



Proceeding Paper

# Bayesian Regularization for Dynamical System Identification: Additive Noise Models <sup>†</sup>

Robert K. Niven <sup>1,2,\*</sup> , Laurent Cordier <sup>3</sup>, Ali Mohammad-Djafari <sup>4</sup>, Markus Abel <sup>5</sup> and Markus Quade <sup>5</sup>

<sup>1</sup> School of Engineering and Technology, The University of New South Wales, Canberra, ACT 2600, Australia

<sup>2</sup> Department of Mechanical Engineering, Auckland University of Technology, Auckland 1142, New Zealand

<sup>3</sup> Institut Pprime, CNRS, Université de Poitiers, ISAE-ENSMA, 86360 Chasseneuil-du-Poitou, France;

laurent.cordier@univ-poitiers.fr

<sup>4</sup> Laboratoire des Signaux et Systèmes (L2S), CentraleSupélec, 91190 Gif-sur-Yvette, France;

ali.mohammad-djafari@l2s.centralesupelec.fr

<sup>5</sup> Ambrosys GmbH, 14482 Potsdam, Germany; markus.abel@ambrosys.de (M.A.);

markus.quade@ambrosys.de (M.Q.)

\* Correspondence: robert.niven@aut.ac.nz

<sup>†</sup> Presented at the 43rd International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Ghent, Belgium, 1–5 July 2024.

## Abstract

Consider the dynamical system  $\dot{x} = f(x)$ , where  $x \in \mathbb{R}^n$  is the state vector,  $\dot{x}$  is the time or spatial derivative, and  $f$  is the system model. We wish to identify unknown  $f$  from its time-series or spatial data. For this, we propose a Bayesian framework based on the maximum a posteriori (MAP) point estimate, to give a generalized Tikhonov regularization method with the residual and regularization terms identified, respectively, with the negative logarithms of the likelihood and prior distributions. As well as estimates of the model coefficients, the Bayesian interpretation provides access to the full Bayesian apparatus, including the ranking of models, the quantification of model uncertainties, and the estimation of unknown (nuisance) hyperparameters. For multivariate Gaussian likelihood and prior distributions, the Bayesian formulation gives a Gaussian posterior distribution, in which the numerator contains a Mahalanobis distance or “Gaussian norm”. In this study, two Bayesian algorithms for the estimation of hyperparameters—the joint maximum a posteriori (JMAP) and variational Bayesian approximation (VBA)—are compared to the popular SINDy, LASSO, and ridge regression algorithms for the analysis of several dynamical systems with additive noise. We consider two dynamical systems, the Lorenz convection system and the Shil’nikov cubic system, with four choices of noise model: symmetric Gaussian or Laplace noise and skewed Rayleigh or Erlang noise, with different magnitudes. The posterior Gaussian norm is found to provide a robust metric for quantitative model selection—with quantification of the model uncertainties—across all dynamical systems and noise models examined.

**Keywords:** Bayesian inverse problem; dynamical systems; system identification; regularization; sparsification



Academic Editor: Geert Verdoolaege

Published: 14 November 2025

**Citation:** Niven, R.K.; Cordier, L.; Mohammad-Djafari, A.; Abel, M.; Quade, M. Bayesian Regularization for Dynamical System Identification: Additive Noise Models. *Phys. Sci. Forum* **2025**, *12*, 17. <https://doi.org/10.3390/psf2025012017>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Consider a dynamical system, commonly represented by

$$\dot{x}(t) = f(x(t)), \quad (1)$$

where  $\mathbf{x} \in \mathbb{R}^n$  is the observable state vector and  $\dot{\mathbf{x}} \in \mathbb{R}^n$  their derivatives, both a function of time  $t$  (or some other parameters), and  $f \in \mathbb{R}^n$  is the system model. Given a set of discrete time series data, how should a user identify the model  $f$ ? In dynamical systems theory, this is referred to as *system identification*. Bayesian practitioners will recognise this as an *inverse problem*, for which the Bayesian inferential framework is eminently suited.

Recently, a number of researchers in dynamical systems have applied regularization methods for system identification from time-series or spatial data (e.g., [1–3]). This is used to determine a matrix of coefficients which, when multiplied by a matrix of functional operations, can reproduce the data. Such methods generally involve a sparsification technique to remove unnecessary coefficients. However, both the regularization term and its coefficient are usually implemented in a heuristic or *ad hoc* manner, without much guidance on how they should be selected for any particular dynamical system. Traditional regularization methods are also unable to provide any further information, for example on the model uncertainty.

In this study, we present a Bayesian framework for dynamical system identification based on the Bayesian maximum a posteriori (MAP) point estimate. This is shown to provide a generalized form of Tikhonov regularization, with the residual and regularization terms corresponding, respectively, to the negative logarithms of the likelihood and prior distributions. The Bayesian interpretation also enables other features of Bayesian inference, including model uncertainty quantification and the estimation of unknown (nuisance) hyperparameters. The present study employs two Bayesian algorithms for hyperparameter estimation, the joint maximum a posteriori (JMAP) and variational Bayesian approximation (VBA), which are compared to the popular SINDy, LASSO, and ridge regression algorithms. We consider two dynamical systems (Lorenz and Shil’nikov) with four choices of noise model (Gaussian, Laplace, Rayleigh, and Erlang), to examine the robustness of the Bayesian algorithms to different additive noise.

## 2. Theoretical Foundations

Traditional regularization methods for system identification (e.g., [1–3]) start from a recorded time (or spatial) series  $\mathbf{x}(t)$  and its derivatives  $\dot{\mathbf{x}}(t)$ , written, respectively, as the  $m \times n$  matrices

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}^\top(t_1) \\ \vdots \\ \mathbf{x}^\top(t_m) \end{bmatrix} = \begin{bmatrix} x_1(t_1) & \dots & x_n(t_1) \\ \vdots & & \vdots \\ x_1(t_m) & \dots & x_n(t_m) \end{bmatrix}, \quad \dot{\mathbf{X}} = \begin{bmatrix} \dot{\mathbf{x}}^\top(t_1) \\ \vdots \\ \dot{\mathbf{x}}^\top(t_m) \end{bmatrix} = \begin{bmatrix} \dot{x}_1(t_1) & \dots & \dot{x}_n(t_1) \\ \vdots & & \vdots \\ \dot{x}_1(t_m) & \dots & \dot{x}_n(t_m) \end{bmatrix}. \quad (2)$$

The user then chooses an alphabet of  $c$  functions based on  $\mathbf{X}$  to populate a  $m \times c$  matrix library; for example,

$$\Theta(\mathbf{X}) = \begin{bmatrix} \mathbf{1} & \mathbf{X} & \mathbf{X}^2 & \mathbf{X}^3 & \dots \end{bmatrix}, \quad (3)$$

using polynomial functions in this case. The dynamical system (1) is then represented by the matrix product:

$$\dot{\mathbf{X}} = \Theta(\mathbf{X})\mathbf{\Xi}, \quad (4)$$

where  $\mathbf{\Xi}$  is a  $c \times n$  matrix of model coefficients  $\xi_{ij} \in \mathbb{R}$ . The computation of each column  $j$  of the coefficient matrix  $\mathbf{\Xi}$  requires the inversion of (4), commonly using a regularization method of the form

$$\hat{\mathbf{\Xi}}_j = \arg \min_{\mathbf{\Xi}_j} J(\mathbf{\Xi}_j), \quad J(\mathbf{\Xi}_j) = \|\dot{\mathbf{X}}_j - \Theta(\mathbf{X})\mathbf{\Xi}_j\|_\beta^\alpha + \kappa_j \|\mathbf{\Xi}_j\|_\delta^\gamma, \quad (5)$$

where  $\|\mathbf{y}\|_p = (\sum_i |y_i|^p)^{1/p}$  for  $p \in \mathbb{R}$  is the  $L_p$  norm,  $\kappa_j \in \mathbb{R}$  is the regularization coefficient, and  $\alpha, \beta, \gamma, \delta \in \mathbb{R}$  are constants. Several methods have been implemented for (5), including

ridge regression [4] for  $\alpha, \beta, \gamma, \delta = 2$ ; the least absolute shrinkage and selection operator (LASSO) [5] for  $\alpha, \beta = 2$  and  $\gamma, \delta = 1$ ; and strong sparsity [3] for  $\alpha, \beta = 2, \gamma = 1$ , and  $\delta = 0$ . Instead of (5), many authors use least squares regression with iterative thresholding, known as the sparse identification of nonlinear dynamics (SINDy) method [1]:

$$J(\Xi_j) = \|\dot{X}_j - \Theta(X)\Xi_j\|_2^2 \quad \text{with} \quad |\zeta_{ij}| \geq \lambda, \forall \zeta_{ij} \in \Xi_j. \tag{6}$$

where  $\lambda$  is the threshold.

In the Bayesian approach to this problem (e.g., [6]), it is recognized that the time series decomposition for each column  $j$  should be written as

$$\dot{X}_j = \Theta(X)\Xi_j + \epsilon_j, \tag{7}$$

where  $\epsilon_j$  is a noise or error term. All variables are considered probabilistic, leading to the posterior probability of  $\Xi_j$  subject to the data, given by Bayes' rule:

$$p(\Xi_j|\dot{X}_j) = \frac{p(\dot{X}_j|\Xi_j)p(\Xi_j)}{p(\dot{X}_j)} \propto p(\dot{X}_j|\Xi_j)p(\Xi_j). \tag{8}$$

The simplest Bayesian method is to calculate the maximum a posteriori (MAP) point estimate of  $\Xi_j$  by maximizing (8). It is convenient to consider the logarithmic maximum, giving

$$\hat{\Xi}_j = \arg \max_{\Xi_j} [\ln p(\Xi_j|\dot{X}_j)] = \arg \max_{\Xi_j} [\ln p(\dot{X}_j|\Xi_j) + \ln p(\Xi_j)]. \tag{9}$$

To reduce (9), we make two assumptions. First, we assume unbiased multivariate Gaussian noise with covariance matrix  $V_{\epsilon_j}$ :

$$p(\epsilon_j|\Xi_j) = \mathcal{N}(\epsilon_j|\mathbf{0}, V_{\epsilon_j}) = \frac{\exp(-\frac{1}{2}\epsilon_j^T V_{\epsilon_j}^{-1}\epsilon_j)}{\sqrt{(2\pi)^m \det V_{\epsilon_j}}} \propto \exp(-\frac{1}{2}\|\epsilon_j\|_{V_{\epsilon_j}^{-1}}^2), \tag{10}$$

where  $\det$  is the determinant, and the second form is written in terms of the Mahalanobis distance  $\|y\|_A := \sqrt{y^T A y}$ , with respect to a symmetric positive semi-definite matrix  $A$ , referred to here as a "Gaussian norm". From (7), this gives the likelihood

$$p(\dot{X}_j|\Xi_j) \propto \exp(-\frac{1}{2}\|\dot{X}_j - \Theta(X)\Xi_j\|_{V_{\epsilon_j}^{-1}}^2). \tag{11}$$

Second, we assume a multivariate unbiased Gaussian prior with covariance matrix  $V_{\Xi_j}$ :

$$p(\Xi_j) = \mathcal{N}(\Xi_j|\mathbf{0}, V_{\Xi_j}) = \frac{\exp(-\frac{1}{2}\|\Xi_j\|_{V_{\Xi_j}^{-1}}^2)}{\sqrt{(2\pi)^c \det V_{\Xi_j}}} \propto \exp(-\frac{1}{2}\|\Xi_j\|_{V_{\Xi_j}^{-1}}^2), \tag{12}$$

The MAP estimator (9)—which for Gaussians, is equivalent to the mean—reduces to [6]:

$$\hat{\Xi} = \arg \min_{\Xi_j} [\|\dot{X}_j - \Theta(X)\Xi_j\|_{V_{\epsilon_j}^{-1}}^2 + \|\Xi_j\|_{V_{\Xi_j}^{-1}}^2]. \tag{13}$$

Equation (13) therefore reduces to an objective function very similar to that used for regularization (5), with the likelihood identified with the residual term and the prior identi-

fied with the regularization term. For Gaussian likelihood (11) and prior (12) distributions, the posterior is also Gaussian, with the analytical solution [7]

$$p(\Xi_j | \dot{X}_j) = \mathcal{N}(\Xi_j | \hat{\Xi}_j, \hat{\Delta}_j) \propto \exp\left(-\frac{1}{2} \|\Xi_j - \hat{\Xi}_j\|_{\hat{\Delta}_j}^2\right) \tag{14}$$

$$\hat{\Xi}_j = \hat{\Delta}_j (\Theta^\top V_{\epsilon_j}^{-1} \dot{X}_j), \quad \hat{\Delta}_j = (\Theta^\top V_{\epsilon_j}^{-1} \Theta + V_{\Xi_j}^{-1})^{-1}.$$

The covariance matrices  $V_{\epsilon_j}$  and  $V_{\Xi_j}$  are unknown and must be determined. In the Bayesian framework, these are incorporated into a joint posterior, which can be simplified to

$$p(\Xi_j, V_{\epsilon_j}, V_{\Xi_j} | \dot{X}_j) \propto p(\dot{X}_j | \Xi_j, V_{\epsilon_j}) p(\Xi_j | V_{\Xi_j}) p(V_{\epsilon_j}) p(V_{\Xi_j}). \tag{15}$$

The covariance priors can be represented by products of inverse gamma distributions:

$$p(V_{\epsilon_j}) = \prod_{i=1}^m p(v_{\epsilon_{ij}}) = \prod_{i=1}^m \text{IG}(v_{\epsilon_{ij}} | \alpha_{\epsilon_{ij}}, \beta_{\epsilon_{ij}}), \quad p(V_{\Xi_j}) = \prod_{\ell=1}^c p(v_{\Xi_{\ell j}}) = \prod_{\ell=1}^c \text{IG}(v_{\Xi_{\ell j}} | \alpha_{\Xi_{\ell j}}, \beta_{\Xi_{\ell j}}) \tag{16}$$

with hyperparameters  $\{\alpha_{\epsilon_{ij}}, \beta_{\epsilon_{ij}}\}$  and  $\{\alpha_{\Xi_{\ell j}}, \beta_{\Xi_{\ell j}}\}$ . Two Bayesian algorithms are used in this study. In the joint maximum a posteriori (JMAP) algorithm, (15) is maximized iteratively with respect to  $\{\Xi_j, V_{\epsilon_j}, V_{\Xi_j}\}$ , to give the estimated parameters  $\hat{\Xi}_j, \hat{V}_{\epsilon_j}$ , and  $\hat{V}_{\Xi_j}$  [8]. In the variational Bayesian approximation (VBA), the posterior in (15) is approximated by  $q(\Xi_j, V_{\epsilon_j}, V_{\Xi_j}) = q_1(\Xi_j)q_2(V_{\epsilon_j})q_3(V_{\Xi_j})$ . The individual MAP estimates are then calculated iteratively, using a Kullback–Leibler distance between  $p(\Xi_j, V_{\epsilon_j}, V_{\Xi_j} | \dot{X}_j)$  and  $q$  as the convergence criterion [8].

### 3. Application

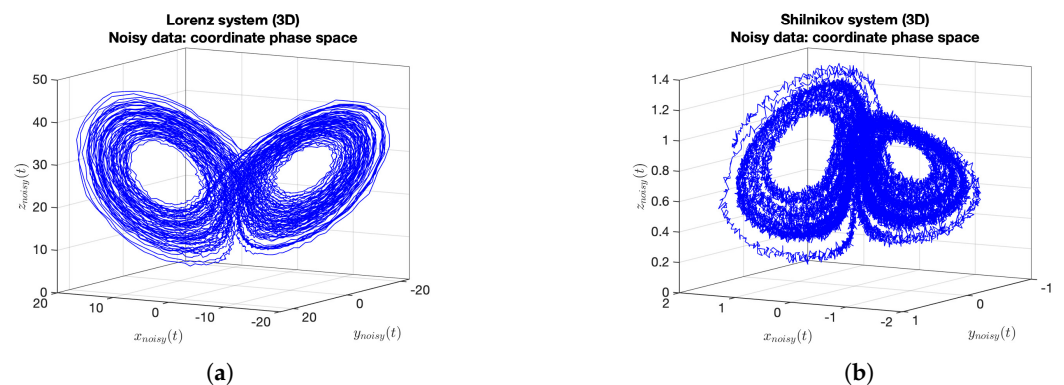
In this study, two dynamical systems are considered. The first is the well-known Lorenz system, a quadratic system given by [9]

$$\dot{x} = [\sigma(y - x), \quad x(\rho - z) - y, \quad xy - \beta z]^\top, \tag{17}$$

which is chaotic for  $[\sigma, \rho, \beta] = [10, \frac{8}{3}, 28]$ . The second is the Shil’nikov system, a cubic system given by [10]

$$\dot{x} = [y, \quad x(1 - z) - Bx^3 - \lambda y, \quad -\alpha(z - x^2)]^\top, \tag{18}$$

which is chaotic for  $[\alpha, \lambda, B] = [\frac{4\sqrt{30}}{135}, \frac{11\sqrt{30}}{90}, \frac{2}{13}]$ . Phase plots of these systems with added Gaussian noise (see below) are shown in Figure 1a,b.



**Figure 1.** Calculated noisy data for added Gaussian noise with  $\mu = 0, \sigma = 1$ , and  $\epsilon = 0.2$ : (a) Lorenz system and (b) Shil’nikov system.

The analyses were conducted in Matlab 2021b on a MacBook Pro with 2.3 GHz Intel Core i9, with numerical integration by the ode45 function. For each system, the data  $X$  were calculated and augmented by additive random noise using the function  $X_{noisy} = X + \epsilon X_{distrib}$ , where  $X_{distrib}$  is sampled from a univariate noise distribution and  $\epsilon$  is a scaling parameter. Four noise models with random variable  $y$  were considered, illustrated in Table 1. These include two symmetric noise models (Gaussian and Laplace with  $\mu = 0$ ) and two non-negative skewed noise models (Rayleigh and Erlang). Only the Gaussian noise model conforms to the assumption of Gaussian error (10), so the other models provide a test of robustness of the algorithms used. The noisy derivatives  $\dot{X}_{noisy}$  were then calculated from the noisy data  $X_{noisy}$  by a dynamical system function call.

**Table 1.** Noise models adopted in this study.

Gaussian noise $\eta = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$ mean $\mu$ , standard deviation $\sigma$	Laplace noise $\eta = \frac{1}{\sigma\sqrt{2}} \exp\left(-\frac{\sqrt{2} y-\mu }{\sigma}\right)$ mean $\mu$ , scale $\sigma/\sqrt{2}$	Rayleigh noise $\eta = \frac{y}{\sigma^2} \exp\left(-\frac{y^2}{2\sigma^2}\right)$ for $y \geq 0$ scale $\sigma$	Erlang noise $\eta = \frac{\lambda^s y^{s-1} e^{-\lambda y}}{(s-1)!}$ for $y \geq 0$ shape $s$ , rate $\lambda$

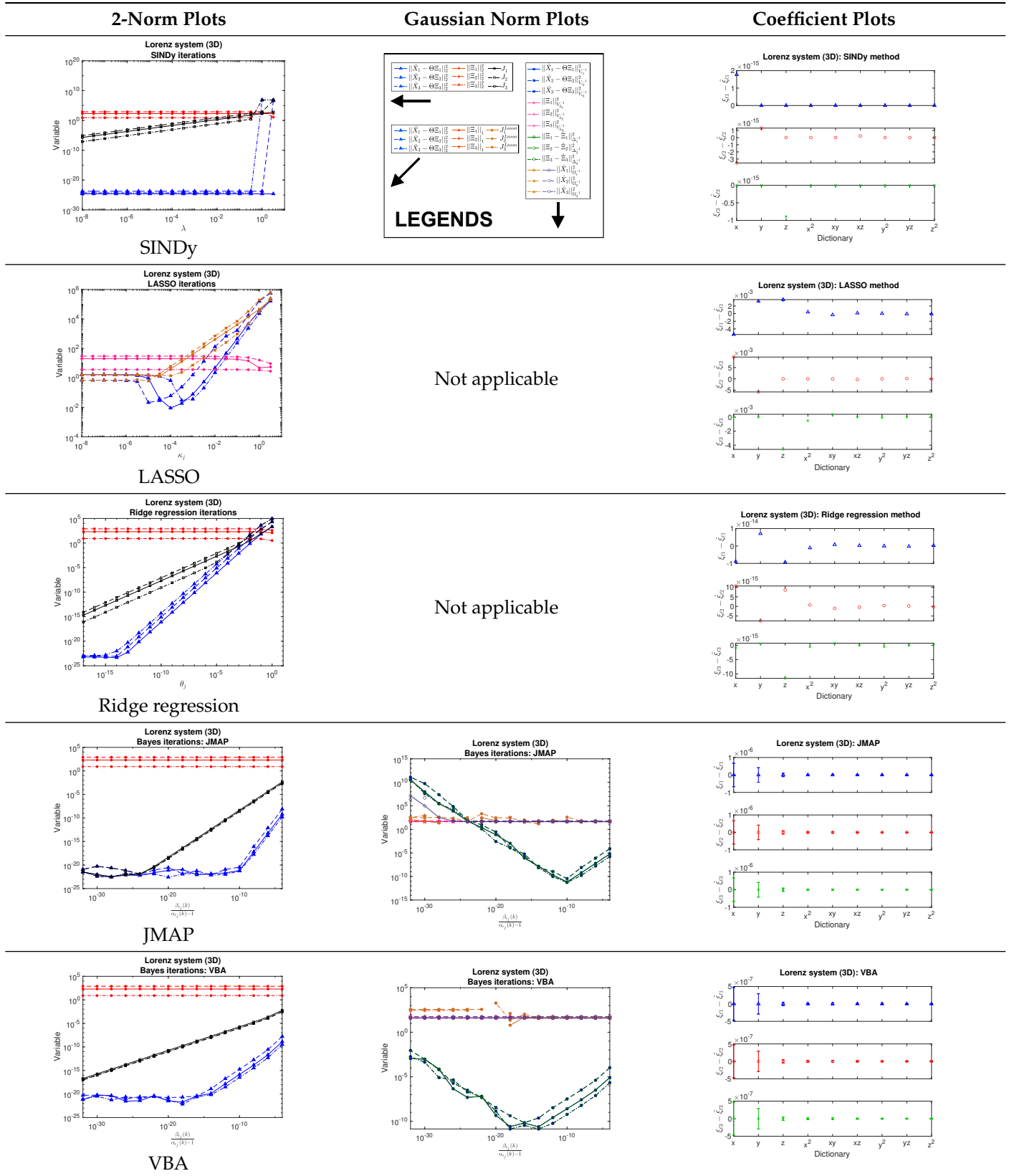
The data were then analyzed by three traditional regularization algorithms (SINDy, LASSO, ridge regression) and two Bayesian algorithms (JMAP and VBA), using inbuilt Matlab functions for LASSO and ridge regression, and modified forms of published codes for SINDy [2], JMAP, and VBA [6]. All analyses were conducted using an inner iteration for fixed hyperparameter(s), and an outer iteration over a sequence of hyperparameter(s). For SINDy, iteration was conducted over  $\lambda$ , proceeding from a high value, while for LASSO and ridge regression, iteration was conducted over each  $\kappa_j$ . For JMAP and VBA, a broad Gaussian model prior and covariance prior with fixed hyperparameters  $\{\alpha_{\Xi_{\ell_j}}, \beta_{\Xi_{\ell_j}}\}$  were selected, to represent the a priori belief that the model coefficients should be symmetric about zero, penalising coefficients approaching  $\pm\infty$ . The outer iteration was then conducted over the error hyperparameters  $\{\alpha_{\epsilon_{ij}}, \beta_{\epsilon_{ij}}\}$ , combined into the inverse gamma expectation  $\beta_{\epsilon_{ij}}/(\alpha_{\epsilon_{ij}} - 1)$ . In each method, the 2-norms and (where available) Gaussian norms were examined.

### 4. Results

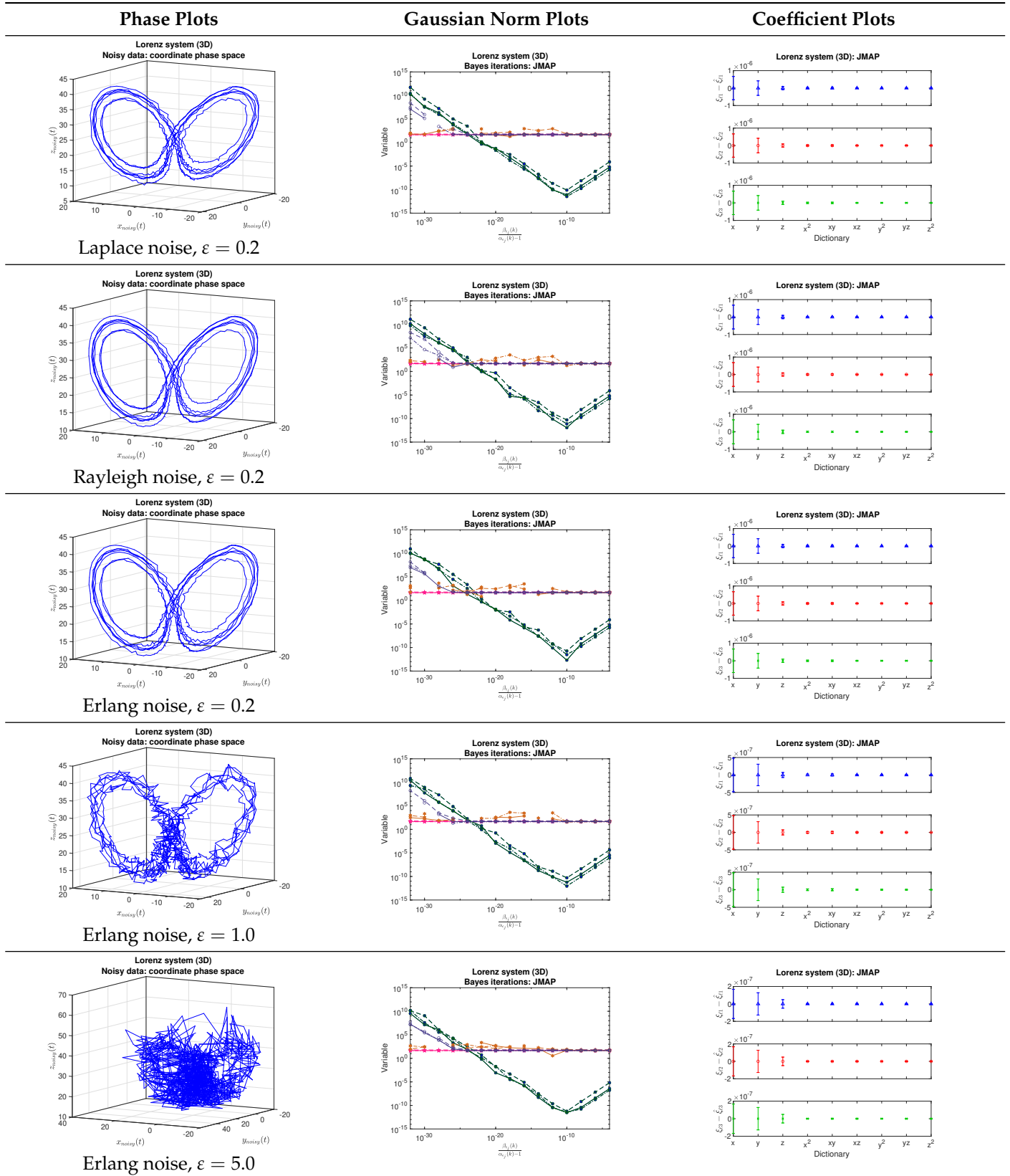
The outputs from several sets of analyses are shown below, including for the Lorenz system with Gaussian noise analyzed by different methods (Table 2), and the Lorenz or Shil’nikov systems with different noise models analyzed by JMAP (respectively, Tables 3 and 4). From these and other analyses, we can interpret the following:

- For JMAP and VBA, the optimum in  $\{\alpha_{\epsilon_{ij}}, \beta_{\epsilon_{ij}}\}$  is identified by a minimum in the posterior Gaussian norm  $\|\Xi_j - \hat{\Xi}_j\|_{\Delta_j}^2$  (green curves, column 1 of Table 2; column 2 of Tables 3 and 4). This optimum corresponds to the maximum in the posterior (14), and is distinct in all examples considered. Both algorithms are fairly efficient in finding the optimum ( $\sim 5$ – $10$  iterations).

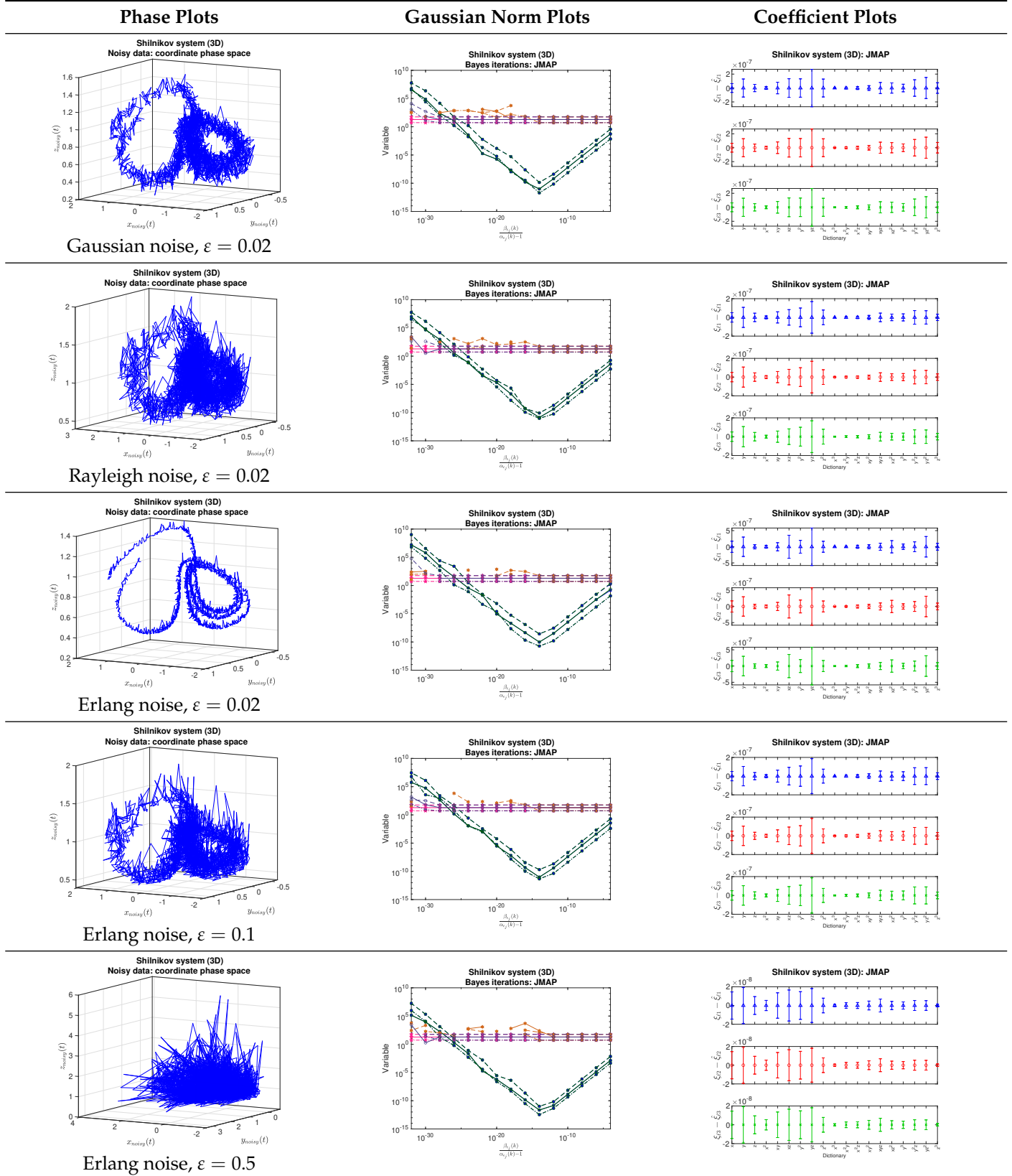
**Table 2.** Analyses of the Lorenz system ( $t = 10, t_{step} = 0.01$ ) with Gaussian noise ( $\mu = 0, \sigma = 1, \varepsilon = 0.2$ ) by SINDy, LASSO, ridge regression, JMAP, and VBA, showing (column 1) plots of residual, regularization term and the objective function; (column 2) plots of prior, likelihood, and posterior Gaussian norms (where available); and (column 3) differences in predicted parameters  $\zeta_{ij} - \hat{\zeta}_{ij}$ , with predicted standard deviations shown as error bars for JMAP and VBA.



**Table 3.** Analyses of the Lorenz system ( $t = 10, t_{step} = 0.01$ ) by JMAP for various noise models ( $\mu = 0, \sigma = s = \lambda = 1$ ) of different scales  $\varepsilon$ , showing (column 1) the phase plot of noisy data; (column 2) plots of prior, likelihood, and posterior Gaussian norms; and (column 3) differences in predicted parameters  $\xi_{ij} - \hat{\xi}_{ij}$ , with predicted standard deviations shown as error bars. The legend for column 2 is for Table 2.



**Table 4.** Analyses of the Shil’nikov system ( $t = 50, t_{step} = 0.02$ ) by JMAP for various noise models ( $\mu = 0, \sigma = s = \lambda = 1$ ) of different scales  $\varepsilon$ , showing (column 1) the phase plot of noisy data; (column 2) plots of prior, likelihood, and posterior Gaussian norms; and (column 3) differences in predicted parameters  $\tilde{\xi}_{ij} - \hat{\xi}_{ij}$ , with predicted standard deviations shown as error bars. The legend for column 2 is for Table 2.



- For SINDy and ridge regression, the optimum in the threshold  $\lambda$  or regularization parameter  $\kappa_j$  is identified by a turning point in the residual  $\|\hat{X}_j - \Theta \Xi_j\|_2^2$  (blue curves, column 1 of Table 2). In contrast, for LASSO, the residuals follow strange curves; the optimum is instead identified by a turning point in the objective function (brown curves, column 1 of Table 2). The SINDy method is very efficient ( $\sim 4$ – $5$  iterations) in finding the optimum, while LASSO and ridge regression require many more iterations.
- The Bayesian methods provide standard deviations on the predicted coefficients, extracted from the posterior covariance  $\hat{\Delta}_j$  in (14) (error bars, column 3 of Tables 2–4). These are of order  $\sim 10^{-8}$ – $10^{-6}$  in all examples. In contrast, SINDy and ridge regression report a coefficient precision of  $\sim 10^{-15}$ – $10^{-14}$ . The inconsistency with JMAP and VBA suggest that this high precision is artificial. LASSO was unsuccessful in all examples examined, with a precision of  $\sim 10^{-3}$  and poor recovery of the dynamical system.
- Contrary to our expectations, JMAP and VBA based on Gaussian noise (13) were highly robust to different choices of additive noise, for all dynamical systems, noise models, and parameter choices examined. Little difference was observed between the symmetric or non-negative noise models, even for such high magnitudes of added noise  $\varepsilon$  that the dynamical system becomes overwhelmed by noise (column 1 of Tables 2 and 3).

## 5. Conclusions

In this study, we consider the problem of system identification of a dynamical system, represented by a nonlinear equation system  $\dot{x} = f(x)$ , from discrete time-series or spatial data. We propose a Bayesian framework based on the maximum a posteriori (MAP) point estimate, to give a generalized form of Tikhonov regularization with the residual and regularization terms identified, respectively, with the negative logarithms of the likelihood and prior distributions. As well as estimates of the model coefficients, the Bayesian interpretation provides access to the full Bayesian apparatus, including the ranking of models, the quantification of model uncertainties, and the estimation of unknown (nuisance) hyperparameters. The present study employs two Bayesian algorithms for hyperparameter estimation, the joint maximum a posteriori (JMAP) and variational Bayesian approximation (VBA), which are compared to the popular SINDy, LASSO, and ridge regression algorithms. For multivariate Gaussian likelihood and prior distributions, the Bayesian formulation gives a Gaussian posterior distribution, in which the numerator contains a Mahalanobis distance or “Gaussian norm”. We consider two dynamical systems (Lorenz and Shil’nikov), with four choices of noise model (Gaussian, Laplace, Rayleigh, and Erlang). Both JMAP and VBA algorithms were successful for model identification and uncertainty quantification, with predicted standard deviations in the model coefficients of  $\sim 10^{-8}$ – $10^{-6}$  in all examples. This compares to the calculated precision of  $\sim 10^{-15}$ – $10^{-14}$  by SINDy and ridge regression, which we consider to be artificially low. LASSO was unsuccessful in all examples examined. For the Bayesian algorithms, the posterior Gaussian norm is found to provide a robust metric for quantitative model selection for all combinations of dynamical system and noise model examined.

**Author Contributions:** Conceptualization and methodology, R.K.N., L.C., M.A. and A.M.-D.; software and analysis, R.K.N., L.C., M.A. and M.Q.; writing—original draft preparation, R.K.N., L.C. and M.A.; writing—review and editing, all authors; funding, R.K.N., L.C. and A.M.-D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by Institute Pprime, CNRS, Poitiers, France, and Centrale-Supélec, Gif-sur-Yvette, France.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The BDSI MATLAB code is available in GitHub (<https://github.com/markusabel/BDSI>).

**Conflicts of Interest:** The authors declare no conflicts of interest. M.A. and M.Q. are from Ambrosys GmbH, while this research is unrelated to the company's primary business activities.

## References

1. Brunton, S.L.; Proctor, J.L.; Kutz, J.N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 3932–3937. [[CrossRef](#)] [[PubMed](#)]
2. Mangan, N.M.; Kutz, J.N.; Brunton, S.L.; Proctor, J.L. Model selection for dynamical systems via sparse regression and information criteria. *Proc. R. Soc. Proc. A* **2017**, *473*, 20170009. [[CrossRef](#)] [[PubMed](#)]
3. Rudy, S.H.; Brunton, S.L.; Proctor, J.L.; Kutz, J.N. Data-driven discovery of partial differential equations. *Sci. Adv.* **2017**, *3*, e1602614. [[CrossRef](#)] [[PubMed](#)]
4. Hoerl, A.E.; Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **1970**, *12*, 55–67. [[CrossRef](#)]
5. Santosa, F.; Symes, W.W. Linear inversion of band-limited reflection seismograms. *SIAM J. Sci. Stat. Comput.* **1986**, *7*, 1307–1330. [[CrossRef](#)]
6. Mohammad-Djafari, A. Inverse problems in signal and image processing and Bayesian inference framework: From basic to advanced Bayesian computation. In Proceedings of the Scube Seminar, L2S, CentraleSupélec, Gif-sur-Yvette, France, 27 March 2015.
7. Tarantola, A. *Inverse Problem Theory and Methods for Model Parameter Estimation*; SIAM: Philadelphia, PA, USA, 2005.
8. Mohammad-Djafari, A.; Dumitru, M. Bayesian sparse solutions to linear inverse problems with non-stationary noise with Student-t priors. *Digit. Signal Proc.* **2015**, *47*, 128–156. [[CrossRef](#)]
9. Lorenz, E.N. Deterministic nonperiodic flow. *J. Atmos. Sci.* **1963**, *20*, 130–141. [[CrossRef](#)]
10. Shil'nikov, A.L.; Shil'nikov, L.P.; Turaev, D.V. Normal forms and Lorenz attractors. *Int. J. Bifurc. Chaos* **1993**, *3*, 1123–1139. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.