

Article

# Bayesian Identification of Dynamical Systems

Robert K. Niven<sup>1</sup>  0000-0001-5781-1259, Ali Mohammad-Djafari<sup>2</sup>, Laurent Cordier<sup>3</sup>, Markus Abel<sup>4,5</sup> and Markus Quade<sup>4</sup>

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

<sup>2</sup> Laboratoire des signaux et systèmes (L2S), CentraleSupélec, Gif-sur-Yvette, France;

<sup>3</sup> Institut Pprime, Poitiers, France;

<sup>4</sup> Ambrosys GmbH, Potsdam, Germany;

<sup>5</sup> Institute for Physics and Astrophysics, University of Potsdam, Potsdam, Germany.

\* Correspondence: r.niven@adfa.edu.au

Version July 10, 2019 submitted to Proceedings

**Abstract:** Many inference problems relate to a dynamical system, as represented by  $dx/dt = f(x)$ , where  $x \in \mathbb{R}^n$  is the state vector and  $f$  is the (in general nonlinear) system function or model. Since the time of Newton, researchers have pondered the problem of *system identification*: how should the user accurately and efficiently identify the model  $f$  – including its functional family or parameter values – from discrete time-series data? For linear models, many methods are available including linear regression, the Kalman filter and autoregressive moving averages. For nonlinear models, an assortment of machine learning tools have been developed in recent years, usually based on neural networks or evolutionary computational methods, or various classification or order reduction schemes. The first group, while very useful, provide “black box” solutions which are not readily adaptable to new situations, while the second group necessarily involve the sacrificing of resolution to achieve order reduction. To address this problem, we propose the use of an inverse Bayesian method for system identification from time-series data. For a system represented by a set of basis functions, this is shown to be mathematically identical to Tikhonov regularization, albeit with a clear theoretical justification for the residual and regularization terms, respectively as the negative logarithms of the likelihood and prior functions. This insight justifies the choice of regularization method, and can also be extended to access the full apparatus of the Bayesian inverse solution. Two Bayesian methods, based on the joint maximum *a posteriori* (JMAP) and variational Bayesian approximation (VBA), are demonstrated for the Lorenz equation system with added Gaussian noise, in comparison to the regularization method of least squares regression with thresholding (the SINDy algorithm). The Bayesian methods are also used to estimate the variances of the inferred parameters, thereby giving the estimated model error, providing an important advantage of the Bayesian approach over traditional regularization methods.

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

## 1. Introduction

Many problems of inference involve a dynamical system, as represented by:

$$\frac{d}{dt}x(t) = f(x(t)) \quad (1)$$

where  $x \in \mathbb{R}^n$  is the observable state vector, a function of time  $t$  (and/or some other parameters), and  $f \in \mathbb{R}^n$  is the (in general nonlinear) system function or model. Given a set of discrete time series

28 data  $[x(t_1), x(t_2), x(t_3), \dots]$  from such a system, how should a user accurately and efficiently identify  
 29 the model  $f$ ? In dynamical systems theory, this is referred to as *system identification*, although for  
 30 many problems of known mathematical structure, it can be simplified into a problem of *parameter*  
 31 *identification*. The question then leads into deeper questions concerning the purpose of the prediction of  
 32  $f$ , and whether it is desired to reproduce a time series exactly, or more simply to extract its important  
 33 mathematical and/or statistical properties.

34 For linear models, many methods are available for identification of the dynamical system (1),  
 35 including linear regression, the Kalman filter and autoregressive moving averages. For nonlinear  
 36 models, an assortment of machine learning tools have been developed in recent years, usually based on  
 37 neural networks or evolutionary computational methods, or various classification or order reduction  
 38 schemes. The first group, while very useful, provide “black box” solutions which are not readily  
 39 adaptable to new situations, while the second group necessarily involve the sacrificing of resolution to  
 40 achieve order reduction.

41 Very recently, a number of researchers in dynamical and fluid flow systems have applied sparse  
 42 regression methods for system identification from time series data [e.g. 1–3]. The regression is used to  
 43 determine a matrix of coefficients which – when multiplied by a matrix of functional operations – can  
 44 be used to reproduce the time series. Such methods generally involve a regularization technique to  
 45 conduct the sparse regression. However, both the regularization term and its coefficient are usually  
 46 implemented in a heuristic or *ad hoc* manner, without much fundamental guidance on how they should  
 47 be selected for any particular dynamical system.

48 In this study, we present a Bayesian framework for the system identification (or parameter  
 49 identification) of a dynamical system using the Bayesian maximum *a posteriori* (MAP) estimate, which  
 50 is shown to be equivalent to a variant of Tikhonov regularization. This Bayesian reinterpretation  
 51 provides a rational justification for the choices of the residual and regularization terms, respectively as  
 52 the negative logarithms of the likelihood and prior functions. The Bayesian approach can be readily  
 53 extended to the full apparatus of the Bayesian inverse solution, for example to quantify the uncertainty  
 54 in the model parameters, or even to explore the functional form of the posterior. In this study, we  
 55 compare the prominent regularization method of least squares regression with thresholding (the  
 56 SINDy algorithm) to two Bayesian methods, by application to the Lorenz system with added Gaussian  
 57 noise. We demonstrate an advantage of the Bayesian methods, in their ability to calculate the variances  
 58 of the inferred parameters, thereby giving the estimated model errors.

## 59 2. Theoretical Foundations

In recent years, a number of researchers have implemented sparse regression methods for the  
 system identification of a variety of dynamical systems [e.g. 1–3]. The method proceeds from a recorded  
 time series, which for  $m$  time steps of an  $n$ -dimensional parameter  $x$  is assembled into the  $m \times n$  matrix:

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

and similarly for the time derivative:

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

The user then chooses an alphabet of  $c$  functions, which are applied to  $\mathbf{X}$  to populate a  $m \times c$  matrix  
 library, for example of the form:

$$\Theta(\mathbf{X}) = [1 \quad \mathbf{X} \quad \mathbf{X}^2 \quad \mathbf{X}^3 \quad \dots \quad \sin(\mathbf{X}) \quad \cos(\mathbf{X}) \quad \dots] \quad (4)$$

in this case based on polynomial and trigonometric functions. The time series data for the dynamical system (1) are then analyzed by the matrix product:

$$\dot{X} = \Theta(X)\Xi \quad (5)$$

in which  $\Xi$  is a  $c \times n$  matrix of coefficients  $\xi_{ij} \in \mathbb{R}$ . The matrix  $\Xi$  is commonly computed by inversion of (5) using sparse regression. This generally involves a minimization equation of the form:

$$\hat{\Xi} = \arg \min_{\Xi} J(\Xi) \quad (6)$$

where  $\hat{\cdot}$  indicates an inferred value, based on an objective function consisting of residual and regularization terms:

$$J(\Xi) = \|\dot{X} - \Theta(X)\Xi\|_{\beta}^{\alpha} + \lambda \|\Xi\|_{\gamma}^{\alpha} \quad (7)$$

where  $\|\cdot\|_p$  is the  $p$  norm,  $\lambda \in \mathbb{R}$  is the regularization coefficient and  $\alpha, \beta, \gamma \in \mathbb{R}$  are constants. For dynamical system identification, (6)-(7) have been variously implemented with  $\alpha \in \{1, 2\}$ ,  $\beta = 2$  and  $\gamma \in \{0, [1, 2]\}$  [e.g. 2–6]. Instead of (7), to enforce a sparse solution, some authors have implemented least squares regression with iterative thresholding, known as the sparse identification of nonlinear dynamics (SINDy) method [1]:

$$J(\Xi) = \|\dot{X} - \Theta(X)\Xi\|_2^2 \quad \text{with} \quad |\xi_{ij}| \geq \lambda, \quad \forall \xi_{ij} \in \Xi \quad (8)$$

60 This has been shown to converge to (7) with  $\alpha = \beta = 2$  and  $\gamma = 0$  [7]. Other authors have implemented  
 61 an objective function containing an information criterion, to preferentially select models with fewer  
 62 parameters [2]. The above methods have been shown to have strong connections to the mathematical  
 63 methods of singular value decomposition (SVD), dynamic mode decomposition (DMD) and Koopman  
 64 analysis using various Koopman operators [e.g. 8–10].

In the Bayesian approach to this problem [e.g. 11–13], it is recognized that instead of (5), the time series decomposition should be written explicitly as:

$$\dot{X} = \Theta(X)\Xi + \epsilon \quad (9)$$

where  $\epsilon$  is a noise or error term, representing the uncertainty in the measurement data. The variables  $\dot{X}$ ,  $X$ ,  $\Xi$  and  $\epsilon$  are considered to be probabilistic, each represented by a probability density function (pdf) defined over their applicable domain. Instead of trying to invert (9), the Bayesian considers the posterior probability of  $\Xi$  given the data, as given by Bayes' rule:

$$p(\Xi|\dot{X}) = \frac{p(\dot{X}|\Xi)p(\Xi)}{p(\dot{X})} \propto p(\dot{X}|\Xi)p(\Xi) \quad (10)$$

The simplest Bayesian method is to consider the maximum *a posteriori* (MAP) estimate of  $\Xi$ , given by maximization of (10):

$$\hat{\Xi} = \arg \max_{\Xi} p(\Xi|\dot{X}) \quad (11)$$

For greater fidelity, it is convenient to consider the logarithmic maximum instead of (11), hence from (10):

$$\hat{\Xi} = \arg \max_{\Xi} [\ln p(\Xi|\dot{X})] = \arg \max_{\Xi} [\ln p(\dot{X}|\Xi) + \ln p(\Xi)] \quad (12)$$

If we now make the simple assumption of unbiased multivariate Gaussian noise with covariance matrix  $\Gamma$ , we have:

$$p(\epsilon|\Xi) = \mathcal{N}(0, \Gamma) = \frac{\exp\left(-\frac{1}{2}\epsilon^T \Gamma^{-1} \epsilon\right)}{\sqrt{(2\pi)^n \det \Gamma}} \quad (13)$$

where  $\det$  is the determinant. The numerator can be written as [13]

$$p(\epsilon|\Xi) \propto \exp\left(-\frac{1}{2}\|\epsilon\|_{\Gamma^{-1}}^2\right) \quad (14)$$

where  $\|\epsilon\|_A^2 = \epsilon^\top A \epsilon$  is the norm defined by the  $A$  bilinear product. From (9), this gives the likelihood

$$p(\dot{X}|\Xi) \propto \exp\left(-\frac{1}{2}\|\dot{X} - \Theta(X)\Xi\|_{\Gamma^{-1}}^2\right) \quad (15)$$

If we also assign a multivariate Gaussian prior with covariance matrix  $\Sigma$

$$p(\Xi) = \mathcal{N}(0, \Sigma) \propto \exp\left(-\frac{1}{2}\|\Xi\|_{\Sigma^{-1}}^2\right) \quad (16)$$

then the MAP estimator (12) becomes [13]:

$$\begin{aligned} \hat{\Xi} &= \arg \max_{\Xi} \left[ \ln \exp\left(-\frac{1}{2}\|\dot{X} - \Theta(X)\Xi\|_{\Gamma^{-1}}^2\right) + \ln \exp\left(-\frac{1}{2}\|\Xi\|_{\Sigma^{-1}}^2\right) \right] \\ &= \arg \max_{\Xi} \left[ -\frac{1}{2}\|\dot{X} - \Theta(X)\Xi\|_{\Gamma^{-1}}^2 - \frac{1}{2}\|\Xi\|_{\Sigma^{-1}}^2 \right] \\ &= \arg \min_{\Xi} \left[ \|\dot{X} - \Theta(X)\Xi\|_{\Gamma^{-1}}^2 + \|\Xi\|_{\Sigma^{-1}}^2 \right] \end{aligned} \quad (17)$$

65 We see that the Bayesian MAP provides a minimization formula based on an objective function, which  
66 is remarkably similar to that used in the regularization method (6)-(7). Indeed, for isotropic variances  
67 of the noise  $\Gamma = \sigma_\epsilon^2 I$  and prior  $\Sigma = \sigma_\Xi^2 I$ , where  $I$  is the identity matrix, (17) reduces to the common  
68 regularization formula (6)-(7) with  $\alpha = \beta = \gamma = 2$  and  $\lambda = \sigma_\epsilon^2/\sigma_\Xi^2$  [11].

In Bayesian inference, any additional parameters can also be incorporated into the inferred posterior pdf. In the present study, the covariance matrices  $\Gamma$  of the noise in (14) and  $\Sigma$  of the prior in (16) are unknown. It is desirable to determine these directly from the Bayesian inversion process. Using the above simple model of isotropic variances, the posterior can be written as:

$$p(\Xi, \sigma_\epsilon^2, \sigma_\Xi^2 | \dot{X}) \propto p(\dot{X}|\Xi) p(\Xi|\sigma_\Xi^2) p(\sigma_\epsilon^2) p(\sigma_\Xi^2) \quad (18)$$

69 In the Bayesian joint maximum *a posteriori* (JMAP) algorithm, (18) is maximized with respect to  $\Xi, \sigma_\epsilon^2$   
70 and  $\sigma_\Xi^2$ , to give the estimated parameters  $\hat{\Xi}, \hat{\sigma}_\epsilon^2$  and  $\hat{\sigma}_\Xi^2$ . In the variational Bayesian approximation  
71 (VBA), the posterior in (18) is approximated by  $q(\Xi, \sigma_\epsilon^2, \sigma_\Xi^2) = q_1(\Xi)q_2(\sigma_\epsilon^2)q_3(\sigma_\Xi^2)$ . The individual  
72 MAP estimates of each parameter are then calculated iteratively, using a Kullback-Leibler divergence  
73  $K = \int q \ln(q/p) d\Xi d\sigma_\epsilon^2 d\sigma_\Xi^2$  as the convergence criterion.

### 74 3. Method

To compare the traditional and Bayesian methods for dynamical system identification, a number of time series of the Lorenz system were generated and analyzed by several regularization methods, including SINDy, JMAP and VBA. The Lorenz system is described by the nonlinear equation [14]:

$$\frac{dx}{dt} = f(x) = [\sigma(y - x), x(\rho - z) - y, xy - \beta z]^\top \quad (19)$$

75 with parameter values  $[\sigma, \rho, \beta]$  commonly assigned to  $[10, \frac{8}{3}, 28]$  to generate chaotic behavior with  
76 a strange attractor. The analyses were conducted in Matlab 2018a on a MacBook Pro with 2.8 GHz  
77 Intel Core i7, with numerical integration by the ode45 function, using a time step of 0.01 and total  
78 time of 100. The calculated position data  $X$  were then augmented by additive random noise, drawn  
79 from the standard normal distribution multiplied by a scaling parameter of 0.2. The regularization  
80 processes were then executed using a modified version of the published SINDy code and other utility

81 functions [2], and modified forms of the JMAP and VBA functions implemented previously [11]. For  
 82 comparisons, the inferred parameters were then used to recalculate the time series and derivatives by  
 83 a further function call. In the Bayesian algorithms, the estimated variances of the parameters and the  
 84 prior were also calculated, assuming inverse gamma distributions for the variance priors; for JMAP  
 85 this has an analytical solution, while for VBA the solution is found iteratively [11].

#### 86 4. Results

87 The calculated noisy data for the Lorenz system are illustrated in Figures 1a-b, respectively for  
 88 the parameter values and their derivatives. The calculated regularization results are then presented  
 89 in Figures 2-4, respectively for the SINDy, JMAP and VBA methods. In each of these plots, the first  
 90 subplot illustrates the difference in each inferred parameter (i.e.,  $\zeta_{ij} - \hat{\zeta}_{ij}$ ), while the second subplot  
 91 gives the inferred time series of the parameters  $X$ , showing the noisy time series, the inferred series  
 92 and their differences.

93 As evident in these plots, the three methods were approximately as effective in selection of the  
 94 coefficients to recreate the Lorenz system. Of the other regularization methods published by [2], the  
 95 iterative hard thresholding least squares and orthogonal matching pursuit also performed well, while  
 96 the LASSO algorithm was unsuccessful for any system examined.

97 As noted, the two Bayesian methods also provided the variances of the predicted parameters,  
 98 shown in Figures 3a and 4a as error bars corresponding to the standard deviations. These calculations  
 99 indicate the inferred parameter errors to be significant, for example  $\pm 0.0116$  in the coefficient of  $x$  in all  
 100 three series predicted by JMAP, and  $\pm 0.0149$  predicted by VBA. These values are substantially higher  
 101 than the error of at most  $|1.8 \times 10^{-14}|$  suggested by the SINDy regularization, giving a more realistic  
 102 estimate of the inherent errors in the system identification method.

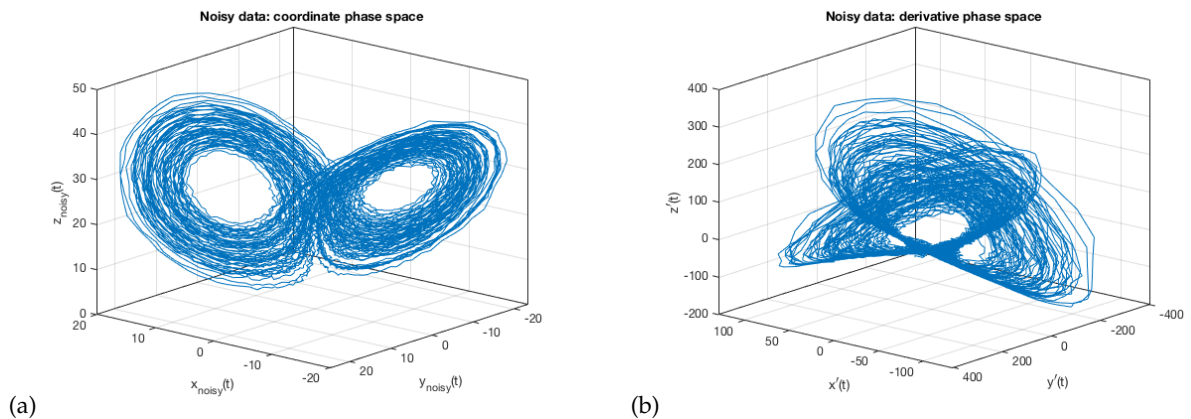


Figure 1. Calculated noisy data for the Lorenz system: (a) parameters  $X$ , and (b) derivatives  $\dot{X}$ .

#### 103 5. Conclusions

104 We examine the problem of system identification of a dynamical system, represented by a  
 105 nonlinear equation system  $dx/dt = f(x)$ , from discrete time series data. For this, we present a Bayesian  
 106 inference framework based on the Bayesian maximum *a posteriori* (MAP) estimate, which for the  
 107 assumption of Gaussian likelihood and prior functions, is shown to be equivalent to a variant of  
 108 Tikhonov regularization. This Bayesian reinterpretation provides a clear theoretical justification for  
 109 the choices of the residual and regularization terms, respectively as the negative logarithms of the  
 110 likelihood and prior functions. The Bayesian approach is readily extended to the full apparatus of the

111 Bayesian inverse solution, for example to quantify the uncertainty in the model parameters, or even to  
112 explore the functional form of the posterior pdf.

113 In this study, we compare the regularization method of least squares regression with thresholding  
114 (the SINDy algorithm) to two Bayesian methods JMAP and VBA, by application to the Lorenz system  
115 with added Gaussian noise. The Bayesian methods are shown to perform almost as effectively as  
116 SINDy for parameter estimation and reconstruction of the Lorenz time series. More importantly,  
117 the Bayesian methods also provide the variances – hence the standard deviations – of the inferred  
118 parameters, thereby giving a mathematical estimate of the system identification error. This is an  
119 important advantage of the Bayesian approach over traditional regularization methods.

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

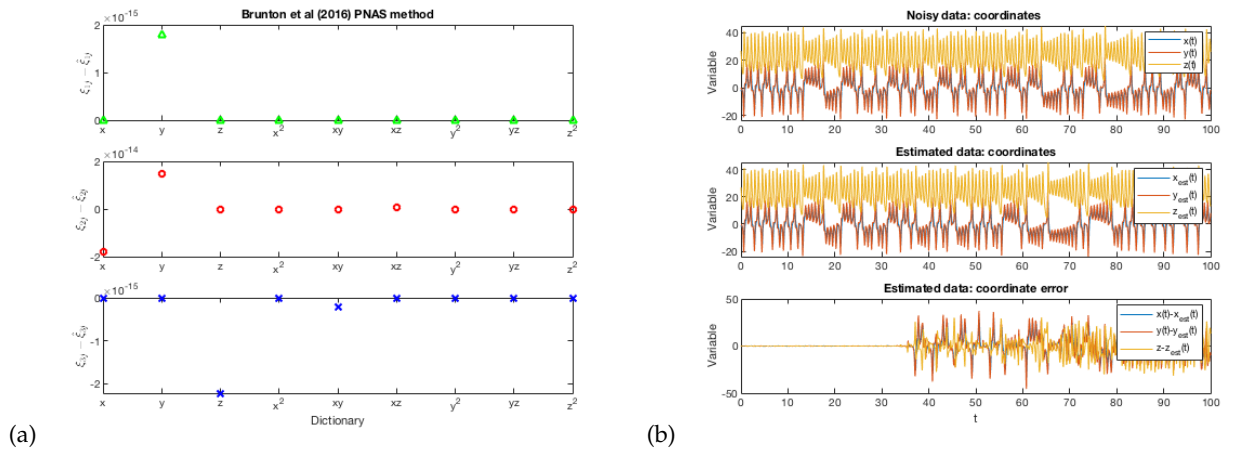
123 **Conflicts of Interest:** The author declares no conflict of interest. The funders had no role in the design of the  
124 study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to  
125 publish the results.

## 126 References

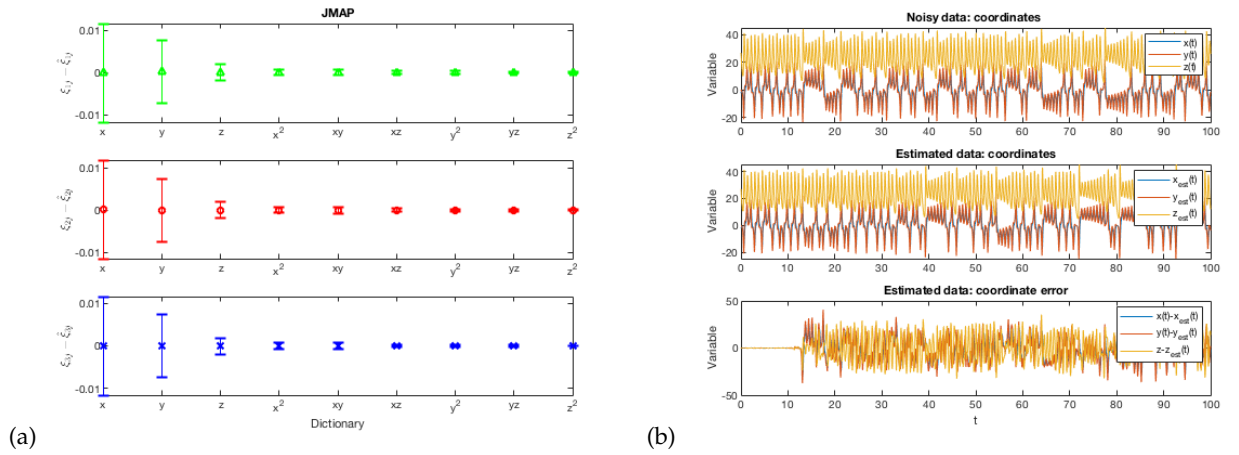
- 127 1. Brunton, SL, Proctor, JL & Kutz, JN (2016), Discovering governing equations from data by sparse  
128 identification of nonlinear dynamical systems, PNAS 113(15), 3932-3937.
- 129 2. Mangan, NM, Kutz, JN, Brunton, SL, Proctor, JL (2017), Model selection for dynamical systems via sparse  
130 regression and information criteria, Roy Soc Proc A 473: 20170009.
- 131 3. Rudy, SH, Brunton, SL, Proctor, JL, Kutz, JN (2017) Data-driven discovery of partial differential equations,  
132 Sci. Adv. 3: e1602614.
- 133 4. Tikhonov, AN (1963), Solution of incorrectly formulated problems and the regularization method, Doklady  
134 Akademii Nauk SSSR. 151: 501-504 (Russian).
- 135 5. Santosa, F & Symes, WW (1986), Linear inversion of band-limited reflection seismograms, SIAM J. Sci. Stat.  
136 Comp. 7 (4): 1307-1330.
- 137 6. Tibshirani, R (1996), Regression shrinkage and selection via the Lasso, J Royal Stat. Soc. B 58 (1): 267-288.
- 138 7. Zhang, L. & Schaeffer, H (2018) On the convergence of the SINDy algorithm, arXiv:1805.06445v1.
- 139 8. Brunton, SL, Brunton, BW, Proctor, JL, Kaiser, E, Kutz, JN (2016), Koopman invariant subspaces and finite  
140 linear representations of nonlinear dynamical systems for control, PLOS One, 11(2): e0150171.
- 141 9. Brunton, SL, Brunton, BW, Proctor, JL, Kaiser, E, Kutz, JN (2017), Chaos as an intermittently forced linear  
142 system, Nature Comm 8: 19.
- 143 10. Taira, K., Brunton, SL, Dawson, STM, Rowley, CW, Colonius, T, McKeon, BJ, Schmidt, OT, Gordeyev, S,  
144 Theofilis, V, Ukeiley, LS, Modal analysis of fluid flows: an overview, AIAA Journal, 55(12), 4013-4041.
- 145 11. Mohammad-Djafari, A (2015), Inverse problems in signal and image processing and Bayesian inference  
146 framework: from basic to advanced Bayesian computation, Scube seminar, L2S, CentraleSupélec, March 27,  
147 2015.
- 148 12. Mohammad-Djafari, A (2016), Approximate Bayesian computation for big data, Tutorial at MaxEnt 2016,  
149 July 10-15, Ghent, Belgium.
- 150 13. Teckentrup, A (2018), Introduction to the Bayesian approach to inverse problems, MaxEnt 2018, July 6, 2018,  
151 Alan Turing Institute, UK.
- 152 14. Lorenz, EN (1963). Deterministic nonperiodic flow, J. Atmospheric Sci, 20 (2): 130-141.

153 © 2019 by the authors. Submitted to *Proceedings* for possible open access publication  
154 under the terms and conditions of the Creative Commons Attribution (CC BY) license  
155 (<http://creativecommons.org/licenses/by/4.0/>).

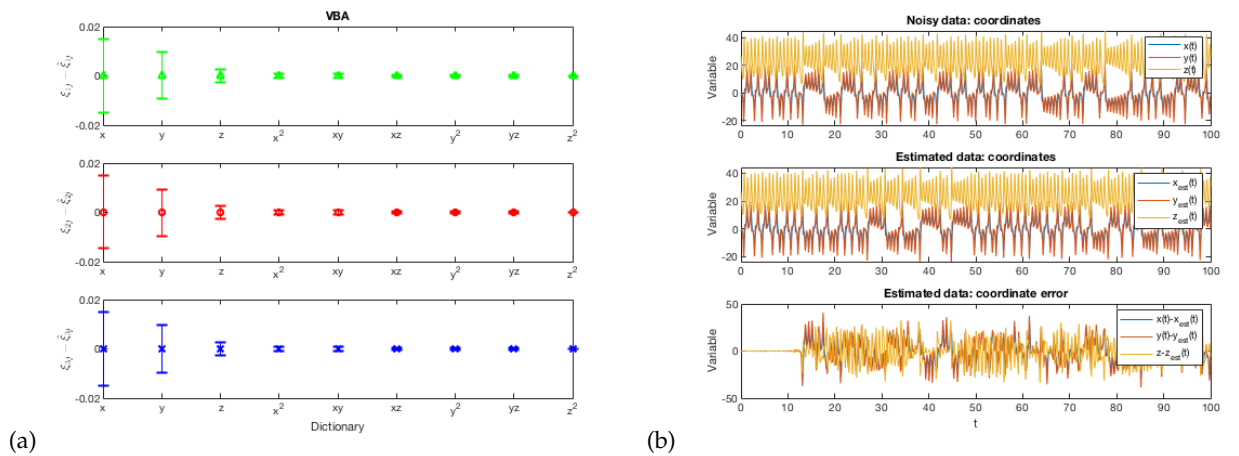




**Figure 2.** Output of SINDy regularization: (a) differences in predicted parameters  $\xi_{ij} - \hat{\xi}_{ij}$ , and (b) comparison of original and predicted time series  $X$ .



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



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