



HAL
open science

Dynamical System Identification by Bayesian Inference

Robert K. Niven, Mohammad-Djafari Ali, Laurent Cordier, Markus Abel,
Markus Quade

► **To cite this version:**

Robert K. Niven, Mohammad-Djafari Ali, Laurent Cordier, Markus Abel, Markus Quade. Dynamical System Identification by Bayesian Inference. 22nd Australasian Fluid Mechanics Conference AFMC2020, Dec 2020, Brisbane, Australia. 10.14264/692fcb8 . hal-03088615

HAL Id: hal-03088615

<https://hal.science/hal-03088615>

Submitted on 26 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Dynamical System Identification by Bayesian Inference

R.K. Niven¹, A. Mohammad-Djafari², L. Cordier³, M. Abel⁴ and M. Quade⁴

¹ School of Engineering and Information Technology, The University of New South Wales, Canberra, ACT, 2600, Australia

² Laboratoire des signaux et systèmes (L2S), CentraleSupélec, Gif-sur-Yvette, France

³ Institut Pprime (CNRS, Université de Poitiers, ISAE-ENSMA), Poitiers, France

⁴ Ambrosys GmbH, Potsdam, Germany

Abstract

Fluid flow systems provide physical expressions of dynamical systems, typically written $\dot{\mathbf{x}} = f(\mathbf{x})$, where \mathbf{x} is the state vector and f is the system function or model. The identification of the dynamical system f from time-series data $\{x_1, \dots, x_n\}$, an inverse problem, is a long-held challenge. Historically, this has been examined by linear or nonlinear regression, convolution methods, neural networks or evolutionary computation, but these mostly lie outside the rigorous framework of Bayesian inference. Here we examine the maximum a-posteriori (MAP) Bayesian method for system identification, which is shown to be equivalent to Tikhonov regularization, and in fact provides sound theoretical justifications for the choices of residual and regularization terms. The joint maximum a-posteriori (JMAP) and variational Bayesian approximation (VBA) are demonstrated by comparison to the popular SINDy regularization method, by application to the Rössler dynamical system.

Keywords

Bayesian inverse problem, dynamical system, system identification, regularization, sparsification

Introduction

A dynamical system such as a fluid flow system is typically represented by the equation:

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

where $\mathbf{x} \in \mathbb{R}^n$ is the observable state vector, $\dot{\mathbf{x}} \in \mathbb{R}^n$ is its time derivative, and f is the system function or model. Usually, a dynamical system is observed at discrete time steps, giving data in the form of a discrete time series $\{\mathbf{x}(t_1), \mathbf{x}(t_2), \mathbf{x}(t_3), \dots\}$. A fundamental question is how to infer the dynamical system model f from such data? This is generally referred to as *system identification*, but if the model is known to arise from a class of models described by a set of parameters, it may reduce to that of *parameter identification*. The inference problem requires the inversion of (1), and so is described as an *inverse problem*. Many methods have been applied to such problems, including linear or nonlinear regression, convolution methods, neural networks or evolutionary computation; however these exhibit a number of deficiencies, in particular an inability to assess (or rank) the appropriateness of the selected model. This arises from the fact that such methods are usually posed outside the framework of Bayesian inference, in which the uncertainty of the model is handled rigorously as part of the inference process.

In recent years, there has been considerable interest in the application of regularization methods for dynamical system identification from time-series or spatial data [1, 2, 3]. Such methods generally apply a sparse regression method, in which a regularization term is imposed as part of the optimization process, to enforce sparsification of the inferred matrix of parameters. Alternatively, the sparsification can be imposed using a parameter

threshold, e.g., in the Sparse Identification of Nonlinear Dynamics (SINDy) method [1]. Such methods have proved to be extremely useful for system (or parameter) identification, but still involve a considerable degree of heuristic or *ad hoc* handling, especially in the choice of regularization method and its regularization parameter.

The aim of this study is to examine the maximum a-posteriori (MAP) Bayesian method for dynamical system identification, based on maximization of the Bayesian posterior probability distribution, the product of the likelihood and prior distributions. Under the assumption of Gaussian distributions, this is shown to reduce to an advanced form of Tikhonov regularization, which underlies many regularization methods used to examine dynamical systems. Furthermore, the Bayesian MAP method provides sound theoretical justifications for the choices of residual and regularization terms. It can also be extended to incorporate additional features of Bayesian inference, including the estimation of uncertainties from the posterior distribution, and if necessary the complete posterior distribution if this is desired. Two Bayesian methods, joint maximum a-posteriori (JMAP) and variational Bayesian approximation (VBA), are here demonstrated by comparison to the Sparse Identification of Nonlinear Dynamics (SINDy) regularization method [1], by application to the Rössler dynamical system with additive noise.

Theory

In sparse regression methods applied to system identification, the data from m time steps of an n -dimensional parameter \mathbf{x} and its time derivative $\dot{\mathbf{x}}$ are assembled into $m \times n$ matrices [1, 2, 3]:

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

$$\dot{\mathbf{X}} = \begin{bmatrix} \dot{\mathbf{x}}^\top(t_1) \\ \vdots \\ \dot{\mathbf{x}}^\top(t_m) \end{bmatrix} = \begin{bmatrix} \dot{x}_1(t_1) & \dots & \dot{x}_n(t_1) \\ \vdots & & \vdots \\ \dot{x}_1(t_m) & \dots & \dot{x}_n(t_m) \end{bmatrix} \quad (3)$$

Considering $\dot{\mathbf{X}}$ as a function of \mathbf{X} , a set or *alphabet* of c functions, such as polynomial or trigonometric functions, are applied to the data to populate a $m \times c$ matrix library, e.g.:

$$\Theta(\mathbf{X}, \dot{\mathbf{X}}) = \begin{bmatrix} 1 & \mathbf{X} & \dot{\mathbf{X}} & \mathbf{X}^2 & \dot{\mathbf{X}}^2 & \mathbf{X}^3 & \dot{\mathbf{X}}^3 \dots \end{bmatrix}, \quad (4)$$

The problem is then formulated as the matrix equation:

$$\dot{\mathbf{X}} = \Theta(\mathbf{X}, \dot{\mathbf{X}})\mathbf{K}, \quad (5)$$

where \mathbf{K} is a $c \times n$ matrix of coefficients $k_{ij} \in \mathbb{R}$. The inverse problem requires inversion of (5) to determine \mathbf{K} . This is commonly posed as the minimization problem:

$$\hat{\mathbf{K}} = \arg \min_{\mathbf{K}} J(\mathbf{K}) = \arg \min_{\mathbf{K}} \left[\|\dot{\mathbf{X}} - \Theta(\mathbf{X}, \dot{\mathbf{X}})\mathbf{K}\|_{\beta}^{\alpha} + \lambda \|\mathbf{K}\|_{\gamma}^{\alpha} \right] \quad (6)$$

where $\hat{\cdot}$ indicates an inferred value, $J(\mathbf{K})$ is the objective function, $\|\cdot\|_p$ is the p norm, $\lambda \in \mathbb{R}$ is the regularization coefficient and $\alpha, \beta, \gamma \in \mathbb{R}$ are constants. Structurally, the objective function in (6) is composed respectively of residual and regularization terms, allowing an interplay between minimization of the residual to extract the solution, and minimization of the regularization term to overcome noise by enforcing a sparse matrix [1, 2, 3]. Eqs. (6) have been applied to a wide range of dynamical systems with $\alpha \in \{1, 2\}$, $\beta = 2$ and $\gamma \in \{0, [1, 2]\}$ [4, 5, 6, 2, 3]. Alternatively, the now-popular SINDy method imposes an iterative thresholding, which can be represented by [1, 7]:

$$J(\mathbf{K}) = \|\dot{\mathbf{X}} - \Theta(\mathbf{X}, \dot{\mathbf{X}})\mathbf{K}\|_2^2 \quad \text{with} \quad |k_{ij}| \geq \lambda, \forall k_{ij} \in \mathbf{K}. \quad (7)$$

Regularization methods have also been shown to have strong connections to the analysis of dynamical systems by singular value decomposition (SVD), dynamic mode decomposition (DMD) and the application of Koopman operators [8, 9, 10].

In the Bayesian approach to the inverse problem, all variables are treated as probabilistic quantities, represented by a probability density functions (pdfs). Rather than inverting (5), the Bayesian seeks the posterior probability of \mathbf{K} given the data, defined by Bayes' rule:

$$p(\mathbf{K}|\dot{\mathbf{X}}) = \frac{p(\dot{\mathbf{X}}|\mathbf{K})p(\mathbf{K})}{p(\dot{\mathbf{X}})} \propto p(\dot{\mathbf{X}}|\mathbf{K})p(\mathbf{K}). \quad (8)$$

The simplest Bayesian method is to extract the \mathbf{K} that maximizes (8), known as the maximum *a posteriori* (MAP) estimate. For higher resolution, it is usual to consider the logarithmic maximum:

$$\hat{\mathbf{K}} = \arg \max_{\mathbf{K}} [\ln p(\mathbf{K}|\dot{\mathbf{X}})] = \arg \max_{\mathbf{K}} [\ln p(\dot{\mathbf{X}}|\mathbf{K}) + \ln p(\mathbf{K})]. \quad (9)$$

Explicitly incorporating the error or noise term $\boldsymbol{\varepsilon}$ in (5) [11, 12, 13, 14, 15]:

$$\dot{\mathbf{X}} = \Theta(\mathbf{X}, \dot{\mathbf{X}})\mathbf{K} + \boldsymbol{\varepsilon}, \quad (10)$$

we then assume a multivariate Gaussian noise distribution with covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$: [14]:

$$p(\boldsymbol{\varepsilon}|\mathbf{K}) = \mathcal{N}(0, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) \propto \exp\left(-\frac{1}{2}\|\boldsymbol{\varepsilon}\|_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}}^2\right), \quad (11)$$

where, making a change in notation, $\|\boldsymbol{\varepsilon}\|_{\mathbf{A}}^2 = \boldsymbol{\varepsilon}^T \mathbf{A} \boldsymbol{\varepsilon}$ based on matrix \mathbf{A} . From (10), we obtain the likelihood:

$$p(\dot{\mathbf{X}}|\mathbf{K}) \propto \exp\left(-\frac{1}{2}\|\dot{\mathbf{X}} - \Theta(\mathbf{X}, \dot{\mathbf{X}})\mathbf{K}\|_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}}^2\right). \quad (12)$$

Secondly we assume a multivariate Gaussian prior with covariance matrix $\boldsymbol{\Sigma}_{\mathbf{K}}$:

$$p(\mathbf{K}) = \mathcal{N}(0, \boldsymbol{\Sigma}_{\mathbf{K}}) \propto \exp\left(-\frac{1}{2}\|\mathbf{K}\|_{\boldsymbol{\Sigma}_{\mathbf{K}}^{-1}}^2\right), \quad (13)$$

From (12)-(13), the MAP estimator (9) becomes [14, 15]:

$$\begin{aligned} \hat{\mathbf{K}} &= \arg \max_{\mathbf{K}} \left[-\frac{1}{2}\|\dot{\mathbf{X}} - \Theta(\mathbf{X}, \dot{\mathbf{X}})\mathbf{K}\|_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}}^2 - \frac{1}{2}\|\mathbf{K}\|_{\boldsymbol{\Sigma}_{\mathbf{K}}^{-1}}^2 \right] \\ &= \arg \min_{\mathbf{K}} \left[\|\dot{\mathbf{X}} - \Theta(\mathbf{X}, \dot{\mathbf{X}})\mathbf{K}\|_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}}^2 + \|\mathbf{K}\|_{\boldsymbol{\Sigma}_{\mathbf{K}}^{-1}}^2 \right]. \end{aligned} \quad (14)$$

The Bayesian MAP estimate thus provides an objective function that is very similar to that for sparse regression method (6). Indeed, ifor isotropic Gaussian distributions for the noise $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \sigma_{\boldsymbol{\varepsilon}}^2 \mathbf{I}$ and prior $\boldsymbol{\Sigma}_{\mathbf{K}} = \sigma_{\mathbf{K}}^2 \mathbf{I}$, where \mathbf{I} is the identity matrix, it can be shown that (14) reduces to the regularization equation

(6) with $\alpha = \beta = \gamma = 2$. The regularization parameter is also obtained explicitly as $\lambda = \sigma_{\boldsymbol{\varepsilon}}^2 / \sigma_{\mathbf{K}}^2$ [11, 12, 15].

In Bayesian inference, any unknown or *nuisance* parameters can be incorporated into the inferred posterior pdf. Here, the covariance properties of the noise and prior are unknown. For isotropic Gaussian distributions, these can be inferred by expanding the posterior as follows:

$$p(\mathbf{K}, \sigma_{\boldsymbol{\varepsilon}}^2, \sigma_{\mathbf{K}}^2 | \dot{\mathbf{X}}) \propto p(\dot{\mathbf{X}}|\mathbf{K})p(\mathbf{K}|\sigma_{\mathbf{K}}^2)p(\sigma_{\boldsymbol{\varepsilon}}^2)p(\sigma_{\mathbf{K}}^2). \quad (15)$$

In the Bayesian joint maximum *a posteriori* (JMAP) algorithm, (15) is maximized with respect to \mathbf{K} , $\sigma_{\boldsymbol{\varepsilon}}^2$ and $\sigma_{\mathbf{K}}^2$, to give the estimated parameters $\hat{\mathbf{K}}$, $\hat{\sigma}_{\boldsymbol{\varepsilon}}^2$ and $\hat{\sigma}_{\mathbf{K}}^2$. In the variational Bayesian approximation (VBA), the posterior in (15) is approximated by $q(\mathbf{K}, \sigma_{\boldsymbol{\varepsilon}}^2, \sigma_{\mathbf{K}}^2) = q_1(\mathbf{K})q_2(\sigma_{\boldsymbol{\varepsilon}}^2)q_3(\sigma_{\mathbf{K}}^2)$. The individual MAP estimates of each parameter are extracted by minimization of a Kullback-Leibler divergence $K = \int q \ln(q/p) d\mathbf{K} d\sigma_{\boldsymbol{\varepsilon}}^2 d\sigma_{\mathbf{K}}^2$. In both cases, an analytical solution is available, from which rapid Bayesian algorithms have been developed without the need for optimization [11, 12].

Application

To compare the traditional and Bayesian methods for dynamical system identification, we examine the Rössler system, the simplest low-dimensional dynamical system with chaotic behaviour, as a proxy for more complex fluid flow systems. This is described by the nonlinear equations [16]:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}) = [-y - z, x + ay, b + z(x - c)]^T, \quad (16)$$

using the parameter values [$a = 0.2, b = 0.2, c = 5.7$] to generate chaotic behaviour. The analyses were conducted in Matlab 2018a, with numerical integration by the ode45 function, using a time step of 0.02 and total time of 350. The position data \mathbf{X} were then modified by additive random noise, drawn from the standard normal distribution multiplied by a scaling parameter of 0.2. The regularization processes were then executed using a modified version of the published SINDy code and other utility functions [2], and modified forms of the JMAP and VBA functions [11, 12] using parameters $a_0 = 10^8$ and $b_0 = 10^{-8}$. For each Bayesian method, the covariance matrix of the posterior can be extracted, from which the variances (hence the standard deviations) of each coefficient k_{ij} can be extracted [11, 12].

Results

The calculated noisy data for the Rössler system are illustrated in Figures 1a-b, showing the raw \mathbf{X} data and the data with added noise. The calculated regularization results are then presented in Figures 2-4, respectively for the SINDy, JMAP and VBA methods. In each plot, the first graph shows the differences between the known and inferred coefficient values ($k_{ij} - \hat{k}_{ij}$) in each dimension, while the second graph shows the noisy and inferred time series, and their differences.

It is clear from these plots that the three methods were fairly similar in their choices of coefficients to recreate the Rössler system. The SINDy method provided coefficient values estimated to a resolution of 10^{-16} . In contrast, the JMAP and VBA give an estimated resolution of the order of 10^{-10} on all coefficients; the standard deviations estimated in these methods are shown as error bars in Figures 3a and 4a. The errors on the coefficients for the 1 and z terms are higher than the others, in each dimension, which accords with the fact that the Rössler system is nonlinear only in the z coordinate [16]. The Bayesian estimates provide a more realistic estimate of the inherent errors in the system identification method than given by the SINDy method.

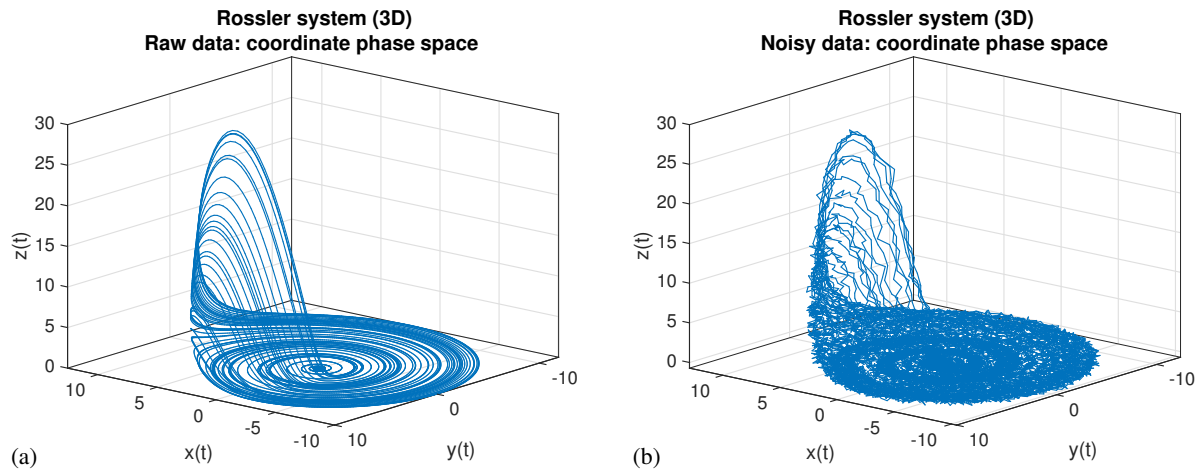


Figure 1. The Rössler system data \mathbf{X} : (a) raw data, and (b) data with added noise.

Conclusions

In this study, we examine regularization methods for dynamical system identification from time-series data. We first show that these can be reinterpreted within the framework of Bayesian inference using the MAP estimate, with the residual term identified with the likelihood distribution, and the regularization term identified with the prior. This provides a rational justification for the choice of residual and regularization terms, and furthermore provides an explicit form of the optimal regularization parameter. The Bayesian approach can also be extended to the full apparatus of the Bayesian inverse solution, for example to quantify the uncertainty in the model parameters, or even to explore the functional form of the posterior pdf.

Two Bayesian methods, JMAP and VBA, are then demonstrated by comparison to the SINDy regularization method, by application to the Rössler dynamical system. All three methods perform similarly; however the Bayesian methods enable the estimation of the model uncertainties, expressed in the form of variances (or standard deviations) of the model coefficients k_{ij} .

Acknowledgements

This research was funded by the Australian Research Council Discovery Projects grant DP140104402, and also supported by French sources including Institute Pprime, CNRS, Poitiers, France, and CentraleSupélec, Gif-sur-Yvette, France.

References

- [1] Brunton, SL, Proctor, JL & Kutz, JN (2016), Discovering governing equations from data by sparse identification of nonlinear dynamical systems, PNAS 113(15), 3932-3937.
- [2] Mangan, NM, Kutz, JN, Brunton, SL, Proctor, JL (2017), Model selection for dynamical systems via sparse regression and information criteria, Roy Soc Proc A 473: 20170009.
- [3] Rudy, SH, Brunton, SL, Proctor, JL, Kutz, JN (2017) Data-driven discovery of partial differential equations, Sci. Adv. 3: e1602614.
- [4] Tikhonov, AN (1963), Solution of incorrectly formulated problems and the regularization method, Doklady Akademii Nauk SSSR. 151: 501-504 (Russian).
- [5] Santosa, F & Symes, WW (1986), Linear inversion of band-limited reflection seismograms, SIAM J. Sci. Stat. Comp. 7 (4): 1307-1330.
- [6] Tibshirani, R (1996), Regression shrinkage and selection via the Lasso, J Royal Stat. Soc. B 58 (1): 267-288.
- [7] Zhang, L. & Schaeffer, H (2018) On the convergence of the SINDy algorithm, arXiv:1805.06445v1.
- [8] Brunton, SL, Brunton, BW, Proctor, JL, Kaiser, E, Kutz, JN (2016), Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control, PLOS One, 11(2): e0150171.
- [9] Brunton, SL, Brunton, BW, Proctor, JL, Kaiser, E, Kutz, JN (2017), Chaos as an intermittently forced linear system, Nature Comm 8: 19.
- [10] Taira, K., Brunton, SL, Dawson, STM, Rowley, CW, Colonius, T, McKeon, BJ, Schmidt, OT, Gordeyev, S, Theofilis, V, Ukeiley, LS, Modal analysis of fluid flows: an overview, AIAA Journal, 55(12), 4013-4041.
- [11] Mohammad-Djafari, A and Dumitru, M (2015), Bayesian sparse solutions to linear inverse problems with non-stationary noise with Student-t priors, Digital Signal Processing 47: 128-156.
- [12] Dumitru, M (2016), Approche bayésienne de l'estimation des composantes périodiques des signaux en chronobiologie, Thèse de Doctorat, L'Université Paris-Saclay préparée à L'Université Paris-Sud, France.
- [13] Mohammad-Djafari, A (2016), Approximate Bayesian computation for big data, Tutorial at MaxEnt 2016, July 10-15, Ghent, Belgium.
- [14] Teckentrup, A (2018), Introduction to the Bayesian approach to inverse problems, MaxEnt 2018, July 6, 2018, Alan Turing Institute, UK.
- [15] Niven, R.K., Mohammad-Djafari, A., Cordier, L., Abel, M., Quade, M. (2020), Bayesian identification of dynamical systems, MDPI Proceedings 2019, 33(1): 33.
- [16] Rössler, OE (1976). An Equation for Continuous Chaos. Physics Letters Vol. 57A (5): 397-398.

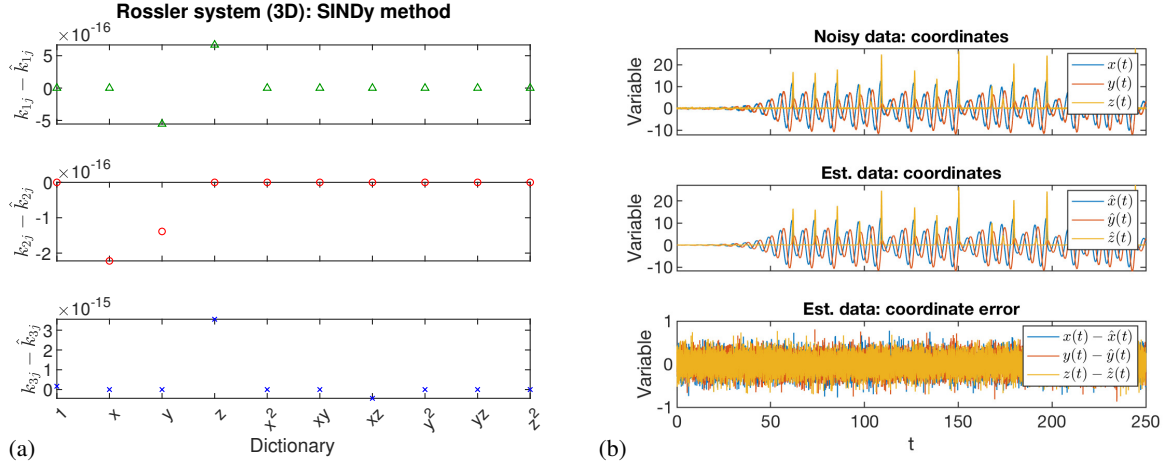


Figure 2. Output of SINDy regularization: (a) differences in predicted parameters $k_{ij} - \hat{k}_{ij}$, and (b) comparison of original and predicted time series \mathbf{X} .

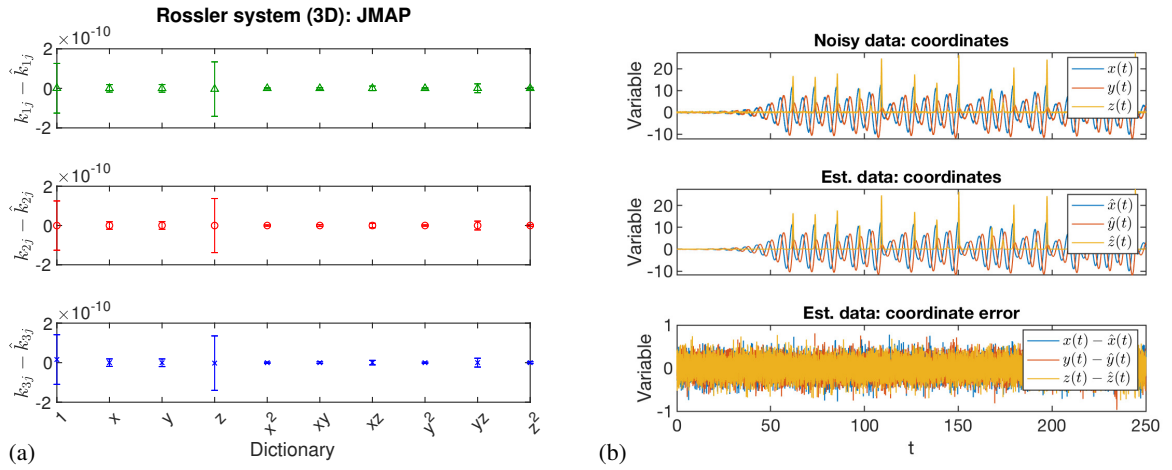


Figure 3. Output of JMAP regularization: (a) differences in predicted parameters $k_{ij} - \hat{k}_{ij}$ (error bars indicate inferred standard deviations from the posterior), and (b) comparison of original and predicted time series \mathbf{X} .

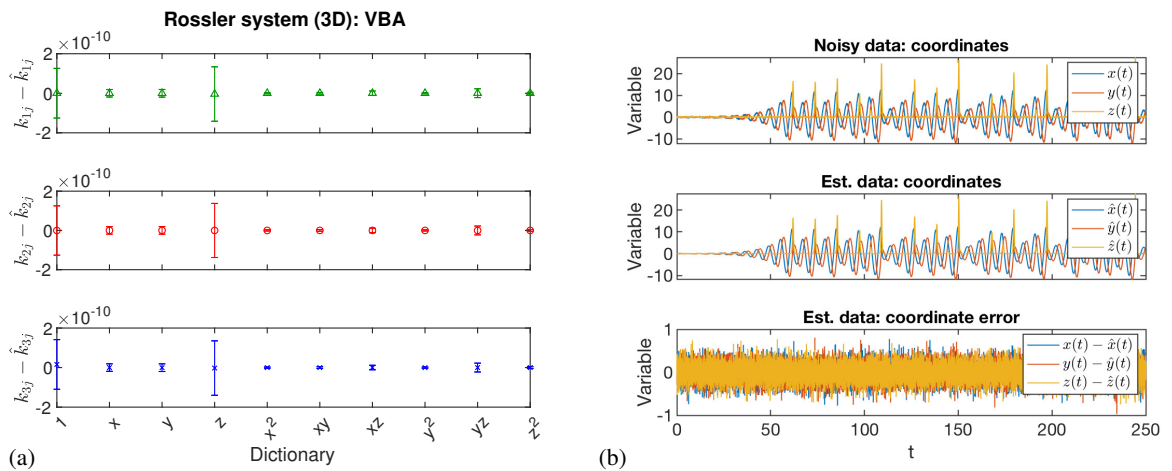


Figure 4. Output of VBA regularization: (a) differences in predicted parameters $k_{ij} - \hat{k}_{ij}$ (error bars indicate inferred standard deviations from the posterior), and (b) comparison of original and predicted time series \mathbf{X} .