, with $z(0) \neq 0$. Then the first two Lorenz equations reduce to

$$0 = 0, \quad 0 = 0,$$

and hence impose no conditions on σ or ρ . The third equation gives $\dot{z} = -\beta z$. Since $z(t) \neq 0$ on the interval under consideration, we obtain

$$\beta = -\frac{\dot{z}(t)}{z(t)}.$$

Thus β is identifiable, while σ and ρ are not identifiable.

- (v) Assume that $\sigma \neq 0$ and $\rho \neq 0$. Let $\mathbf{x}(t) = (x(t), y(t), z(t))$ be a trajectory of the Lorenz system that is neither an equilibrium trajectory nor the trajectory described in (iv). We claim that there exists a time t_* such that

$$x(t_*) \neq 0, \quad y(t_*) \neq x(t_*), \quad z(t_*) \neq 0.$$

Suppose, for contradiction, that no such time exists. Then

$$x(t)(y(t) - x(t))z(t) = 0$$

for all t in the time interval under consideration. Since the Lorenz vector field is polynomial, its solutions are real analytic in time. Hence the function

$$t \mapsto x(t)(y(t) - x(t))z(t)$$

is real analytic and vanishes identically. By [70, Corollary 1.2.6], a real analytic function whose zero set has an accumulation point in an interval must vanish identically on that interval. Therefore, if none of the factors $x(t)$, $y(t) - x(t)$, and $z(t)$ vanished identically on any nontrivial subinterval, each factor would have only isolated

zeros. The union of three isolated zero sets cannot contain an interval. Hence, on some nontrivial subinterval, one of the following alternatives holds:

$$x(t) \equiv 0, \quad y(t) \equiv x(t), \quad \text{or} \quad z(t) \equiv 0.$$

We examine these cases. First, suppose $x(t) \equiv 0$ on a nontrivial interval. Then $\dot{x}(t) \equiv 0$, and the first Lorenz equation gives

$$0 = \sigma(y(t) - x(t)) = \sigma y(t).$$

Since $\sigma \neq 0$, it follows that $y(t) \equiv 0$ on this interval. The third equation then reduces to

$$\dot{z}(t) = -\beta z(t).$$

Thus

$$z(t) = e^{-\beta(t-t_0)} z(t_0),$$

and the trajectory is precisely of the type described in (iv), which is excluded.

Second, suppose $y(t) \equiv x(t)$ on a nontrivial interval. Then

$$\dot{x}(t) = \sigma(y(t) - x(t)) = 0,$$

so $x(t) \equiv c$ and $y(t) \equiv c$ on that interval. The second equation becomes

$$0 = c(\rho - z(t)) - c = c(\rho - z(t) - 1).$$

If $c = 0$, then $x(t) = y(t) = 0$, and the trajectory reduces to the case described in (iv). If $c \neq 0$, then

$$z(t) \equiv \rho - 1.$$

The third equation then gives

$$0 = c^2 - \beta z.$$

Thus z is constant, and hence x , y , and z are all constant on this interval. Therefore, this is an equilibrium trajectory and thus excluded by assumption.

Third, suppose $z(t) \equiv 0$ on a nontrivial interval. Then $\dot{z}(t) \equiv 0$, and the third Lorenz equation gives

$$0 = x(t)y(t).$$

Since x and y are continuous, either $x(t) \equiv 0$ or $y(t) \equiv 0$ on a nontrivial sub-interval. If $x(t) \equiv 0$, then the first case applies. If $y(t) \equiv 0$, then the second Lorenz equation gives

$$0 = x(t)(\rho - z(t)) - y(t) = \rho x(t),$$

because $z(t) \equiv 0$ and $y(t) \equiv 0$. Since $\rho \neq 0$, it follows that $x(t) \equiv 0$, and we again reduce to the first case.

In each case, the trajectory is an equilibrium trajectory or the trajectory described in (iv), both of which are excluded. Therefore there must exist a time t_* such that

$$x(t_*) \neq 0, \quad y(t_*) \neq x(t_*), \quad z(t_*) \neq 0.$$

By continuity, these inequalities persist on a sufficiently small interval $\mathcal{I}$ containing t_* . Hence there exists $\delta > 0$ such that

$$\{\mathbf{x}(t) : t \in \mathcal{I}\} \subset \mathcal{S}_\delta.$$

The well-posedness estimate (22) then follows from part (ii). □

We now pass from full-state information to partial-state information. The previous result shows that, away from the degeneracy set $\mathcal{S}$, the full Lorenz trajectory determines the parameter vector in a well-conditioned way. However, the parameter-estimation framework developed in this paper uses only partial observations, and in the numerical experiments we primarily use the x -component. Theorem 3 shows that this partial observation is sufficient: if two admissible Lorenz trajectories remain away from the degeneracy set, then the distance between their parameter vectors can be controlled directly by the difference between their observed x -components. The proof follows by combining Theorem 1 and Theorem 2.

Theorem 3 (Well-posedness under partial observation of the state). *Let*

$$\mathcal{S} = \{(x, y, z) : y = x \text{ or } x = 0 \text{ or } z = 0\}$$

and for $\delta > 0$, let

$$\mathcal{S}_\delta = \{\mathbf{x} \in \mathbb{R}^3 : d(\mathbf{x}, \mathcal{S}) \geq \delta\}.$$

Let $\Theta_{\text{ad}} \subset \mathbb{R}^3$ be a compact set of admissible parameter values, and let

$$\theta_i = (\sigma_i, \rho_i, \beta_i) \in \Theta_{\text{ad}}, \quad i = 1, 2$$

be two parameter vectors for the Lorenz system. Assume that the corresponding trajectories

$$X_i(\cdot) = (x_i(\cdot), y_i(\cdot), z_i(\cdot)), \quad i = 1, 2,$$

belong to $\mathcal{S}_\delta$ for all $t \in \mathcal{I}$, where $\mathcal{I} \subset \mathbb{R}$ is any nontrivial interval if $X_i(\cdot) \in \mathcal{A}, i = 1, 2$, or else contains the interval $[t_0, \infty)$ for some $t_0 > 0$. Then, there exists a constant $C_{\delta, \Theta_{\text{ad}}} > 0$ such that we have the (well-posedness) estimate

$$\|\theta_1 - \theta_2\|_{\mathbb{R}^3} \leq C_{\delta, \Theta_{\text{ad}}} \|x_1 - x_2\|_{L^\infty(\mathcal{I})}, \quad (23)$$

where the trajectories satisfy (11).

Proof. Let $e_x = \|x_1 - x_2\|_{L^\infty(\mathcal{I})}$. Since x is a determining functional for the Lorenz system on the class of trajectories satisfying (11), there exists a constant $C_D > 0$, depending only on the absorbing bound and on the admissible parameter set Θ_{ad} , such that

$$\|X_1 - X_2\|_{L^\infty(\mathcal{I})} \leq C_D \|x_1 - x_2\|_{L^\infty(\mathcal{I})} = C_D e_x.$$

In particular,

$$\|y_1 - y_2\|_{L^\infty(\mathcal{I})} + \|z_1 - z_2\|_{L^\infty(\mathcal{I})} \leq C_D e_x. \quad (24)$$

Choose a compact subinterval $[a, b] \subset \mathcal{I}$ and set $L = b - a > 0$. Since $X_i(t) \in \mathcal{S}_\delta$, we have

$$|y_i(t) - x_i(t)| \geq \delta, \quad |x_i(t)| \geq \delta, \quad |z_i(t)| \geq \delta, \quad t \in [a, b].$$

By continuity, each of $y_i - x_i$, x_i , and z_i has a fixed sign on $[a, b]$. Hence

$$\left| \int_a^b (y_i - x_i) dt \right| \geq \delta L, \quad \left| \int_a^b x_i dt \right| \geq \delta L, \quad \left| \int_a^b z_i dt \right| \geq \delta L. \quad (25)$$

We estimate the three parameters separately. From the first Lorenz equation, $\dot{x}_i = \sigma_i(y_i - x_i)$. Integrating over $[a, b]$ gives

$$x_i(b) - x_i(a) = \sigma_i \int_a^b (y_i - x_i) dt.$$

Therefore,

$$\sigma_i = \frac{x_i(b) - x_i(a)}{\int_a^b (y_i - x_i) dt}.$$

Using (25), (24), and the uniform bound (11), we obtain

$$|\sigma_1 - \sigma_2| \leq C_{\delta, \Theta_{\text{ad}}} e_x.$$

Next, from the second Lorenz equation,

$$\dot{y}_i = \rho_i x_i - y_i - x_i z_i.$$

Integrating over $[a, b]$ gives

$$y_i(b) - y_i(a) = \rho_i \int_a^b x_i dt - \int_a^b y_i dt - \int_a^b x_i z_i dt.$$

Hence

$$\rho_i = \frac{y_i(b) - y_i(a) + \int_a^b y_i dt + \int_a^b x_i z_i dt}{\int_a^b x_i dt}.$$

Again using (25), the uniform bound (11), and (24), we find

$$|\rho_1 - \rho_2| \leq C_{\delta, \Theta_{\text{ad}}} e_x.$$

Finally, from the third Lorenz equation,

$$\dot{z}_i = x_i y_i - \beta_i z_i.$$

Integrating over $[a, b]$ gives

$$z_i(b) - z_i(a) = \int_a^b x_i y_i dt - \beta_i \int_a^b z_i dt.$$

Therefore,

$$\beta_i = \frac{\int_a^b x_i y_i dt - (z_i(b) - z_i(a))}{\int_a^b z_i dt}.$$

Using (25), (11), and (24), we obtain

$$|\beta_1 - \beta_2| \leq C_{\delta, \Theta_{\text{ad}}} e_x.$$

Combining the three estimates yields

$$\|\theta_1 - \theta_2\|_{\mathbb{R}^3} \leq C_{\delta, \Theta_{\text{ad}}} \|x_1 - x_2\|_{L^\infty(I)}.$$

This proves (23). □

4.3 Nudging-Augmented Optimization for Lorenz '63 Parameter Estimation

Theorem 2 shows that, in principle, the parameters of the Lorenz system can be recovered from pointwise measurements of the state and its time derivatives. This corresponds to the classical setting of parameter estimation, where both the full state and derivative information are assumed to be available. However, this setting is rarely available in practice. First, frequently we only have partial information of the state, either, for instance, some components of the state variables or a functional of them. Secondly, even when full state is observed, derivative information is typically not directly accessible from observed data and must be approximated numerically, introducing additional error. This might introduce large errors in derivative estimates for a chaotic system. We therefore employ the nudging-augmented optimization framework outlined in Section 3.2. For the Lorenz–63 experiments, Algorithm 1 is applied with $\theta = (\sigma, \rho, \beta)$, and with observation of the x -component. Thus,

$$\mathcal{O}(t) = x(t), \quad H(x, y, z) = x.$$

The general cost functional (7) therefore reduces to

$$C(\tilde{\theta}) = C(\tilde{\sigma}, \tilde{\rho}, \tilde{\beta}) = \frac{1}{T} \int_{\tau}^{\tau+T} |\tilde{x}(t) - x(t)|^2 dt, \tau \geq 0, T > 0 \text{ and } \tilde{\theta} = (\tilde{\sigma}, \tilde{\rho}, \tilde{\beta}). \quad (26)$$

Let $\theta^* = (\sigma^*, \rho^*, \beta^*)$ and $X_{\theta^*} = (x^*, y^*, z^*)$ be the true trajectory of Lorenz on the attractor corresponding to the parameter θ^* . Our parameter estimation strategy is to minimize this cost functional with respect to $\tilde{\theta}$. However, our numerical experiments suggest that this cost functional is highly nonconvex and therefore it is conceivable that it might have multiple minima. Due to part (ii) of Theorem 1, we must have $C(\theta^*) = C(\sigma^*, \rho^*, \beta^*) = 0$. In other words, the minimum value of the cost functional is zero. The question then arises as to whether this minimum is unique. In other words, if $C(\tilde{\theta}) = 0$, must it imply $\tilde{\theta} = \theta^*$. We prove a partial uniqueness result of the minimizer in Theorem 4.

Theorem 4 (Uniqueness of parameter recovery with σ^* fixed). *Let $X(t) = (x(t), y(t), z(t)), t \in \mathbb{R}$ be a non-equilibrium solution of the Lorenz system with parameters $\theta^* = (\sigma^*, \rho^*, \beta^*)$ and assume that the trajectory lies on the global attractor. Then, for sufficiently large μ ,*

$$C(\sigma^*, \rho, \beta) = 0 \iff (\rho, \beta) = (\rho^*, \beta^*).$$

Proof. Assume $C(\sigma^*, \rho, \beta) = 0$. Then for all $t \in [0, T]$, $\tilde{x}(t) = x(t)$. On $[0, T]$, substituting $\tilde{x} = x$ into the $\tilde{x}$ -equation gives $\dot{x} = \dot{\tilde{x}} = \sigma^*(\tilde{y} - x)$, while the true x -equation yields $\dot{x} = \sigma^*(y - x)$. Since $\sigma^* > 0$, this immediately yields $\tilde{y}(t) = y(t)$ on $[0, T]$. Comparing the $\tilde{y}$ and y equations, we have

$$\dot{\tilde{y}} = x(\rho - \tilde{z}) - y, \quad \dot{y} = x(\rho^* - z) - y,$$

one obtains

$$0 = \dot{\tilde{y}} - \dot{y} = x(t)(\rho - \tilde{z}(t) - \rho^* + z(t)) \quad \text{for } t \in [0, T].$$

Because x is continuous and not identically zero on $[0, T]$, we get

$$\tilde{z}(t) - z(t) = \rho - \rho^* = \text{constant}, \quad \text{for all } t \in [0, T]. \quad (27)$$

Subtracting the $\tilde{z}$ and z equations gives

$$\dot{\tilde{z}} - \dot{z} = -\beta\tilde{z} + \beta^*z.$$

Since $\tilde{z} - z$ is constant on $[0, T]$, its derivative vanishes and hence

$$-\beta\tilde{z} + \beta^*z = 0, \quad \text{for all } t \in [0, T]. \quad (28)$$

Then, Eq. (27) and (28) gives:

$$\beta(\rho - \rho^*) + z(\beta - \beta^*) = 0, \quad \text{for all } t \in [0, T]. \quad (29)$$

If $\beta \neq \beta^*$, then Eq. (29) gives

$$z(t) = -\frac{\beta(\rho - \rho^*)}{\beta - \beta^*}, \quad t \in [0, T],$$

so z is constant on $[0, T]$. By Lemma 1, the trajectory $(x(\cdot), y(\cdot), z(\cdot))$ then must be an equilibrium point, contradicting the hypothesis of the theorem. Thus $\beta = \beta^*$ and consequently from (29), $\rho = \rho^*$ as well.

Conversely, assume $(\rho, \beta) = (\rho^*, \beta^*)$. Since the true trajectory (x, y, z) lies on the global attractor, then, for μ chosen as in the Theorem 1, the nudged system converges to the true system and all variables synchronize: $(\tilde{x}, \tilde{y}, \tilde{z}) \equiv (x, y, z)$. Consequently, $C(\sigma^*, \rho, \beta) = 0$. Therefore, $C(\sigma^*, \rho, \beta) = 0$ if and only if $(\rho, \beta) = (\rho^*, \beta^*)$. $\square$

We now derive a quantitative time-delayed stability estimate for the recovery of the full parameter vector $\theta = (\sigma, \rho, \beta)$. The resulting bound controls the full parameter error in terms of the time-delayed cost functional (i.e. the value of cost functional at the approximate minimum) and an exponentially decaying synchronization residual. We will first need the following two somewhat technical propositions. These propositions separate the argument into two steps: first establishing an abstract parameter-error estimate from the mismatch map, and then expressing the required nondegeneracy condition in terms of computable partial derivative equations. In what follows, the $L^2(\tau, \tau + T)$ -norm is taken with respect to the normalized Lebesgue measure dt/T . Thus,

$$\|f\|_{L^2(\tau, \tau+T)}^2 := \frac{1}{T} \int_{\tau}^{\tau+T} |f(t)|^2 dt.$$

Proposition 5. *Let $X(t) = (x(t), y(t), z(t))$ be a trajectory of (8) corresponding to the parameter vector. $\theta^* = (\sigma^*, \rho^*, \beta^*)$. For each $\theta = (\sigma, \rho, \beta)$ in a neighborhood $B(\theta^*)$ of θ^* , let $\tilde{X}(t; \theta) = (\tilde{x}(t; \theta), \tilde{y}(t; \theta), \tilde{z}(t; \theta))$ denote the solution of the nudged system (12). The nudged system is initialized at an arbitrary initial condition, not necessarily equal to $X(0)$. For $\tau \geq 0$ and $T > 0$, define the observation map*

$$\mathcal{F}(\theta)(t) = x(t) - \tilde{x}(t; \theta), \quad t \in [\tau, \tau + T],$$

viewed as a map $\mathcal{F} : B(\theta^) \subset \mathbb{R}^3 \rightarrow L^2(\tau, \tau + T)$. Note that, the cost functional, given in Eq. (26), can now be written as*

$$C(\theta) = \|\mathcal{F}(\theta)\|_{L^2(\tau, \tau+T)}^2 = \frac{1}{T} \int_{\tau}^{\tau+T} |x(t) - \tilde{x}(t; \theta)|^2 dt.$$

Assume further that the conditions on μ and γ are as in Theorem 1.

(i) *We have the estimate*

$$C(\theta) \leq e^{-2\gamma\tau} \left(2\mathfrak{K} + \frac{\|h\|^2}{2\gamma} \right) \left(\frac{1 - e^{-2\gamma T}}{2\gamma T} \right), \quad \text{where } h = \theta - \theta^*. \quad (30)$$

(ii) *$\mathcal{F}$ is continuously differentiable in $B(\theta^*)$. Let D denote the differentiation with respect to the parameter vector θ and $U \subset B(\theta^*)$ be a convex neighborhood of θ^* . Thus, for any $\theta \in U$, the line segment $\theta_s = \theta^* + s(\theta - \theta^*)$, $0 \leq s \leq 1$, is contained in U . Assume that there exists $\alpha > 0$ such that*

$$\|D\mathcal{F}(\theta^*)h\|_{L^2(\tau, \tau+T)} \geq \alpha\|h\|_{\mathbb{R}^3} \quad \text{for all } h \in \mathbb{R}^3,$$

and

$$\|(D\mathcal{F}(\eta) - D\mathcal{F}(\theta^*))h\|_{L^2(\tau, \tau+T)} \leq \frac{\alpha}{2}\|h\|_{\mathbb{R}^3} \quad \text{for all } \eta \in U \text{ and all } h \in \mathbb{R}^3.$$

Then, for every $\theta \in U$,

$$\|\theta - \theta^*\|_{\mathbb{R}^3} \leq \frac{2}{\alpha}C(\theta)^{1/2} + \frac{2}{\alpha} \left(\frac{1 - e^{-2\gamma T}}{2\gamma T} \right)^{1/2} e^{-\gamma\tau} \|X(0) - \tilde{X}(0; \theta^*)\|_{\mathbb{R}^3}.$$

Proof. (i) Applying Theorem 1 with $\theta = \theta^* + h$, and using the fact that the observed mismatch is only the x -component of the full synchronization error, we obtain

$$|x(t) - \tilde{x}(t; \theta)|^2 \leq \|X(t) - \tilde{X}(t; \theta)\|_{\mathbb{R}^3}^2 \leq e^{-2\gamma t} \left(2\mathfrak{K} + \frac{\|h\|^2}{2\gamma} \right).$$

Therefore, integrating this estimate over the time-delayed window $[\tau, \tau + T]$ gives

$$C(\theta) = \frac{1}{T} \int_{\tau}^{\tau+T} |x(t) - \tilde{x}(t; \theta)|^2 dt \leq e^{-2\gamma\tau} \left(2\mathfrak{K} + \frac{\|h\|^2}{2\gamma} \right) \left(\frac{1 - e^{-2\gamma T}}{2\gamma T} \right),$$

which is precisely (30).

(ii) *Proof of Part (ii):* Let $h = \theta - \theta^*$. By the fundamental theorem of calculus in Banach spaces,

$$\mathcal{F}(\theta) - \mathcal{F}(\theta^*) = \int_0^1 D\mathcal{F}(\theta^* + sh)h ds.$$

Hence,

$$\mathcal{F}(\theta) = \mathcal{F}(\theta^*) + D\mathcal{F}(\theta^*)h + \int_0^1 [D\mathcal{F}(\theta^* + sh) - D\mathcal{F}(\theta^*)]h ds.$$

Taking the $L^2(\tau, \tau + T)$ -norm and using the reverse triangle inequality gives

$$\|\mathcal{F}(\theta)\|_{L^2(\tau, \tau+T)} \geq \|D\mathcal{F}(\theta^*)h\|_{L^2(\tau, \tau+T)} - \|\mathcal{F}(\theta^*)\|_{L^2(\tau, \tau+T)} - \int_0^1 \|[D\mathcal{F}(\theta^* + sh) - D\mathcal{F}(\theta^*)]h\|_{L^2(\tau, \tau+T)} ds.$$

By the assumed lower bound on $D\mathcal{F}(\theta^*)$,

$$\|D\mathcal{F}(\theta^*)h\|_{L^2(\tau, \tau+T)} \geq \alpha\|h\|_{\mathbb{R}^3}.$$

Moreover, since $\theta^* + sh \in U$ for $0 \leq s \leq 1$, the continuity assumption on $D\mathcal{F}$ implies

$$\int_0^1 \|[D\mathcal{F}(\theta^* + sh) - D\mathcal{F}(\theta^*)]h\|_{L^2(\tau, \tau+T)} ds \leq \frac{\alpha}{2}\|h\|_{\mathbb{R}^3}.$$

Therefore,

$$\|\mathcal{F}(\theta)\|_{L^2(\tau, \tau+T)} \geq \frac{\alpha}{2}\|h\|_{\mathbb{R}^3} - \|\mathcal{F}(\theta^*)\|_{L^2(\tau, \tau+T)}.$$

Rearranging, we obtain

$$\|h\|_{\mathbb{R}^3} \leq \frac{2}{\alpha}\|\mathcal{F}(\theta)\|_{L^2(\tau, \tau+T)} + \frac{2}{\alpha}\|\mathcal{F}(\theta^*)\|_{L^2(\tau, \tau+T)}.$$

Since

$$C(\theta) = \|\mathcal{F}(\theta)\|_{L^2(\tau, \tau+T)}^2,$$

we have

$$\|\mathcal{F}(\theta)\|_{L^2(\tau, \tau+T)} = C(\theta)^{1/2}.$$

It remains to estimate $\|\mathcal{F}(\theta^*)\|_{L^2(\tau, \tau+T)}$. At the true parameter value, Theorem 1 gives

$$|\tilde{x}(t; \theta^*) - x(t)| \leq e^{-\gamma t} \|X(0) - \tilde{X}(0; \theta^*)\|_{\mathbb{R}^3}.$$

Thus, using the normalized L^2 -norm,

$$\begin{aligned}\|\mathcal{F}(\theta^*)\|_{L^2(\tau, \tau+T)} &= \left(\frac{1}{T} \int_{\tau}^{\tau+T} |x(t) - \tilde{x}(t; \theta^*)|^2 dt \right)^{1/2} \\ &\leq \left(\frac{1}{T} \int_{\tau}^{\tau+T} e^{-2\gamma t} dt \right)^{1/2} \|X(0) - \tilde{X}(0; \theta^*)\|_{\mathbb{R}^3} \\ &= e^{-\gamma\tau} \left(\frac{1 - e^{-2\gamma T}}{2\gamma T} \right)^{1/2} \|X(0) - \tilde{X}(0; \theta^*)\|_{\mathbb{R}^3}.\end{aligned}$$

Combining the above estimates yields

$$\|\theta - \theta^*\|_{\mathbb{R}^3} \leq \frac{2}{\alpha} C(\theta)^{1/2} + \frac{2}{\alpha} e^{-\gamma\tau} \left(\frac{1 - e^{-2\gamma T}}{2\gamma T} \right)^{1/2} \|X(0) - \tilde{X}(0; \theta^*)\|_{\mathbb{R}^3}.$$

This proves the result. $\square$

Proposition 5 shows that the cost functional controls the parameter error, provided the derivative $D\mathcal{F}(\theta^*)$ is bounded below and this lower bound persists locally. The remaining issue is how this nondegeneracy condition can be checked for the Lorenz nudged system. Proposition 6 derives the partial derivatives of the nudged trajectory with respect to σ , ρ , and β . It also shows that the required lower bound is equivalent to a positive-definiteness condition on a Gram matrix built from the observed x -components of these partial derivatives.

Proposition 6. *Let $\theta^* = (\sigma^*, \rho^*, \beta^*)$ be the true parameter vector, and let $\tilde{\theta} = (\tilde{\sigma}, \tilde{\rho}, \tilde{\beta})$ denote the parameter vector used in the nudged Lorenz system (12). Fix $\tau \geq 0$ and $T > 0$, and set $I_{\tau, T} := [0, \tau + T]$. Assume that the observed trajectory $x(t)$ is continuous on $I_{\tau, T}$, and that the initial condition for the nudged system is independent of $\tilde{\theta}$. Let U be a compact neighborhood of θ^* such that $\tilde{\sigma} > 0, \tilde{\beta} > 0, \tilde{\theta} \in U$. Assume further that the nudging coefficient μ satisfies the condition (13) for all $\tilde{\theta} \in U$. Let $\tilde{X}^*(t) = (\tilde{x}^*(t), \tilde{y}^*(t), \tilde{z}^*(t)) := \tilde{X}(t; \theta^*), \theta^* = (\sigma^*, \rho^*, \beta^*)$. Define*

$$(p_{\sigma}, q_{\sigma}, r_{\sigma}) := \partial_{\tilde{\sigma}}(\tilde{x}, \tilde{y}, \tilde{z})|_{\tilde{\theta}=\theta^*},$$

$$(p_{\rho}, q_{\rho}, r_{\rho}) := \partial_{\tilde{\rho}}(\tilde{x}, \tilde{y}, \tilde{z})|_{\tilde{\theta}=\theta^*}, \quad (p_{\beta}, q_{\beta}, r_{\beta}) := \partial_{\tilde{\beta}}(\tilde{x}, \tilde{y}, \tilde{z})|_{\tilde{\theta}=\theta^*}.$$

(i) (Equations for the partial derivatives) The partial derivatives satisfy the following systems on $I_{\tau, T}$.

For $\tilde{\sigma}$,

$$\begin{cases} p'_{\sigma} = (\tilde{y}^* - \tilde{x}^*) - (\sigma^* + \mu)p_{\sigma} + \sigma^*q_{\sigma}, \\ q'_{\sigma} = (\rho^* - \tilde{z}^*)p_{\sigma} - \tilde{x}^*r_{\sigma} - q_{\sigma}, \\ r'_{\sigma} = \tilde{y}^*p_{\sigma} + \tilde{x}^*q_{\sigma} - \beta^*r_{\sigma}. \end{cases}$$

For $\tilde{\rho}$,

$$\begin{cases} p'_{\rho} = -(\sigma^* + \mu)p_{\rho} + \sigma^*q_{\rho}, \\ q'_{\rho} = \tilde{x}^* + (\rho^* - \tilde{z}^*)p_{\rho} - \tilde{x}^*r_{\rho} - q_{\rho}, \\ r'_{\rho} = \tilde{y}^*p_{\rho} + \tilde{x}^*q_{\rho} - \beta^*r_{\rho}. \end{cases}$$

For $\tilde{\beta}$,

$$\begin{cases} p'_{\beta} = -(\sigma^* + \mu)p_{\beta} + \sigma^*q_{\beta}, \\ q'_{\beta} = (\rho^* - \tilde{z}^*)p_{\beta} - \tilde{x}^*r_{\beta} - q_{\beta}, \\ r'_{\beta} = \tilde{y}^*p_{\beta} + \tilde{x}^*q_{\beta} - \beta^*r_{\beta} - \tilde{z}^*. \end{cases}$$

Moreover, since the initial condition of the nudged system is independent of $\tilde{\theta}$,

$$(p_j, q_j, r_j)(0) = (0, 0, 0), \quad j \in \{\sigma, \rho, \beta\}.$$

(ii) (Gram matrix lower bound) Define $\mathcal{F}(\tilde{\theta})(t) = x(t) - \tilde{x}(t; \tilde{\theta})$, $t \in [\tau, \tau + T]$. Let

$$\langle f, g \rangle_{\tau, T} := \int_{\tau}^{\tau+T} f(t)g(t) dt.$$

Define the Gram matrix

$$G_{\tau, T} = \begin{pmatrix} \langle p_{\sigma}, p_{\sigma} \rangle_{\tau, T} & \langle p_{\sigma}, p_{\rho} \rangle_{\tau, T} & \langle p_{\sigma}, p_{\beta} \rangle_{\tau, T} \\ \langle p_{\rho}, p_{\sigma} \rangle_{\tau, T} & \langle p_{\rho}, p_{\rho} \rangle_{\tau, T} & \langle p_{\rho}, p_{\beta} \rangle_{\tau, T} \\ \langle p_{\beta}, p_{\sigma} \rangle_{\tau, T} & \langle p_{\beta}, p_{\rho} \rangle_{\tau, T} & \langle p_{\beta}, p_{\beta} \rangle_{\tau, T} \end{pmatrix},$$

where $p_{\sigma}, p_{\rho}, p_{\beta}$ are the x -components of the partial derivatives from part (i). Assume that $G_{\tau, T}$ is positive definite. Then, for every $h = (h_{\sigma}, h_{\rho}, h_{\beta}) \in \mathbb{R}^3$,

$$\|D\mathcal{F}(\theta^*)h\|_{L^2(\tau, \tau+T)} \geq \sqrt{\lambda_{\min}(G_{\tau, T})} \|h\|_{\mathbb{R}^3}.$$

Proof. We defer the proof to Appendix A.1. □

We now combine the quantitative estimate in Proposition 5 with the computable nondegeneracy condition in Proposition 6. This gives the main post-transient result for the nudging-based cost functional used in the optimization procedure. The theorem states that the true parameter gives zero post-transient mismatch in the noise-free case, that multiplicative observational noise produces a controlled increase in the mismatch at the true parameter, and that, under the stated nondegeneracy assumptions, the post-transient cost functional controls the full parameter error. Denote

$$C_{\lim}(\theta) = \limsup_{\tau \rightarrow \infty} C(\theta).$$

We refer to $C_{\lim}(\theta)$ as the *post-transient cost functional*.

Theorem 7 (Uniqueness and stability of the post-transient cost-functional minimizer). *Let $X(t) = (x(t), y(t), z(t))$, $t \in [t_0, \infty)$ be a trajectory of (8) corresponding to the parameter $\theta^* = (\sigma^*, \rho^*, \beta^*)$. Let $\tilde{X}_{\theta}(t) = (x_{\theta}(t), y_{\theta}(t), z_{\theta}(t))$ be the solution of the (12) corresponding to the parameter $\theta = (\sigma, \rho, \beta)$.*

- (i) *We have the equality $C_{\lim}(\theta^*) = 0$.*
- (ii) *Let $\mathcal{O}(t) = x(t)(1 + \varepsilon(t))$, where $\varepsilon(t)$ is the observational error. Let $\tilde{X}_{\theta}$ be the solution of (12) where the true observation $x(t)$ is replaced by the contaminated observation $\mathcal{O}(t)$. Then, we have the estimate $C_{\lim}(\theta^*) \leq C_{\mathbb{R}} \|\varepsilon\|_{\infty}^2$.*
- (iii) *Assume that there exists $\alpha > 0$ such that*

$$\liminf_{\tau \rightarrow \infty} \|D\mathcal{F}(\theta^*)h\|_{L^2(\tau, \tau+T)} \geq \alpha \|h\|_{\mathbb{R}^3} \quad \text{for all } h \in \mathbb{R}^3,$$

and

$$\limsup_{\tau \rightarrow \infty} \|(D\mathcal{F}(\eta) - D\mathcal{F}(\theta^*))h\|_{L^2(\tau, \tau+T)} \leq \frac{\alpha}{2} \|h\|_{\mathbb{R}^3} \quad \text{for all } \eta \in U \text{ and all } h \in \mathbb{R}^3.$$

Then, for every $\theta \in U$, we have the estimate

$$\|\theta - \theta^*\|_{\mathbb{R}^3} \leq \frac{2}{\alpha} (C_{\lim}(\theta))^{1/2}.$$

Proof. The proof follows readily from Theorem 1 and Proposition 5. □

Remark 2 (Interpretation of the nondegeneracy assumptions in Theorem 7). The assumptions in part (iii) of Theorem 7 express a local nondegeneracy condition for the post-transient observation mismatch map

$$\mathcal{F}(\theta)(t) = x(t) - \tilde{x}(t; \theta), \quad t \in [\tau, \tau + T].$$

The first condition requires that the linearized map $D\mathcal{F}(\theta^*)$ be bounded below as $\tau \rightarrow \infty$. In other words, nonzero perturbations of the parameter vector must produce a nonzero first-order change in the observed component of the nudged trajectory.

Proposition 6 provides a concrete way to verify this condition. Indeed, the derivative $D\mathcal{F}(\theta^*)h$ is determined by the x -components of the partial derivative equations,

$$p_\sigma = \partial_{\tilde{\sigma}} \tilde{x}|_{\tilde{\theta}=\theta^*}, \quad p_\rho = \partial_{\tilde{\rho}} \tilde{x}|_{\tilde{\theta}=\theta^*}, \quad p_\beta = \partial_{\tilde{\beta}} \tilde{x}|_{\tilde{\theta}=\theta^*}.$$

Thus, the relevant information is encoded in the Gram matrix $G_{\tau,T}$. If $G_{\tau,T}$ is positive definite on the observation window, then Proposition 6 gives

$$\|D\mathcal{F}(\theta^*)h\|_{L^2(\tau,\tau+T)} \geq \sqrt{\lambda_{\min}(G_{\tau,T})} \|h\|_{\mathbb{R}^3}.$$

Therefore, the lower-bound constant in Theorem 7 may be chosen in terms of the smallest eigenvalue of $G_{\tau,T}$, provided this eigenvalue remains bounded away from zero along the post-transient windows.

The second condition in part (iii) of Theorem 7 requires this linearized nondegeneracy to persist uniformly for parameters $\eta \in U$ near θ^* . This prevents the parameter-to-observation map from becoming nearly singular in a neighborhood of the true parameter. Together, these two assumptions ensure that the post-transient cost functional controls the parameter error, leading to the estimate

$$\|\theta - \theta^*\|_{\mathbb{R}^3} \leq \frac{2}{\alpha} (C_{\lim}(\theta))^{1/2}.$$

5 Numerical Results

In this section, we present numerical results for the Lorenz system in both chaotic and non-chaotic regimes. To represent a chaotic regime, we choose the true parameter values

$$\sigma^* = 10, \quad \rho^* = 30, \quad \beta^* = \frac{8}{3} \approx 2.67,$$

whereas to represent the non-chaotic regime, we choose

$$\sigma^* = 10, \quad \rho^* = 14, \quad \beta^* = \frac{8}{3} \approx 2.67.$$

Unless otherwise stated, the initial condition for the reference Lorenz system is

$$(x_0, y_0, z_0) = (1, 1, 1),$$

and, in the nudging and parameter estimation experiments, the initial condition for the nudged system is

$$(\tilde{x}_0, \tilde{y}_0, \tilde{z}_0) = (1, 0, 0).$$

Noisy observations are generated using the noise model described in Eq. (2). For parameter estimation, the cost functional in Eq. (26) is minimized using the Nelder–Mead algorithm unless another optimization method is explicitly specified.

5.1 Parameter Identifiability and Sensitivity for the Lorenz System

This subsection examines structural identifiability, practical identifiability, and parameter sensitivity for the Lorenz system, following the framework discussed in Section 2. The analysis is performed in both the chaotic and non-chaotic regimes using either x - or y -observations. For the sensitivity analysis, the parameters are sampled within a 5% neighborhood of their true values using a Sobol sequence. The Lorenz system is then simulated for each sample, and the resulting output variance is used to estimate the sensitivity indices.

5.1.1 Chaotic Regime

We begin with the identifiability and sensitivity analysis in the chaotic regime. The results are organized into two parts: structural and practical identifiability, followed by Sobol sensitivity analysis.

Parameter Identifiability. Structural identifiability is assessed through the input–output equation obtained from $x(t)$ -observations. For the Lorenz system in Eq. (8), the variables $y(t)$ and $z(t)$ can be eliminated from the governing equations. This yields the following input–output differential equation involving only $x(t)$ and its derivatives:

$$x \ddot{x} - \dot{x} \ddot{x} + (\sigma + \beta + 1)x \ddot{x} - (\sigma + 1)\dot{x}^2 + x^3 \dot{x} + \sigma x^4 + \beta(\sigma + 1)x \dot{x} + \sigma\beta(1 - \rho)x^2 = 0. \quad (31)$$

Both DAISY and SIAN confirm that all parameters are globally structurally identifiable. *However, we note that $x(t) = y(t) = 0, z(t) = z_0 e^{-\beta^* t}, t \in \mathbb{R}$ is a global trajectory of the Lorenz system, and for this trajectory, none of the parameters is identifiable if one observes x , which validates our discussion in Theorem 2.* Theorem 2 identifies the exceptional set

$$\mathcal{S} = \{(x, y, z) : y = x \text{ or } x = 0 \text{ or } z = 0\}.$$

The trajectory above belongs to this set for all time. Thus, the identifiability conclusion should be interpreted away from such degenerate trajectories, while Theorem 2 identifies the exceptional dynamical situations that must be excluded in order to obtain a well-posed data-to-parameter inverse map.

Practical identifiability is assessed on the time interval $[0, 30]$. The results in Table 1 show that the ARE values for σ and β exceed the corresponding noise levels for both x -only and y -only observations. For ρ , the ARE values are also above the noise level at lower noise levels and remain around 6% across the experiments. According to the criterion described in Section 2.3, these results indicate that the full parameter vector (σ, ρ, β) is not practically identifiable from either x -only or y -only observations in this setting.

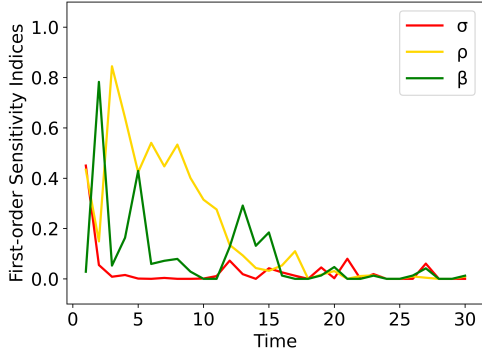
Noise Level (%)	$ARE_\sigma(\%)$		$ARE_\rho(\%)$		$ARE_\beta(\%)$	
	x - observ.	y - observ.	x - observ.	y - observ.	x - observ.	y - observ.
2	16.02	13.25	6.01	6.10	18.84	20.29
4	16.11	12.89	6.03	6.12	18.85	20.47
6	16.06	13.17	6.07	6.11	18.76	20.37
8	16.20	12.84	6.04	6.10	18.96	20.28
10	16.16	13.32	6.04	6.13	18.77	20.41

Table 1: Average Relative Percentage Error (ARE) for Lorenz system in chaotic regime using only x -observations or only y -observations.

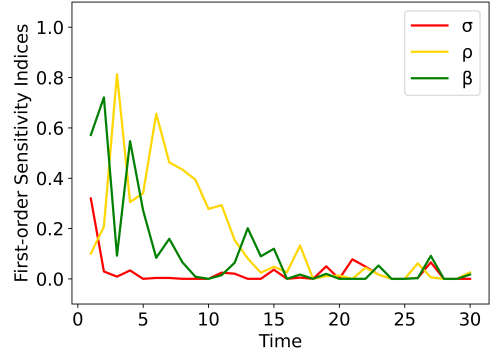
Parameter Sensitivity. The influence of each parameter is quantified using Sobol sensitivity indices. This analysis helps explain how the relative importance of σ , ρ , and β changes along the chaotic trajectory.

Figure 1 presents the Sobol sensitivity indices for the Lorenz system in the chaotic regime. Figures 1a and 1b show the first-order Sobol sensitivity indices with respect to the x -observation and the y -observation, respectively. For sensitivity with respect to the x -observation, ρ and β exhibit noticeable early-time effects, while the sensitivity with respect to σ remains small throughout the interval. After the initial transient phase, the first-order indices for ρ and β also decrease substantially and remain close to zero for most of the remaining time interval. A similar pattern is observed for sensitivity with respect to the y -observation. Thus, for both observation choices, the first-order effects are mainly transient, with ρ and β contributing during the early dynamics and σ showing weak individual influence over the full interval.

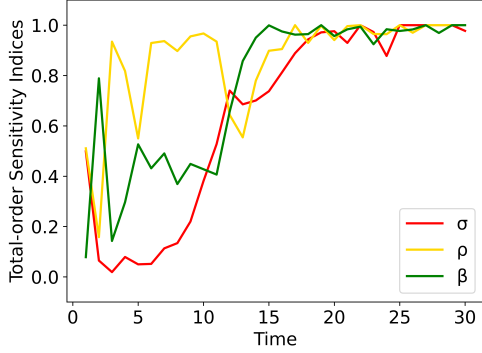
Figures 1c and 1d show the total-order Sobol sensitivity indices with respect to the x -observation and the y -observation, respectively. For sensitivity with respect to the x -observation, ρ and β have large total-order effects during the early and intermediate time intervals, while the total-order sensitivity with respect to σ increases after the transient phase. A similar behavior is observed for sensitivity with respect to the y -observation. At later times, the total-order indices for all three parameters are close to one, indicating that the parameters remain influential collectively even when their individual first-order effects are small.



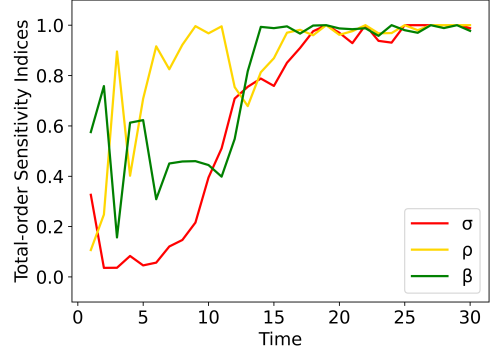
(a) First-order sensitivity indices – x observable.



(b) First-order sensitivity indices – y observable.



(c) Total-order sensitivity indices – x observable.



(d) Total-order sensitivity indices – y observable.

Figure 1: Plots of the first-order and total-order Sobol sensitivity indices in the chaotic regime using x - and y -observations. Figures (a) and (b) show the first-order indices for x - and y -observations, respectively, while Figures (c) and (d) show the corresponding total-order indices. The abscissa denotes time, and the ordinate represents the sensitivity index for each parameter.

5.1.2 Non-Chaotic Regime

The same identifiability and sensitivity analysis is repeated in the non-chaotic regime. In this case, the trajectory approaches an equilibrium, which leads to different practical identifiability and sensitivity behavior.

Parameter Identifiability. The structural identifiability result is the same as in the chaotic regime, since structural identifiability depends on the model equations and the observation scheme rather than on the particular parameter regime used in the numerical experiment. Practical identifiability is then evaluated over the time interval $[0, 30]$, using the same procedure as in the chaotic regime. The results in Table 2 show that all three parameters are practically identifiable from either x -observations or y -observations alone.

Noise Level (%)	$ARE_{\sigma}(\%)$		$ARE_{\rho}(\%)$		$ARE_{\beta}(\%)$	
	x - observ.	y - observ.	x - observ.	y - observ.	x - observ.	y - observ.
2	0.57	0.74	0.06	0.07	0.07	0.10
4	1.11	1.37	0.11	0.13	0.15	0.20
6	1.60	2.03	0.17	0.20	0.22	0.32
8	2.06	2.63	0.23	0.27	0.28	0.41
10	2.47	3.18	0.28	0.33	0.35	0.51

Table 2: Average Relative Percentage Error (ARE) for Lorenz system in non-chaotic regime using only x -observations or only y -observations.

Parameter Sensitivity. The first-order and total-order Sobol sensitivity indices are then computed in the non-chaotic regime. Figure 2 shows the corresponding indices using both x - and y -observations. Figures 2a and 2b present the

first-order indices for x - and y -observations, respectively. In both cases, the indices rapidly level off after a short transient. The parameters ρ and β exhibit persistently larger first-order effects, while the sensitivity index for σ remains close to zero throughout the time interval. This suggests that, in the non-chaotic regime, the direct contribution of σ to the observed dynamics is negligible compared with the effects of ρ and β .

Figures 2c and 2d show the corresponding total-order sensitivity indices for x - and y -observations, respectively. The total-order indices display the same overall behavior: ρ has the largest influence, followed by β , while σ remains small. In contrast to the chaotic regime, the sensitivity profiles in the non-chaotic regime are smoother and approach nearly constant values over time.

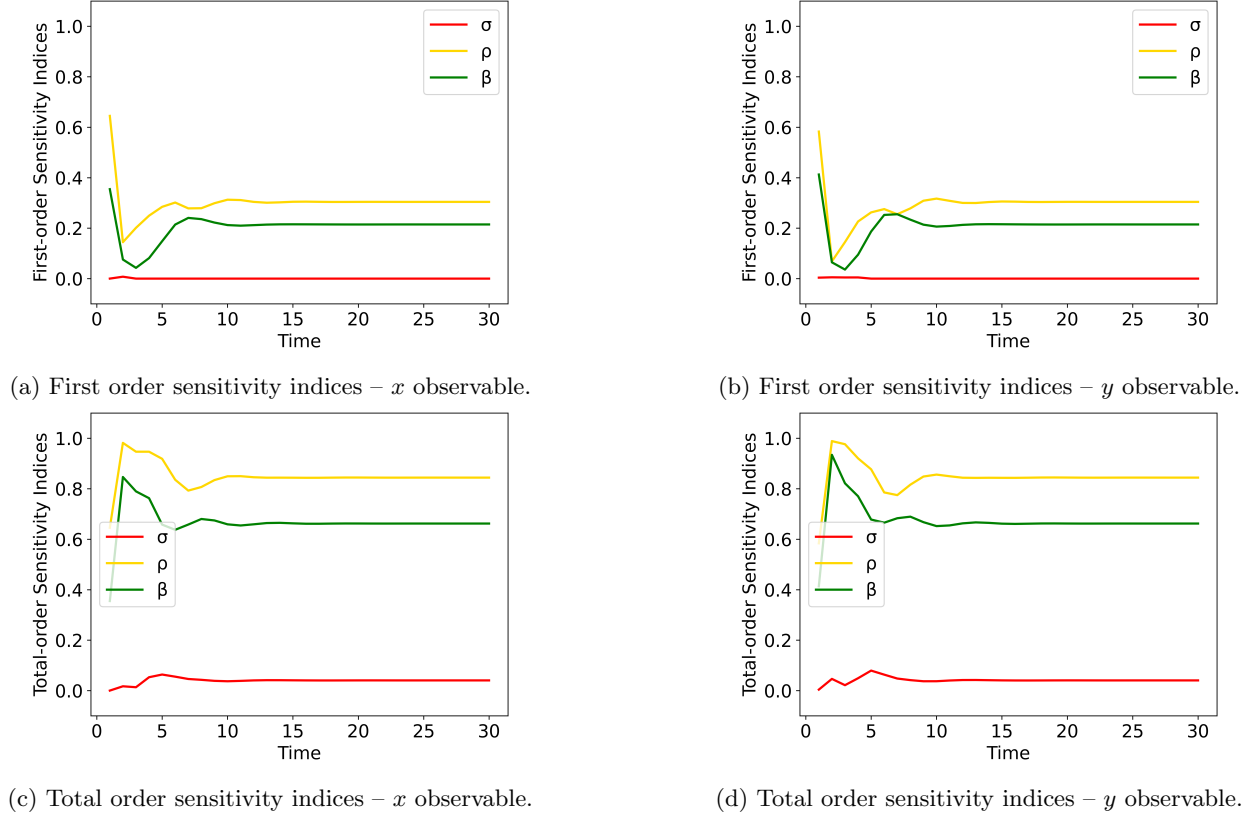


Figure 2: Plots of the first-order and total-order Sobol sensitivity indices in the non-chaotic regime using x - and y -observations. Panels (a) and (b) show the first-order indices for x - and y -observations, respectively, while panels (c) and (d) show the corresponding total-order indices. The abscissa denotes time, and the ordinate represents the sensitivity index for each parameter, scaled from 0 to 1.

Remark 3. The identifiability and sensitivity results provide a useful context for the parameter estimation problem studied in Section 5.3. In particular, we find that based on observing x or y , in the chaotic regime, all parameters are structurally identifiable while none of the parameters are practically identifiable. However, all parameters are highly sensitive in terms of total sensitivity in the chaotic regime. We show subsequently that incorporating data assimilation allows us to reliably recover the parameters in the chaotic regime.

On the other hand, all parameters in the non-chaotic regime are both structurally and practically identifiable. The sensitivity analysis shows that σ is not sensitive, while the other parameters are highly sensitive. In this case also, data assimilation allows us to reliably estimate the parameters.

5.2 Forward Prediction Using Nudging

This subsection presents forward prediction results for both chaotic and non-chaotic regimes, and include noise-free as well as noisy observations. We begin with x -observations and then compare the effect of observing different state variables.

For the chaotic regime, the nudged system is driven by noise-free x -observations with nudging coefficient $\mu = 100$. Figure 3 shows the logarithm (base 10) of the absolute relative errors between the reference and nudged state variables, $|x - \tilde{x}|$, $|y - \tilde{y}|$, $|z - \tilde{z}|$. The errors decay exponentially, showing that the nudged trajectory synchronizes with the reference

trajectory. This behavior is consistent with the estimate in Eq. (14). The chaotic-regime experiment is then repeated using noisy x -observations. Figure 4 shows that, across the noise levels considered, the nudging approach maintains an accuracy of approximately 10^{-2} in forward prediction.

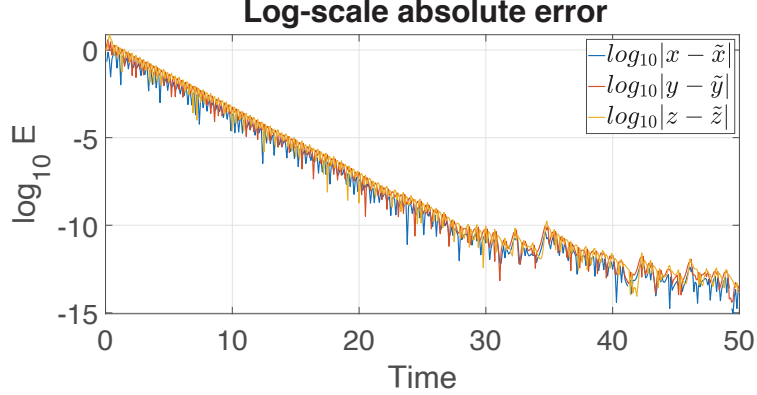


Figure 3: Logarithm (base 10) of the absolute relative errors between the state variables of the Lorenz system and the nudged system in the chaotic regime using noise-free x -observations.

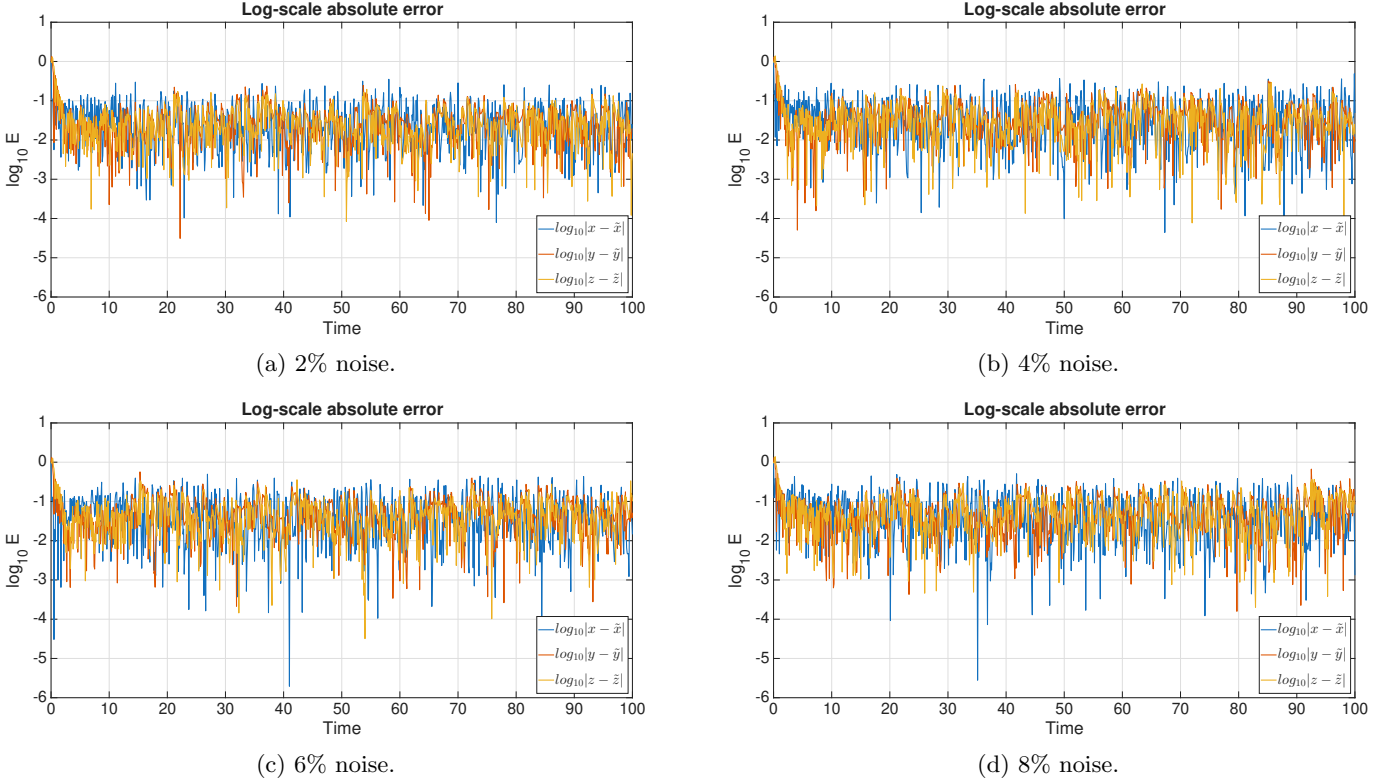


Figure 4: Logarithm (base 10) of the absolute relative errors between the state variables of the Lorenz system and the nudged system in the chaotic regime using noisy x -observations. Figures (a)–(d) correspond to different noise levels.

The non-chaotic regime is considered next. Figure 5 shows the forward prediction errors for noise-free x -observations. As in the chaotic regime, the errors decay exponentially for all state variables, even though only the x -component is observed. Notably, in the non-chaotic regime, the convergence occurs approximately 1.5 times faster than in the chaotic regime. This behavior is expected, as the non-chaotic regime is characterized by stable dynamics, which reduces sensitivity to initial conditions and facilitates faster synchronization compared to the chaotic regime.

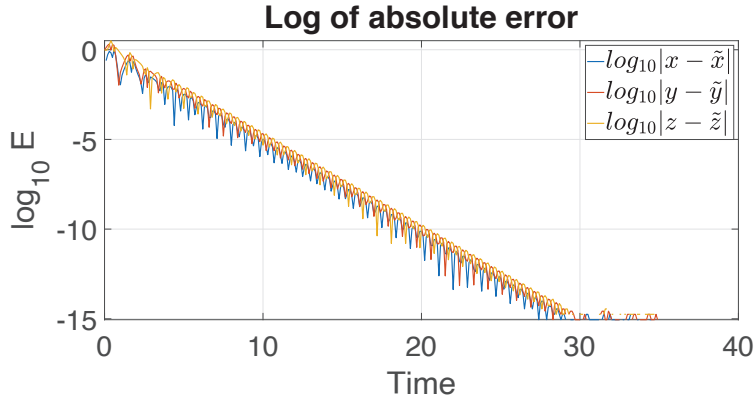


Figure 5: Logarithm (base 10) of the absolute relative errors between the state variables of the Lorenz system and the nudged system in the non-chaotic regime using noise-free x -observations.

We next consider the case of noisy observations in the non-chaotic regime. Figure 6 shows that, across different noise levels, the proposed approach maintains an accuracy of approximately 10^{-2} in forward prediction.

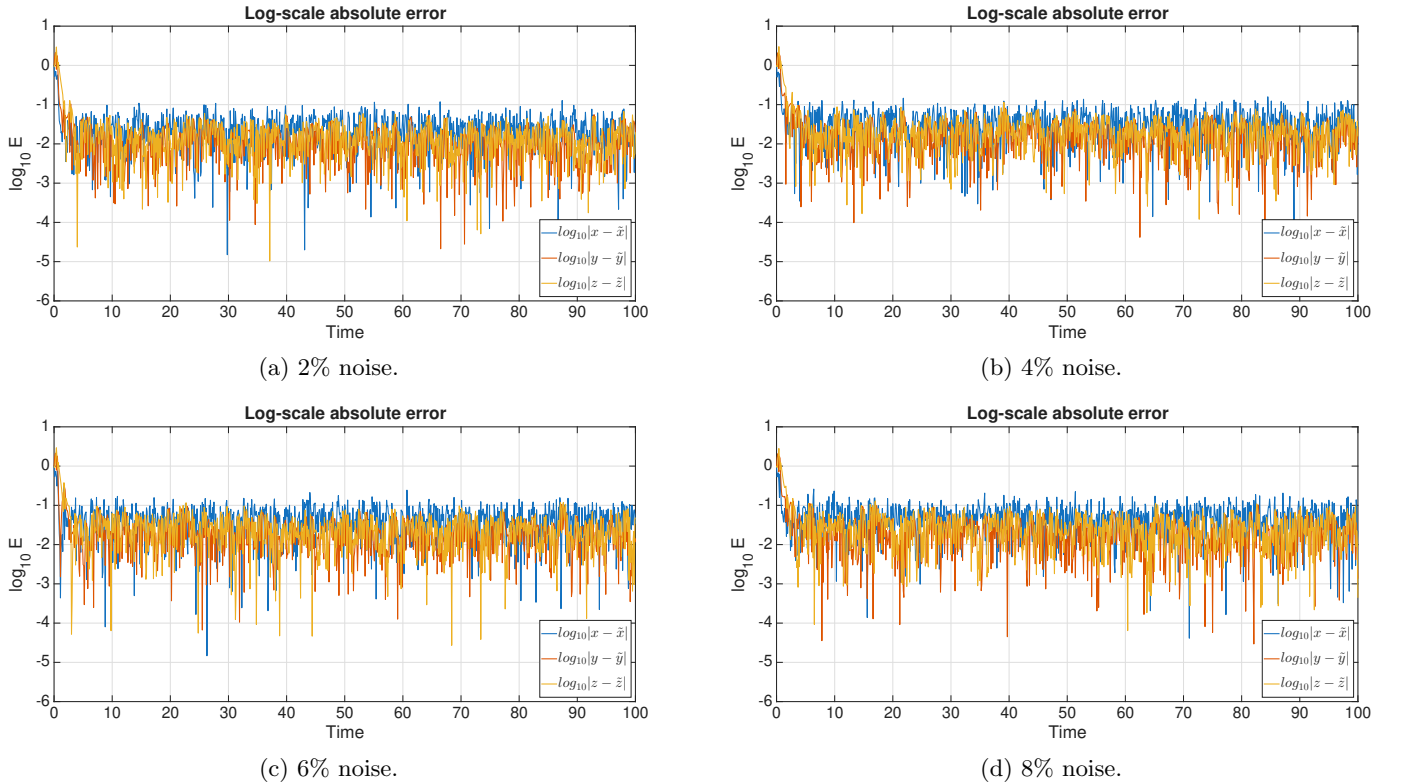


Figure 6: Logarithm (base 10) of the absolute relative errors between the state variables of the Lorenz system and the nudged system in the non-chaotic regime using noisy x -observations. Figures *a*, *b*, *c*, *d* illustrate the results for different noise levels.

Finally, we compare the effect of observing different state variables. For the Lorenz system, observing either the x - or y -component is sufficient to synchronize the nudged system with the reference trajectory and recover the full state. In contrast, observing only the z -component fails to produce synchronization. This behavior is illustrated in Figure 7.

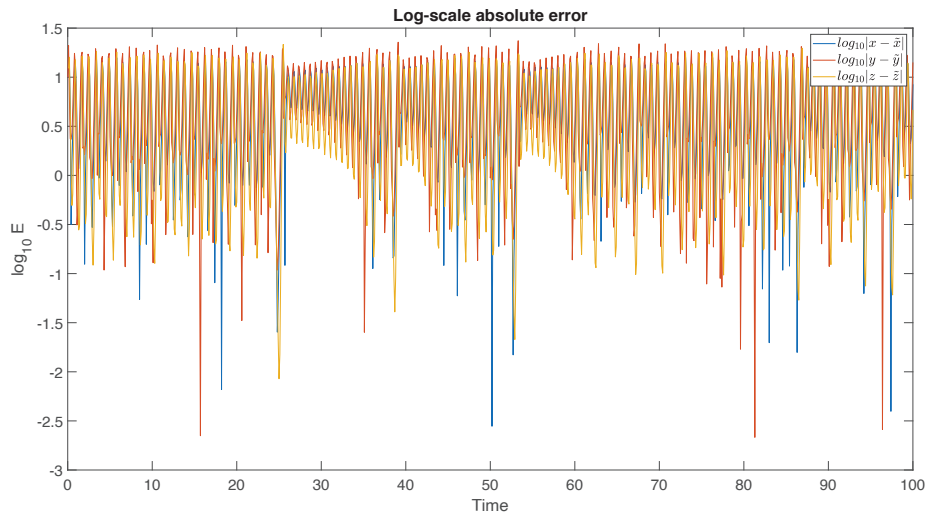


Figure 7: Failure of convergence when only the z -component is observed. The nudged system does not synchronize with the reference Lorenz trajectory. The abscissa represents time, and the ordinate represents the logarithm of the absolute error between the true and nudged states.

5.3 Parameter Estimation

Algorithm 1 is now used to estimate the Lorenz parameters from partial observations. Results are reported for both chaotic and non-chaotic regimes and for different levels of observational noise. Nelder–Mead is also compared with Particle Swarm Optimization and Simulated Annealing, and the effect of different initial parameter guesses is examined.

5.3.1 Chaotic Regime

Parameter estimation in the chaotic regime is considered first. This case is especially of interest because the practical identifiability results from Section 5.1.1 indicate that the parameters are difficult to recover directly from noisy observations.

The initial parameter guesses are set to $\tilde{\sigma} = 12$, $\tilde{\rho} = 35$, and $\tilde{\beta} = 5$. Table 3 reports the relative percentage errors in the estimated parameters for different noise levels. Across all cases, the estimation error remains smaller than the magnitude of the noise introduced in the observations, indicating that the nudging-based estimation procedure is robust in the chaotic regime.

Noise Level (%)	$RE_{\sigma}(\%)$		$RE_{\rho}(\%)$		$RE_{\beta}(\%)$	
	x - observ.	y - observ.	x - observ.	y - observ.	x - observ.	y - observ.
0	0.11	0.09	0.04	0.15	0.20	0.17
2	0.38	0.94	0.26	0.41	0.56	0.22
4	0.29	1.28	0.31	0.88	0.33	1.80
6	0.65	0.68	0.53	1.50	0.41	1.74
8	0.97	0.19	0.82	0.38	0.71	0.92
10	1.83	1.22	0.93	4.40	0.08	5.69

Table 3: Relative percentage errors in parameter estimation across different noise levels in the chaotic regime using x - and y -observations.

The Nelder–Mead method is also compared with Particle Swarm Optimization (PSO) and simulated annealing (SA). These methods are used to minimize the same nudging-based cost functional in Eq. (26). For background on SA and PSO, we refer to [71] and [72], respectively. Table 4 presents the relative percentage errors obtained using the three optimization algorithms. The estimates are similar across the methods. Among the algorithms considered, Nelder–Mead was the fastest in terms of computational time, while simulated annealing was the slowest.

Algorithm	$RE_\sigma(\%)$	$RE_\rho(\%)$	$RE_\beta(\%)$
Particle Swarm	0.0477	0.0447	0.0541
Nelder–Mead	0.1096	0.0435	0.0784
Simulated Annealing	0.0965	0.0311	0.0353

Table 4: Relative percentage errors for parameter estimation in the chaotic regime of the Lorenz system using different optimization algorithms.

The effect of the initial parameter guesses is examined by selecting ten initial guesses from the boundary of a sphere centered at the true parameter vector $(10, 30, 8/3)$. This experiment is repeated for several values of the radius r . For each initial guess, the cost functional in Eq. (26) is minimized, and the average relative percentage errors are reported in Table 5. The results show that the errors remain small across the different radii considered. In all cases, the average relative errors remain below 0.53%, indicating that the method is not highly sensitive to the initial parameter guess in this experiment.

Radius	$RE_\sigma(\%)$	$RE_\rho(\%)$	$RE_\beta(\%)$
0.5	0.5274	0.1367	0.3973
3	0.4620	0.0969	0.2967
5	0.4524	0.1100	0.3203
10	0.4515	0.1078	0.3172

Table 5: Average relative percentage errors for different initial-guess radii in the chaotic regime.

5.3.2 Non-Chaotic Regime

Parameter estimation in the non-chaotic regime is considered next. The initial parameter guesses are set to $\tilde{\sigma} = 12$, $\tilde{\rho} = 16$, and $\tilde{\beta} = 5$. The relative percentage errors for different noise levels are reported in Table 6. The error in the estimate of σ^* is noticeably larger than the errors for ρ^* and β^* , even at low noise levels. This behavior is consistent with the weak sensitivity of the observed dynamics to σ in the non-chaotic regime. Near equilibrium, where $x \approx y$, the equation $\dot{x} = \sigma(y - x)$ provides limited information about σ . Thus, accurate recovery of σ is less reliable in this regime. Nevertheless, accurate estimation of σ may not be essential for accurate prediction, as it is not a sensitive parameter in the non-chaotic regime.

Noise Level (%)	$RE_\sigma(\%)$		$RE_\rho(\%)$		$RE_\beta(\%)$	
	x - observ.	y - observ.	x - observ.	y - observ.	x - observ.	y - observ.
0	1.07	3.55	0.61	1.28	0.76	1.41
2	3.93	16.3	0.14	0.10	0.64	0.06
4	12.6	16.1	0.73	2.29	1.52	3.88
6	4.85	8.70	1.23	2.38	1.41	3.59
8	15.5	18.28	1.44	1.92	2.32	3.20
10	19.4	21.36	3.07	4.21	3.71	5.24

Table 6: Relative percentage errors in parameter estimation across different noise levels in the non-chaotic regime using x - and y -observations.

The three optimization algorithms are also compared in the non-chaotic regime, and the effect of randomly chosen initial parameter guesses is examined. The results are presented in Tables 7 and 8. From Table 7, PSO gives the most accurate estimates among the three methods, while simulated annealing gives the least accurate estimates in this case.

Algorithm	$RE_\sigma(\%)$	$RE_\rho(\%)$	$RE_\beta(\%)$
Particle Swarm	0.8817	0.4367	0.4560
Nelder–Mead	1.0703	0.6077	0.6407
Simulated Annealing	1.5522	1.8644	1.8677

Table 7: Relative percentage errors for parameter estimation in the non-chaotic regime of the Lorenz system using different optimization algorithms.

Radius	RE_σ (%)	RE_ρ (%)	RE_β (%)
0.5	0.8923	0.4314	0.4503
3	0.8786	0.4260	0.4435
5	0.8874	0.4301	0.4486
10	0.8835	0.4318	0.4516

Table 8: Average relative percentage errors for different initial-guess radii in the non-chaotic regime.

6 Discussion

We have presented a data-assimilation-based optimization framework for parameter estimation in nonlinear dynamical systems, using the Lorenz-63 system as a test case. The theoretical analysis establishes the main mechanisms that support the proposed method. Theorem 1 shows that, when the parameters in the nudged system match the true parameters, the nudged solution synchronizes with the true trajectory from arbitrary initial conditions. The same result also shows that, under parameter mismatch, the nudged solution remains close to the true trajectory, with an asymptotic error controlled by the size of the parameter mismatch. This justifies the use of the nudged-system cost functional for parameter estimation and explains why the method can reduce sensitivity to inaccurate or unknown initial conditions, which is especially important in chaotic regimes.

The theoretical results also clarify why partial observations can be sufficient for parameter recovery. In the Lorenz-63 system, the x -component has a determining property, so the full state can be recovered through nudging even when the full state is not observed. The well-posedness results for full and partial observations identify conditions under which the data-to-parameter inverse map is well-conditioned, while also showing that parameter recovery can fail or become ill-conditioned on degeneracy sets, such as equilibrium or otherwise non-informative trajectories. Proposition 5 and Theorem 7 further connect the cost functional to parameter recovery by showing that, under suitable nondegeneracy assumptions, the cost functional controls the parameter error. Finally, Proposition 6 makes this nondegeneracy condition computable through the sensitivity equations and the associated Gram matrix $G_{\tau,T}$. Thus, the theoretical analysis links synchronization, partial observation, well-posedness, and local identifiability to the practical optimization procedure used in the numerical experiments.

The computational efficiency of the proposed framework provides another important advantage. All numerical simulations were performed on an Intel Core i7 2.80 GHz processor with 16GB of RAM, and each experiment required less than one minute of runtime, with slightly higher runtimes observed only in noisy chaotic cases. An important practical advantage of the proposed framework is that it does not require observations from every equation in which unknown parameters appear. This distinguishes it from existing on-the-fly parameter estimation approaches such as [44–46], which typically require full state observations to recover all unknown parameters. In many real-world applications, such complete observations may not be available. The proposed framework instead uses only the observed component to drive the nudged system and constructs the parameter-estimation cost functional from the corresponding observation mismatch.

To further examine this issue, we implemented the method proposed in [44] in both chaotic and non-chaotic regimes, using noise-free observations and multiplicative noisy observations of all state variables. The noisy data were generated according to the multiplicative noise model in Eq. (2). We use this noise model for consistency with the rest of the present work, noting that Carlson et al. [44] instead consider additive noise with a very small noise amplitude. Table 9 shows that the method performs very well in the chaotic noise-free case, with MREs below 1% for all three parameters. As the noise level increases, however, the parameter estimates become more sensitive to observational perturbations, especially for higher noise levels.

In comparison, this proposed framework uses only x -observations to estimate all three parameters and maintains small relative errors across the same noise levels while [44] necessitates observing all the components. In the chaotic regime, the relative errors remain below 2% for all parameters, while in the non-chaotic regime the errors remain substantially smaller in the noisier cases. These results suggest that, for the test cases considered here, the proposed nudging-augmented optimization framework remains accurate under measurement noise while requiring only partial observations.

Noise Level (%)	Chaotic			Non-chaotic		
	$\text{MRE}_\sigma(\%)$	$\text{MRE}_\rho(\%)$	$\text{MRE}_\beta(\%)$	$\text{MRE}_\sigma(\%)$	$\text{MRE}_\rho(\%)$	$\text{MRE}_\beta(\%)$
0	0.38	0.05	0.64	0.50	0.13	0.68
2	15.11	1.19	18.66	22.49	0.84	12.99
4	33.37	6.16	1.01×10^3	22.38	0.46	34.75
6	21.83	4.65	81.93	362.21	7.06	28.65
8	22.61	11.03	102.71	26.52	7.37	123.74
10	4.7×10^4	1.4×10^4	83	1.3×10^3	448.04	65.38

Table 9: Mean relative error percentages for the method proposed in [44] in the chaotic and non-chaotic Lorenz–63 regimes under different observational noise levels.

We further compare the proposed framework with Bayesian approaches, which are widely used for parameter estimation in dynamical systems [32–35]. To this end, we conducted a Bayesian MCMC-based parameter estimation experiment using x -observations for both chaotic and non-chaotic regimes under noise-free, 2%, and 8% noise settings. The resulting posterior density estimates are shown in Figures 8 and 9.

In the chaotic regime, Figure 8 shows that the posterior distributions can have a noticeable spread, and in some cases the densities exhibit more than one mode. This behavior indicates that the available x -observations may support multiple plausible parameter values under the Bayesian formulation, particularly as the observational noise level increases. A similar effect is observed in the non-chaotic regime in Figure 9, where several posterior densities remain broad and some display multimodal structure. These features reflect increased uncertainty in the estimated parameters and show that the posterior mean alone may not fully represent the structure of the parameter uncertainty.

Since Bayesian MCMC methods require repeated forward simulations in order to explore the posterior distribution, they can become computationally expensive for nonlinear dynamical systems. In contrast, the proposed nudging-based optimization framework directly computes point estimates by minimizing the observation mismatch. Using only x -observations, the proposed method produces accurate estimates across both regimes and noise levels, as shown in Tables 3 and 6. Thus, in the experiments considered here, the proposed framework provides accurate point estimates from partial observations, while the Bayesian posterior densities illustrate the uncertainty, spread, and possible multimodality associated with the corresponding MCMC-based inference problem.

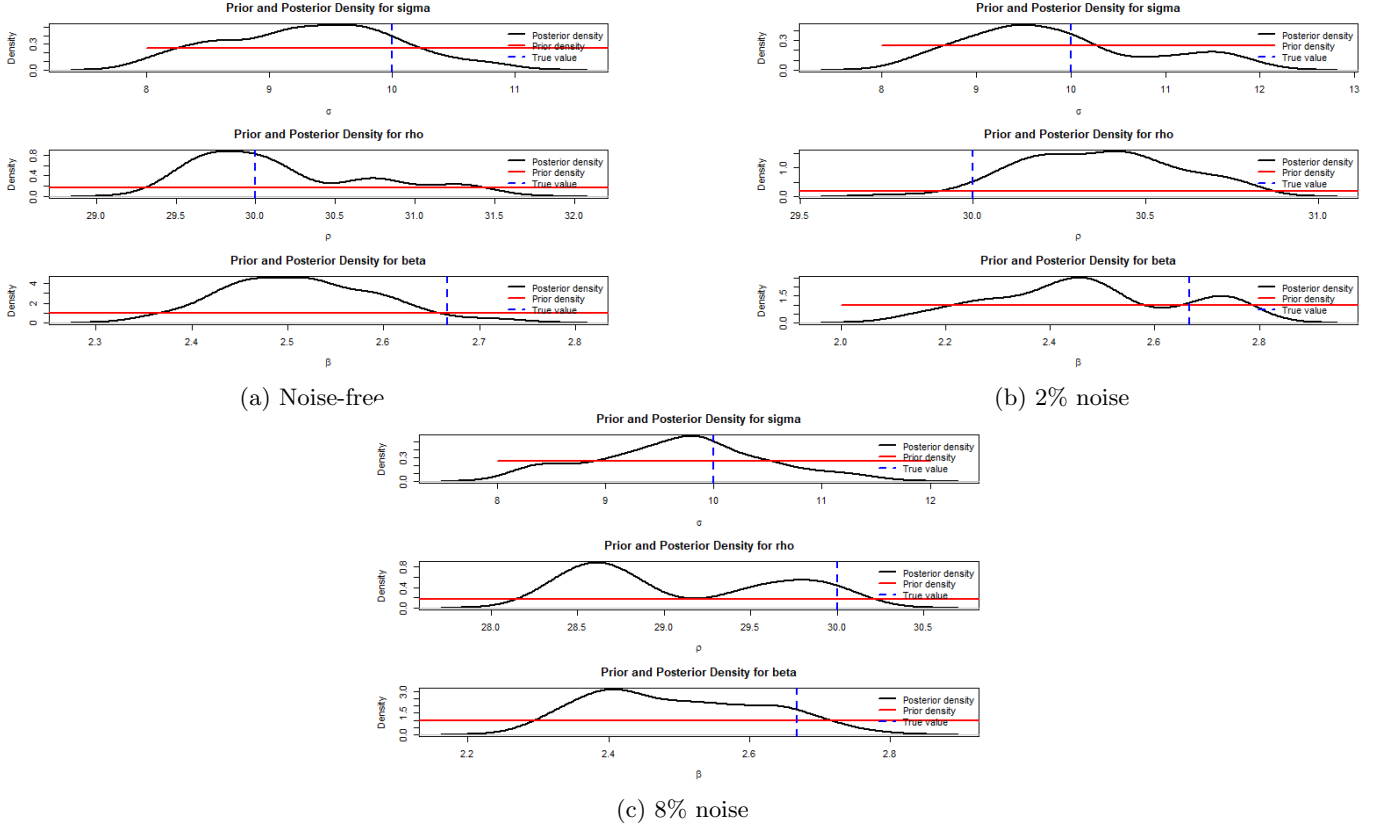


Figure 8: Posterior density estimates obtained using Bayesian MCMC parameter estimation in the chaotic regime with x observations under varying noise levels. The dashed blue lines represent the true parameter values.

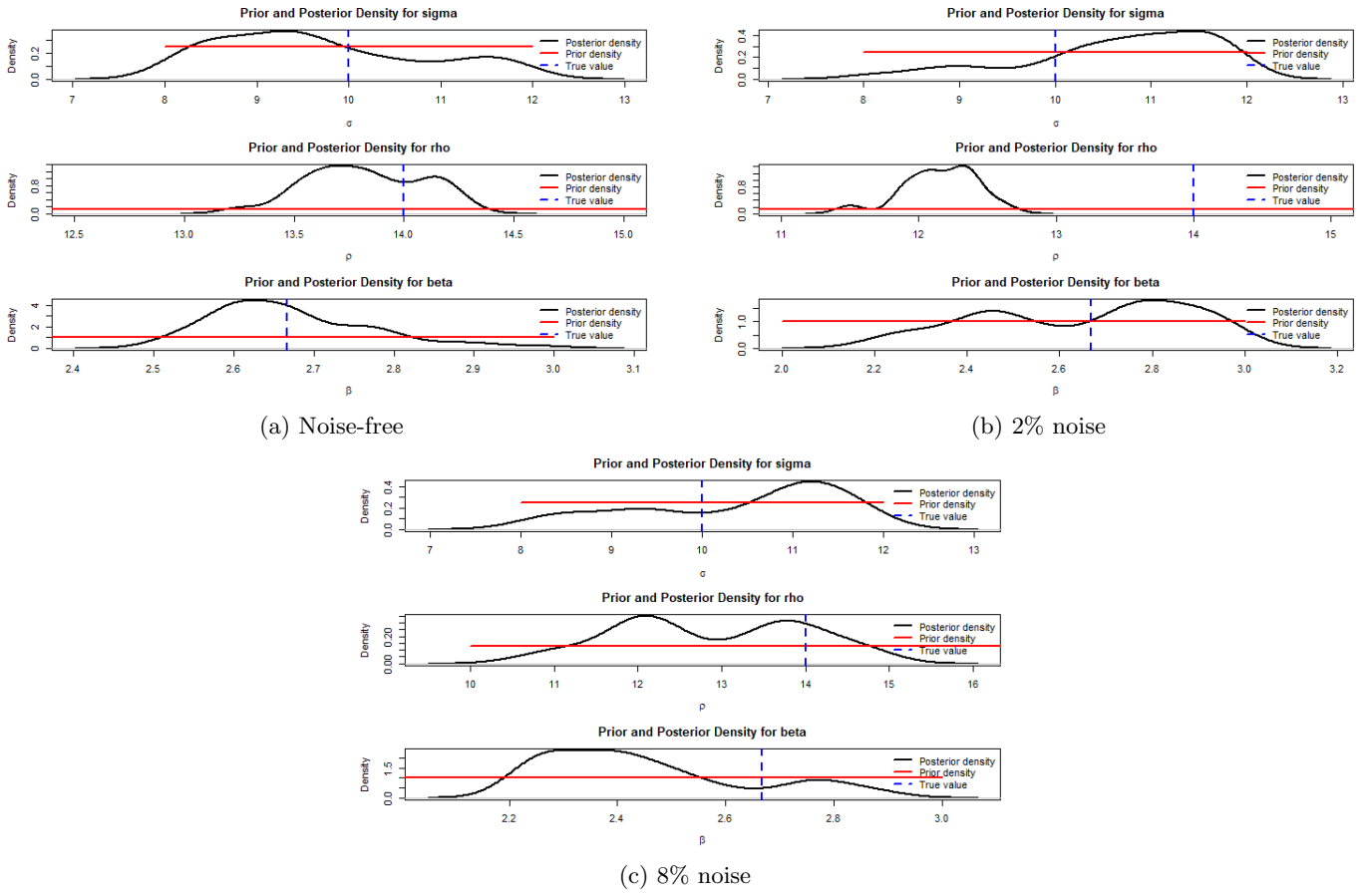


Figure 9: Posterior density estimates obtained using Bayesian MCMC parameter estimation in the non-chaotic regime with x observations under varying noise levels. The dashed blue lines represent the true parameter values.

There are several promising directions for future research. One important direction involves models with time-varying parameters, which arise frequently in practical applications and would require additional methodological development. Finally, although the present Lorenz-63 experiments were computationally efficient, the proposed optimization procedure requires repeated evaluations of the cost functional, and each such evaluation involves solving the nudged dynamical system. For larger or higher-dimensional systems, this repeated ODE-solving step may increase the computational cost. One possible way to reduce this cost is to replace repeated forward solves by learned surrogate models, such as physics-informed neural networks, neural ODEs, or related scientific machine-learning approaches [73–75]. In addition to this repeated-solve cost, larger parameter spaces may introduce further computational challenges for optimization, especially when a broad search over admissible parameter values is required. One may explore quantum optimization and hybrid quantum-classical optimization methods to address this issue, since quantum optimization algorithms have shown promise for optimization problems [76, 77]. Initial work in this direction is being pursued in our manuscript, *A Data-Assimilation-Informed Quantum Optimization Framework for Parameter Estimation in Dynamical Systems* [78].

Credit authorship contribution statement

Muhammad Jalil Ahmad: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis; **Animikh Biswas:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Funding acquisition, Investigation, Formal analysis, Conceptualization; **Kathleen Hoffman:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Conceptualization.

Acknowledgments

The authors gratefully acknowledge support from the UMBC Strategic Awards for Research Transitions (START) Program. A. Biswas was also partially supported by NSF grant DMS-2529382. The authors also thank Dr. Katherine Gurski and Boris Alemi for their helpful assistance with the computational and numerical aspects of this work.

A Technical Proofs

Lemma 1 (Non-constancy of the z -component away from equilibria). *Let $(x(t), y(t), z(t))$ be the solution of the Lorenz system*

$$\begin{cases} \dot{x} = \sigma^*(y - x), \\ \dot{y} = x(\rho^* - z) - y, \\ \dot{z} = xy - \beta^*z, \end{cases}$$

with true parameters $\sigma^ > 0$, $\beta^* > 0$, and $\rho^* \neq 0$. If the trajectory is not at an equilibrium, then $z(t)$ cannot be constant on any open time interval.*

Proof. Assume, for contradiction, that $z(t) \equiv \bar{z}$ is constant on some open interval I . From the third equation, $\dot{z} = 0$ implies

$$xy - \beta^*\bar{z} = 0, \quad \text{i.e.,} \quad xy \equiv c := \beta^*\bar{z}.$$

Case 1: $c \neq 0$. Since x and y never vanish on I , we may write $y = c/x$. From the first Lorenz equation:

$$\dot{x} = \sigma^*\left(\frac{c}{x} - x\right) = \frac{\sigma^*(c - x^2)}{x}.$$

Differentiating $y = c/x$ gives

$$\dot{y} = -\frac{c}{x^2}\dot{x} = -\frac{\sigma^*c(c - x^2)}{x^3}. \quad (32)$$

From the second Lorenz equation:

$$\dot{y} = x(\rho^* - \bar{z}) - \frac{c}{x}. \quad (33)$$

Equating Eq. (32) and (33) gives:

$$-\frac{\sigma^*c(c - x^2)}{x^3} = x(\rho^* - \bar{z}) - \frac{c}{x}. \quad (34)$$

Multiplying Eq. (34) by x^3 yields:

$$-\sigma^*c(c - x^2) = (\rho^* - \bar{z})x^4 - cx^2.$$

Expanding and rearranging gives:

$$0 = (\rho^* - \bar{z})x^4 - c(1 + \sigma^*)x^2 + \sigma^*c^2.$$

Let $u = x^2 > 0$. Then

$$(\rho^* - \bar{z})u^2 - c(1 + \sigma^*)u + \sigma^*c^2 = 0.$$

This is a constant-coefficient quadratic equation in $u(t)$ that holds for all $t \in I$. Hence $u(t)$ is constant, so $x(t)$ is constant (up to sign), and $y(t) = c/x(t)$ is constant as well. With x, y, z constant, the first Lorenz equation gives $y = x$. The second Lorenz equation then yields $0 = x(\rho^* - \bar{z}) - x$, so $\bar{z} = \rho^* - 1$. From $xy = c = \beta^*\bar{z}$ and $y = x$, we get $x^2 = \beta^*(\rho^* - 1)$. Thus,

$$(x, y, z) = \left(\pm\sqrt{\beta^*(\rho^* - 1)}, \pm\sqrt{\beta^*(\rho^* - 1)}, \rho^* - 1 \right),$$

which are exactly the nontrivial Lorenz equilibria, contradicting the assumption.

Case 2: $c = 0$. Then $\beta^*\bar{z} = 0$, so $\bar{z} = 0$ and $xy \equiv 0$. Since x and y are real analytic functions of time, the identity $xy \equiv 0$ on I implies that either $x \equiv 0$ or $y \equiv 0$ on I . If $x \equiv 0$, the first equation gives $y \equiv 0$, yielding the origin (equilibrium). If $y \equiv 0$, the second equation gives $x\rho^* = 0$, hence $x \equiv 0$, again the origin. Thus the solution is at equilibrium. This contradicts the assumption. $\square$

Lemma 2 (Well-posedness and C^1 dependence of the nudged system on $\tilde{\theta}$). *Let $\tilde{\theta} = (\tilde{\sigma}, \tilde{\rho}, \tilde{\beta})$ and let $\tilde{X}(t; \tilde{\theta}) = (\tilde{x}(t; \tilde{\theta}), \tilde{y}(t; \tilde{\theta}), \tilde{z}(t; \tilde{\theta}))$ denote the solution of the nudged system (12). Fix $\tau \geq 0$ and $T > 0$, and set $I_{\tau, T} := [0, \tau + T]$. Assume that the observed trajectory $x(t)$ is continuous on $I_{\tau, T}$, and that the initial condition for the nudged system is independent of $\tilde{\theta}$. Let U be a compact neighborhood of $\theta^* = (\sigma^*, \rho^*, \beta^*)$ such that $\tilde{\sigma} > 0, \tilde{\beta} > 0, \tilde{\theta} \in U$. Assume also that μ satisfies the condition (13) for all $\tilde{\theta} \in U$. Then:*

- (i) for each $\tilde{\theta} \in U$, the nudged system (12) admits a unique solution on $I_{\tau,T}$;
- (ii) the solution depends continuously on the initial data and on $\tilde{\theta}$;
- (iii) the solution map $\tilde{\theta} \mapsto \tilde{X}(\cdot; \tilde{\theta})$ is C^1 from U into $C(I_{\tau,T}; \mathbb{R}^3)$. In particular, the partial derivatives

$$(p_\sigma, q_\sigma, r_\sigma) := \partial_{\tilde{\sigma}}(\tilde{x}, \tilde{y}, \tilde{z}),$$

$$(p_\rho, q_\rho, r_\rho) := \partial_{\tilde{\rho}}(\tilde{x}, \tilde{y}, \tilde{z}), \quad (p_\beta, q_\beta, r_\beta) := \partial_{\tilde{\beta}}(\tilde{x}, \tilde{y}, \tilde{z})$$

exist and are continuous on $I_{\tau,T}$.

Proof. Set $\tilde{X} = (\tilde{x}, \tilde{y}, \tilde{z})^\top$. The nudged system (12) can be written as the non-autonomous ordinary differential equation $\dot{\tilde{X}} = f(t, \tilde{X}, \tilde{\theta})$, where

$$f(t, \tilde{X}, \tilde{\theta}) = \begin{pmatrix} \tilde{\sigma}(\tilde{y} - \tilde{x}) + \mu(x(t) - \tilde{x}) \\ \tilde{x}(\tilde{\rho} - \tilde{z}) - \tilde{y} \\ \tilde{x}\tilde{y} - \tilde{\beta}\tilde{z} \end{pmatrix}.$$

Since $x(t)$ is continuous on $I_{\tau,T}$, the map $t \mapsto f(t, \tilde{X}, \tilde{\theta})$ is continuous. Moreover, f is polynomial in $\tilde{X}$ and smooth in $\tilde{\theta} = (\tilde{\sigma}, \tilde{\rho}, \tilde{\beta})$. Hence, for each fixed $\tilde{\theta} \in U$, f is locally Lipschitz in $\tilde{X}$. By the Picard–Lindelöf theorem, the system has a unique maximal solution on some interval $[0, T_{\max})$.

It remains to show that $T_{\max} \geq \tau + T$. Since $x(t)$ is continuous on the compact interval $I_{\tau,T}$, there exists $K > 0$ such that $|x(t)| \leq K, t \in I_{\tau,T}$. Define $V(t) = \frac{1}{2}(\tilde{x}^2 + \tilde{y}^2 + \tilde{z}^2)$. Differentiating V along solutions of (12) gives

$$\begin{aligned} \dot{V} &= \tilde{x}\dot{\tilde{x}} + \tilde{y}\dot{\tilde{y}} + \tilde{z}\dot{\tilde{z}} \\ &= -(\tilde{\sigma} + \mu)\tilde{x}^2 - \tilde{y}^2 - \tilde{\beta}\tilde{z}^2 + (\tilde{\sigma} + \tilde{\rho})\tilde{x}\tilde{y} + \mu x(t)\tilde{x}. \end{aligned}$$

By Young's inequality,

$$|(\tilde{\sigma} + \tilde{\rho})\tilde{x}\tilde{y}| \leq (|\tilde{\sigma} + \tilde{\rho}| + K)^2 \tilde{x}^2 + \frac{1}{4}\tilde{y}^2,$$

and

$$\mu|x(t)\tilde{x}| \leq \frac{K^2}{\tilde{\beta}}\tilde{x}^2 + \frac{\mu^2\tilde{\beta}}{4}.$$

Therefore,

$$\dot{V} \leq -\left[\tilde{\sigma} + \mu - (|\tilde{\sigma} + \tilde{\rho}| + K)^2 - \frac{K^2}{\tilde{\beta}}\right]\tilde{x}^2 - \frac{3}{4}\tilde{y}^2 - \tilde{\beta}\tilde{z}^2 + \frac{\mu^2\tilde{\beta}}{4}.$$

By the assumption (13),

$$\mu > (|\tilde{\sigma} + \tilde{\rho}| + K)^2 + \frac{K^2}{\tilde{\beta}} + \frac{1}{4} - \tilde{\sigma},$$

and hence

$$\tilde{\sigma} + \mu - (|\tilde{\sigma} + \tilde{\rho}| + K)^2 - \frac{K^2}{\tilde{\beta}} > \frac{1}{4}.$$

Thus

$$\dot{V} \leq -\frac{1}{4}\tilde{x}^2 - \frac{3}{4}\tilde{y}^2 - \tilde{\beta}\tilde{z}^2 + \frac{\mu^2\tilde{\beta}}{4}.$$

Let

$$c := \min\left\{\frac{1}{4}, \frac{3}{4}, \tilde{\beta}\right\} > 0, \quad C := \frac{\mu^2\tilde{\beta}}{4}.$$

Then

$$\dot{V} \leq -c(\tilde{x}^2 + \tilde{y}^2 + \tilde{z}^2) + C = -2cV + C.$$

By Grönwall's inequality,

$$V(t) \leq e^{-2ct}V(0) + \frac{C}{2c}(1 - e^{-2ct}), \quad t \in I_{\tau,T}.$$

Hence $\tilde{X}(t)$ remains bounded on $I_{\tau,T}$. Therefore the maximal solution cannot blow up before time $\tau + T$, and so the solution exists uniquely on the entire interval $I_{\tau,T}$.

We now prove continuous dependence. Let $\tilde{X}_1$ and $\tilde{X}_2$ be two solutions corresponding to initial data $\tilde{X}_{1,0}, \tilde{X}_{2,0}$ and trial parameters $\tilde{\theta}_1, \tilde{\theta}_2 \in U$. The boundedness estimate above shows that both solutions remain in a bounded subset of $\mathbb{R}^3$ on $I_{\tau,T}$. On this bounded set, f is uniformly Lipschitz in $\tilde{X}$ and uniformly continuous in $\tilde{\theta}$. Therefore,

$$\begin{aligned} \|\tilde{X}_1(t) - \tilde{X}_2(t)\| &\leq \|\tilde{X}_{1,0} - \tilde{X}_{2,0}\| + L \int_0^t \|\tilde{X}_1(s) - \tilde{X}_2(s)\| ds \\ &\quad + \int_0^t \|f(s, \tilde{X}_2(s), \tilde{\theta}_1) - f(s, \tilde{X}_2(s), \tilde{\theta}_2)\| ds. \end{aligned}$$

The last integral tends to zero as $\tilde{\theta}_1 \rightarrow \tilde{\theta}_2$, uniformly for $t \in I_{\tau,T}$. Applying Grönwall's inequality gives

$$\sup_{t \in I_{\tau,T}} \|\tilde{X}_1(t) - \tilde{X}_2(t)\| \rightarrow 0, \quad \text{whenever} \quad \tilde{X}_{1,0} \rightarrow \tilde{X}_{2,0}, \quad \tilde{\theta}_1 \rightarrow \tilde{\theta}_2.$$

Thus the solution depends continuously on the initial data and on $\tilde{\theta}$.

Finally, since f is C^1 in $(\tilde{X}, \tilde{\theta})$, the standard C^1 -dependence theorem for ordinary differential equations, (see [79, Theorem 2]) implies that the solution map $\tilde{\theta} \mapsto \tilde{X}(\cdot; \tilde{\theta})$ is C^1 from U into $C(I_{\tau,T}; \mathbb{R}^3)$. Since the initial condition of the nudged system is independent of $\tilde{\theta}$, differentiating the initial condition gives zero initial data for each partial derivative. Therefore the partial derivatives with respect to $\tilde{\sigma}$, $\tilde{\rho}$, and $\tilde{\beta}$ exist and are continuous on $I_{\tau,T}$. $\square$

A.1 Proof of Proposition 6

Proof of Part (i): By Lemma 2, the solution of the nudged system (12) is C^1 with respect to $\tilde{\theta} = (\tilde{\sigma}, \tilde{\rho}, \tilde{\beta})$. Therefore the partial derivatives with respect to each component of $\tilde{\theta}$ exist and are obtained by differentiating (12). First, differentiating $\dot{\tilde{x}} = \tilde{\sigma}(\tilde{y} - \tilde{x}) + \mu(x - \tilde{x})$ with respect to $\tilde{\sigma}$ gives

$$p'_\sigma = (\tilde{y} - \tilde{x}) + \tilde{\sigma}(q_\sigma - p_\sigma) - \mu p_\sigma.$$

Evaluating at $\tilde{\theta} = \theta^*$ yields

$$p'_\sigma = (\tilde{y}^* - \tilde{x}^*) - (\sigma^* + \mu)p_\sigma + \sigma^* q_\sigma.$$

Next, differentiating $\dot{\tilde{y}} = \tilde{x}(\tilde{\rho} - \tilde{z}) - \tilde{y}$ with respect to $\tilde{\sigma}$ gives

$$q'_\sigma = p_\sigma(\tilde{\rho} - \tilde{z}) - \tilde{x}r_\sigma - q_\sigma,$$

and hence, at $\tilde{\theta} = \theta^*$,

$$q'_\sigma = (\rho^* - \tilde{z}^*)p_\sigma - \tilde{x}^*r_\sigma - q_\sigma.$$

Similarly, differentiating $\dot{\tilde{z}} = \tilde{x}\tilde{y} - \tilde{\beta}\tilde{z}$ with respect to $\tilde{\sigma}$ gives

$$r'_\sigma = \tilde{y}p_\sigma + \tilde{x}q_\sigma - \tilde{\beta}r_\sigma,$$

and therefore, at $\tilde{\theta} = \theta^*$

$$r'_\sigma = \tilde{y}^*p_\sigma + \tilde{x}^*q_\sigma - \beta^*r_\sigma.$$

The equations for (p_ρ, q_ρ, r_ρ) are obtained in the same way. Since the first equation of (12) has no explicit $\tilde{\rho}$ -dependence,

$$p'_\rho = \tilde{\sigma}(q_\rho - p_\rho) - \mu p_\rho.$$

Evaluating at $\tilde{\theta} = \theta^*$ gives

$$p'_\rho = -(\sigma^* + \mu)p_\rho + \sigma^* q_\rho.$$

The second equation contains the explicit factor $\tilde{\rho}$. Thus

$$q'_\rho = \tilde{x} + (\tilde{\rho} - \tilde{z})p_\rho - \tilde{x}r_\rho - q_\rho,$$

and hence, at $\tilde{\theta} = \theta^*$

$$q'_\rho = \tilde{x}^* + (\rho^* - \tilde{z}^*)p_\rho - \tilde{x}^*r_\rho - q_\rho.$$

The third equation gives

$$r'_\rho = \tilde{y}p_\rho + \tilde{x}q_\rho - \tilde{\beta}r_\rho,$$

and therefore, at $\tilde{\theta} = \theta^*$

$$r'_\rho = \tilde{y}^*p_\rho + \tilde{x}^*q_\rho - \beta^*r_\rho.$$

Finally, differentiating with respect to $\tilde{\beta}$, the first equation has no explicit $\tilde{\beta}$ -dependence, so

$$p'_\beta = \tilde{\sigma}(q_\beta - p_\beta) - \mu p_\beta.$$

At $\tilde{\theta} = \theta^*$, this becomes

$$p'_\beta = -(\sigma^* + \mu)p_\beta + \sigma^* q_\beta.$$

The second equation gives

$$q'_\beta = (\tilde{\rho} - \tilde{z})p_\beta - \tilde{x}r_\beta - q_\beta,$$

and hence, at $\tilde{\theta} = \theta^*$

$$q'_\beta = (\rho^* - \tilde{z}^*)p_\beta - \tilde{x}^*r_\beta - q_\beta.$$

For the third equation, the parameter $\tilde{\beta}$ appears explicitly:

$$\dot{\tilde{z}} = \tilde{x}\tilde{y} - \tilde{\beta}\tilde{z}.$$

Differentiating with respect to $\tilde{\beta}$ gives

$$r'_\beta = \tilde{y}p_\beta + \tilde{x}q_\beta - \tilde{\beta}r_\beta - \tilde{z}.$$

Evaluating at $\tilde{\theta} = \theta^*$ yields

$$r'_\beta = \tilde{y}^*p_\beta + \tilde{x}^*q_\beta - \beta^*r_\beta - \tilde{z}^*.$$

Since the initial condition of the nudged system is fixed independently of the trial parameter vector $\tilde{\theta}$, we have $\tilde{X}(0; \tilde{\theta}) = (\tilde{x}_0, \tilde{y}_0, \tilde{z}_0)$ for all $\tilde{\theta} \in U$. Hence the map $\tilde{\theta} \mapsto \tilde{X}(0; \tilde{\theta})$ is constant. Differentiating this identity with respect to $\tilde{\sigma}$, $\tilde{\rho}$, and $\tilde{\beta}$ gives

$$\partial_{\tilde{\sigma}}\tilde{X}(0; \tilde{\theta}) = 0, \quad \partial_{\tilde{\rho}}\tilde{X}(0; \tilde{\theta}) = 0, \quad \partial_{\tilde{\beta}}\tilde{X}(0; \tilde{\theta}) = 0.$$

Evaluating at $\tilde{\theta} = \theta^*$, this yields

$$(p_j, q_j, r_j)(0) = (0, 0, 0), \quad j \in \{\sigma, \rho, \beta\}.$$

This completes the proof.

Proof of Part (ii): By Lemma 2, the map $\tilde{\theta} \mapsto \tilde{x}(\cdot; \tilde{\theta})$ is C^1 from U into $C(I_{\tau, T})$. Restricting to $[\tau, \tau + T]$ and using the continuous embedding

$$C([\tau, \tau + T]) \hookrightarrow L^2(\tau, \tau + T),$$

it follows that $\mathcal{F} : U \rightarrow L^2(\tau, \tau + T)$ is C^1 . Since $x(t)$ is the fixed observed trajectory and does not depend on $\tilde{\theta}$, we have

$$D\mathcal{F}(\theta^*)h = -D_{\tilde{\theta}}\tilde{x}(t; \theta^*)h.$$

For $h = (h_\sigma, h_\rho, h_\beta)$, linearity of the derivative gives

$$D_{\tilde{\theta}}\tilde{x}(t; \theta^*)h = h_\sigma p_\sigma(t) + h_\rho p_\rho(t) + h_\beta p_\beta(t).$$

Therefore,

$$D\mathcal{F}(\theta^*)h = -(h_\sigma p_\sigma + h_\rho p_\rho + h_\beta p_\beta).$$

The sign does not affect the L^2 -norm, and so

$$\|D\mathcal{F}(\theta^*)h\|_{L^2(\tau, \tau+T)}^2 = \int_\tau^{\tau+T} |h_\sigma p_\sigma(t) + h_\rho p_\rho(t) + h_\beta p_\beta(t)|^2 dt.$$

Expanding the square gives

$$\begin{aligned} \|D\mathcal{F}(\theta^*)h\|_{L^2(\tau, \tau+T)}^2 &= h_\sigma^2 \langle p_\sigma, p_\sigma \rangle_{\tau, T} + h_\rho^2 \langle p_\rho, p_\rho \rangle_{\tau, T} + h_\beta^2 \langle p_\beta, p_\beta \rangle_{\tau, T} \\ &\quad + 2h_\sigma h_\rho \langle p_\sigma, p_\rho \rangle_{\tau, T} + 2h_\sigma h_\beta \langle p_\sigma, p_\beta \rangle_{\tau, T} + 2h_\rho h_\beta \langle p_\rho, p_\beta \rangle_{\tau, T}. \end{aligned}$$

By the definition of $G_{\tau, T}$, the right-hand side is exactly

$$h^\top G_{\tau, T} h.$$

Thus

$$\|D\mathcal{F}(\theta^*)h\|_{L^2(\tau, \tau+T)}^2 = h^\top G_{\tau, T} h.$$

Since $G_{\tau,T}$ is positive definite, its smallest eigenvalue is positive. By the Rayleigh quotient characterization,

$$h^\top G_{\tau,T} h \geq \lambda_{\min}(G_{\tau,T}) \|h\|_{\mathbb{R}^3}^2.$$

Therefore,

$$\|D\mathcal{F}(\theta^*)h\|_{L^2(\tau, \tau+T)}^2 \geq \lambda_{\min}(G_{\tau,T}) \|h\|_{\mathbb{R}^3}^2.$$

Taking square roots yields

$$\|D\mathcal{F}(\theta^*)h\|_{L^2(\tau, \tau+T)} \geq \sqrt{\lambda_{\min}(G_{\tau,T})} \|h\|_{\mathbb{R}^3}.$$

This completes the proof.

References

- [1] H. Miao, X. Xia, A. S. Perelson, and H. Wu. On identifiability of nonlinear ode models and applications in viral dynamics. *SIAM Review*, 53(1):3–39, 2011. doi:10.1137/090757009.
- [2] L. Ljung and T. Glad. On global identifiability for arbitrary model parametrizations. *Automatica*, 30(2):265–276, 1994.
- [3] E. Walter and L. Pronzato. *Identification of Parametric Models: From Experimental Data*. Communications and Control Engineering. Springer, 1997.
- [4] F. Anstett-Collin, L. Denis-Vidal, and G. Millérioux. A priori identifiability: An overview on definitions and approaches. *Annual Reviews in Control*, 50:139–149, 2020. doi:10.1016/j.arcontrol.2020.10.006.
- [5] N. Meshkat and A. Shiu. Structural identifiability of compartmental models: Recent progress and future directions. *Current Opinion in Systems Biology*, page 100559, 2025.
- [6] O. Saucedo, A. Laubmeier, T. Tang, B. Levy, L. Asik, T. Pollington, and O. P. Feldman. Comparative analysis of practical identifiability methods for an SEIR model. *AIMS Mathematics*, 9(9):24722–24761, 2024. doi:10.3934/math.20241204.
- [7] N. Tuncer and T. T. Le. Structural and practical identifiability analysis of outbreak models. *Math. Biosci.*, 299:1–18, 2018. doi:10.1016/j.mbs.2018.02.004.
- [8] M. Rodriguez-Fernandez, P. Mendes, and J. R. Banga. A hybrid approach for efficient and robust parameter estimation in biochemical pathways. *BioSystems*, 83(2–3):248–265, 2006. doi:10.1016/j.biosystems.2005.06.016.
- [9] H. T. Banks, S. Hu, and W. C. Thompson. *Modeling and Inverse Problems in the Presence of Uncertainty*. CRC Press, 2014. doi:10.1201/b16760.
- [10] D. Venzon and S. Moolgavkar. A method for computing profile-likelihood-based confidence intervals. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 37(1):87–94, 1988. doi:10.2307/2347496.
- [11] J. A. Jacquez and T. Perry. Parameter estimation: local identifiability of parameters. *American Journal of Physiology-Endocrinology and Metabolism*, 258(4):E727–E736, 1990. doi:10.1152/ajpendo.1990.258.4.E727.
- [12] M. C. Eisenberg and M. A. Hayashi. Determining identifiable parameter combinations using subset profiling. *Mathematical Biosciences*, 256:116–126, 2014. doi:10.1016/j.mbs.2014.08.008.
- [13] F. G. Wieland, A. L. Hauber, M. Rosenblatt, C. Tönsing, and J. Timmer. On structural and practical identifiability. *Current Opinion in Systems Biology*, 25:60–69, 2021.
- [14] A. F. Villaverde, A. Barreiro, and A. Papachristodoulou. Structural identifiability of dynamic systems biology models. *PLoS Computational Biology*, 12(10):e1005153, 2016. doi:10.1371/journal.pcbi.1005153.
- [15] S. Marino, I. B. Hogue, C. J. Ray, and D. E. Kirschner. A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of Theoretical Biology*, 254(1):178–196, 2008.
- [16] X.-Y. Zhang, M. N. Trame, L. J. Lesko, and S. Schmidt. Sobol sensitivity analysis: A tool to guide the development and evaluation of systems pharmacology models. *CPT: Pharmacometrics & Systems Pharmacology*, 4(2):69–79, 2015.
- [17] S. Bidah, O. Zakary, and M. Rachik. Stability and global sensitivity analysis for an agree-disagree model: Partial rank correlation coefficient and Latin hypercube sampling methods. *International Journal of Differential Equations*, 2020:1–14, 2020.
- [18] I. M. Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280, 2001.
- [19] A. Hoare, D. G. Regan, and D. P. Wilson. Sampling and sensitivity analyses tools (SaSAT) for computational modelling. *Theoretical Biology and Medical Modelling*, 5(1):4, 2008.
- [20] A. Saltelli and J. Marivoet. Nonparametric statistics in sensitivity analysis for model output: A comparison of selected techniques. *Reliability Engineering & System Safety*, 28(2):229–253, 1990.

- [21] A. Saltelli, M. Ratto, S. Tarantola, and F. Campolongo. Sensitivity analysis for chemical models. *Chemical Reviews*, 105(7):2811–2827, 2005.
- [22] J. P. C. Kleijnen and J. C. Helton. Statistical analyses of scatterplots to identify important factors in large-scale simulations, 1: Review and comparison of techniques. *Reliability Engineering & System Safety*, 65(2):147–185, 1999.
- [23] Y. Yang and H. Li. Neural ordinary differential equations for robust parameter estimation in dynamic systems with physical priors. *Applied Soft Computing*, 169:112649, 2025. doi:10.1016/j.asoc.2024.112649.
- [24] Q. He, L. Wang, and B. Liu. Parameter estimation for chaotic systems by particle swarm optimization. *Chaos, Solitons & Fractals*, 34(2):654–661, 2007. doi:10.1016/j.chaos.2006.03.079.
- [25] M.J. Ahmad and K. Günel. Can neural networks estimate parameters in epidemiology models using real observed data? *Applied Intelligence*, 55(2):133, 2024. doi:10.1007/s10489-024-06012-w.
- [26] H. Jo, K. Josić, and J. K. Kim. Neural network-based parameter estimation for nonautonomous differential equations with discontinuous signals. *SIAM Journal on Applied Mathematics*, 86(1):327–347, 2026. doi:10.1137/25M1741340.
- [27] V. V. Alavani, P. Danumjaya, and P. Seshaiyer. Fractional-order prey- predator models with parameter estimation via fractional physics-informed neural networks (FPINNS). *Journal of Machine Learning for Modeling and Computing*, 7(1), 2026.
- [28] Y. Bard. *Nonlinear Parameter Estimation*. Academic Press, 1974.
- [29] R. C. Aster, B. Borchers, and C. H. Thurber. *Parameter Estimation and Inverse Problems*. Academic Press, 2005.
- [30] H. Liang and H. Wu. 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, 2008. PMID: 19956350. doi:10.1198/016214508000000797.
- [31] J. O. Ramsay, G. Hooker, D. Campbell, and J. Cao. Parameter estimation for differential equations: A generalized smoothing approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(5):741–796, 2007. doi:10.1111/j.1467-9868.2007.00610.x.
- [32] P. H. Boersch-Supan, S. J. Ryan, and L. R. Johnson. deBInfer: Bayesian inference for dynamical models of biological systems. *Methods in Ecology and Evolution*, 8:511–518, 2017. doi:10.1111/2041-210X.12679.
- [33] O. Ghasemi, M. L. Lindsey, T. Yang, N. Nguyen, Y. Huang, and Y.-F. Jin. Bayesian parameter estimation for nonlinear modelling of biological pathways. *BMC Systems Biology*, 5(Suppl 3):S9, 2011. doi:10.1186/1752-0509-5-S3-S9.
- [34] S. Springer, H. Haario, J. Susiluoto, A. Bibov, A. Davis, and Y. Marzouk. Efficient bayesian inference for large chaotic dynamical systems. *Geoscientific Model Development*, 14(7):4319–4333, 2021. doi:10.5194/gmd-14-4319-2021.
- [35] Y. Huang, D. Liu, and H. Wu. Hierarchical bayesian methods for estimation of parameters in a longitudinal HIV dynamic system. *Biometrics*, 62:413–423, 2006.
- [36] K. J. H. Law, A. M. Stuart, and K. C. Zygalakis. *Data Assimilation: A Mathematical Introduction*. Springer, Cham, 2015.
- [37] E. Kalnay. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 2003.
- [38] M. Asch, M. Bocquet, and M. Nodet. *Data Assimilation: Methods, Algorithms, and Applications*, volume 11 of *Fundamentals of Algorithms*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2016.
- [39] A. Carrassi, M. Bocquet, L. Bertino, and G. Evensen. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *WIREs Climate Change*, 9(5):e535, 2018. doi:10.1002/wcc.535.
- [40] D. R. Stauffer and N. L. Seaman. Use of four-dimensional data assimilation in a limited-area mesoscale model. part I: Experiments with synoptic-scale data. *Monthly Weather Review*, 118(6):1250–1277, 1990.

- [41] R. A. Anthes. Data assimilation and initialization of hurricane prediction models. *Journal of Atmospheric Sciences*, 31(3):702 – 719, 1974. doi:10.1175/1520-0469(1974)031<0702:DAAIOH>2.0.CO;2.
- [42] E. Blayo, J. Verron, and J.-M. Molines. Assimilation of TOPEX/POSEIDON altimeter data into a circulation model of the north atlantic. *Journal of Geophysical Research: Oceans*, 99(C12):24691–24705, 1994.
- [43] A. Azouani, E. Olson, and E. S. Titi. Continuous data assimilation using general interpolant observables. *Journal of Nonlinear Science*, 24(2):277–304, 2014. doi:10.1007/s00332-013-9189-y.
- [44] E. Carlson, J. Hudson, A. Larios, V. R. Martinez, E. Ng, and J. P. Whitehead. Dynamically learning the parameters of a chaotic system using partial observations. *Discrete and Continuous Dynamical Systems*, 42(8):3809–3839, 2022. doi:10.3934/dcds.2022033.
- [45] V. R. Martinez, J. Murri, and J. P. Whitehead. Relaxation-based schemes for on-the-fly parameter estimation in dissipative dynamical systems. *Inverse Problems*, 41(5):055001, Apr 2025. doi:10.1088/1361-6420/adc766.
- [46] J. Newey, J. P. Whitehead, and E. Carlson. Model discovery on the fly using continuous data assimilation. *Journal of Computational Physics*, 537:114121, 2025. doi:10.1016/j.jcp.2025.114121.
- [47] A. Biswas and J. Hudson. Determining the viscosity of the Navier–Stokes equations from observations of finitely many modes. *Inverse Problems*, 39(12):125012, 2023.
- [48] A. F. Villaverde, D. Pathirana, F. Fröhlich, J. Hasenauer, and J. R. Banga. A protocol for dynamic model calibration. *Briefings in Bioinformatics*, 2022.
- [49] Y. R. Liyanage, O. Saucedo, N. Tuncer, and G. Chowell. A tutorial on structural identifiability of epidemic models using structural identifiability. *Journal of Mathematical Biology*, 88(6):102, 2026.
- [50] S. Audoly, G. Bellu, L. D’Angio, M. P. Saccomani, and C. Cobelli. Global identifiability of nonlinear models of biological systems. *IEEE Trans. Biomed. Eng.*, 48:55–65, 2001.
- [51] S. M. V. Lunel. Parameter identifiability of differential delay equations. *Internat. J. Adapt. Control Signal Process.*, 15:655–678, 2001.
- [52] F. Ollivier. *Le problème de l’identifiabilité globale: étude théorique, méthodes effectives et bornes de complexité*. PhD thesis, École Polytechnique, Paris, France, 1990.
- [53] X. Xia and C. H. Moog. Identifiability of nonlinear systems with application to HIV/AIDS models. *IEEE Transactions on Automatic Control*, 48(2):330–336, 2003.
- [54] H. Wu, H. Zhu, H. Miao, and A. S. Perelson. Parameter identifiability and estimation of HIV/AIDS dynamic models. *Bulletin of Mathematical Biology*, 70:785–799, 2008.
- [55] G. Bellu, M. P. Saccomani, S. Audoly, and L. D’Angio. Daisy: A new software tool to test global identifiability of biological and physiological systems. *Comput. Methods Programs Biomed.*, 88:52–61, 2007.
- [56] H. Hong, A. Ovchinnikov, G. Pogudin, and C. E. Yap. SIAN: software for structural identifiability analysis of ODE models. *Bioinformatics*, 35:2873–2874, 2019.
- [57] S. Wang and W. Hao. A systematic computational framework for practical identifiability analysis in mathematical models arising from biology. *Advanced Science*, 12(35):e04346, 2025. doi:10.1002/advs.202504346.
- [58] J. A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:308–313, 1965.
- [59] H. Yuan, H. Dai, X. Wei, and P. Ming. Model-based observers for internal states estimation and control of proton exchange membrane fuel cell system: A review. *Journal of Power Sources*, 468:228376, 2020. doi:10.1016/j.jpowsour.2020.228376.
- [60] C. Foias and E. S. Titi. Determining nodes, finite difference schemes and inertial manifolds. *Nonlinearity*, 4(1):135–153, 1991.
- [61] D. A. Jones and E. S. Titi. Upper bounds on the number of determining modes, nodes, and volume elements for the navier–stokes equations. *Indiana University Mathematics Journal*, 42(3):875–887, 1993.

- [62] E. Olson and E. S. Titi. Determining modes for continuous data assimilation in 2D turbulence. *Journal of Statistical Physics*, 113(5–6):799–840, 2003. doi:10.1023/A:1027312703252.
- [63] C. Foias, M. S. Jolly, R. Kravchenko, and E. S. Titi. A determining form for the two-dimensional navier–stokes equations: The fourier modes case. *Journal of Mathematical Physics*, 53(11):115623, 2012. doi:10.1063/1.4766459.
- [64] C. Foias and R. Temam. Determination of the solutions of the navier–stokes equations by a set of nodal values. *Mathematics of Computation*, 43(167):117–133, 1984.
- [65] E. D. Sontag. *Mathematical control theory: deterministic finite dimensional systems*, volume 6. Springer Science & Business Media, 2013.
- [66] E. N. Lorenz. Deterministic nonperiodic flow. *Journal of Atmospheric Sciences*, 20(2):130–141, 1963.
- [67] C. Sparrow. *The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors*. Springer, 1982.
- [68] M. W. Hirsch, S. Smale, and R. L. Devaney. *Differential Equations, Dynamical Systems, & An Introduction to Chaos*. Academic Press, Boston, MA, 2 edition, 2003.
- [69] C. R. Doering and J. D. Gibbon. *Applied Analysis of the Navier-Stokes Equations*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 1995.
- [70] S. G. Krantz and H. R. Parks. *A Primer of Real Analytic Functions*. Birkhäuser, 1992.
- [71] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983. doi:10.1126/science.220.4598.671.
- [72] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of ICNN’95 - International Conference on Neural Networks*, volume 4, pages 1942–1948, 1995. doi:10.1109/ICNN.1995.488968.
- [73] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.
- [74] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, 2018.
- [75] C. Rackauckas, Y. Ma, J. Martensen, C. Warner, K. Zubov, R. Supekar, D. Skinner, A. Ramadhan, and A. Edelman. Universal differential equations for scientific machine learning. *arXiv preprint arXiv:2001.04385*, 2020.
- [76] E. Farhi, J. Goldstone, and S. Gutmann. A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028*, 2014.
- [77] K. Blekos, D. Brand, A. Ceschini, C.-H. Chou, R.-H. Li, K. Pandya, and A. Summer. A review on quantum approximate optimization algorithm and its variants. *Physics Reports*, 1068:1–66, 2024. doi:10.1016/j.physrep.2024.03.002.
- [78] M.J. Ahmad, M. Mohammadisiahroudi, A. Biswas, and K. Hoffman. A data-assimilation-informed quantum optimization framework for parameter estimation in dynamical systems. Manuscript in preparation, 2026.
- [79] L. Perko. *Differential Equations and Dynamical Systems*. Springer, New York, 3 edition, 2001.