Quantifying the Predictability of Noisy Nonlinear Biogeochemical Systems

Barbara A. Bailey
Department of Mathematics and Statistics
San Diego State University

Collaborators: Scott C. Doney and Ivan D. Lima, WHOI

Outline

- The Biogeochemical Model
- Nonlinear Models and Systems
- Global and Local Lyapunov Exponents
- The Noisy Biogeochemical System
- Conclusions and Future Work

Motivation

- Need to understand and predict the ocean response to and feedbacks on anthropogenic perturbations.
- Currently, coupling of biogeochemical models to the large-scale ocean physical circulation.
- Marine ecosystems strongly influenced by physical environmental forcing (storms, mesoscale turbulence). Both intrinsic and extrinsic variability are important.

Question: "What are the effects of different types of noise on the dynamics and predictability of aquatic biogeochemical systems?"

Goal: Improve understanding of the relationships between internal (biological) and external (physically forced) variability.

The NPZD Biogeochemical Model

- Coupled, nonlinear system of first-order differential equations:

$$\begin{aligned}
\frac{dN}{dt} &= a(1-m)gG(P,D)Z - U(I,N)IP + eD \\
\frac{dP}{dt} &= U(I,N)P - gG(P,D)Z - sP \\
\frac{dZ}{dt} &= amgG(P,D)Z - dZ^2 \\
\frac{dD}{dt} &= (1-a)gG(P,D)Z - gG(P,D)Z + sP + dZ^2 - eD \\
N_o &= N + P + Z + D
\end{aligned}$$

- a, g, s, m, d, and e are model parameters
- I is light intensity

- $U(I, N)$ is phytoplankton growth: controlled by light (I) and nutrient concentration (N) and losses through grazing ($-gG(P, D)Z$) and natural mortality ($-sP$)
- Zooplankton growth is a function of total food availability, $G(P, D)$ functional response
- N_o and I are the primary factors controlling the dynamics.
  - Asymptotically stable equilibrium points at low N_o and high to moderate I.
  - Oscillates in a limit cycle at low I and high N_o.
- We chose I and N_o in a range where system is unstable and oscillates in a limit cycle.
- Details for choices of parameters see Lima, et al. (2002) Intrinsic dynamics and stability properties of size-structured pelagic ecosystem models. *J. Plankton Res.*, 24, 533-556.

Nonlinear Models and Nonlinear Autoregressive Processes

- A general class of nonlinear models is from data of observed univariate responses Y_i, dependent on corresponding d-dimensional inputs $\boldsymbol{x}_i$:

  $$Y_i = f(\boldsymbol{x}_i; \theta) + e_i$$

  where θ is a p-dimensional vector of unknown parameters and $\{e_i\}$ is a sequence of i.i.d random variables.

- A nonlinear autoregressive process is a univariate time series:

  $$x_t = f(x_{t-1}, x_{t-2}, \ldots, x_{t-d}; \theta) + e_t$$

  where $\{e_t\}$ is a sequence of i.i.d random variables.

Example: Lorenz System

- Coupled, nonlinear system of three first order differential equations:

$$\begin{aligned}\frac{dx}{dt} &= -s(x-y)\\ \frac{dy}{dt} &= -xz + rx - y\\ \frac{dz}{dt} &= xy - bz\end{aligned}$$

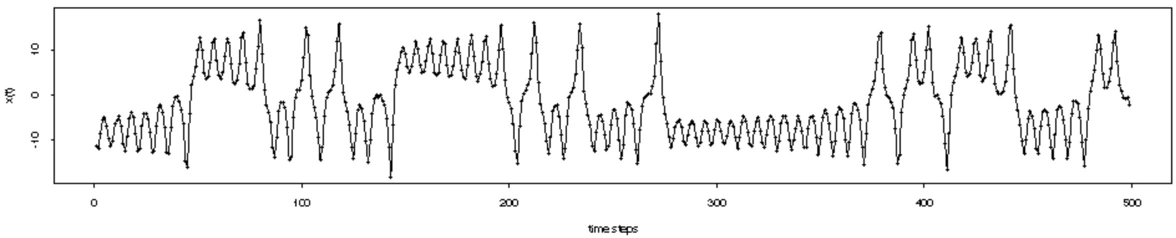
  for $s = 10, r = 28$ and $b = 8/3$ get famous "butterfly"

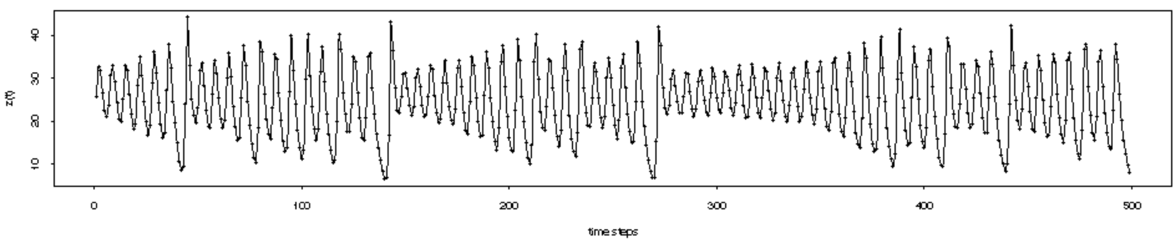
- Data: numerically integrate and add noise at every integration time-step.

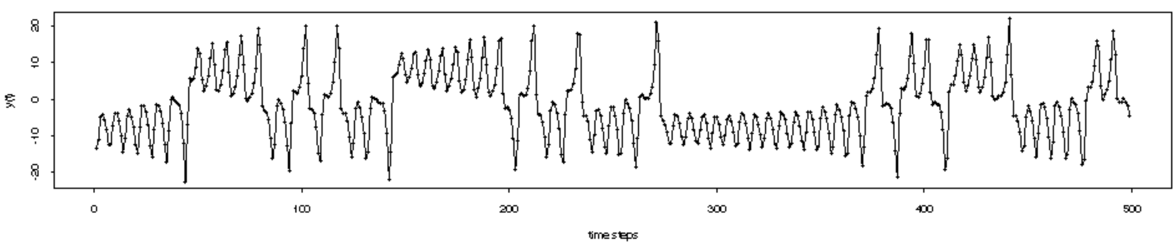
- State-Space System:

$$\begin{aligned}X_t &= F_1(X_{t-1}, Y_{t-1}, Z_{t-1}) + e_{1,t}\\ Y_t &= F_2(X_{t-1}, Y_{t-1}, Z_{t-1}) + e_{2,t}\\ Z_t &= F_3(X_{t-1}, Y_{t-1}, Z_{t-1}) + e_{3,t}\end{aligned}$$

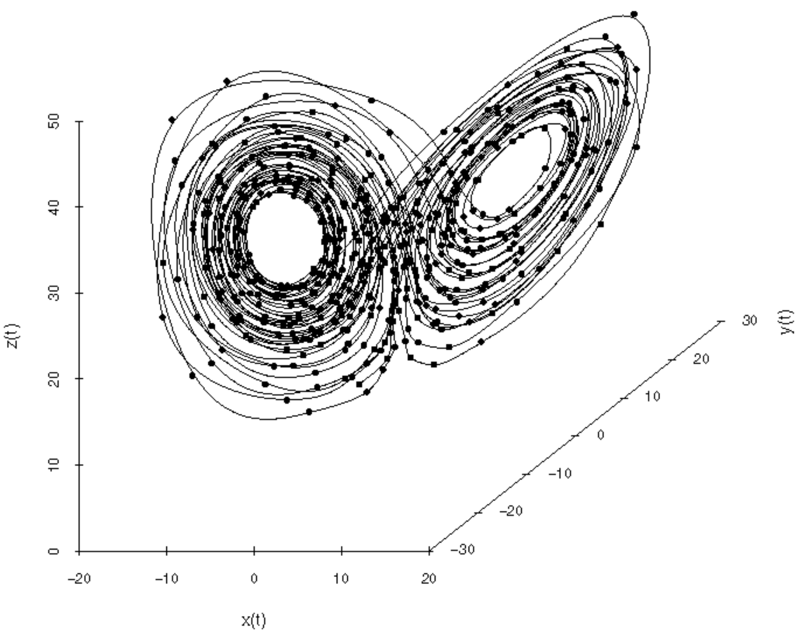
Lorenz Time Series, $X(t), Z(t), Y(t)$







Phase Space for Noisy Lorenz System



Lyapunov Exponents: Measuring Sensitivity to Initial Conditions

Taylor Series expansion to approximate the action of the map F on two initial state vectors X_1,Y_1

$$\begin{aligned} X_2 - Y_2 &= F(X_1) - F(Y_1) \\ &\approx DF(X_1)(X_1 - Y_1) \end{aligned}$$

Let $J_t = DF(X_t)$, the Jacobian matrix of F. By the chain rule for differentiation

$$X_n - Y_n \approx J_n \cdot J_{n-1} \cdots J_1(X_1 - Y_1)$$

Global Lyapunov Exponent:

$$\lambda = \lim_{n\to\infty} \frac{1}{n} \ln \|J_n J_{n-1} \ldots J_1\|$$

Local Lyapunov Exponent:

$$\lambda_n(t) = \frac{1}{n} \ln \|J_{n+t-1} J_{n+t-2} \cdots J_t\|$$

Chaos Facts

- If X_t is ergodic, stationary and bounded. Then λ exists and is independent of the trajectory (Multiplicative Ergodic Theorem of Oseledec)
- A system with $\lambda > 0$ has the property of "sensitive dependence on initial conditions" and is *chaotic* .

Neural Networks

- Class of nonlinear models:

$$Y = f(\boldsymbol{x};\theta) + e$$

$$Y_t = F(Y_{t-1}) + e_t$$

- The form of the model:

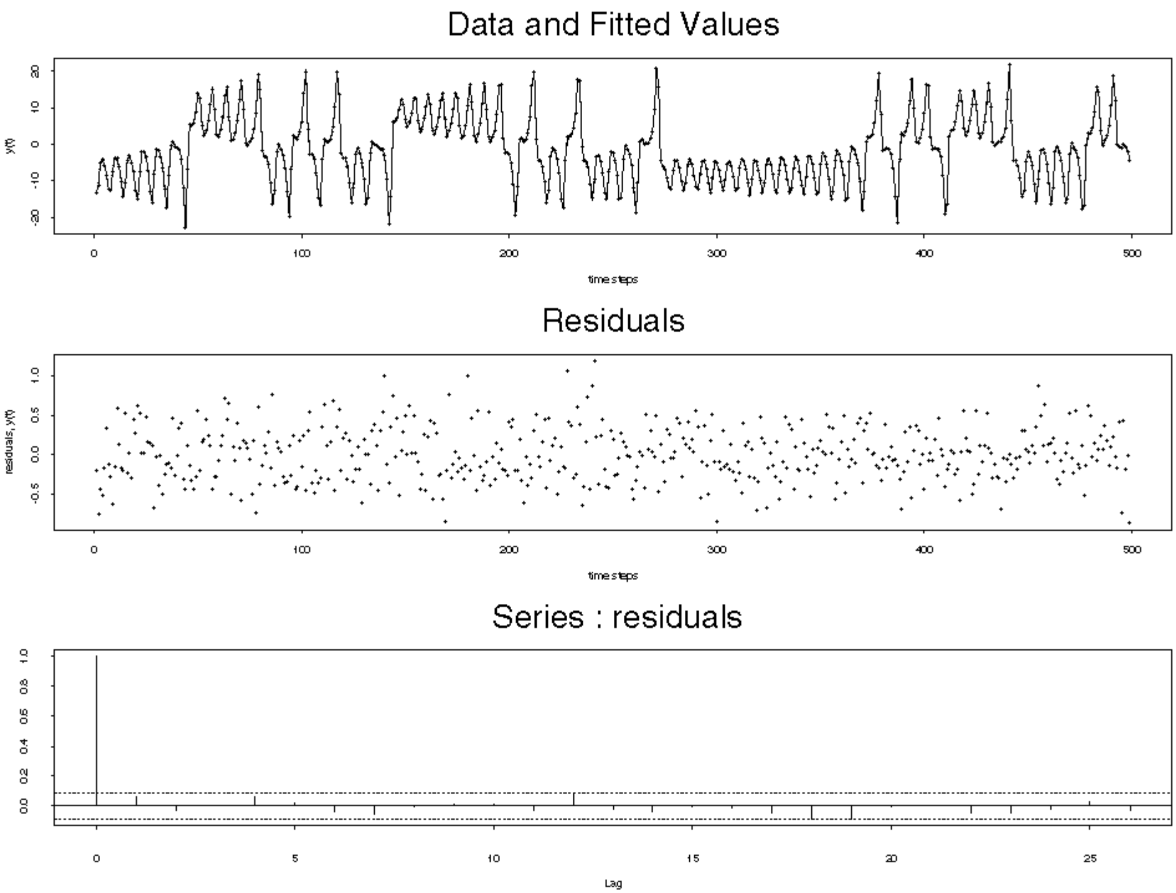
$$f(\boldsymbol{x}) = \beta_0 + \sum_{i=1}^{k} \beta_i \varphi(\boldsymbol{x}^T \gamma_i + \mu_i)$$

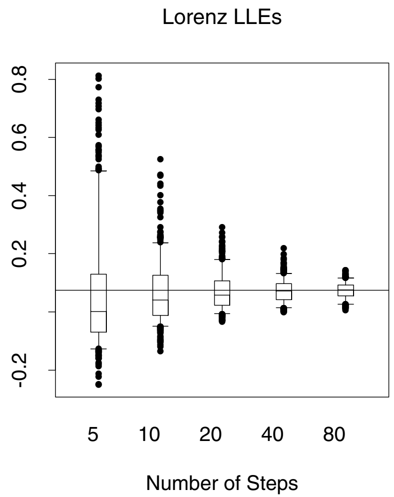
  where $\varphi(u) = e^u/(1+e^u)$

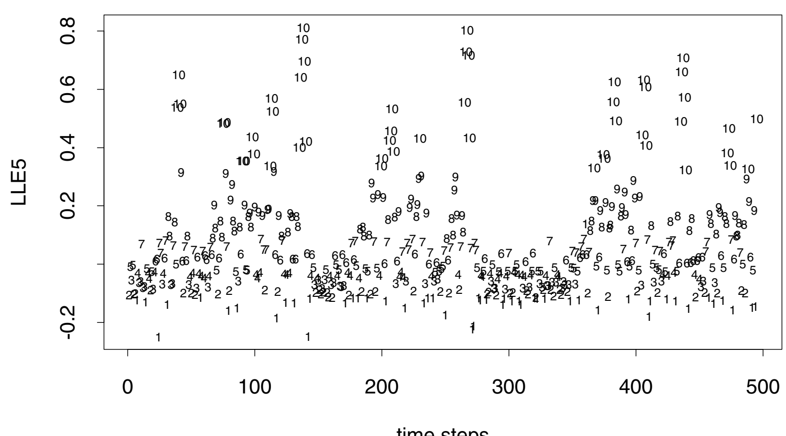
  - Net parameters are estimated by nonlinear least squares.
  - Total number of parameters is $p = 1 + k(d+2)$ where d is dimension of $\boldsymbol{x}$.
  - Complexity of the model chosen by Cross Validation:

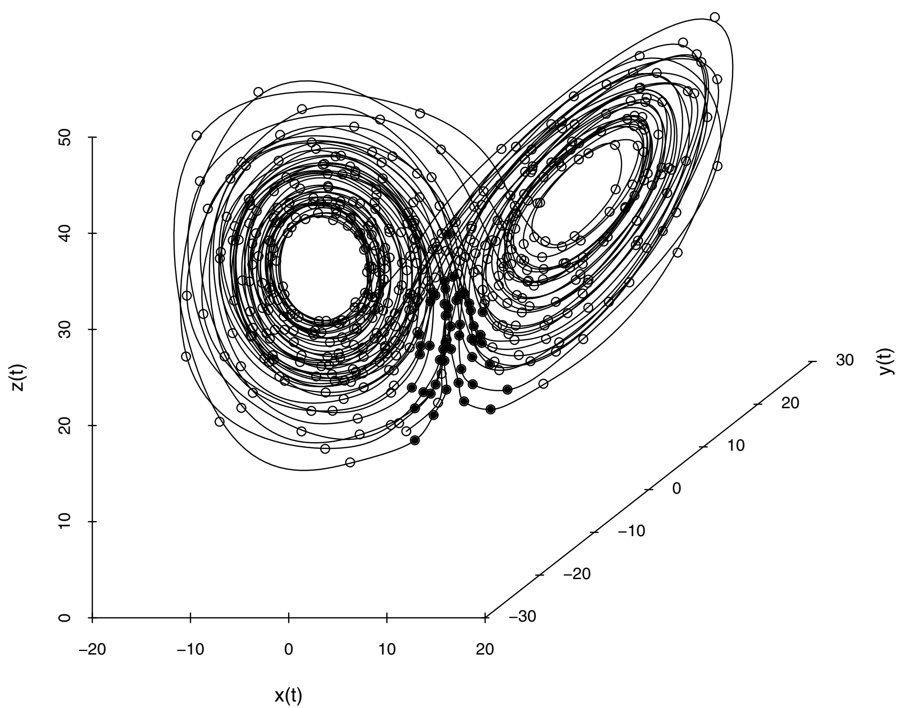
$$V_c = \frac{\frac{1}{n} RSS}{\left(1 - p\frac{c}{n}\right)^2}$$

Neural Network Fits to Noisy Lorenz System, Y(t)




Lorenz LLEs
0.8
0.6
0.4
0.2
-0.2
5
10
20
40
80
Number of Steps
(a)


LLE5
0.8
0.6
0.4
0.2
-0.2
0
100
200
300
400
500
time steps
(b)



z(t)
0
10
20
30
40
50
x(t)
−20
−10
0
10
20
y(t)
−30
−20
−10
0
10
20
30

Confidence Intervals for Paremeters

Based on asymptotics of MLE

Assumptions:

A1 X_t is stationary and ergodic.

A2 e_t is $i.i.d.$ $N(0, \sigma^2)$.

A3 $\theta \in \Theta$,and Θ is compact.

A4 F, first, second and third partials of F exist and are continuous and uniformly bounded for all $\theta \in \Theta$.

Theorem: Under Assumptions A1-A4, there exists a ML estimator $\hat{\theta}_n$ of θ for which $\hat{\theta}_n \xrightarrow{a.s.} \theta$ and
$I_n^{1/2}(\theta)(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, 1)$ as $n \to \infty$.

Corollary:

$$-2\ln(L(\theta)/L(\hat{\theta})) \xrightarrow{d} \chi^2(p)$$

Confidence Intervals for Parameters of Nonlinear Models

- Approximate Confidence set for θ:

$$\mathcal{A}_\theta = \{\theta : -2ln(L(\theta)/L(\hat{\theta})) \leq c\}$$

- Applications:

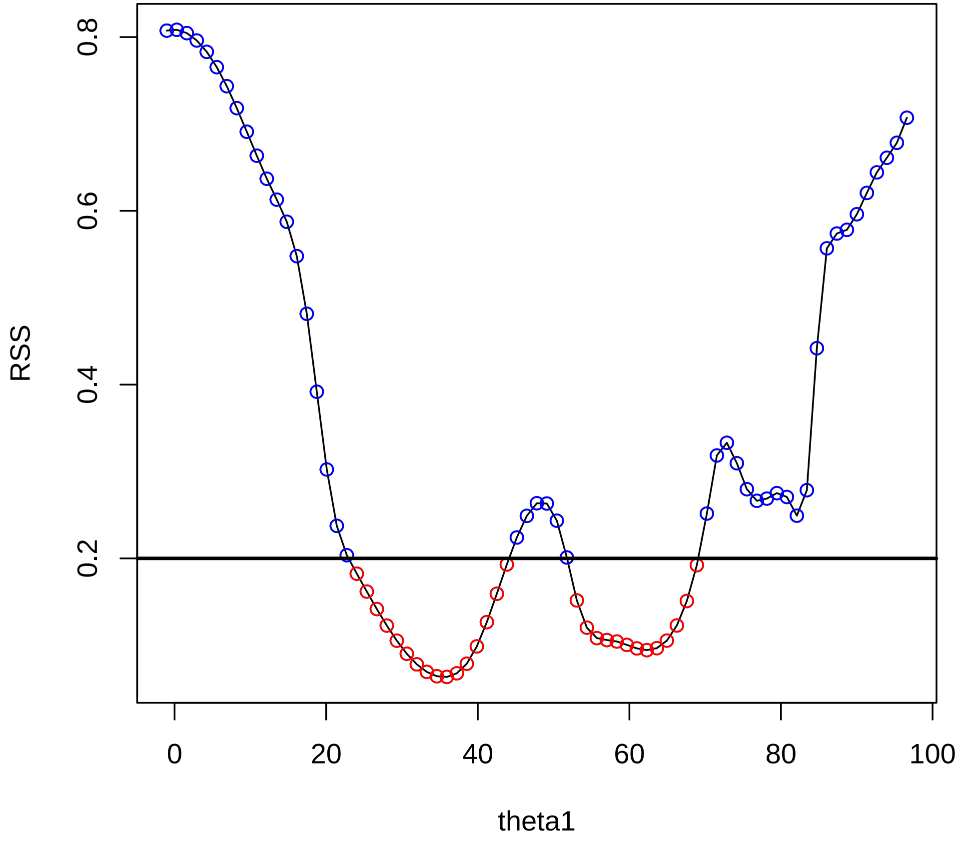
$$\mathcal{A}_\theta = \left\{\theta : S(\theta) \leq S(\hat{\theta}) \left[1 + \frac{p}{n-p}\mathrm{F}(p, n-p, \alpha)\right]\right\}$$

  where $S(\theta)$ is the residual sum of squares and $\hat{\theta}$ is the least-squares estimate of θ.

- CI for a functional of the parameters, $\varphi(\theta)$ is the min and max of the set:

$$\mathcal{A}_{\varphi(\theta)} = \{\varphi(\theta) : \theta \in \mathcal{A}_\theta\}$$

RSS Surface, $S(\theta)$



The NPZD Biogeochemical Model

- Coupled, nonlinear system of first-order differential equations:

$$
\begin{aligned}
\frac{dN}{dt} &= a(1-m)gG(P,D)Z - U(I,N)IP + eD \\
\frac{dP}{dt} &= U(I,N)P - gG(P,D)Z - sP \\
\frac{dZ}{dt} &= amgG(P,D)Z - dZ^2 \\
\frac{dD}{dt} &= (1-a)gG(P,D)Z - gG(P,D)Z + sP + dZ^2 - eD \\
N_o &= N + P + Z + D
\end{aligned}
$$

- a, g, s, m, d, and e are model parameters
- I is light intensity

- $U(I, N)$ is phytoplankton growth: controlled by light (I) and nutrient concentration (N) and losses through grazing ($-gG(P, D)Z$) and natural mortality ($-sP$)
- Zooplankton growth is a function of total food availability, $G(P, D)$ functional response
- N_o and I are the primary factors controlling the dynamics.
  - Asymptotically stable equilibrium points at low N_o and high to moderate I.
  - Oscillates in a limit cycle at low I and high N_o.
- We chose I and N_o in a range where system is unstable and oscillates in a limit cycle.
- Details for choices of parameters see Lima, et al. (2002) Intrinsic dynamics and stability properties of size-structured pelagic ecosystem models. *J. Plankton Res.*, 24, 533-556.

Generating Noisy NPZD Data

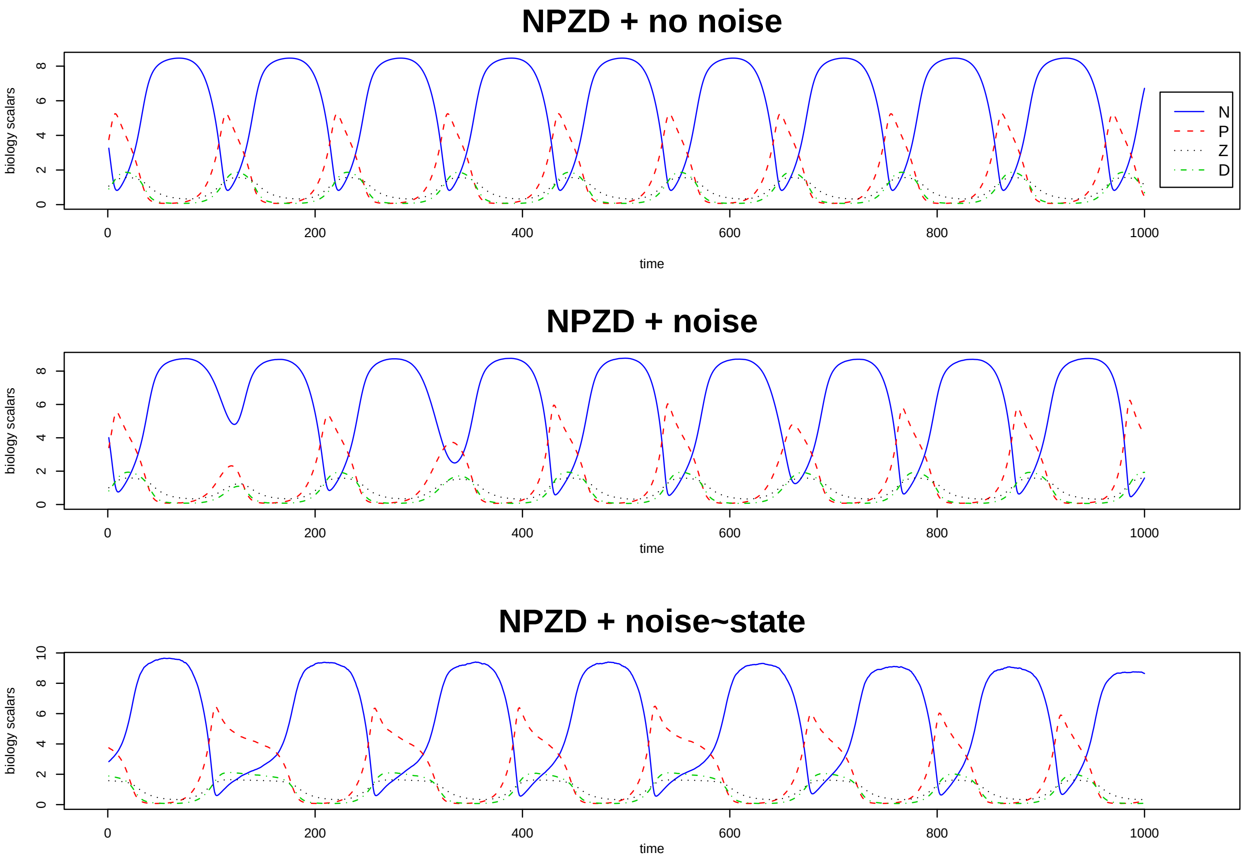
Data: numerically integrate and add noise at every integration time-step. Generalized as follows:

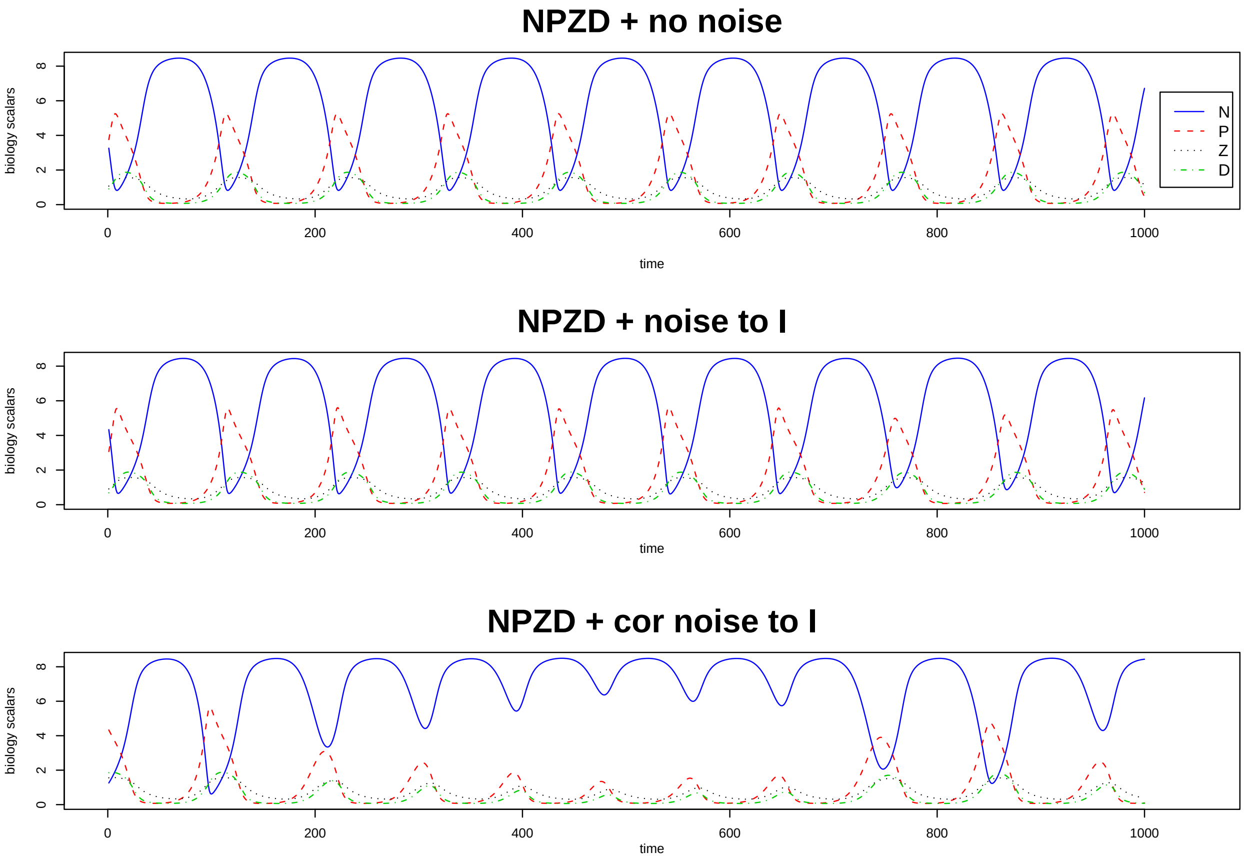
$$Y_t = f(Y_{t-1}, \eta + g_t) + \varepsilon_t$$

- Y_t is the state vector (N_t, P_t, Z_t, D_t)
- η is the vector of model parameters
- f is the dynamical operator for the discretized set of equations
- Model noise term ε_t is assumed to be an independent, Gaussian random variable with mean zero and variance σ_ε^2
  - more realistic to adjust the noise so that it is proportional to the magnitude of each state variable
- Parameter noise term more realistic if $g_t = \phi g_{t-1} + a_t$

Data:

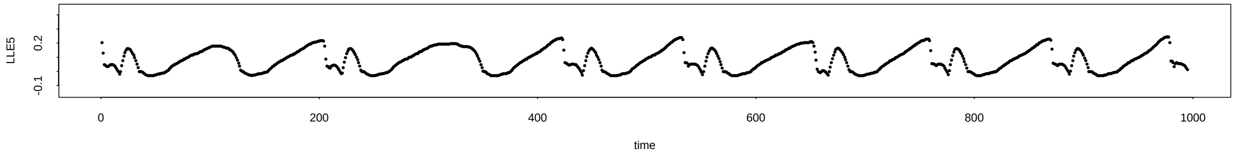
1. Noise is added to each state variable (N, P, Z, D). The noise, ε_t is i.i.d. $N(0, \sigma_\varepsilon^2)$ with $\sigma_\varepsilon = 0.002$.
2. Noise is added to each state variable. The noise is proportional to the magnitude of the state variable. The noise, is $\varepsilon_t Y_{t-1}$, where ε_t is i.i.d. $N(0, \sigma_\varepsilon^2)$ with $\sigma_\varepsilon = 0.002$.
3. Noise is added to the parameter I. The noise, $g_t = a_t$ $(\phi = 0)$ is i.i.d. $N(0, \sigma_a^2)$ with $\sigma_a = 0.025$.
4. Noise is added to the parameter I. The correlated noise is $g_t = \phi g_{t-1} + a_t$ with $\phi = 0.7$ and a_t is i.i.d. $N(0, \sigma_a^2)$ with $\sigma_a = .0178$ $(\sigma_g = 0.025)$.
5. Identical to Data Case 1 except $\sigma_\varepsilon = 0.010$.


NPZD + no noise
biology scalars
time
N
P
Z
D
0
200
400
600
800
1000
0
2
4
6
8
NPZD + noise
biology scalars
time
NPZD + noise~state
biology scalars
time
10

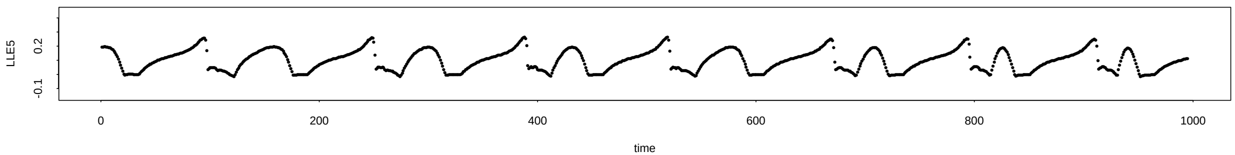


NPZD + no noise
biology scalars
0
2
4
6
8
0
200
400
600
800
1000
time
N
P
Z
D
NPZD + noise to I
biology scalars
0
2
4
6
8
0
200
400
600
800
1000
time
NPZD + cor noise to I
biology scalars
0
2
4
6
8
0
200
400
600
800
1000
time

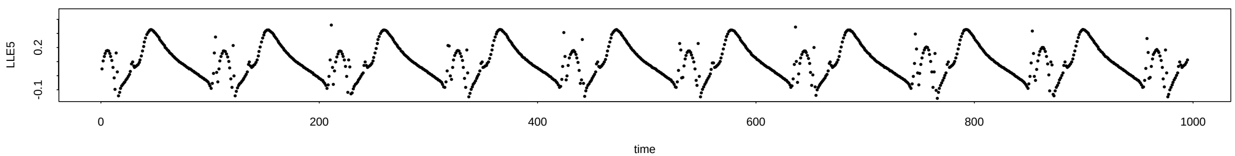
Case 1: LLE5s



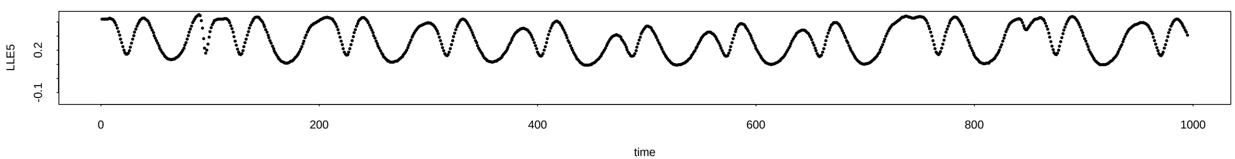
Case 2: LLE5s

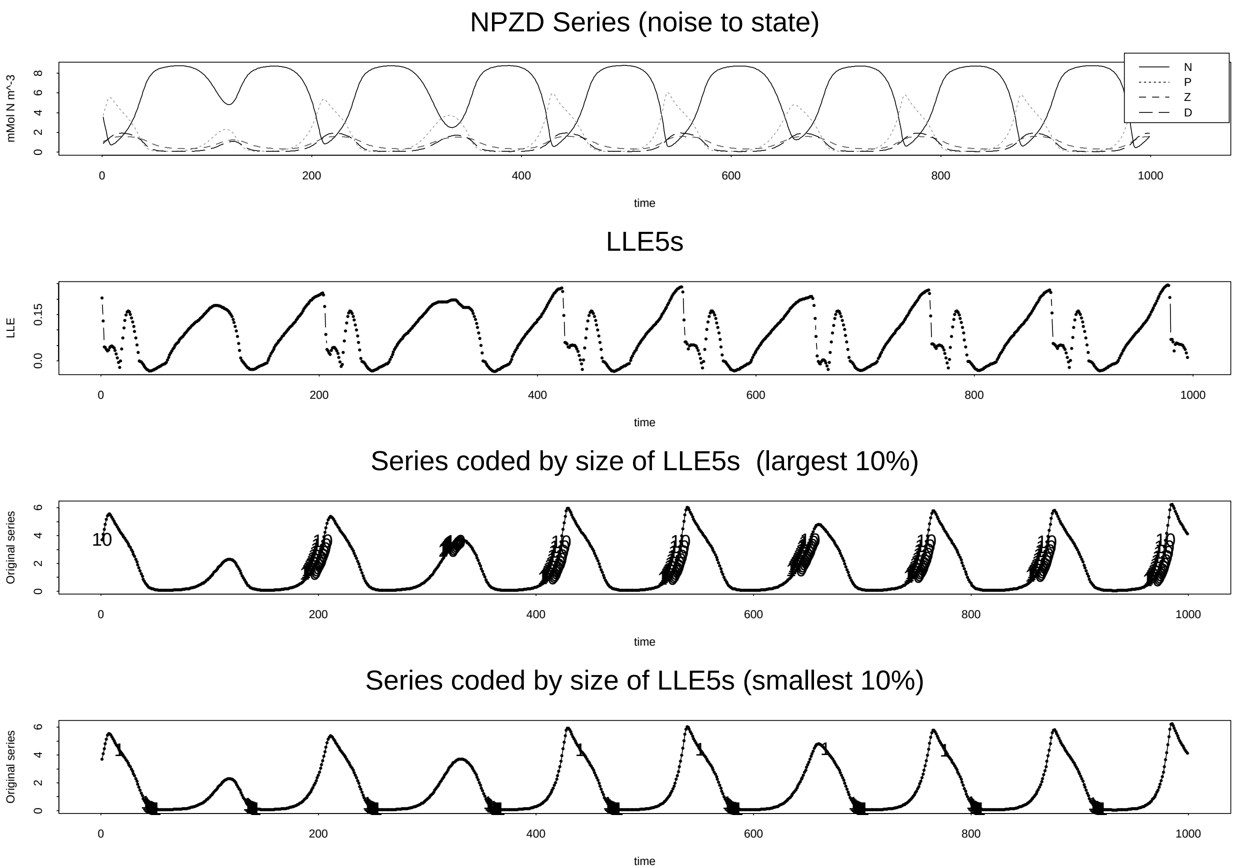


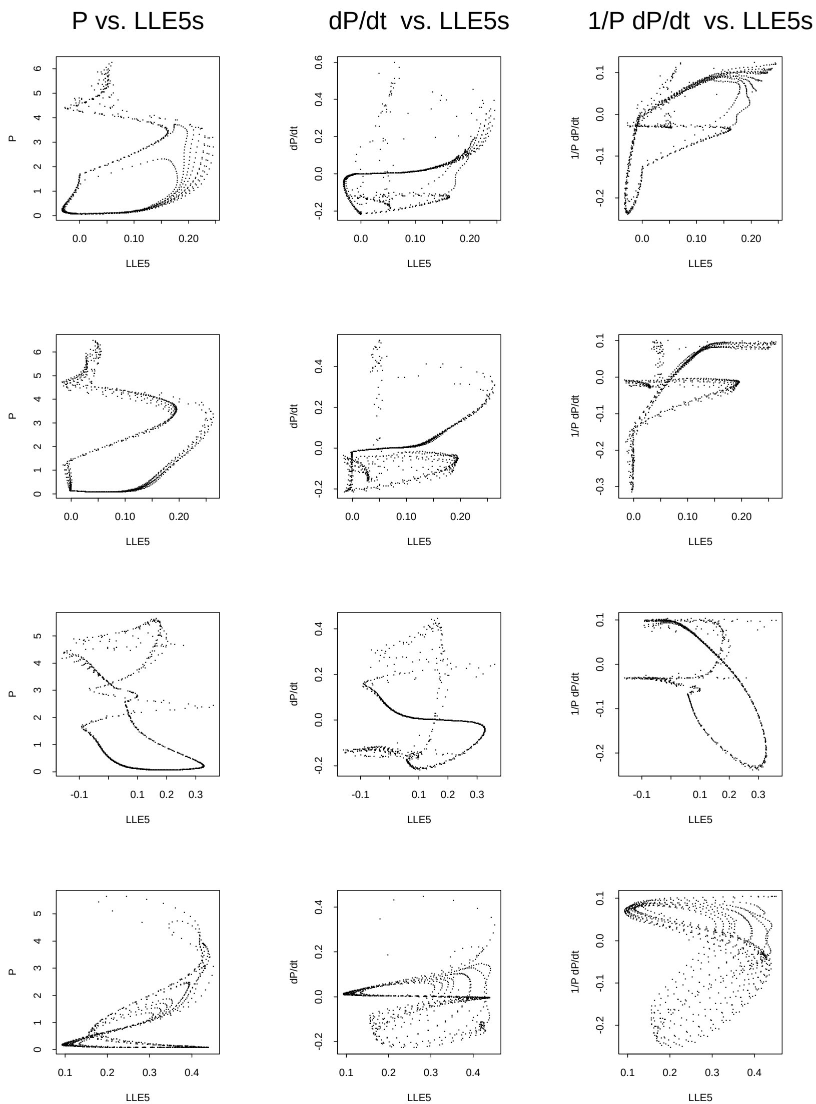
Case 3: LLE5s



Case 4: LLE5s




NPZD Series (noise to state)
mMol N m^-3
N
P
Z
D
time
LLE5s
LLE
time
Series coded by size of LLE5s (largest 10%)
Original series
time
Series coded by size of LLE5s (smallest 10%)
Original series
time


P vs. LLE5s
dP/dt vs. LLE5s
1/P dP/dt vs. LLE5s
P
dP/dt
1/P dP/dt
LLE5
0.0
0.10
0.20
-0.1
0.1
0.2
0.3
0.4

Short-term Predictability of Phytoplankton (P)

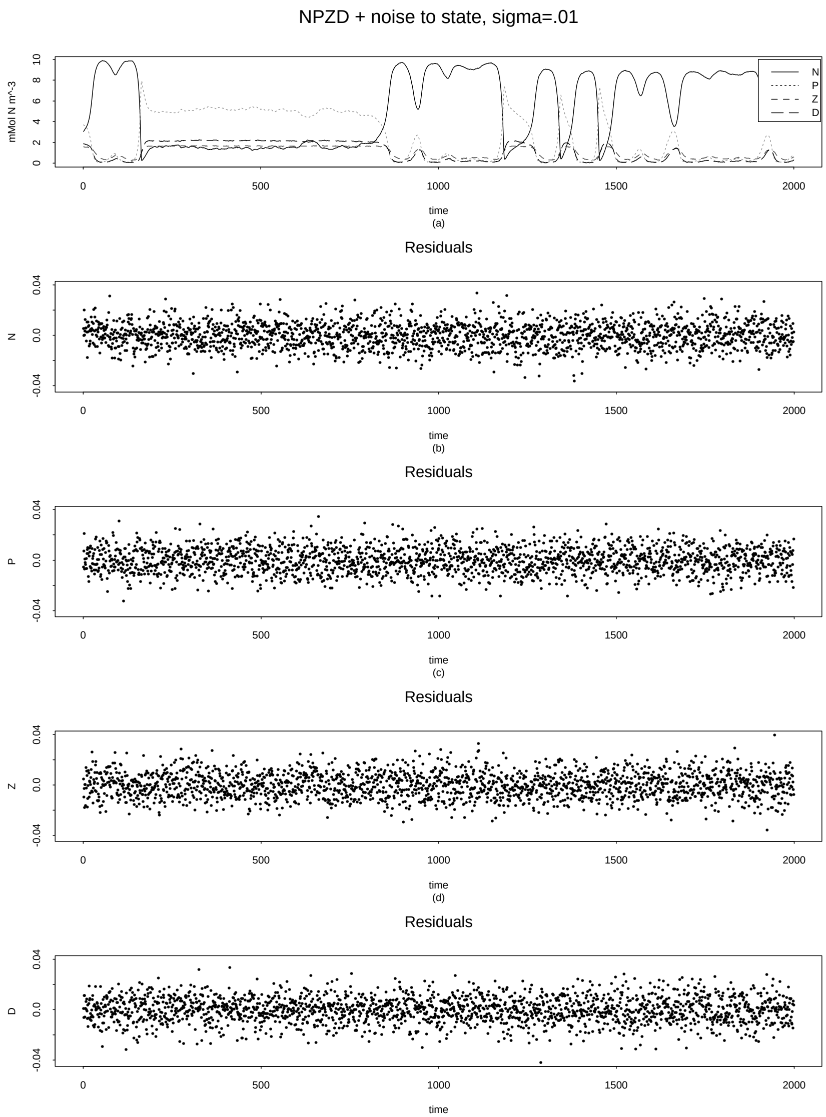
- Case 1 and 2 are similar.
- Quantity $1/P\ dP/dt$ is the relative or fractional growth rate.
  - Shape of phytoplankton limit cycle outlined by P vs $1/P\ dP/dt$, is similar across all four data cases.
  - Local predictablitiy patterns are similar in that the LLE5 values are typically large during the growth phase of the bloom (when $1/P\ dP/dt$ is near a maximum).
- Presence of a secondary period of large LLE5s during the collapse of the bloom is case dependent.

Conclusions and Future Work

- Statistical Framework for Noisy Nonlinear Systems (LEs and LLEs are useful diagnositc tools!)
- Different types of noise can generate very different dynamics (LLEs can vary significantly)
- Need to compare noisy systems with ocean field data
- Development of a class of stochastic models is proposed as a "complement" to the need for more complex deterministic biogeochemical models.

Bailey, B.A., S.C. Doney, and I.D. Lima, Quantifying the effects of dynamical noise on the predictability of a simple ecosystem model, *Environmetrics*, **15**, 337–355, 2004.

Appendix


NPZD + noise to state, sigma=.01
mMol N m^-3
N
P
Z
D
time
(a)
Residuals
N
time
(b)
Residuals
P
time
(c)
Residuals
Z
time
(d)
Residuals
D
time
(e)