Data-Driven Approach to Nonlinear Dynamic Equation Discovery

Joshua S. North   Erin M. Schliep   Christopher K. Wikle

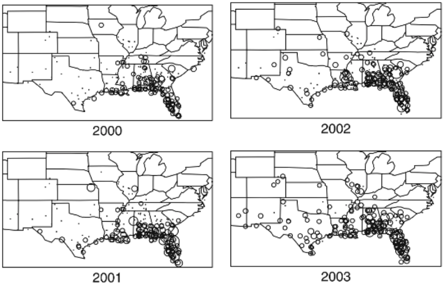
University of Missouri Columbia
Department of Statistics

August 9, 2021

Motivation

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(\delta \frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left(\delta \frac{\partial u}{\partial y}\right)$$

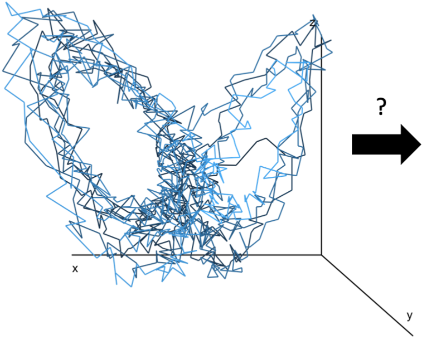
- Differential equations (DE) in models are based on an understanding of the governing dynamics of the physical systems
- Approximate the dynamics
- e.g., reaction diffusion for the spread of avian species (Wikle, 2003; Hooten and Wikle, 2008)



Spread of Eurasian Collared-Dove across the United States (Hooten and Wikle, 2008).

Motivation

- Modeling physical processes using DE, while generally sufficient at capturing system dynamics, suffer in that they are only an approximation of the true physical process (Holmes et al., 1994)
- Instead of modeling the process using DE, we want to **discover the governing equation(s) that define dynamic system**



$$\frac{dx}{dt} = \sigma \cdot (y - x)$$
$$\frac{dy}{dt} = x \cdot (\rho - z) - y$$
$$\frac{dz}{dt} = xy - \beta z$$

What Has Been Done

- Originally Bongard and Lipson (2007); Schmidt and Lipson (2009) using symbolic regression
  - Able to discover dynamics
  - Symbolic regression is computationally expensive
- Brunton et al. (2016) shift the focus of dynamic system discovery to sparse identification, proposing Sparse Identification of Nonlinear Dynamics (SINDy)
  - SINDy involves three major steps: (1) numerical differentiation and denoising, (2) determining the candidate functions, termed the "feature library", and (3) sparse regression
  - Extensions include PDEs (Rudy et al., 2017, 2019), stochastic processes (Boninsegna et al., 2018), numerical improvement (Schaeffer, 2017; Schaeffer et al., 2018; Lagergren et al., 2020), and improved uncertainty quantification (Zhang and Lin, 2018; Niven et al., 2020)
  - Developed into a Python package (pysindy; de Silva et al., 2020)

Our Approach

- Bayesian hierarchical modeling (BHM) approach to data-driven discovery of dynamic equations
    - Compartmentalize uncertainty
    - Incorporate process dependence
    - Borrow dependence across processes
- Compute derivatives analytically through a basis expansion
    - Make inference on the derivative when only the process is observed
    - Forces the latent process to be smooth
- Feature library
    - Impart system dynamics
- Missing/imperfect data
    - BHM allows for complete latent space

Dynamic System

Consider the dynamic system

$$\frac{d}{dt}\mathbf{x}_t = \dot{\mathbf{x}}_t = M(\mathbf{x}_t), \tag{1}$$

where the vector $\mathbf{x}_t \in \mathbb{R}^n$ denotes the realization of the system at time $t = 1, ..., T$, and the function $M(\cdot)$ represents the, potentially nonlinear, evolution function.

Reparameterizing Eqn. 1, and accounting for potential stochastic forcing,

$$\dot{\mathbf{x}}_t = \mathbf{M}\mathbf{f}(\mathbf{x}_t) + \boldsymbol{\eta}_t, \tag{2}$$

where $\mathbf{M}$ is a $n \times p$ *sparse* matrix of coefficients, $\mathbf{f}(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^p$ is a vector-valued nonlinear transformation function, and $\boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q})$ is a mean zero Gaussian process.

Dynamic System

In general, only the process $\mathbf{X} = \{\mathbf{x}_t\}_{t=1,\ldots,T}$ is observed and measured. To make inference on the derivative of the process, we decompose our observed system using temporal basis functions, and use the basis functions to analytically obtain the derivatives. Let $\mathbf{X}' = \mathbf{\Phi}\mathbf{A}$ and $\dot{\mathbf{X}}' = \dot{\mathbf{\Phi}}\mathbf{A}$, where $\mathbf{\Phi}$ and $\dot{\mathbf{\Phi}}$ are $T \times p_a$ matrices of the basis functions and derivative of the basis functions, respectively, and $\mathbf{A}$ is a $p_a \times n$ matrix of basis coefficients. This results in the process equations,

$$\mathbf{A}'\dot{\phi}_t' = \mathbf{M}\mathbf{f}(\mathbf{A}'\phi_t') + \boldsymbol{\eta}_t,$$

where $\mathbf{M}$ and $\mathbf{f}(\cdot)$ are the same as in Eqn. 2, but now $\boldsymbol{\eta}_t$ accounts for basis truncation error and stochastic forcing.

Dynamic Model

For time points $t = 1, ..., T$, our general model is

$$\begin{aligned} \mathbf{y}_t &= \mathbf{H}_t \mathbf{A}' \phi_t' + \boldsymbol{\epsilon}_t \\ \mathbf{A}' \dot{\phi}_t' &= \mathbf{M}\mathbf{f}(\mathbf{A}' \phi_t') + \boldsymbol{\eta}_t \end{aligned} \quad (3)$$

- $\mathbf{y}_t \in \mathbb{R}^m$ is the observed process
- $\mathbf{H}_t$ is the $m \times n$ matrix mapping the latent to observed process
- $\mathbf{A}' \phi_t' = \mathbf{x}_t \in \mathbb{R}^n$ is the latent observation vector
- $\mathbf{f}(\cdot) : \mathbb{R}^n \to \mathbb{R}^p$ is the nonlinear function, $n << p$,
- $\mathbf{M}$ is the $n \times p$ coefficient matrix
- $\boldsymbol{\epsilon}_t \sim N_m(\mathbf{0}, \mathbf{R})$ is the measurement error
- $\boldsymbol{\eta}_t \sim N_n(\mathbf{0}, \mathbf{Q})$ is the process error

Parameter Specification

- $\mathbf{R} = diag(\sigma^2_{r_1}, ..., \sigma^2_{r_m})$, with $\sigma^2_{r_1}, ..., \sigma^2_{r_m} \sim$ Half-t$(2, 1e^5)$ (Huang and Wand, 2013)
    - Non-informative prior
    - Assumes measurement noise is independent
- $\mathbf{Q} \sim$ matrix Half-t$(2, 1e^5)$
    - Non-informative prior
    - Accounts for process dependence structure
- **M** SSVS (George et al., 1993)
    - Inclusion probability
    - Variable selection
- **A** Elastic Net (Li and Lin, 2010)
    - Coefficient shrinkage
    - Stochastic gradient

Importance of Basis Expansion

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \boldsymbol{\epsilon}_t$$
$$\dot{\mathbf{x}}_t = \mathbf{M}\mathbf{f}(\mathbf{x}_t) + \boldsymbol{\eta}_t$$

- Approximate $\dot{\mathbf{x}}_t$ (e.g., finite difference)
- Update $[\dot{\mathbf{x}}_t|\cdot]$ and $[\mathbf{x}_t|\cdot]$
  - Potentially very complex
  - Dependence between $[\dot{\mathbf{x}}_t|\cdot]$ and $[\mathbf{x}_t|\cdot]$
  - Costly update step $O(T)$

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{A}' \phi_t' + \boldsymbol{\epsilon}_t$$
$$\mathbf{A}' \dot{\phi}_t' = \mathbf{M}\mathbf{f}(\mathbf{A}'\phi_t') + \boldsymbol{\eta}_t$$

- Analytic derivative $\phi_t \rightarrow \dot{\phi}_t$
- Update $[\mathbf{A}|\cdot]$
  - Stochastic gradient descent (Mandt et al., 2016)
  - Automatic differentiation
  - Estimates derivative and system jointly
  - $O(N)$ update step, $N << T$

Simulations and Examples

Simulations - 4th-order Runge-Kutta

1. Lotka-Volterra System ($\Delta t = 0.05, t = 0, ..., 50$) - no noise, measurement noise
2. Lorenz-63 Attractor ($\Delta t = 0.01, t = 0, ..., 10$) - no noise, measurement noise, missing data

Examples

1. Hare-Lynx Predator Prey - Hudson Bay Company circa 1845-1935

Lotka-Volterra System

$$dx/dt = \alpha x - \beta xy \qquad \Rightarrow \qquad \alpha = 1.1 \qquad \Rightarrow \qquad \mathbf{X}_t = [x_t, y_t] \quad \text{no noise}$$
$$dy/dt = \delta xy - \gamma y \qquad\qquad \beta = 0.4$$
$$\delta = 0.1 \qquad\qquad \mathbf{Z}_t = \mathbf{X}_t + N(\mathbf{0}, 0.5\mathbf{I}_2) \quad \text{noise}$$
$$\gamma = 0.4$$

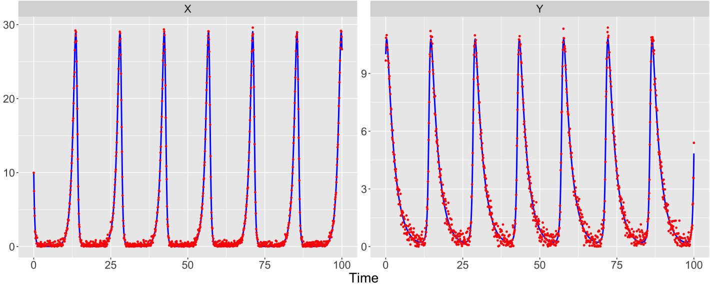


Figure: Data simulated from Lotka Volterra system without noise ($\mathbf{X}_t$, blue) and with measurement noise ($\mathbf{Z}_t$, red).

Lotka-Volterra System

System	Truth
dx/dt	$1.1x - 0.4xy$
dy/dt	$-0.4y + 0.1xy$

Table: True solution with correct parameter values.

System	**X**	**Z**
dx/dt	$1.105x - 0.454xy$	$-0.269 + 1.317x - 0.392xy + 0.033yy$
dy/dt	$-0.4y + 0.113xy$	$0.13 - 0.482y + 0.151xy$

Table: Recovered equations for the Lotka-Volterra simulation for data simulated with no noise (**X**, left) and with measurement noise (**Z**, right). All parameter values are the point-wise posterior mean estimates and rounded to three significant figures. Library included polynomials up to the third order, all possible interactions, and an intercept.

Lorenz-63 Attractor

$$dx/dt = \sigma \cdot (y - x)$$
$$dy/dt = x \cdot (\rho - z) - y$$
$$dz/dt = xy - \beta z$$

$$\Rightarrow \quad \sigma = 10, \quad \rho = 28, \quad \beta = 8/3 \quad \Rightarrow$$

$\mathbf{X}_t = [x_t, y_t, z_y]$ no noise

$\mathbf{Z}_t = \mathbf{X}_t + N(\mathbf{0}, \mathbf{I}_3)$ noise

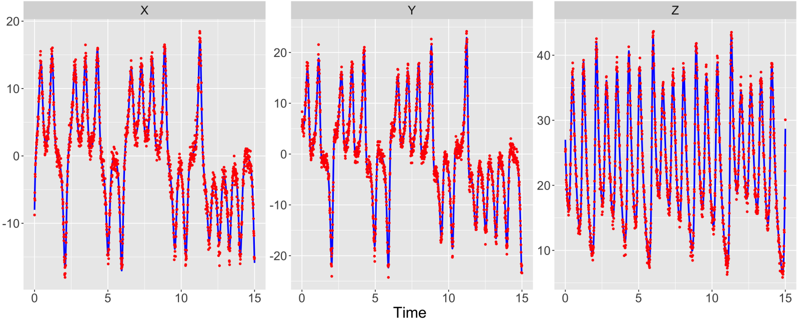


Figure: Data simulated from Lorenz-63 without noise ($\mathbf{X}_t$, blue) and with measurement noise ($\mathbf{Z}_t$, red).

Lorenz-63 Attractor

Randomly remove 5% of data from each system $\Rightarrow \mathbf{ZM}_t$

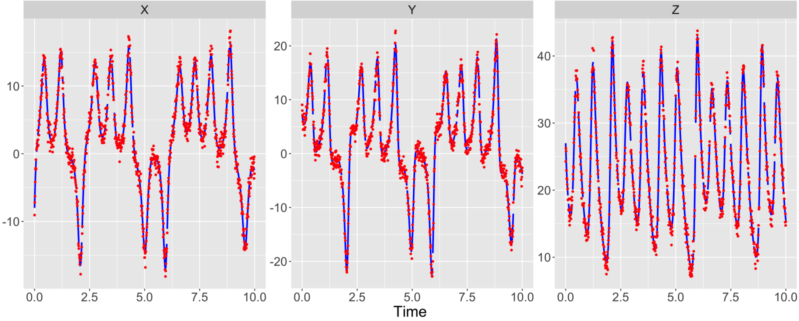


Figure: Data simulated from Lorenz-63 with measurement noise and missing data ($\mathbf{ZM}_t$, red). The blue lines are to help show where there is missing data.

Lorenz-63 Attractor

System	Truth	**X**
dx/dt	$-10x + 10y$	$-9.999x + 10y$
dy/dt	$28x - 1y - 1xz$	$27.997x - 0.998y - 1xz$
dz/dt	$-2.667z + 1xy$	$-2.667z + 1xy$

(a) True solution (left) and recovered solution for data simulated with no noise (**X**, right).

System	**Z**	**ZM**
$dx/dt =$	$-8.892x + 9.381y$	$-9.103x + 9.697y$
$dy/dt =$	$26.571x - 0.883xz - 0.086yz$	$24.078x - 0.816xz$
$dz/dt =$	$-2.67z + 0.99xy$	$-2.585z + 0.963xy$

(b) Recovered solutions for data simulated with measurement noise (**Z**, left) and with measurement noise and missing data (**ZM**, right).

Table: Recovered equations for the Lorenz-63 simulations. All parameter values are the point-wise posterior mean estimates and rounded to three significant figures. Library included polynomials up to the third order, all possible interactions, and an intercept.

Hare-Lynx Predator Prey

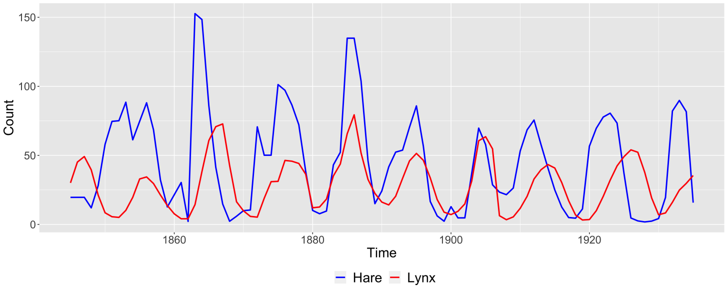


Figure: Canadian Lynx and Snowshoe Hare data set[1]

System	
dH/dt	$1.373 + 2.326H - 0.667L - 0.086HL$
dL/dt	$-0.75 - 5.452L + 0.154HL$

Table: Recovered parameters from the Hare-Lynx System. Library included polynomials up to the third order, all possible interactions, and an intercept.

[1] https://tuvalabs.com/datasets/lynx_and_snowshoe_hare_in_canada/activities

References I

Bongard, J. and Lipson, H. (2007). Automated reverse engineering of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 104(24):9943–9948.

Boninsegna, L., Nüske, F., and Clementi, C. (2018). Sparse learning of stochastic dynamical equations. *The Journal of Chemical Physics*, 148(24):241723.

Brunton, S. L., Proctor, J. L., and Kutz, J. N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937.

de Silva, B., Champion, K., Quade, M., Loiseau, J.-C., Kutz, J., and Brunton, S. (2020). PySINDy: A Python package for the sparse identification of nonlinear dynamical systems from data. *Journal of Open Source Software*, 5(49):2104.

References II

George, E. I., McCulloch, R. E., George, E. I., and McCulloch, R. E. (1993). Variable Selection Via Gibbs Sampling. *Journal of the American Statistical Association*, 88(423):881–889.

Holmes, E. E., Lewis, M. A., Banks, J. E., and Veit, R. R. (1994). Partial Differential Equations in Ecology: Spatial Interactions and Population Dynamics. *Ecology*, 75(1):17–29.

Hooten, M. B. and Wikle, C. K. (2008). A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian Collared-Dove. *Environmental and Ecological Statistics*, 15(1):59–70.

Huang, A. and Wand, M. P. (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis*, 8(2):439–452.

References III

Lagergren, J. H., Nardini, J. T., Michael Lavigne, G., Rutter, E. M., and Flores, K. B. (2020). Learning partial differential equations for biological transport models from noisy spatio-temporal data. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 476(2234):20190800.

Li, Q. and Lin, N. (2010). The Bayesian elastic net. *Bayesian Analysis*, 5(1):151–170.

Mandt, S., Hoffman, M., and Blei, D. (2016). A Variational Analysis of Stochastic Gradient Algorithms. *Proceedings of The 33rd International Conference on Machine Learning*, 48:354–363.

Niven, R., Mohammad-Djafari, A., Cordier, L., Abel, M., and Quade, M. (2020). Bayesian Identification of Dynamical Systems. *Proceedings*, 33(1):33.

References IV

Rudy, S., Alla, A., Brunton, S. L., and Kutz, J. N. (2019). Data-Driven Identification of Parametric Partial Differential Equations. *SIAM Journal on Applied Dynamical Systems*, 18(2):643–660.

Rudy, S. H., Brunton, S. L., Proctor, J. L., and Kutz, J. N. (2017). Data-driven discovery of partial differential equations. *Science Advances*, 3(4):e1602614.

Schaeffer, H. (2017). Learning partial differential equations via data discovery and sparse optimization. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2197):20160446.

Schaeffer, H., Tran, G., and Ward, R. (2018). Extracting Sparse High-Dimensional Dynamics from Limited Data. *SIAM Journal on Applied Mathematics*, 78(6):3279–3295.

Schmidt, M. and Lipson, H. (2009). Distilling Free-Form Natural Laws from Experimental Data. *Science*, 324(5923):81–85.

References V

Wikle, C. K. (2003). Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology*, 84(6):1382–1394.

Zhang, S. and Lin, G. (2018). Robust data-driven discovery of governing physical laws with error bars. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 474(2217):20180305.