

Quantifying the predictability of noisy space-time dynamical processes

BARBARA A. BAILEY

Many environmental processes are complex space-time dynamical systems and the predictability of the system is an important feature of its dynamics. The extension of local Lyapunov exponents, the quantity that measures the short-term growth of a perturbation in time to include implicit spatial dependence is developed in this paper. A nonlinear modeling approach using flexible neural network models is used to describe the space-time dynamics and quantify the predictability of data from nonlinear stochastic systems. This allows for estimation of dynamical system quantities from data, along with measures of uncertainty for these estimates. The evolution of cloud cover over time and its space-time relationship to other climate variables is an interesting dynamical system that is very important in climate modeling. In the spirit of a cloud parameterization, a nonlinear nearest-neighbor model to describe grid cell relationships is fit to data. The estimation of the space-time local Lyapunov exponents are used to quantifying the stability and predictability of the space-time cloud process.

AMS 2000 SUBJECT CLASSIFICATIONS: Primary 37M25; secondary 62M45.

KEYWORDS AND PHRASES: Nonlinear time series, Neural network models, Local Lyapunov exponents.

1. INTRODUCTION

Many environmental processes evolve over space and time creating a complex dynamical system. Typically there is interest in the time evolution of important variables over specific spatial domains. Such dynamical processes involve complicated temporal and spatial structure.

Statistical modeling of dynamical systems makes the estimation and construction of confidence intervals for interesting quantities from data possible. When the noise or stochastic component is an integral part of the system's dynamics, a nonlinear space-time modeling approach can be used to quantify the dynamics and predictability of the system. This involves fitting nonlinear models, estimating dynamical systems quantities of interest such as global and local Lyapunov exponents, along with measures of uncertainty for these estimates.

The characteristic feature of a chaotic system is that initially nearby trajectories diverge in time [9]. The exponential growth of initially small perturbations indicates “sensitive dependence on initial conditions” and is responsible for the unpredictability of a chaotic system. This unpredictable time evolution of nonlinear systems occurs in a variety of fields such as physics and meteorology [e.g., 2, 13, 18, 23].

The Lyapunov exponent measures the asymptotic average exponential divergence (or convergence) of nearby trajectories in phase space and has proven to be a useful diagnostic to detect and quantify chaos [9]. A positive Lyapunov exponent or exponential divergence implies that the system is unpredictable. Similarly, a negative Lyapunov exponent or exponential convergence implies that the system is predictable. The predictability of a system is an important feature of its dynamics. Predictability, not only quantified by the Lyapunov exponent, which by definition is a “global” quantity, but by the identification times in the series or regions “locally” in the phase space that have different short term sensitivity to small perturbations in time. The examination of the fluctuations of these finite-time local Lyapunov exponents, both positive and negative, are useful in describing the dynamics and heterogeneity of a system.

The goal of this paper is to extend the definition of local Lyapunov exponents to include implicit spatial dependence, called space-time local Lyapunov exponents. A nonlinear time series and space-time modeling approach is used to estimate these important dynamical system quantities from data. The estimation and examination of space-time local Lyapunov exponents provides a natural framework to describe and quantify space-time dynamical systems. Local Lyapunov exponents will be the focus of this paper since they describe the short-term behavior which will be most useful in applications. For example, the evolution of cloud cover is an important atmospheric process and dynamical system. The space-time modeling of cloud cover is an important step in developing a sub-grid scale parameterization for the amount of cloud cover (or cloud fraction) in a grid box of a numerical model that predicts the time evolution of atmospheric variables. This is relevant since clouds represent a significant source of potential error in climate simulations [29]. Space-time Lyapunov exponents can be used to quantify the space-time stability properties and give important information about the modeled noisy dynamical sys-

tem. The spatial map or image plot of the exponents is innovative and useful in visualizing the information. In practice, interest lies in both the fit of the space-time model as well as the space-time stability properties of the system.

This paper is organized as follows. First, the nonlinear modeling and computational aspects of the nonlinear stochastic problem will be discussed in Section 2. Section 3 is a series of three progressive examples. Section 3.2 is an example of the use of a nonlinear time series approach involving data generated from three first-order differential equations with additive dynamic noise. Section 3.3 will extend the analysis to a neighborhood dependence. Finally, Section 3.4 will demonstrate the use of space-time Lyapunov exponents with a space-time model for cloud cover data. Section 4 will contain some concluding remarks.

2. NOISY DYNAMICAL SYSTEMS AND NEURAL NETWORK MODELING

2.1 Background

This section will provide an introduction to noisy dynamical systems and define global and local Lyapunov exponents. We consider a discrete dynamical system perturbed by noise. Assume that $\mathbf{X}_t, \mathbf{e}_t \in \mathbb{R}^d$ and that the time evolution of the $\{\mathbf{X}_t\}$ stochastic process follows the following dynamics,

$$(1) \quad \mathbf{X}_t = F(\mathbf{X}_{t-1}) + \mathbf{e}_t,$$

where $\{\mathbf{e}_t\}$, each component of the vector \mathbf{e}_t , is a sequence of i.i.d. random variables. It is important to note that the error in (1) is not measurement error, but dynamic noise, an inherent part of the dynamics of the system. The stochastic process of (1), called a nonlinear autoregressive process is a natural nonlinear statistical model for estimating dynamical system quantities from data. This is a nonlinear version of an autoregressive processes commonly used in the analysis of time series data.

The local Lyapunov exponent (LLE) is defined by making an infinitesimal perturbation of \mathbf{X}_t at time t and following forward the perturbed and unperturbed trajectories or sample paths. The LLE measures the log-difference in the norms of the two trajectories after n time steps. For small perturbations, the linear approximation will be defined by the derivatives of the map F then

$$(2) \quad \lambda_n(t) = \frac{1}{n} \ln \|\mathbf{J}_{n+t-1} \mathbf{J}_{n+t-2} \cdots \mathbf{J}_t\|,$$

where \mathbf{J}_t is the Jacobian matrix of F evaluated at \mathbf{X}_t . The matrix norm $\|\cdot\|$ used in this paper is the largest singular value of the matrix. Since $\lambda_n(t)$ is a function of time, the LLE depends on the trajectory and can be thought of as an “ n -step ahead” local Lyapunov exponent process.

The global Lyapunov exponent for (1) is

$$(3) \quad \lambda = \lim_{n \rightarrow \infty} \frac{1}{n} \ln \|\mathbf{J}_n \mathbf{J}_{n-1} \cdots \mathbf{J}_1\|.$$

The global Lyapunov exponent, often referred to as the Lyapunov exponent, measures sensitive dependence on initial conditions, that is, if λ is positive then initially close trajectories will diverge exponentially over time.

Comparing (2) and (3), the LLE is a finite-time version of the global Lyapunov exponent. The Lyapunov exponent is a single number and LLEs are a distribution of numbers. The analysis using LLEs involves examining the short-term predictability of the system and identifying the parts of the time series or regions in state space where they occur [e.g., 1, 4, 36].

Lyapunov exponents for the nonlinear system are defined to be the Lyapunov exponents of the tangent map system described by (6), below. To see that the time evolution of a small perturbation to the state of the nonlinear system of (1) is approximated by a linear system let $\mathbf{X}_t^{(1)}$ and $\mathbf{X}_t^{(2)}$ be solutions of (1) that differ only in their initial values. If the initial perturbation $\mathbf{Y}_0 = (\mathbf{X}_0^{(1)} - \mathbf{X}_0^{(2)})$ is sufficiently small, then the subsequent separation $\mathbf{Y}_t = (\mathbf{X}_t^{(1)} - \mathbf{X}_t^{(2)})$ can be approximated by the linearization

$$(4) \quad \mathbf{Y}_t = F(\mathbf{X}_{t-1}^{(1)}) - F(\mathbf{X}_{t-1}^{(2)}) \approx \mathbf{J}_{t-1} \mathbf{Y}_{t-1}.$$

It is important to note that the iteration of the model (1) assumes the same random shock \mathbf{e}_t . Now, let $\mathbf{T}_t = \mathbf{J}_{t-1} \mathbf{J}_{t-2} \cdots \mathbf{J}_0$, then the Lyapunov exponent in (3) can also be written as

$$(5) \quad \lambda = \lim_{t \rightarrow \infty} \frac{1}{t} \ln \|\mathbf{T}_t \mathbf{Y}_0\|,$$

where $\mathbf{T}_t \mathbf{Y}_0$ are solutions to the linear system

$$(6) \quad \mathbf{Y}_t = \mathbf{J}_t \mathbf{Y}_{t-1}.$$

Therefore a small perturbation of the state of the nonlinear system (1) is approximated by the linear system (6). The approximation becomes exact in the formal limit of infinitesimally small \mathbf{Y}_0 . See McCaffrey et al. [20] and Nychka et al. [21] for a more detailed description of Lyapunov exponents for nonlinear stochastic systems.

2.2 Estimating Lyapunov exponents from data

There are two classes of methods for estimating Lyapunov exponents (and LLEs) from data: direct methods and Jacobian methods. In direct methods, the growth rate of an infinitesimal perturbation is estimated by tracking the evolution of differences between initially close observed data points [15, 35]. This procedure is sensitive to noise. In noisy systems, the trajectories being compared will not have the same sequence of random shocks and therefore the divergence might simply be due to the random component [21].

Jacobian methods use the data to estimate the Jacobian matrices. When a random component is present as in (1), nonlinear regression can be used to estimate the Jacobian

matrices. The approach in which Jacobian estimates are derived from nonlinear estimation of F have been shown to be successful [e.g., 4, 12, 20, 21, 28, 37]. McCaffrey et al. [20] considered nonparametric estimators such as thin-plate splines, radial basis functions, neural networks, and projection pursuit methods. A simulation study of nonlinear maps (e.g., Hénon map [26]) both with and without added noise, indicates that neural networks are feasible and therefore will be used in this paper.

The calculation of LLEs in (2) and global Lyapunov exponents in (3) requires an estimation of the Jacobian matrix \mathbf{J}_t . The map F is approximated by a flexible nonlinear model, a feedforward single hidden layer network with a single output. For each time t , the form of the map is

$$(7) \quad F(\mathbf{X}) = \beta_0 + \sum_{i=1}^k \beta_i \varphi(\mathbf{X}'\boldsymbol{\gamma}_i + \mu_i),$$

where $\varphi(u) = e^u/(1 + e^u)$ and the parameter vectors $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\mu}$ are unknown. The parameters β_k are known as connection weights and μ_k are the biases [24]. If \mathbf{X} is a d dimensional vector, $\boldsymbol{\beta}$ has length $k + 1$, $\boldsymbol{\mu}$ has length k and the $\boldsymbol{\gamma}_i$ are k vectors each of length d . The total number of parameters in the model is equal to $1 + k(d + 2)$.

Model selection will consist of choosing the number of hidden units k in (7) and will be based on generalized cross validation (GCV). Cross validation is a standard approach for selecting smoothing parameters in nonparametric regression [34]. The GCV function is defined

$$(8) \quad GCV(p) = \frac{\frac{1}{n}RSS_p}{(1 - p\frac{c}{n})^2},$$

where n is the number of data points used to fit the model, p is number of parameters in the model, and RSS_p is the residual sum of squares for the best fitting p -parameter model. The cost parameter, c , is an ad hoc adjustment to increase $GCV(p)$ for larger values of p . The standard GCV function is with $c = 1$, but to guard against over fitting, a cost parameter value of $c = 2$ is preferred. The model fitting strategy is to fit models with a range of hidden units k . The model order p is chosen to be the p that minimizes the $GCV(p)$ function and $p = 1 + k(d + 2)$.

The parameter vectors $\boldsymbol{\beta}$; $\boldsymbol{\gamma}_i, i = 1, \dots, k$; and $\boldsymbol{\mu}$ are unknown and will be estimated by nonlinear least squares. From a very large collection (250,000) of starting parameter sets, the parameter sets with the lowest root mean squared error (RMSE) are used as initial points in an iterative procedure for minimization of the RMSE with a moderate convergence tolerance. A specified number (20) of these parameter sets are then used as starting points in the minimization procedure, but with a smaller convergence tolerance. The parameter set with the smallest RMSE is taken to be the least-squares estimate. The above procedure is repeated for each value of k in the range of interest. For more details of

the minimization procedure and software utilized see Nychka et al. [22].

The following is a description of the construction of confidence intervals for parameter estimates of the nonlinear autoregressive process of (1), where the nonlinear map is estimated by a neural network model of (7). It is necessary to have a model with stochastic components in order to make a confidence statement about parameter estimates. For the purpose of asymptotic properties of neural network model parameters and the construction of confidence intervals it will be necessary to make assumptions about the noise and the map F of (1). Following the conditions in Bailey, Ellner and Nychka [4], the noise $\{e_t\}$ is assumed to be a sequence of independent Gaussian random variables with mean zero and variance σ^2 . For the map F , the first, second and third partial derivatives of F are assumed to be continuous and uniformly bounded for parameter sets in a compact set. The neural network model of (7) satisfies these conditions on F . The confidence interval is based on the identification of a joint confidence region for the model parameters $\boldsymbol{\theta}$, using the likelihood ratio statistic. The approximate confidence set \mathcal{A}_θ will be all the values of $\boldsymbol{\theta}$ such that

$$(9) \quad RSS(\boldsymbol{\theta}) \leq RSS(\hat{\boldsymbol{\theta}}) \left[1 + \frac{p}{n-p} F(p, n-p, \alpha) \right],$$

where $RSS(\boldsymbol{\theta})$ is the residual sum of squares and $\hat{\boldsymbol{\theta}}$ is the least-squares estimate of $\boldsymbol{\theta}$ [27]. For the F distribution, p is the number of parameters in the model, n is the sample size and α is the confidence level. The computation of the confidence interval is a two step process. First, a neural net is fit to the data to obtain the least-squares estimate for $\boldsymbol{\theta}$. Then parameter sets (500) are sampled from the confidence set \mathcal{A}_θ . For all $\boldsymbol{\theta}$'s in the confidence set, an appropriate functional of the parameters can be evaluated (e.g., LEs and LLEs). The minimum and maximum value of the functional is the confidence interval (CI). Simulations showed that the coverage probabilities of the confidence intervals generated with this procedure were close to the corresponding significance level [4]. For a more detailed description of constructing confidence intervals for estimated global and local Lyapunov exponents see Bailey, Ellner and Nychka [4].

3. EXAMPLES TOWARDS SPACE-TIME LOCAL LYAPUNOV EXPONENTS

This section defines and presents a series of three examples to illustrate the important information that can be obtained from estimating space-time local Lyapunov exponents from data. These examples are a progression from a time series model and estimating LLEs to a space-time model and estimating space-time LLEs. In the first example data is generated from the Lorenz three-variable system with additive noise. The second example includes spatial dependence by the use of the Lorenz forty-variable system with additive noise. In the first two examples, additive noise

will generate data with a stochastic component so that the data is described by (1). The second example is the extension of time series data to time series data with a spatial nearest-neighbor component. The third example is a space-time model of cloud cover developed as a parameterization and the estimated space-time LLEs. The data for this example are satellite and assimilated observational data.

3.1 Space-time local Lyapunov exponents

The nonlinear autoregressive process described in (1) is a state space vector system and can include the typical univariate time series model as a special case. For example, consider the univariate time series

$$(10) \quad x_t = f(x_{t-1}, x_{t-2}, \dots, x_{t-d}) + \varepsilon_t,$$

where $x_t \in \mathbb{R}^1$ and $\{\varepsilon_t\}$ is a sequence of i.i.d random variables. To express the system as (1), let $\mathbf{X}_t = (x_t, x_{t-1}, \dots, x_{t-d+1})'$ and the error vector $\mathbf{e}_t = (\varepsilon_t, 0, \dots, 0)'$. However, in the example using the Lorenz three-variable system we have access to all the state variables in the system and not just a single time series. So, we will use the three variables to construct a state space vector system.

In dynamical systems theory for the system in (10) with no noise, Takens' embedding Theorem states that for sufficiently large d there exists the map f and we can extract important characteristics of the dynamical system, such as Lyapunov exponents [31]. We will assume our data follows the process in (1) and that the noise is an integral part of the system, so Takens' Theorem is not directly applicable. Intuitively, in Takens' Theorem the time lags of a single variable act as surrogates for the unobserved variables of the system. It is possible to use only time lags of a single variable in the estimation of the map of F in (1), but if other predictor variables are used, it can result in a simpler model. For example, in the case of modeling the dynamics of measles epidemics, incorporating an estimate of the susceptibles in the population resulted in simpler model with better forecast accuracy [10]. A simpler model is one with a smaller number of neural network parameters. In this spirit, we are willing to include other predictor or state variables and use the model selection criteria to determine the best model to estimate F and the matrix of partial derivatives of the map F . The first example using the Lorenz three-variable system with additive noise demonstrates the use of all three variables in the state space system and the ability to use these state variables successfully in LLE estimation. The second example extends to the case of implicit spatial dependence. In the cloud model example, the other predictor variables are very important and in fact, are necessary for the parameterization.

To define space-time LLEs it will be necessary to extend the general class of nonlinear time series models of (1) to space-time processes. A more detailed description is given in

the example of Section 3.4. Let $X_{s,t}$ denotes the value of the process of interest at location s at time t , where $(s, t) \in \mathcal{D}$ and \mathcal{D} is a lattice or grid in space-time. The space-time dynamics are described by

$$(11) \quad X_{s,t+1} = F(\mathbf{X}_{\mathcal{N}(s),t}, \boldsymbol{\theta}) + e_{s,t+1},$$

where $e_{s,t}$ are random errors assumed to be mutually independent Gaussian random variables with mean zero and common (unknown) variance σ^2 , across both space and time. The $\boldsymbol{\theta}$ is some collection of model parameters. The vector \mathbf{X} can contain values of X_t , along with other predictor variables. In general, $\mathcal{N}(s)$ represents a neighborhood or a collection of sites defined to be neighbors of the site s . The space-time LLE at site s will be defined as

$$(12) \quad \lambda_n^{(s)}(t) = \frac{1}{n} \ln \|\mathbf{J}_{n+t-1} \mathbf{J}_{n+t-2} \cdots \mathbf{J}_t\|,$$

where \mathbf{J}_t is the Jacobian matrix of F in (11) evaluated at $X_{s,t}$. This exponent is an extension of the LLE to a process with implicit spatial dependence. This exponent can be visualized with a spatial map of the LLEs on the grid \mathcal{D} .

3.2 The Lorenz system

In the first example, noisy data is generated from a “well-known” system. Noise is added to the system to generate data from the nonlinear autoregressive process of (1) and demonstrate the methods described in Section 2. As described in Section 2, a nonlinear time series approach is used to estimate the nonlinear map and estimate Lyapunov exponents along with confidence intervals for the estimates. This demonstrates statistical model fitting to data and gives results that are consistent with what is known about the Lorenz system. For other examples of adding noise to systems and estimating quantities from data see Bailey, Ellner and Nychka [4] and Bailey, Doney and Lima [3].

The Lorenz system is a coupled, nonlinear system of three first-order differential equations:

$$(13) \quad \begin{aligned} \frac{dx}{dt} &= -s(x - y) \\ \frac{dy}{dt} &= -xz + rx - y \\ \frac{dz}{dt} &= xy - bz, \end{aligned}$$

where s, r and b are constants that determine the behavior of the system [17]. To this day, the system models cellular convection. To obtain the famous “butterfly” attractor and for this example we let $s = 10, r = 28$ and $b = 8/3$.

Data is generated from the Lorenz system by numerically integrating with a fixed time step and sampling at every time step. A fourth-order Runge-Kutta scheme with step size of 0.1 is used to generate the data. Dynamic noise is added to each of the variables of the system at each sampling

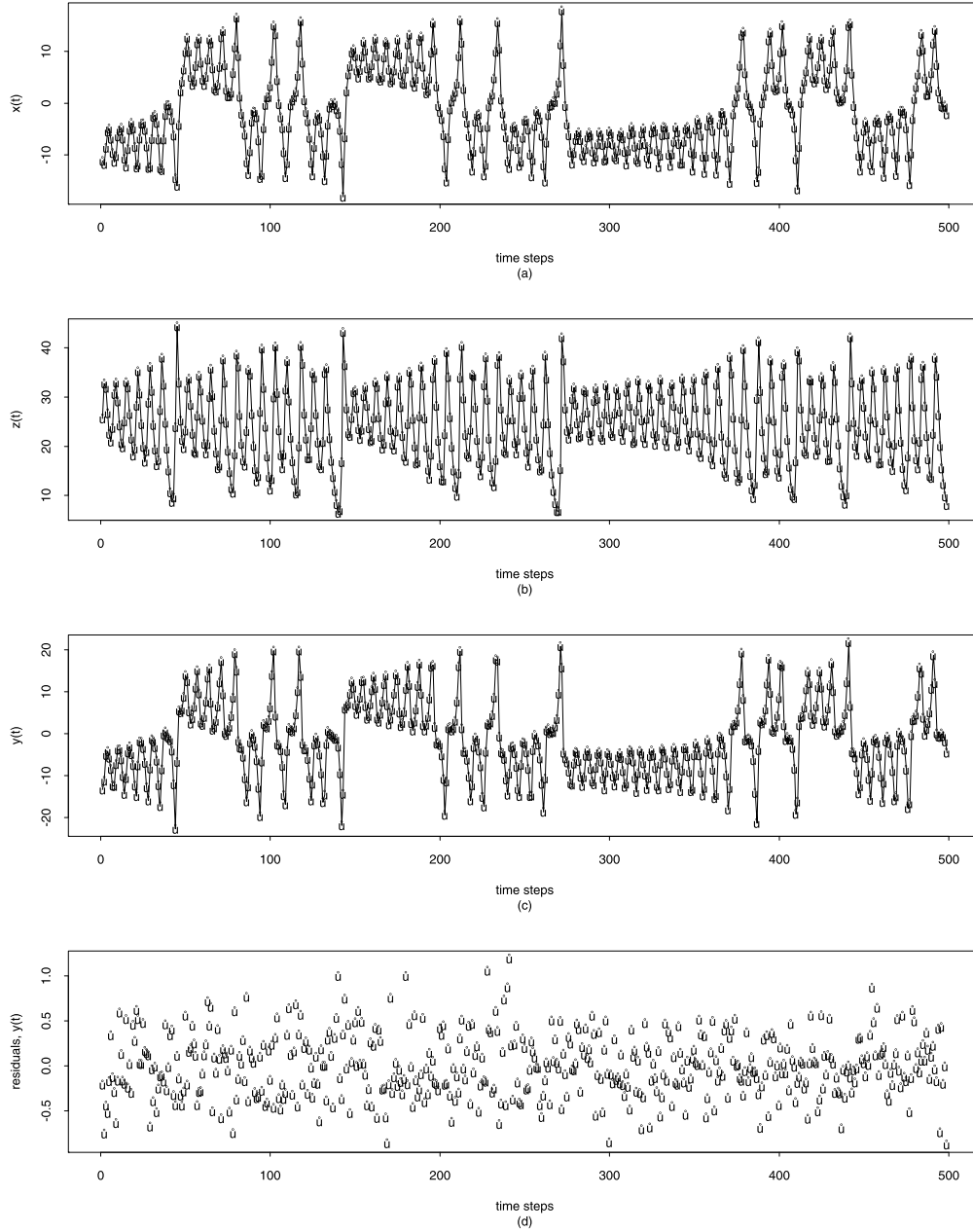


Figure 1. (a) Lorenz time series X_t of 500 data points. (b) Lorenz time series Z_t of 500 data points. (c) Lorenz time series Y_t of 500 data points. (d) Residuals over time to the fit of Y_t .

time. A $\sigma^2 = 0.35^2$ is chosen for the variance of $\{e_{i,t}\}$ ($i = 1, \dots, 3$), where $\{e_{i,t}\}$ are mutually independent Gaussian random variables with mean zero and common variance σ^2 . The magnitude of noise is chosen so that the attractor or attracting set is not destroyed which can be verified by visual inspection (see Figure 2).

To be consistent with the example in the next Section 3.3 by following Berliner, Lu and Snyder [6], the first 6,000 time steps are discarded as transients and the next 500 time steps are saved as the dataset. The transients are discarded to

ensure the system has evolved toward the attractor or attracting set as seen in Figure 2. The Lorenz time series of $\{X_t\}$, $\{Y_t\}$, and $\{Z_t\}$ gives state-space dynamics of the form:

$$\begin{aligned} X_t &= F_1(X_{t-1}, Y_{t-1}, Z_{t-1}, \boldsymbol{\theta}_1) + e_{1,t} \\ Y_t &= F_2(X_{t-1}, Y_{t-1}, Z_{t-1}, \boldsymbol{\theta}_2) + e_{2,t} \\ Z_t &= F_3(X_{t-1}, Y_{t-1}, Z_{t-1}, \boldsymbol{\theta}_3) + e_{3,t}. \end{aligned} \quad (14)$$

The time series plots for the Lorenz system with noise described by (14) can be seen in Figure 1(a)–(c). The solid

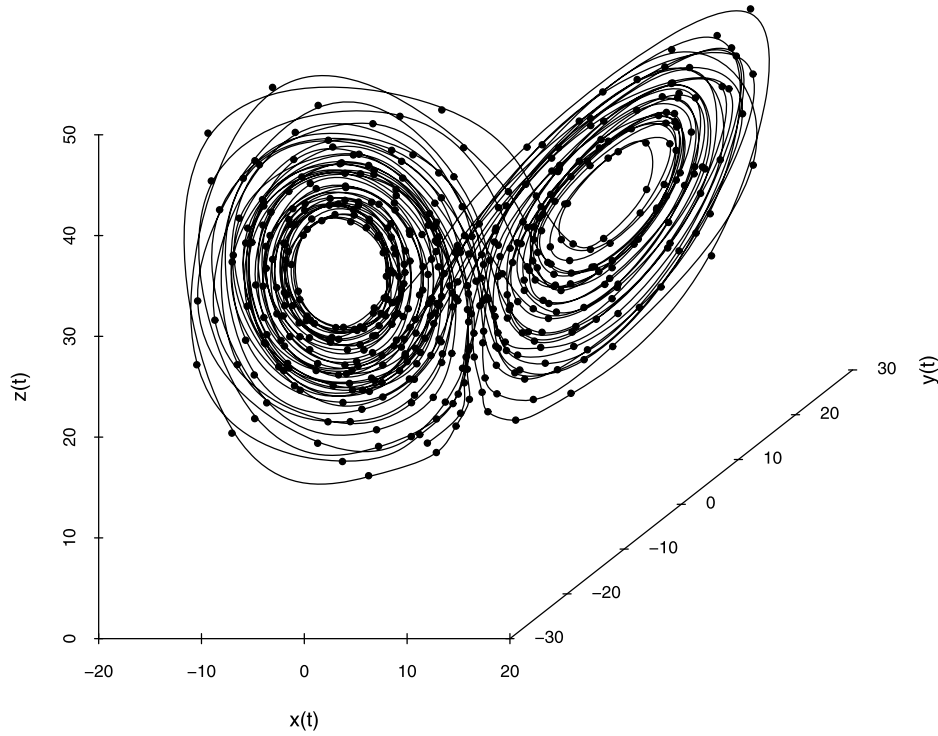


Figure 2. 3-d phase space for the Lorenz system.

dots represent the actual data points. For the choice of parameters and amount of noise, the familiar “butterfly” can be seen in the phase space plot of Figure 2.

The estimation of each of the maps F_i and corresponding model parameters θ_i of (14) will be necessary in order to estimate the quantities of interest. The estimated maps F_i will be used to estimate the Jacobian matrices in (2) and (3). A neural network with $k = 4, 5$, and 5 hidden units is chosen based on the minimum $GCV(p)$ for F_1 , F_2 , and F_3 , respectively. An examination and diagnostics of the fits and residuals indicates that the neural network is fitting the data well. The RMSE of the residuals is 0.368, 0.356, and 0.348, all very close to $\sigma = .35$. For example, Figure 1(d) is a time series plot of the residuals to the fit to Y_t indicating no systematic pattern.

Following the methods described in Section 2, Figure 3(a) shows the distribution of estimated local Lyapunov exponents as products of Jacobians increase from 5, 10, 20, 40, to 80. As n increases, the mean of the local Lyapunov exponents appears to converge to the global Lyapunov exponent (solid line). The estimated global Lyapunov exponent is $\hat{\lambda} = 0.075$, with a 95% CI (0.05, 0.08).

In this example, the 5-step ahead local Lyapunov exponents (LLE5s) are chosen for further analysis. There are 496 products of five Jacobians possible. The range of the LLE5s is divided into ten equally spaced groups or deciles. The smallest 10% of the LLE5s are labeled 1 and the largest 10% or tenth decile are labeled 10. Figure 3(b) is the time

series plot of the LLE5s coded by their size. The largest local exponents are of interest because they represent the times when the series is most unpredictable. The 10’s often appear at the bottom of the butterfly, just before the trajectories separate to one of the two “wings”. The Lorenz attractor has been extensively studied and this is consistent with previous results [14, 30]. The largest 10% of the LLE5s are indicated by a solid dot on the 3-d plot of Figure 4(b). There is also almost no correlation (0.005) between the residuals in Figure 1(d) and the LLE5s in Figure 3(b).

It is important to note that if there is interest in the times when the series is most predictable, a corresponding analysis of the smallest LLE5s is possible. The LLE5s labeled with 1 in Figure 3(b) are negative and represent the places in the time series (and regions in the phase space) where there is short-term trajectory convergence.

In this example, the 5-step ahead local Lyapunov exponents (LLE5s) are used to demonstrate the usefulness of LLEs. Examining the distribution of LLEs for different values of n is also useful in describing the changes in predictability of a system. Figure 3(a) demonstrates a Central Limit Theorem and that an $n = 5$ and $n = 10$ describe the heterogeneity of the system, but for an $n = 80$, the distribution of the LLE80’s have a small variance. Some applications have a natural time step that is important. For example, when quantifying the predictability of quarterly log transformed measles cases, the interest may be in a 4-step ahead or one-year ahead predictability since there is seasonality in incidence of measles cases [4].

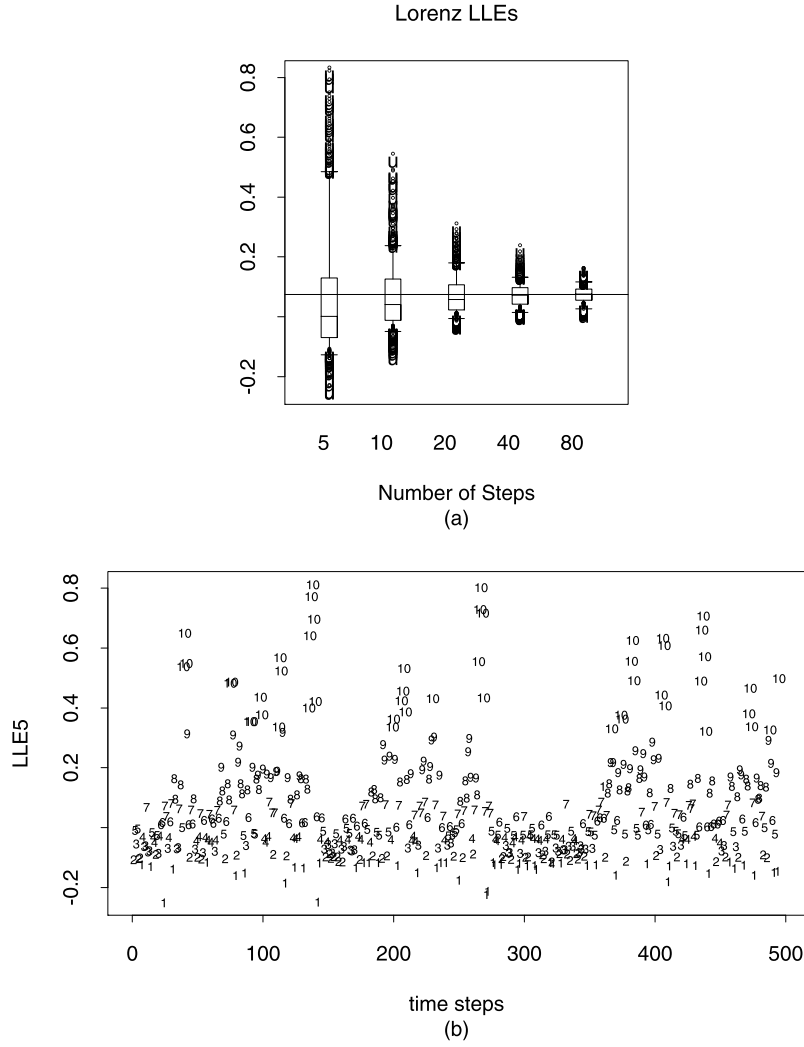


Figure 3. (a) Distribution of LLEs for the Lorenz system. (b) Time series of the 5-step ahead local Lyapunov exponents (LLE5s) for the Lorenz system coded by size.

This example is encouraging and demonstrates that a nonlinear time series approach to estimate local Lyapunov exponents is useful in detecting unpredictable behavior in data generated from a noisy or stochastic nonlinear system.

3.3 The Lorenz forty-variable system

The Lorenz system described in this section is a low-order model that resembles some aspects of the atmosphere. The model consists of variables defined at J points and evolving according to the governing equations:

$$(15) \quad \frac{dx_j}{dt} = -x_{j-2}x_{j-1} + x_{j-1}x_{j+1} - x_j + f,$$

for $j = 1, \dots, J$ and where f is a forcing. The variable x_j is periodic:

$$(16) \quad x_{j+J} = x_j.$$

The periodic condition of (16) means that the variables form a cyclic chain and may represent the values of some unspecified scalar meteorological quantity, for example, vorticity or temperature [18]. The J equally spaced sites can be thought of as extending around a latitude circle. It is a simple system that treats all variables alike and shares certain properties with many atmospheric models. The physics of the atmosphere is present to the extent that the constant terms represent external forcing, the linear terms represent internal dissipation, and the quadratic terms simulate advection. A more complete description of the system can be found in Lorenz [18] and Lorenz and Emanuel [19].

Following Lorenz and Emanuel [19], for this example, $J = 40$ and $f = 8$. The complexity of the system is controlled by the value of f . Lorenz and Emanuel [19] deduced that for values of f tending up towards 8, the system undergoes a bifurcation from periodic to chaotic behavior. Lorenz and Emanuel [19] assumed the time unit is the dissipative

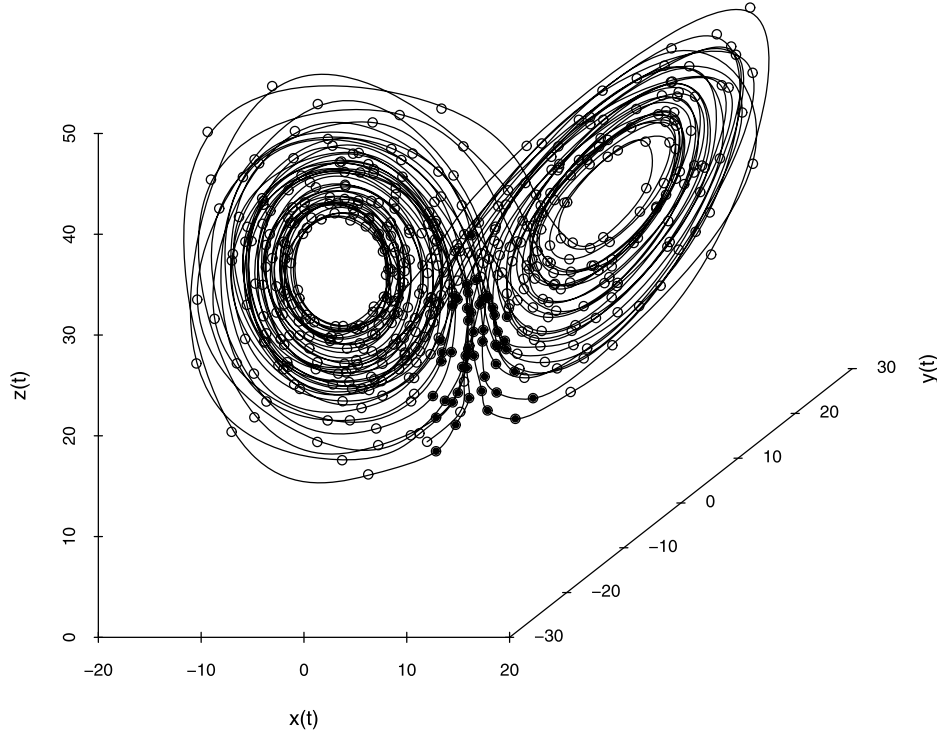


Figure 4. 3-d phase space for the Lorenz system coded to identify the times at which the LLE5 is in the top 10% of the distribution of LLE5s, indicated by solid dots.

decay time equal to 5 days. The system was found to be computationally stable using a fourth-order Runge-Kutta scheme with a time-step of 0.05 units (or 6 hours) and the global Lyapunov exponent corresponded to a doubling time of 0.42 units, or 2.1 days. These findings are used to generate the following data and to compare the results.

Data is generated from the Lorenz forty-variable system by numerically integrating with a fixed time step and sampling at every time step. A fourth-order Runge-Kutta scheme with step size of 0.05 is used to generate the data. Dynamic noise with a $\sigma^2 = 0.3^2$ is chosen for the variance of $\{e_{j,t}\}$ ($j = j, \dots, 40$) and is added to each variable of the system at each sampling time. As in the previous example, the noise $\{e_{j,t}\}$ are mutually independent Gaussian random variables with mean zero and common variance σ^2 . The first 6,000 time steps are discarded as transients and the next 1,000 time steps are saved as the dataset. Berliner, Lu and Snyder [6] used this model to generate data for a Bayesian experimental design study. Observational noise with a $\sigma^2 = 0.3^2$, which is assumed to be Gaussian noise was added to the data after a certain number of interactions or integration steps of the model. Here we choose this magnitude of noise for dynamic noise and the residuals and analysis of the fits described later indicates that it is not too large.

Since all variables are treated alike, X_{10} is chosen arbitrarily as the site to model. The discrete time dynamical

system corresponding to X_{10} using the the governing equation of (16) is

$$(17) \quad X_{10,t} = F(X_{8,t-1}, X_{9,t-1}, X_{10,t-1}, X_{11,t-1}, \theta_{10}) + e_{10,t}.$$

To construct the corresponding state-space system, it will be necessary to model the dynamics of $\{X_{8,t}\}$, $\{X_{9,t}\}$, $\{X_{10,t}\}$, and $\{X_{11,t}\}$ giving state-space dynamics of the form:

$$(18) \quad \begin{aligned} X_{8,t} &= F_8(X_{6,t-1}, X_{7,t-1}, X_{8,t-1}, X_{9,t-1}, \theta_8) + e_{8,t} \\ X_{9,t} &= F_9(X_{7,t-1}, X_{8,t-1}, X_{9,t-1}, X_{10,t-1}, \theta_9) + e_{9,t} \\ X_{10,t} &= F_{10}(X_{8,t-1}, X_{9,t-1}, X_{10,t-1}, X_{11,t-1}, \theta_{10}) + e_{10,t} \\ X_{11,t} &= F_{11}(X_{9,t-1}, X_{10,t-1}, X_{11,t-1}, X_{12,t-1}, \theta_{11}) + e_{11,t}. \end{aligned}$$

The system can be thought of to represent equally spaced points on a latitude circle, therefore there is a “neighborhood” structure defined and therefore implicit spatial dependence in the model. Since the dependence is known from the equations of (15), we choose to model this spatial dependence with (18). Alternatively one could model all forty-variables, but this would increase the dimension of the Jacobian matrix making numerical computations more difficult.

As in the Lorenz three-variable example of Section 3.2 it will be necessary to estimate each of the maps F_j of (18). A neural network with $k = 6, 5, 5$, and 4 hidden units is chosen

Lorenz96 LLEs

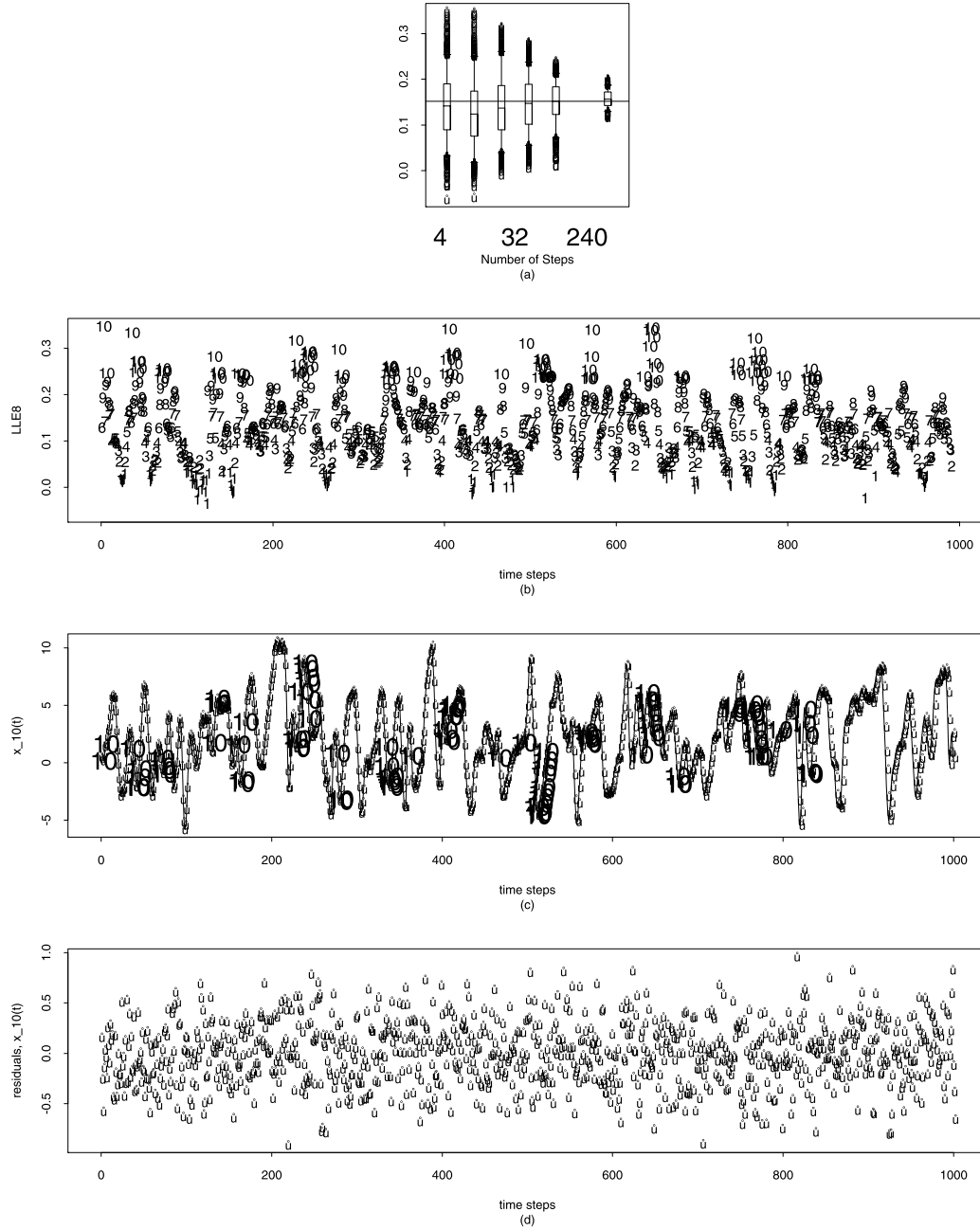


Figure 5. (a) Distribution of LLE8s for the Lorenz forty-variable system. (b) Time series of the LLE8s coded by size. (c) $X_{10,t}$ time series coded to identify the times at which the LLE8 is in the top 10% of the distribution, indicated by 10's. (d) Residuals over time to the fit of $X_{10,t}$.

based on the minimum $GCV(p)$ for F_8 , F_9 , F_{10} , and F_{11} , respectively. Again, an examination and diagnostics of the fits and residuals indicates that the neural network is fitting the data well. For example, Figure 5(d) is a time series plot of the residuals to the fit to $X_{10,t}$ indicating no systematic pattern. The RMSE of the residuals is 0.323, 0.316, 0.3183, and 0.326, all close to $\sigma = 0.3$.

Figure 5(a) shows the distribution of estimated local Lyapunov exponents as products of Jacobians increase from 4, 8, 16, 32, 64, to 240. The estimated global Lyapunov exponent is $\hat{\lambda} = 0.152$ with a 95% CI (0.1396, 0.1784). This is close to the value of $\lambda = 0.119$ obtained by Lorenz and Emanuel [19], for a system without noise. There is a difference due to the fact that noise is added to the system.

In this example, the 8-step ahead space-time local Lyapunov exponents (LLE8s) are chosen for further analysis. These exponents correspond to predictability two-days ahead. Figure 5(b) shows the range of these exponents divided into ten equally spaced groups or deciles. The smallest 10% of the LLE8s have been labeled 1 and the largest 10% have been labeled 10. Figure 5(b) is the time series plot of the LLE8s coded by their size. The top 10% of the distribution of LLE8s indicated by a 10 are overlaid on the time series in Figure 5(c). There are regions where there are clusters of 10's. For this system, those are the times where $X_{10,t}$ are most unpredictable two-days ahead. A close examination of Figure 5(b) indicates there are times when the 1's will be clustered and since the 1's are close to 0, these are the times when the system is most predictable two-days ahead. This example is encouraging and demonstrates that if there is known implicit spatial dependence of a neighborhood structure in the nonlinear stochastic system, a flexible neural network can be used to estimate space-time LLEs from data.

3.4 The cloud model

The examples in Sections 3.2 and 3.3 of the three-variable Lorenz system and the forty-variable Lorenz system with implicit spatial dependence demonstrated the usefulness of estimated local Lyapunov exponents in describing the predictability of a system. In these two examples, since the data was generated from known systems, we had the luxury of using the other known state variables in the neural network model and use the model selection criteria to determine the complexity in the neural network model. For the cloud model, we will build a model that predicts clouds based on our knowledge and understanding of the cloud process, which will include other state or predictor variables. We are interested in quantifying the short term predictability of the model, whether that be short term convergence or divergence of the system and use a spatial map to quantify and visualize the local Lyapunov exponents from a space-time process.

Clouds are visible, condensed water vapor and they play a fundamental and complex role in determining the amounts of solar and infrared radiation available to the climate system. Clouds reflect solar radiation and absorb infrared radiation emitted from the surface of the earth. Clouds transfer heat from the oceans to the atmosphere through evaporation and precipitation. Also, they are important in a variety of chemical processes in the atmosphere. See Trenberth [32], Emanuel [11], and Salby [25] for further discussion. This example will involve fitting nonlinear models to satellite and assimilated observational data. The estimates of space-time local Lyapunov exponents from a nonlinear regression model that describes the space-time evolution of cloud cover are used to characterize the stability of the nonlinear model.

The use of nonlinear regression models as parameterizations has been developed and studied in Bailey et al. [5]. A cloud parameterization is needed for atmospheric general circulation models when cloud processes vary over spatial scales that are smaller than the spatial grid that is used for computing the solutions of the numerical model of the atmosphere. Cloud parameterizations are important because clouds and their interactions with radiation represent the greatest uncertainty in climate projections [29]. Here is a brief description of nonlinear regression models in the spirit of a parameterization. For example, let C_{t+1} be the variable of interest to construct a parameterization:

$$(19) \quad C_{t+1} \approx F(\mathbf{x}_t, \boldsymbol{\theta}),$$

where F is a map and $\boldsymbol{\theta}$ is some collection of parameters. Let \mathbf{x}_t be the collection of physical variables that are resolved by the atmospheric general circulation model. The relationship in (19) is an approximation and the development of this approximation will be viewed as a statistical regression problem by considering the stochastic model

$$(20) \quad C_{t+1} = F(\mathbf{x}_t, \boldsymbol{\theta}) + e_{t+1}.$$

An alternative parameterization in the spirit of time series analysis is to build a time series regression model

$$(21) \quad C_{t+1} = F(\mathbf{x}_t, C_t, \boldsymbol{\theta}) + e_{t+1}.$$

For a more detailed description of the development of a parameterization from dynamical systems for selected atmospheric variables see Berliner et al. [7] and Bailey et al. [5].

A first-order time-lagged, spatial nearest-neighbor model is used. "First-order time-lagged" means that only information at time t is used to predict cloud cover amount at time $(t+1)$. "Spatial nearest-neighbor" means that predictions of cloud cover in a particular grid box (or site) at time $(t+1)$ are based on information available at time t at that site as well as nearby sites.

Consider a regular (equal area boxes), gridded region in two dimensions, with r rows and c columns. Define $S = r \times c$ and let $\mathbf{X}_t = (x_{1t}, \dots, x_{St})'$ and $\mathbf{C}_t = (C_{1t}, \dots, C_{St})'$ be S -vectors of values at time t . (Each of the x values may be vectors themselves but this is not reflected in the notation.) In general, $\mathcal{N}(s)$ represents a *neighborhood* or collection of sites defined to be neighbors of site s . For example, if we chose $\mathcal{N}(s)$ to be the site s itself and its four nearest neighbors, then the nearest-neighborhood for site s , with row-column index (i, j) , is the set

$$(22) \quad \mathcal{N}(s) = \{(i, j), ((i+1), j), ((i-1), j), (i, (j+1)), (i, (j-1))\}.$$

Note that there is a problem near the boundaries of the region in that some sites do not have modeled neighbors. For the purposes of this article, we take the most expedient solution and only fit the model for sites with complete neighbor

information. For example, if the neighborhood is described as in (22), we only fit the model at $S^* = (r - 2) \times (c - 2)$ sites.

A first-order lagged near-neighbor prediction model for each grid site s is

$$(23) \quad C_{s,t+1} = F(\mathbf{X}_{\mathcal{N}(s),t}, \boldsymbol{\theta}) + e_{s,t+1},$$

where $\boldsymbol{\theta}$ is a collection of model parameters and $e_{s,t+1}$ is a random model error. We assume that each $e_{s,t+1}$ has mean zero, and, for simplicity, that they are mutually uncorrelated random variables with common (unknown) variance σ^2 across both space and time.

The primary dataset used here consists of three months of hourly infrared radiation (IR) temperature satellite data collected during the TOGA COARE (Tropical Ocean Global Atmosphere Coupled Ocean-Atmosphere Response Experiment) [33]. From the IR temperature data, individual clouds have been identified by a numerical algorithm that starts from the coldest pixel center of each cloud in the image and then searches outward from the center. It stops at a specified threshold temperature and the pixels for that individual cloud are identified. Every pixel in the satellite image is identified as being a cloud or no cloud. Our cloud cover variable is the 0–1 (no cloud/cloud) pixel count of the image. The resolution of the cloud count data is a 0.045° or a 5km square grid. At a specified grid resolution for site s at time t , we define $C_{s,t}$ to be the cloud cover count.

We consider development of a first-order lagged near-neighbor neural network model. First, we consider three large-scale variables (described below) as predictors. In addition, we also consider inclusion of $C_{s,t}$'s in predicting $C_{s,t+1}$'s. This provides a time series or dynamical systems interpretation of the model.

To a significant extent cloud formation and evolution are controlled by (i) water vapor available for cloud formation and growth and (ii) the potential energy available in the atmosphere to enable convection (rising moisture). The variables are as follows:

- column mean relative humidity (RHM).
- time rate of change of generalized convective available potential energy (DCAPE).
- Richardson number (RICHN).

Relative humidity is defined as the ratio between actual and saturation vapor pressure. Saturation refers to the critical combination of temperature and pressure at which water vapor changes to liquid. Generalized convective available potential energy (CAPE) is a measure of the available (i.e., potential) energy which can be converted into kinetic energy, leading to convection. The derived variable, DCAPE, used here is actually a coarse measure (averaged over six-hour time steps) of the time rate of change of CAPE. The Richardson number also provides information relevant to the presence of convection, but involves physical reasoning (i.e.,

aspects of turbulence) beyond our scope here. Detailed descriptions and discussion of these variables can be found in Emanuel [11].

Information regarding these three predictor variables coinciding with our three-month cloud dataset are available as model output from Florida State University (FSU) [16]. The FSU data are available on a six-hourly time scale: four times/day at 0Z, 6Z, 12Z, and 18Z, on a 0.7° grid. These data can be used to calculate our three predictors variables RHM, DCAPE, and RICHN on the same space-time grid.

To assess the potential value of neural network statistical models for capturing the space-time evolution of cloud cover, we used the FSU grid (0.7°) in defining sites (as opposed to a coarser resolution corresponding to a typical AGCM). Cloud cover amount is defined based on the IR data, as described above. Recall that the cloud cover data is on a 0.045° (5km square) grid, so aggregating to the FSU grid yields approximately 15^2 cloud count pixels in each grid site.

Let $C_{s,t}$ be the cloud cover at FSU-grid site s at time t . Let $DCAPE_{s,t}$, $RICHN_{s,t}$, and $RHM_{s,t}$ be the values of the three explanatory variables, also at grid site s at time t . For each site s , a first-order time-lagged spatial near-neighbor model for $C_{s,t+1}$ is

$$(24) \quad C_{s,t+1} = F(\mathbf{RHM}_{\mathcal{N}(s),t}, \mathbf{DCAPE}_{\mathcal{N}(s),t}, \mathbf{RICHN}_{\mathcal{N}(s),t}, \boldsymbol{\theta}) + e_{s,t+1}.$$

The errors are all assumed to be mutually uncorrelated random variables, with common mean zero and unknown variance σ^2 . If we chose $\mathcal{N}(s)$ to be the site s itself and its four nearest neighbors, the set $\mathbf{X}_{\mathcal{N}(s),t}$ in (23) is a fifteen dimensional vector.

Though the sampling time-step for the FSU data is six hours, the time step of the cloud data and our prediction model is one hour. We hold $RHM_{s,t}$, $DCAPE_{s,t}$, and $RICHN_{s,t}$ constant in the six-hour period. (Alternatively, one might consider interpolation schemes to create the explanatory variables at an hourly rate; for brevity, we do not do so here.) Hence, the model results are presented in six-hour increments.

Next, we consider models that incorporate $C_{s,t}$'s in predicting $C_{s,t+1}$'s. Specifically, the models for $C_{s,t+1}$ are of the form

$$(25) \quad C_{s,t+1} = F(\mathbf{RHM}_{\mathcal{N}(s),t}, \mathbf{DCAPE}_{\mathcal{N}(s),t}, \mathbf{RICHN}_{\mathcal{N}(s),t}, \mathbf{C}_{\mathcal{N}(s),t}, \boldsymbol{\theta}) + e_{s,t+1},$$

with the same assumptions regarding the errors as in (24). We considered the case where $\mathbf{C}_{\mathcal{N}(s),t} = C_{s,t}$. That is, only clouds at the site will be included in the model which is the form of (21).

In the previous examples of the Lorenz three-variable and forty-variable systems, the systems are numerically integrated and it is possible to discard the first 6,000 time steps

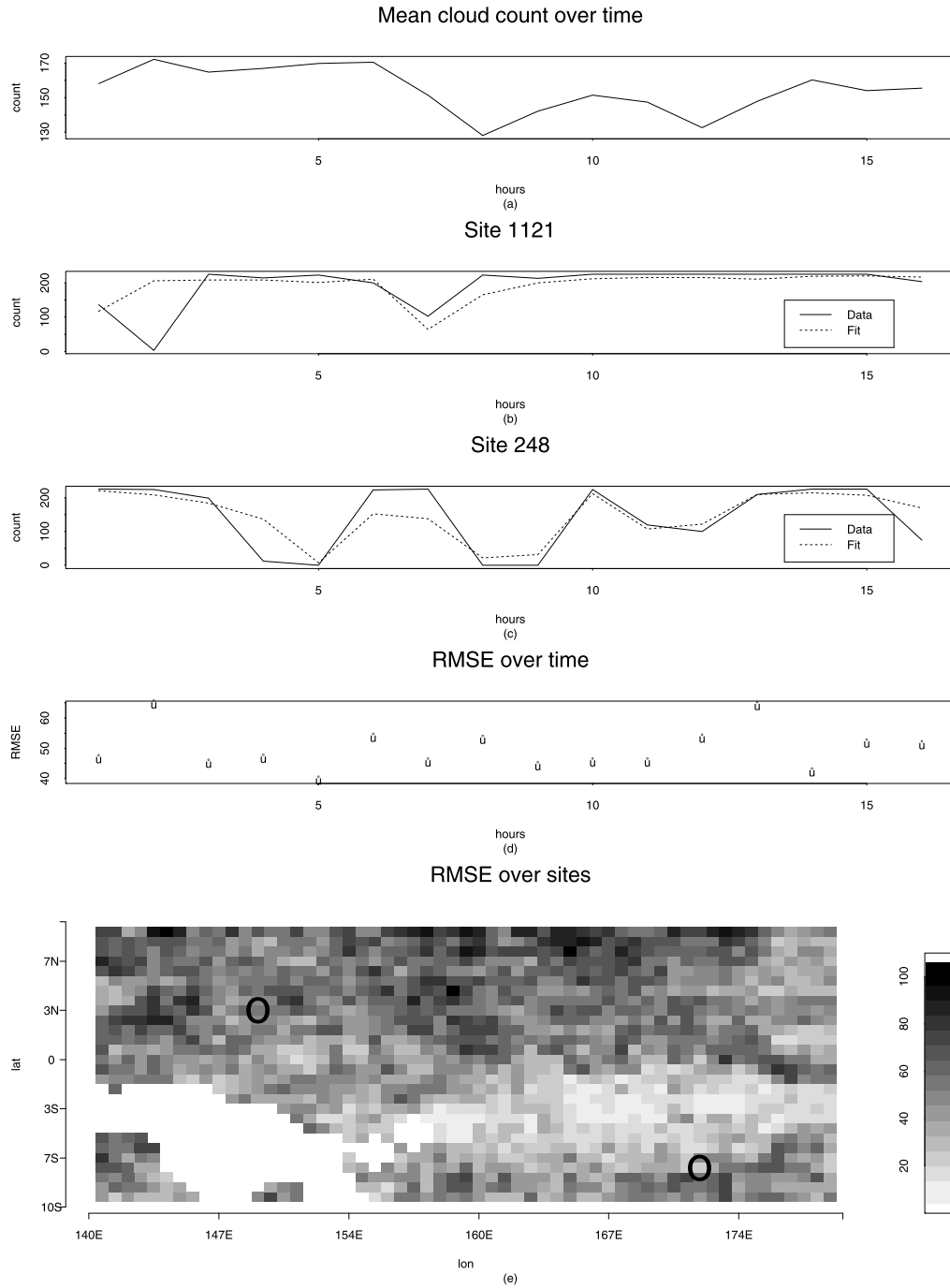


Figure 6. (a) Mean cloud cover over time for convective time period. (b) and (c) Cloud cover and predicted cloud cover for two selected sites, circled on (e). (d) Plot of RMSE of the sites over time. (e) Image plot of the RMSE over sites for the convective time period.

to assure that the system has converged to the attracting set or stationary distribution. If observational data is used, visual inspection of time series plots are useful in checking whether assumptions of stationarity are reasonable. Based on the average values of outgoing longwave radiation (OLR) over the region, the convective period 12/21/92–12/24/92 is chosen for the analysis [8]. Recall the sampling time for the

FSU data is 6-hourly, but the time step for the prediction (cloud cover data) is one hour. Therefore it is important to note for the following model results that the time is in 6-hourly increments and the time period of 16 6-hour increments spans four days. Examination of time series plots indicates that the convective period (or heavy clouding) is reasonably stationary. Figure 6(a) is a time series plot of

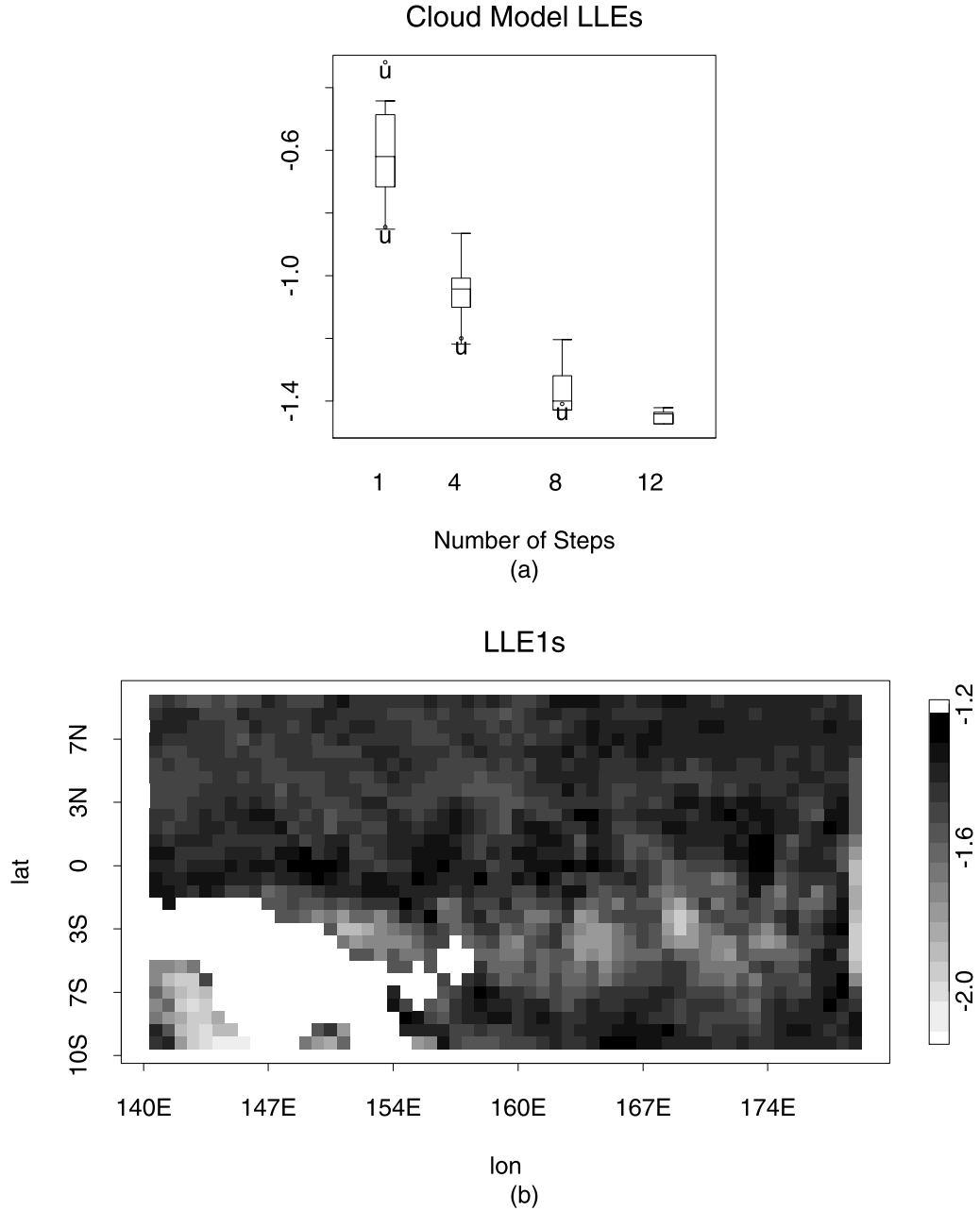


Figure 7. (a) Distribution of the local Lyapunov exponent of the cloud model. (b) Image plot of the mean LLE1s over the convective time period for the cloud model.

the mean cloud cover over the convective time period and Figure 6(b) and (c) are time series plots of cloud cover (solid line) at a two selected sites over the convective time period described below.

Figure 6(b)–(e) is a summary of the fit to (24) with eight hidden units. An examination of the fit and residuals indicates that the neural network model is capturing how cloud cover evolves reasonably well. Figure 6(d) is a plot of RMSE, averaged over sites, for each hour of the convective time period. The neural net fit resulted in an average of 22% RMSE over the convective time period. In comparison, Figure 6(e)

is an image plot of RMSE for all sites averaged over time. Two sites with contrasting RMSE have been chosen for comparison. The location of the sites have been circled on image plot of Figure 6(e). Sites 248 is in the upper left corner and Site 1121 is in the lower right corner. The white areas on the image plot correspond to land where no analyses were performed. Figure 6(b) and (c) are the time series plots of the data and fitted (or predicted) values for the two individual sites.

Figure 7(a) is a boxplot of the 1,4,8, and 12-step ahead space-time LLEs for the cloud system for the 16 hour con-

vective time period. These LLEs are computed from the estimated map of (24). They are all < 0 indicating the system is stable for the selected ahead time step. This weak stability is acceptable for the cloud cover model.

As a summary plot, Figure 7(b) is an image plot of the mean over time for the one-step ahead LLEs for the cloud model. This can be interpreted as making a perturbation at a grid cell and quantifying the convergence of the perturbation for each time t , and then averaging these values over the time period. The image plot indicates there are some regions of greater stability than others, indicated by the darker areas. Overall, the range is small, -2 to -1.2 , indicating there are not large spatial differences in the local stability. Notably, there is practical value in comparing Figure 6(e) and Figure 7(b), which indicate both the model fit and model stability.

4. CONCLUSION

A nonlinear autoregressive process and a nonlinear nearest-neighbor space-time process are useful models for data. The nonlinear space-time modeling approach can be used to estimate important dynamical system quantities from data, along with measures of uncertainty. The extension of a finite-time local Lyapunov exponent to include implicit spatial dependence, as a space-time local Lyapunov exponent, is useful in describing the space-time dynamics and predictability of data from atmospheric processes.

In the first example, a flexible neural network model fit to data generated from a nonlinear stochastic system produced results consistent with what is known about the Lorenz system. The estimated Lyapunov exponent and LLEs describe the unpredictable behavior of the system and identified the most unpredictable (short-term trajectory divergence) places in the time series and regions in phase.

In the second example, the data is generated from a stochastic nonlinear system with known spatial dependence. The neural network model residuals showed no systematic pattern and the estimated LLEs identified the most predictable and unpredictable places in the time series. This example shows the useful information obtained by extending the LLE to a setting with implicit spatial dependence.

The evolution of cloud cover over time and its space-time relationship to other climate variables is another interesting dynamical system and very important in the development of a cloud parameterization and climate modeling. The use of a spatial nearest-neighbor neural network to model cloud cover gives acceptable stability properties quantified by space-time local Lyapunov exponents.

The space-time dynamical system described in (11) is general and applicable to many nonlinear processes. The extension of the LLE to a spatial setting is an important step in understanding space-time stability. If there is a map with implicit spatial dependence on a lattice or grid over time, the interpretation of perturbing a grid cell and quantifying the

error growth in space and time is important in quantifying the stability properties of space-time processes. Summarizing and interpreting the space-time local Lyapunov process is an area of exciting future research.

Received 31 March 2011

REFERENCES

- [1] ABARBANEL, H. D. I., BROWN, R. and KENNEL, B. (1992). Local Lyapunov exponents computed from observed data. *Journal of Nonlinear Science* **2** 343–365. [MR1186765](#)
- [2] ABARBANEL, H. D. I., LEVINE, H., MACDONALD, G. and ROTHBAUS, O. (1990). Statistics of extreme events with application to climate. Technical Report No. JSR-90-305, JASON The MITRE Corporation, 7525 Colshire Drive, McLean, VA 22102-3481.
- [3] BAILEY, B. A., DONEY, S. C. and LIMA, I. (2004). Quantifying the effects of dynamical noise on the predictability of a simple ecosystem model. *Environmetrics* **15** 337–355.
- [4] BAILEY, B. A., ELLNER, S. and NYCHKA, D. W. (1997). Chaos with confidence: Asymptotics and applications of local Lyapunov exponents. In *Proceedings of the Fields/CRM Workshop on Nonlinear Dynamics and Time Series: Building a Bridge Between the Natural and Statistical Sciences* **11** 115–133. [MR1426617](#)
- [5] BAILEY, B. A., BERLINER, L. M., COLLINS, W., NYCHKA, D. and KIEHL, J. T. (2000). Neural networks: Cloud parameterizations. In *Case Studies in Statistics and the Atmospheric Sciences* (L. M. Berliner, D. Nychka and T. Hoar, eds.) 97–116. Springer-Verlag, New York.
- [6] BERLINER, L. M., LU, Z. Q. and SNYDER, C. (1999). Statistical design for adaptive weather observations. *Journal of the Atmospheric Sciences* **56** 2536–2552.
- [7] BERLINER, L. M., ROYLE, J. A., WIKLE, C. and MILLIFF, R. F. (1999). Bayesian methods in the atmospheric sciences. In *Bayesian Statistics 6* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) 83–100. Oxford University Press, New York. [MR1723494](#)
- [8] CHEN, S. S., HOUZE, R. A. and MAPES, B. E. (1996). Multiscale variability of deep convection in relation to large-scale circulation in TOGA COARE. *Journal of the Atmospheric Sciences* **53** 1380–1409.
- [9] ECKMANN, J. P. and RUELLE, D. (1985). Ergodic theory of chaos and strange attractors. *Review of Modern Physics* **57** 617–656. [MR0800052](#)
- [10] ELLNER, S. P., BAILEY, B. A., BOBASHEV, G. V., GALLANT, A. R., GRENFELL, B. T. and NYCHKA, D. W. (1998). Noise and nonlinearity in measles epidemics: Combining mechanistic and statistical approaches to population modeling. *American Naturalist* **15** 425–440.
- [11] EMANUEL, K. A. (1994). *Atmospheric Convection*. Oxford University Press, New York.
- [12] GALLANT, A. R. and WHITE, H. (1992). On learning the derivatives of an unknown mapping with multilayer feedforward networks. *Neural Networks* **5** 129–138.
- [13] GOLLUB, J. P. and BENSON, S. V. (1980). Many routes to turbulent convection. *Journal of Fluid Mechanics* **100** 449–470.
- [14] GUCKENHEIMER, J. (1976). A strange, strange attractor. In *The Hopf Bifurcation and its Applications* (J. E. Marsden and M. McCracken, eds.) Springer-Verlag, New York. [MR0494309](#)
- [15] GUCKENHEIMER, J. (1982). Noise in chaotic systems. *Nature* **298** 358–361.
- [16] KRISHNAMURTI, T. N., JHA, B., RASCH, P. J. and RAMANATHAN, V. (1997). A high resolution global reanalysis highlighting the winter monsoon. Part I, reanalysis fields. *Meteorology and Atmospheric Physics* **64** 123–150.

- [17] LORENZ, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences* **20** 282–293.
- [18] LORENZ, E. N. (1996). Predictability: A problem partially solved. In *Proceedings Seminar on Predictability*, Vol. 1, 1–18. ECMWF, Reading, Berkshire, UK.
- [19] LORENZ, E. N. and EMANUEL, K. A. (1998). Optimal sites for supplementary weather observations: Simulations with a small model. *Journal of the Atmospheric Sciences* **55** 399–414.
- [20] MCCAFFREY, D. F., ELLNER, S., GALLANT, A. R. and NYCHKA, D. W. (1992). Estimating the Lyapunov exponent of a chaotic system with nonparametric regression. *Journal of the American Statistical Association* **87** 682–695. [MR1185190](#)
- [21] NYCHKA, D. W., ELLNER, S., MCCAFFREY, D. F. and GALLANT, A. R. (1992). Finding chaos in noisy systems. *Journal of the Royal Statistical Society, Series B* **54** 399–426. [MR1160477](#)
- [22] NYCHKA, D. W., HAALAND, P., OCONNEL, M., BAILEY, B. A. and ELLNER, S. (1998). FUNFITS data analysis and statistical tools of estimating functions. Statlib, Carnegie Mellon University.
- [23] PORCHER, R. and THOMAS, G. (2001). Estimating Lyapunov exponents in biomedical time series. *Physical Review E* **64** 010902–010904.
- [24] RIPLEY, B. D. (1996). *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge, UK. [MR1438788](#)
- [25] SALBY, K. A. (1996). *Fundamentals of Atmospheric Physics*. Academic Press, New York.
- [26] SCHUSTER, H. G. (1992). *Deterministic Chaos: An Introduction*, 2nd ed. VCH, Weinheim. [MR0935128](#)
- [27] SEBER, G. A. F. and WILD, C. J. (1989). *Nonlinear Regression*. John Wiley & Sons. [MR0986070](#)
- [28] SHINTANI, M. and OLIVER, L. (2004). Nonparametric neural network estimation of Lyapunov exponents and a direct test for chaos. *Journal of Econometrics* **120** 1–22. [MR2047779](#)
- [29] SOLOMON, S., QIN, D., MANNING, M., CHEN, Z., MARQUI, M., AVERYT, M., TIGNOR, M. and MILLER, H. L., eds. (2007). *Contributions of Working Group I to the Fourth Assessment Report on the Intergovernmental Panel on Climate Change, 2007*. Cambridge University Press, Cambridge.
- [30] SPARROW, C. (1982). *The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors*. Springer-Verlag, New York. [MR0681294](#)
- [31] TAKENS, F. (1981). *Lecture Notes in Mathematics, Dynamical Systems and Turbulence*, Vol. 898. Springer, Berlin. [MR0654900](#)
- [32] TRENBERTH, K. E., ed. (1992). *Climate System Modeling*. Cambridge University Press, London.
- [33] VELDEN, C. S. and YOUNG, J. A. (1994). Satellite observations during TOGA COARE: Large-scale descriptive overview. *Monthly Weather Review* **122** 2426–2441.
- [34] WAHBA, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia. [MR1045442](#)
- [35] WOLF, A., SWIFT, J. B., SWINNEY, H. L. and VASTANO, J. A. (1985). Determining Lyapunov exponents from a times series. *Physica* **16D** 285–315. [MR0805706](#)
- [36] WOLFF, R. C. L. (1992). Local Lyapunov exponents: Looking closely at chaos. *Journal of the Royal Statistical Society Series B* **54** 353–371. [MR1160475](#)
- [37] YONEMOTO, K. and YANAGAWA, T. (2007). Estimating the Lyapunov exponent for chaotic time series with dynamic noise. *Statistical Methodology* **4** 461–480. [MR2432449](#)

Barbara A. Bailey
 Department of Mathematics and Statistics
 San Diego State University
 San Diego, CA 92182
 USA
 E-mail address: babailey@sciences.sdsu.edu
 url: www.rohan.sdsu.edu/~babailey