

# Chaos with Confidence: Asymptotics and Applications of Local Lyapunov Exponents

**Barbara A. Bailey**

**Stephen Ellner**

**Douglas W. Nychka**

Biomathematics Graduate Program

Department of Statistics

North Carolina State University, Raleigh, NC 27695-8203

**Abstract.** In this paper we define a version of Local Lyapunov Exponents (LLEs) for discrete-time dynamical systems perturbed by noise, which are related to short-term predictability and noise amplification and thus can be used to detect regions of state space where the dynamics may be more predictable. We present a Central Limit Theorem for fluctuations of LLEs about the global LE, as the number of forward time steps used in computing LLEs is increased. We also present some large-sample statistical properties when LLEs are estimated by fitting a statistical model to a time series, that justify the construction of confidence intervals for maximum likelihood estimates.

As an application of these results we analyze time series of measles monthly case reports, using a neural network time series model. The results show large changes in dynamic structure and predictability over the course of an epidemic.

Because LLEs provide a detailed description of a systems' nonlinearity, they are well suited to compare mechanistic models with data. Comparison of observed with model LLEs shows that current models for measles do not correctly predict the regions of state space with the highest short-term sensitivity to initial conditions. This disagreement is not apparent in more common measures used to assess these models.

## 1 Introduction

The global Lyapunov exponent of a dynamical system measures the average rate at which nearby trajectories diverge. A positive exponent indicates "sensitive dependence on initial conditions", hence chaotic dynamics and unpredictability, but there may still be short-term trajectory convergence and hence higher predictability in some regions of state space. Conversely, a non-chaotic system may have short-term trajectory divergence in some regions of state space. One way of quantifying these fluctuations is by calculating finite-time "local" Lyapunov exponents (LLEs).

Local Lyapunov exponents provide a more detailed description of a system's dynamics than just the global Lyapunov exponent. In many real data sets, such as case reports of measles, and field and laboratory population dynamics, the global

---

1991 *Mathematics Subject Classification.* Primary 62M10, 58F13; Secondary 60F05.

exponent is close to zero and the system is near the “transition to chaos” (Ellner and Turchin [1995]). In these cases, it may be more useful to quantify the short-term convergence (or divergence) of trajectories and thereby determine whether or not the system exhibits “local” chaos. Local Lyapunov exponents therefore provide a useful tool for characterizing nonlinear dynamics. The analysis involves examining the short term predictability (or unpredictability) of the system and identifying the parts of the time series or regions in state space when they occur (Abarbanel, Brown, and Kennel [1992], Smith [1992], Wolff [1992]).

The paper is organized as follows. A local Lyapunov exponent is defined and estimation procedures presented in section 2. Section 3 is a discussion of the Central Limit Theorem behavior of local Lyapunov exponents. Construction of confidence intervals based on maximum likelihood estimates is described in section 4, including simulation results of coverage probabilities. Finally, section 5 contains applications analyzing time series of measles cases and model output. The analysis includes the comparison of measles dynamics between Copenhagen and New York City case data and the comparison between Copenhagen case data and output from a stochastic model.

The theory is presented mostly for the case of scalar data  $x_t$ , fitted by a model involving past values  $x_{t-1}, x_{t-2}, \dots$ , which is the most common in applications, but exactly parallel results hold for “state-space” models with vector data  $X_t$  fitted by a model involving only  $X_{t-1}$ . Our technical assumptions were motivated by the time-series model that we have found most useful for estimating LLEs from data, a feedforward neural network with Gaussian noise. The neural network model has the advantage that Lyapunov Exponent estimates are robust to mis-estimation of the dimension of the model (McCaffrey et al. [1992]). For measles data in particular, we found that neural net models had the highest out-of-sample forecasting accuracy of a suite of statistical and mechanistic models (Ellner et al. [1996]).

## 2 Definition of Local Lyapunov Exponents

We assume that the data  $\{x_t\}$  are a time series generated by a nonlinear autoregressive model

$$x_{t+1} = f(x_t, x_{t-1}, \dots, x_{t-d+1}) + \varepsilon_t \quad (2.1)$$

where  $x_t \in \mathfrak{R}$  and  $\{\varepsilon_t\}$  is a sequence of independent random variables or perturbations with  $E(\varepsilon_t) = 0$  and  $\text{Var}(\varepsilon_t) = \sigma^2$ . It is important to note that the error in (2.1) is not measurement error, but dynamic noise, an inherent part of the dynamics of the system. The state-space representation of the system is

$$X_{t+1} = F(X_t) + e_t \quad (2.2)$$

where  $X_t = (x_t, x_{t-1}, \dots, x_{t-d+1})$  and  $e_t = (\varepsilon_t, 0, 0, \dots, 0)$  are in  $\mathfrak{R}^d$ .

Intuitively, the LLE is defined by making a small perturbation of  $X_t$  to  $X_t^*$ , and following forward the perturbed and unperturbed trajectories. The LLE measures the difference between the two trajectories after  $m$  time steps,

$$\lambda_m(t) = \frac{1}{m} \ln \frac{d(X_{t+m}, X_{t+m}^*)}{d(X_t, X_t^*)},$$

where  $d$  is a measure of distance.

Different versions of local Lyapunov exponents can be defined, depending on the choice of perturbation, the distance measure used, and the choice of which trajectories to follow forward (e.g., the average trajectory at  $X_t$ , the noise-free trajectory

starting at  $X_t$ , or the actual sample path from times  $t$  to  $t+m$ , in which case  $\lambda_m(t)$  is a random variable). We would not want to argue that there is a single “right” definition (though there can be “wrong” definitions, in which the local exponents do not converge to the global exponent as  $m \rightarrow \infty$ ). Rather, the choice should depend on what description of the dynamics is of interest. We choose to perturb only the first component of  $X_t$  and follow the growth of that perturbation along the actual sample path. This corresponds to a perturbation of the system (2.1) at time  $t$ , i.e. perturbing  $x_t$  and then following the growth of that single perturbation along the actual sample path. In the state-space representation,  $x_t$  is the first component of  $X_t$ . One advantage of perturbing only the first component is that if (2.1) is regarded as a reconstruction in time-delay coordinates from one state variable  $x_t$  of a higher-dimensional system, then the local exponents for (2.1) correspond exactly to those for the true system when  $x_t$  is used as the first component of the state vector, for both finite and infinite  $m$ .

In addition, results of Yao and Tong [1994] show that this definition of the LLE is related to noise amplification by the system, and thus to predictability. Let  $\sigma_m^2(x) = \text{Var}(X_{t+m} | X_t = x)$ . In a scalar system with small additive noise, Yao and Tong [1994] showed

$$\sigma_m^2(x) = \sigma^2 \mu_m(x) (1 + o(1))$$

where  $\sigma^2 = \text{Var}(\varepsilon(t))$  and

$$\mu_m(x) = 1 + \sum_{j=1}^{m-1} \left\{ \prod_{k=j}^{m-1} f'(f^{(k)}(x)) \right\}^2.$$

This can be re-written as

$$\mu_m(x) = 1 + \sum_{j=1}^{m-1} \exp \left\{ 2(m-j) \lambda_{m-j}(f^{(j)}(x)) \right\} \quad (2.3)$$

where  $\lambda_j$  are the LLEs along the “skeleton” system with the noise deleted. However to leading order in  $\sigma$  in this expansion, the skeleton LLEs agree with those along the sample path (for any fixed finite  $m$ ). For a system with dynamic noise the value of  $\mu_m$  sets the limits to large-sample forecasting accuracy (Tong [1995]), so equation (2.3) shows a close tie (for small noise) between LLEs and local unpredictability.

As perturbations become small,  $\lambda_m(t)$  will depend on derivatives of the map  $F$ . Then

$$\lambda_m(t) = \frac{1}{m} \ln \| J_{m+t-1} J_{m+t-2} \cdots J_t U_0 \|, \quad (2.4)$$

where  $J_t$  (sometimes written  $J(X_t)$ ) is the Jacobian matrix of  $F$  evaluated at  $X_t$ ,  $U_0$  is a unit vector and  $\|\cdot\|$  is the Euclidean vector norm. Since  $\lambda_m(t)$  is a function of time, the LLE depends on the trajectory and can be thought of as an “m-step ahead” local Lyapunov exponent process. If  $U_0$  is chosen at random with respect to the uniform measure on the unit sphere, then with probability 1 as  $m \rightarrow \infty$ ,  $\lambda_m(t)$  converges to the global Lyapunov exponent, because  $U_0$  has zero probability of falling into the subspace corresponding to subdominant exponents. In practice, however, we usually take  $U_0 = (1, 0, 0, \dots, 0)$ .

### 3 The Limiting Distribution of Local Lyapunov Exponents

Associated with a local Lyapunov exponent is the number of forward time steps  $m$  in equation (2.4) that are used to calculate the quantity. As  $m$  becomes large the

distribution of local exponents will converge to a normal distribution. Figure 1 is a series of histograms of the LLEs from the Ikeda map which is described in section 4. As the number of steps  $m$  increases from 1 to 64, the histograms become closer to a normal distribution. The last plot in the series is a plot of the interquartile range vs the number of steps  $m$ , on a log scale. The slope of the line is approximately one-half, indicating  $1/\sqrt{m}$  scaling. In this section, a central limit theorem for local Lyapunov exponents will be stated.

The key to deriving the CLT result is the construction of a vector Markov chain that includes both the state vector and the Jacobian products. Let  $X_t$  be a Markov chain and consider the vector Markov chain  $W_t$  defined as

$$W_t = \begin{bmatrix} X_t \\ \cdots \\ U_t \end{bmatrix}$$

where  $X_t$  is the state vector and the dynamics of  $U_t$  are given by

$$U_t = \frac{J_t U_{t-1}}{\|U_{t-1}\|}.$$

Define  $h(W_t) = \ln(|U_t|)$ . Then it follows from equation (2.4) that

$$\lambda_m(t) = \frac{1}{m} \sum_{j=1}^m h(W_{t+j}).$$

Thus  $\lambda_m(t)$  is the average of a functional of this process. Let  $\lambda_m$  denote a LLE starting at any time  $t$ . A CLT for  $\lambda_m$  can therefore be derived by applying standard CLT results for functionals of Markov chains (Doob [1953]). The dependence on time  $t$  has been dropped because the CLT is independent of time.

### Assumption 3.1

1.  $F$  is a bounded map.
2.  $e_t$  is i.i.d. and has a bounded probability density function with respect to Lebesgue measure in  $\mathbb{R}^d$ .
3. Define  $\mathcal{J}_l = \prod_{j=1}^l J(X_j)$  and let  $\omega(z | x)$  be the probability density function of  $\text{vec}(\mathcal{J}_l^\top)$  conditional on  $X_0 = x$ . Assume that  $\omega(z | x)$  exists with respect to Lebesgue measure on  $\mathbb{R}^{d^2}$  and that it is uniformly bounded with respect to  $x$  in  $L^{1+\epsilon}(\mathbb{R}^{d^2})$  norm, for some  $\epsilon > 0$  and  $l > 0$ .
4.  $J(x)$  is uniformly bounded.
5. There is a single ergodic set for  $\{W_t\}$ .

**Theorem 3.1** (CLT) *If Assumptions 3.1 are satisfied, then*

$$\frac{\sqrt{m}(\lambda_m - \lambda)}{\sqrt{\text{Var}(\lambda_m)}} \rightarrow N(0, 1) \text{ as } m \rightarrow \infty.$$

The details of the proof of Theorem 3.1 are presented in Bailey [1996]. If there is more than one ergodic set then the limiting distribution is still normal but may depend on  $W_0$ . In the Appendix we give conditions for state space models to have a single ergodic set for  $W_t$ .

The role of nonlinearity and the noise  $e_t$  in CLT behavior is complex. If there is not much nonlinearity, i.e. the system is close to linear, then the LLEs will be nearly identical and therefore a CLT is not expected. Theiler and Smith [1995] show that if the map  $f$  is conjugate to a constant-slope map and the system is

noise free, then the Lyapunov exponents scale as  $1/m$  not  $1/\sqrt{m}$ . The meaning of Theorem 3.1 is that if the system has “enough” noise and nonlinearity, then the Lyapunov exponent does scale as  $1/\sqrt{m}$  and converge to Gaussian.

#### 4 Estimating Local Lyapunov Exponents and Construction of Confidence Intervals

The calculation of global and local Lyapunov exponents uses equation (2.4) and requires an estimate of the Jacobian matrix  $J_t$ . In all our examples the map  $f$  in (2.1) or each component of  $F$  in a state-space model is estimated as a feed-forward neural network with a single layer of hidden units (Ellner, Nychka, and Gallant [1992]). The form of the model is

$$F(X) = \beta_0 + \sum_{i=1}^k \beta_i \varphi(X^T \gamma_i + \mu_i) \quad (4.1)$$

where  $\varphi(u) = e^u/(1 + e^u)$ . The parameters  $\beta, \gamma$  and  $\mu$  are found by least squares. The complexity of the model, i.e. the embedding dimension and the number of hidden units  $k$ , is chosen based on generalized cross validation (GCV). The estimated Jacobian is then the derivative of the estimated  $F$ .

The construction of confidence intervals for the estimates is based on the identification of a joint confidence region for the model parameters  $\theta$ , using the likelihood ratio statistic. The approximate confidence set  $\mathcal{A}_\theta$  will be all the values of  $\theta$  such that

$$S(\theta) \leq S(\hat{\theta}) \left[ 1 + \frac{p}{n-p} F(p, n-p, \alpha) \right] \quad (4.2)$$

where  $S(\theta)$  is the residual sum of squares and  $\hat{\theta}$  is the least-squares estimate of  $\theta$ . For the F distribution,  $p$  is the number of parameters in the model,  $n$  is the sample size and  $\alpha$  is the probability level. This confidence set is based on the asymptotic normality of the nonlinear regression model parameters. Using a standard Taylor series argument, asymptotic normality of the nonlinear regression model parameters implies

$$-2 \log(L(\theta)/L(\hat{\theta})) \xrightarrow{d} \chi^2(p) \quad (4.3)$$

where  $L(\theta)$  is the likelihood function and  $\chi^2(p)$  denotes the chi-squared distribution with  $p$  degrees of freedom. The distributional result of (4.3) implies the finite sample size result of (4.2) (Seber and Wild [1989]).

In order to insure the asymptotic normality of the nonlinear autoregression parameters and therefore justify the confidence set defined in equation (4.2), the following assumptions are needed.

##### Assumption 4.1

1.  $X_t$  is stationary and ergodic.
2.  $e_t$  is i.i.d.  $N(0, \sigma^2)$ .
3.  $\theta \in \Theta$  and  $\Theta$  is compact.
4.  $F$ , first, second and third partials of  $F$  exist and are continuous and uniformly bounded for all  $\theta \in \Theta$ .

**Theorem 4.1** *Under Assumptions 4.1, there exists a maximum likelihood (ML) estimator  $\hat{\theta}_n$  of  $\theta$  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  $n \rightarrow \infty$ , where  $I_n(\theta)$  is the conditional information matrix.*

The proof of Theorem 4.1 is presented in Bailey [1996] and it involves showing Assumptions 4.1 imply assumptions stated by Hall and Heyde [1980]. Most importantly, Theorem 4.1 implies equation (4.3) which justifies the cutoff value used in equation (4.2) for a nonlinear autoregressive process.

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  $\theta$ . The residual sum of squares from the least squares estimate is then  $S(\hat{\theta})$  in equation (4.2). The goal of the second step is to find  $\theta$ 's in the confidence set. The procedure is to randomly generate many parameter starting values and perform on each a partial optimization or "rough" fit. The generation of parameter sets and the partial optimization is repeated until 500 parameter sets are in the confidence set. To make sure that the parameter sets cover the entire confidence region, 100 of the parameter sets are forced to be within 20% of the cutoff (4.2). For all  $\theta$ 's in the confidence region, an appropriate functional  $\varphi$  of the parameters can be evaluated. For example, the functional  $\varphi(\theta)$  could be the global Lyapunov exponent or the 10th or 90th percentile of the LLEs. The confidence interval (CI) for  $\varphi$  is then given by the minimum and maximum of  $\{\varphi(\theta_i)\}_{i=1}^{500}$  where  $\{\theta\}_{i=1}^{500}$  are the  $\theta$  values found in the confidence set  $\mathcal{A}_\theta$ .

We conducted a simulation study to examine the actual coverage probabilities of the approximate confidence intervals. Confidence intervals for the global Lyapunov exponent and functionals of the five step ahead local Lyapunov exponents (LLE5s) were generated by the techniques described above for the Ikeda map.

The Ikeda Map is a 2-dimensional nonlinear map (Smith [1995]):

$$\begin{aligned} x_t &= 1 + \mu x_{t-1} \cos \left( 0.4 - \frac{6}{x_{t-1}^2 + y_{t-1}^2 + 1} \right) \\ &\quad - \mu y_{t-1} \sin \left( 0.4 - \frac{6}{x_{t-1}^2 + y_{t-1}^2 + 1} \right) \end{aligned} \quad (4.4)$$

$$\begin{aligned} y_t &= \mu x_{t-1} \sin \left( 0.4 - \frac{6}{x_{t-1}^2 + y_{t-1}^2 + 1} \right) \\ &\quad + \mu y_{t-1} \cos \left( 0.4 - \frac{6}{x_{t-1}^2 + y_{t-1}^2 + 1} \right) \end{aligned} \quad (4.5)$$

For the simulation study we used  $\mu = 0.75$ . Dynamic noise was added to both  $x_t$  and  $y_t$ , giving state-space dynamics of the form

$$X_t = f_1(X_{t-1}, Y_{t-1}) + e_{1,t} \quad (4.6)$$

$$Y_t = f_2(X_{t-1}, Y_{t-1}) + e_{2,t} \quad (4.7)$$

The noise terms  $e_{i,t}$  were both given uniform distributions on  $(-0.1, 0.1)$ . The noise was uniform to make sure that the system stays in a locally attracting set. Figure 2(d) shows the distribution of points on the attracting set generated by this system.

A sample size of 400 was chosen for the simulation. For this sample size, 100  $(x_t, y_t)$  realizations were generated and a neural net was fit to each realization. Since there are two variables, a neural net was fit to each data set. For each pair of neural net fits there is a corresponding confidence interval for the estimate. Confidence intervals were generated for the global Lyapunov exponent and the 10th and 90th percentile of the five step ahead local Lyapunov exponents (LLE5s).

The results of the simulation study are summarized in "coverage probability" plots (Figure 2). The axes of this plot are the lower and upper limit of the 95% confidence interval. Each plotted point ( $\bullet$ ) represents the confidence interval from

one of the 100 data sets simulated. The solid lines on the plot are the true value of the quantity being estimated. The dotted lines on the plots are the expected value of the estimate, i.e. the mean of the 100 estimates corresponding to the best fit parameter values for each realization of  $(x_t, y_t)$ . The intersection of the solid lines divides the plot into four quadrants. The top left quadrant is the quadrant where the lower endpoint of the confidence interval is smaller than the true value, and the upper endpoint of the confidence interval is larger than the true value. This is the quadrant where the confidence interval is covering the true value. If points are in the bottom left quadrant then the lower endpoint of the confidence interval is below the true value but the upper endpoint is also smaller than the true value. Similarly, if points are in the top right quadrant then the upper endpoint of the confidence interval is above the true value but the lower endpoint is larger than the true value. It is impossible for points to be in the bottom right quadrant. A coverage probability of 95% corresponds to 95 out of the 100 points lying in the top left quadrant.

Figure 2(a) is the coverage probability plot for the 10th percentile of LLE5s. This plot shows that all confidence intervals cover the true value. The 90th percentile of LLE5s are in Figure 2(b). The coverage is not as good for this estimate. Fifteen confidence intervals do not cover the true value. However, all but one of the fifteen are in the bottom left quadrant where the intervals are “too small”, indicating that the model produces conservative estimates. Figure 2(c) is the coverage probability for the global LE. Three intervals do not cover the true value. The coverage probability plots from the Ikeda map simulation indicate that it is computationally feasible to construct confidence intervals using a reasonable sample of the likelihood surface. Also, when confidence intervals do not cover the true value they indicate that the model produces conservative estimates guarding against false positives. Other simulation studies to examine the coverage probabilities of the confidence intervals using different test systems showed similar results (Bailey [1996]).

## 5 Applications

The dynamics of infectious diseases with recurrent epidemics is an area where the LLE analysis can be useful. Because of the variable cycle of outbreaks, measles exhibits very complicated dynamics. A comparison of LLEs and their fluctuations in measles data from different cities may provide insight into the similarities and differences in the dynamics of disease incidence. Stochastic models have attempted to capture the dynamics of measles. Local Lyapunov exponents can be used as a “probe” into the adequacy and inadequacy of these models.

**5.1 Comparison between Cities.** The time series of quarterly measles cases from two cities, New York City and Copenhagen, were chosen for comparison. The model fit to the log transformed case data ( $\log C_t$ ) involved not only lagged values of the log cases, but also a component for the number of susceptibles in the population at a given time. The susceptibles from the population are reconstructed from the reported cases data following the procedure described in Bobashev et al. [1995]. The mass balance equation for the number of susceptibles ( $S_t$ ) in a population at time  $t$  is a function of the number of cases ( $C_t$ ) and the birth rate ( $b_t$ ). The following set of equations describe the number of cases and susceptibles at a given

time.

$$\log C_{t+1} = f(\log C_t, S_t, \sin(2\pi t/4), \cos(2\pi t/4)) + e_t \quad (5.1)$$

$$S_{t+1} = S_t - C_{t+1} + b_{t+1} \quad (5.2)$$

In the first equation,  $f$  is estimated by a neural net and  $\sin(2\pi t/4)$  and  $\cos(2\pi t/4)$  are seasonal forcing terms with periods corresponding to quarterly data. The distribution of LLEs for both of the cities is seen in Figure 3. The solid line represents the estimated global Lyapunov exponent. The distribution of LLEs for both cities is similar.

Figure 4 shows a series of two plots to describe the dynamics of measles. Figure 4(b) is the estimated global Lyapunov exponent for each city and the corresponding 95% confidence interval. New York City has an estimated  $\lambda = -0.06$  with a 95% CI  $(-0.13, -0.02)$  and Copenhagen has an estimated  $\lambda = -0.006$  with a 95% CI  $(-0.15, 0.09)$ . The global LE for New York City cases has a confidence interval with endpoints that are both negative, indicating that the series is not chaotic with 95% confidence. The confidence interval for the global LE for Copenhagen cases covers zero, so it is inconclusive. The time series from both cities have a  $\lambda$  near 0 and are thus near the transition to chaos. Investigation of the LLEs shows that both systems vary between short term stable and chaotic behavior. A four step-ahead local Lyapunov exponent (LLE4) was chosen to characterize the local behavior. The LLE4 will describe how well the system can be predicted one year ahead. Figure 4(a) show confidence intervals for the positive fraction of LLE4s. It can be seen that the confidence intervals for the positive fraction of LLE4s and consequently the negative fraction of LLE4s are both greater than zero for the Copenhagen series. This indicates that with 95% confidence there are some regions of state space where the LLE4 is positive, but there is a much larger percentile of LLE4s that is negative. The results for New York City are less conclusive. Our best estimate is that about 7% of the LLE4s are positive, but the 95% confidence interval for the positive fraction includes 0. Thus we can conclude that the Copenhagen series varies between finite time stability and chaos, (with the finite time chaos being rare), while it is possible that the New York City series has only finite time stability.

**5.2 Comparison of Measles Incidence Data and Model Output.** A variety of mechanistic models have proposed to describe the dynamics of measles. These range in complexity from simple deterministic models (such as the SEIR model, a system of 3 coupled ordinary differential equations) up to elaborate stochastic simulation models incorporating finite-population effects, age structure, and immigration (Anderson and May [1992], Grenfell, Bolker, and Kleczkowski [1995]). SEIR-type models have proved to be inadequate, in both deterministic and stochastic formulations (Bolker and Grenfell [1993]). One current candidate is an age-structured stochastic model, implemented as a Monte Carlo simulation with a large but finite population classified by age and disease state (Bolker and Grenfell [1993]). Here we illustrate the use of LLEs to compare this model with data on measles incidence in Copenhagen. For this comparison the model used a population of one million individuals (roughly the population of Copenhagen over the time period of the data) and was “tuned” to produce output that resembled the dynamics of the Copenhagen series. Figure 5 is a plot of the Copenhagen data (quarterly case totals) and corresponding model output.

LLEs for the two time series were estimated by fitting equation (2.1), where  $f$  is estimated by a neural net. Figure 6 is a plot of the estimated global Lyapunov exponents and the 10th and 90th percentiles of the LLE4s with corresponding 95% confidence intervals. In Figure 6(c), the global Lyapunov exponent of the real data is  $\lambda = -0.06$  with 95% CI  $(-0.28, -0.004)$  while that of the Monte Carlo model is  $\lambda = -0.05$  with 95% CI  $(-0.16, 0.03)$ . The confidence interval for the global Lyapunov exponent of the real data is below zero indicating that the series is not chaotic. The Monte Carlo model’s estimated global Lyapunov exponent is very close to that of the real data.

In Figure 6(a), the 10th percentile of the LLE4s for the real data and the model indicate that the model is capturing the smallest LLE4s. The confidence intervals for the 10th percentiles are all negative. Figure 6(b) shows that the 90th percentile confidence intervals for the data and model are positive, indicating the model is also capturing the largest LLE4s. From examining the global and the 10th and 90th percentile of the LLE4s, the Monte Carlo model seems to match the real case data very well.

Figures 7-8 are a series of three plots for the data and the model. The top plot (a) is the time series of the LLE4s. The middle plot (b) is the original time series coded to identify the times at which the LLE4 was in the top 20% of the distribution of LLE4s, i.e. indicated by a 5 overlaid on the time series. The bottom plot (c) is a scatter plot of all LLE4s plotted against the number of log cases at the time  $t$  when the LLE4s are calculated. Figures 7(b) and 7(c) show that in the real data, the largest positive LLE4s occur when there are a small number of cases, i.e. it is very hard to predict the start of an epidemic. The Monte Carlo model does not seem to capture this aspect of the dynamics very well, as seen in Figures 8(b) and 8(c). The model has the feature of unpredictability or large LLE4s when there are a large number of cases, which is not present in the real data. This can be seen in the pattern of the scatter plots of Figure 8(c).

Even though the Monte Carlo model has comparable distribution of local and global Lyapunov exponents, it does not seem to capture the positive local Lyapunov exponents or the short term unpredictability of the system when there are small numbers of infectives. The model indicates the need to investigate the critical number of infectives or susceptibles needed to start a measles outbreak and the need to look closer into the dynamics when the number of cases is small.

## 6 Discussion

A LLE analysis of a dynamical system provides useful information in understanding the system behavior. Measles incidence data is an example of a system where the global LE is close to zero (Ellner et al. [1996]). Even though it is possible to estimate the global LE, most likely the confidence interval will cover zero and it is not possible to make a statement about whether the system is chaotic or not. However, the examination of the behavior of LLEs can determine whether the system has short term instability. i.e. “local” chaos.

The identification of the parts of the time series that have large (or small) LLEs and consequently less (or more) predictability provides information about the heterogeneity of the system. The results of the LLE analysis of measles incidence data show large changes in dynamic structure and predictability over the course of an epidemic.

Local Lyapunov exponents are also useful as a probe into the adequacy or inadequacy of a mechanistic model. Our examination of the stochastic model for measles indicates that the model is not producing the same short term unpredictability as the data in particular regions of the state space. By “localizing” the discrepancy between model and data, these findings indicate where to focus efforts in improving the model so that it matches observed dynamics.

### References

- Abarbanel, H. D. I., Brown, R., and Kennel, M. B. [1992] *Local Lyapunov exponents computed from observed data*, J. Nonlinear Sci., **2**, 343–365.
- Anderson, R. M. and May, R. M. [1992] *Infectious Diseases of Humans*, Oxford University Press, Oxford.
- Bailey, B. A. [1996] *Asymptotics and Applications of Local Lyapunov Exponents*, PhD Thesis, North Carolina State University, Raleigh, NC.
- Bobashev, G., Ellner, S., Nychka, D. W., and Grenfell, B. [1995] *Susceptibles reconstruction from measles epidemic data*, Institute of Statistics Mimeo Series, Statistics Department, North Carolina State University, Raleigh, NC.
- Bolker, B. M. and Grenfell, B. T. [1993] *Chaos and biological complexity in measles dynamics*, Proc. Roy. Soc. Lond. B, **251**, 75–81.
- Doob, J. L. [1953] *Stochastic Processes*, John Wiley & Sons, Inc., New York, 221–232 .
- Ellner, S., Nychka, D. W., and Gallant, A. R. [1992] *LENNS, a program to estimate the dominant Lyapunov exponent of noisy nonlinear systems from time series data*, Institute of Statistics Mimeo Series, Statistics Department, North Carolina State University, Raleigh, NC.
- Ellner, S. and Turchin, P. [1995] *Chaos in a noisy world: New methods and evidence from time-series analysis*, Amer. Natur., **145**, 343–375.
- Ellner, S., Bailey, B. A., Bobashev, G., Gallant, A. R., Grenfell, B., and Nychka, D. W. [1996] *Noise vs. determinism in measles epidemics dynamics: estimates from nonlinear forecasting*, Manuscript.
- Grenfell, B., Bolker, B., and Kleczkowski, A. [1995] *Seasonality, demography, and the dynamics of measles in developed countries* in D. Mollison (ed.) *Epidemic Models: Their Structure and Relation to Data*, Cambridge University Press, Cambridge.
- Hall, P. and Heyde, C. C. [1980] *Martingale Limit Theory and Its Application*, Academic Press, 155–161.
- McCaffrey, D., Ellner, S., Nychka, D. W., and Gallant, A. R. [1992] *Estimating the Lyapunov exponent of a chaotic system with nonlinear regression*, J. Amer. Statist. Assoc., **87**, 682–695.
- Meyn, S. P. and Tweedie, R. L. [1993] *Markov Chains and Stochastic Stability*, Springer-Verlag, London.
- Seber, G. A. F. and Wild, C. J. [1989] *Nonlinear Regression*, John Wiley & Sons, 228–232.
- Smith, L. A. [1992] *Identification and prediction of low dimensional dynamics*, Physica D, **58**, 50–76.
- Smith, L. A. [1995] *Locally optimized prediction of nonlinear systems: stochastic and deterministic*, Chaos and Forecasting (ed. H. Tong), World Scientific, Singapore, 87–108.
- Theiler, J. and Smith, L. A. [1995] *Anomalous convergence of Lyapunov exponent estimates*, Phys. Rev. E, **51**, 3738–3741.
- Tong, H. [1995] *A personal overview of nonlinear time series analysis from a chaos perspective*, Scand. J. Statist., **22**, 399–445 .
- Wolff, R. C. L. [1992] *Local Lyapunov exponents: looking closely at chaos*, J. R. Statist. Soc. B, **54**, 353–371.
- Yao, Q. and Tong, H. [1994] *Quantifying the influence of initial values on non-linear prediction*, J. R. Statist. Soc. B, **56**, 701–725.

### Appendix: Single Ergodic set for state-space systems.

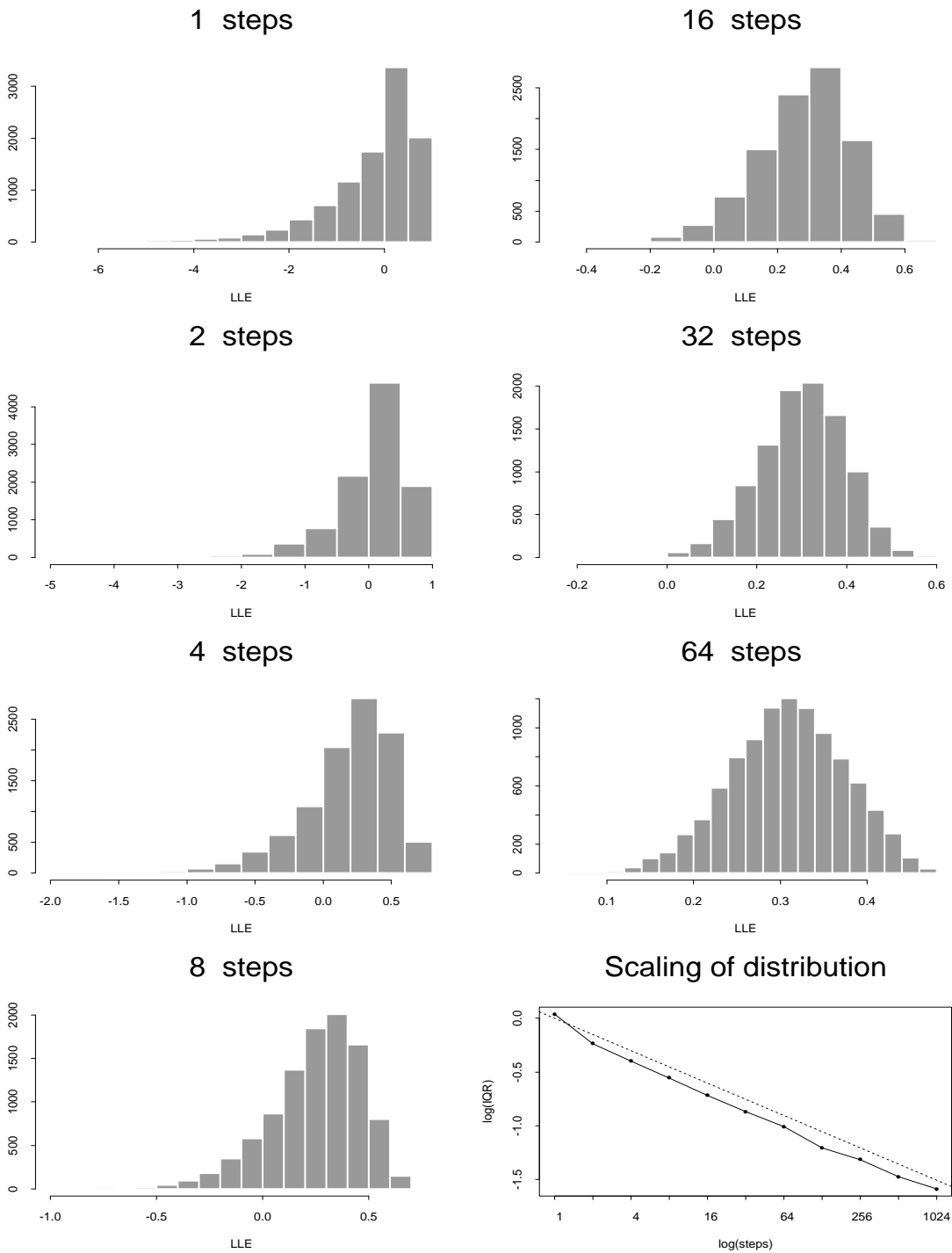
Consider a state-space system (such as the Ikeda map example) defined by (2.2) in which the noise vector  $e_t$  is i.i.d. with continuous positive density on all of  $\mathbb{R}^d$ , and the first derivative  $DF$  is continuous. Because the value of  $h(W_t)$  is unaffected if we replace  $U_t$  by  $-U_t$ , the state space for  $U_t$  in the definition of  $W_t$  can be taken to be the set of unordered vector pairs  $\Delta = \{\{u, -u\}, u \in \mathbb{R}^d\}$ . To establish that  $W_t$  has a single ergodic set we require an assumption similar to the third part of Assumption 3.1:

3\*. (a) There exists  $m > 0$  such that the distribution of  $(X_m, U_m)$  conditional on  $(X_0, U_0)$  has a nonsingular component with respect to Lebesgue measure whose density is positive on an open set and continuous in  $(X_m, U_m, X_0, U_0)$ . (b) There is a sequence of points  $\Sigma = (X_1^*, X_2^*, \dots, X_k^*)$  such that the Jacobian product  $J_k^* = DF(X_k^*)DF(X_{k-1}^*) \cdots DF(X_1^*)$  has a strictly dominant eigenvalue.

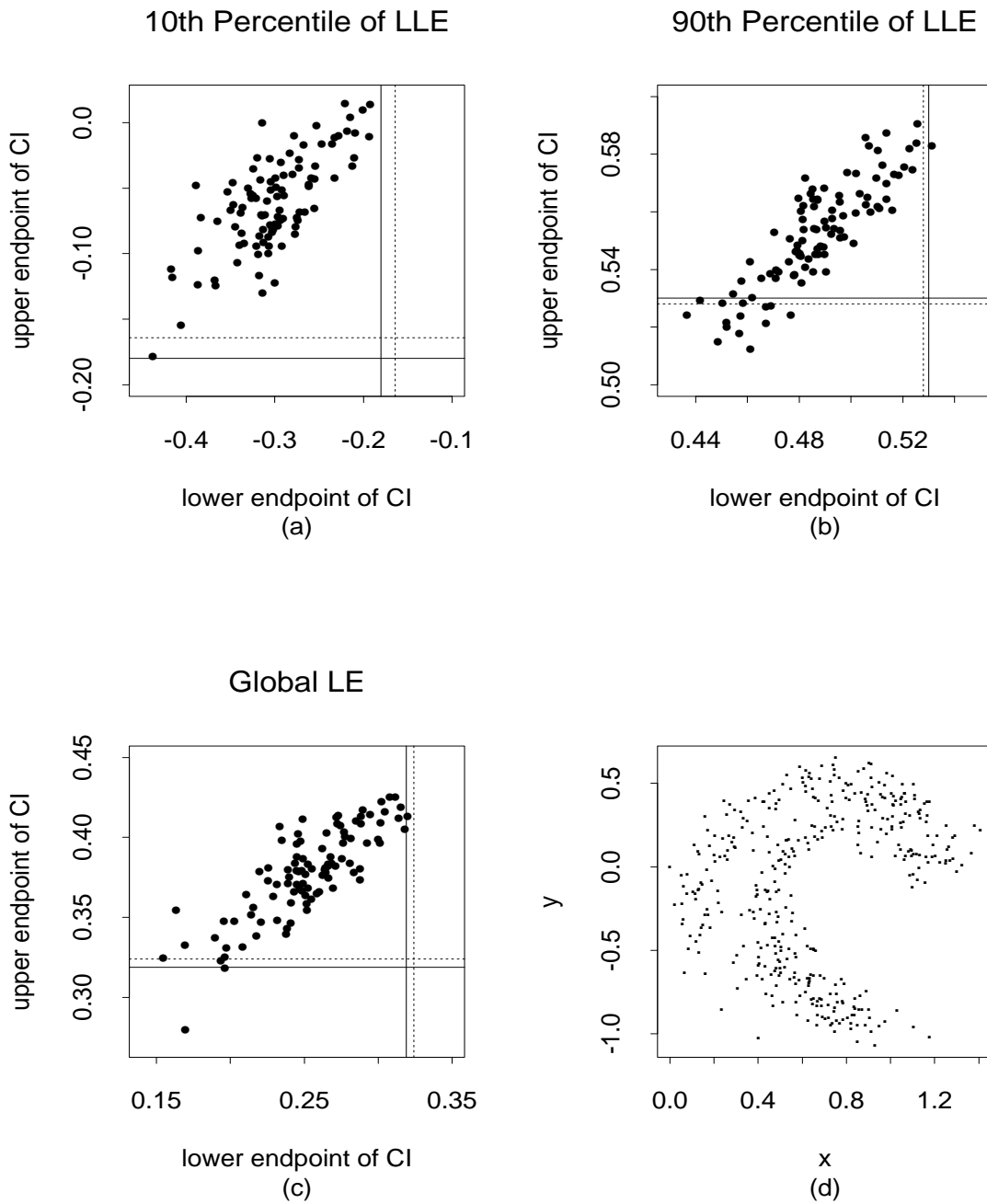
The intuitive meaning of 3\* is that different sample paths must generate different Jacobian products. The sample path  $(X_1, X_2, \dots, X_{m-1})$  has a smoothly varying density conditional on  $(X_0, X_m)$ ; if this translates into a density on  $\mathbb{R}^{d \times d}$  for the product of Jacobians taking  $U_0$  to  $U_m$ , this will generate a density for  $U_m$  and guarantee that at least one point in the product's support has a strictly dominant eigenvalue. We expect that  $m \geq d$  will be necessary, because the system receives only  $d$  independent random shocks at each time step and therefore at least  $d$  time steps would be required to generate a density on  $\mathbb{R}^{d \times d}$ .

Assumption 3\*(a) implies that  $W$  is a T-chain via the “forward accessibility” criterion for nonlinear state space models (Meyn and Tweedie [1993], Chapter 6 and Section 7.1). It will be shown that  $W$  has a reachable point  $W^*$  (for a Markov chain transition probability  $P(x, A)$  a point  $x^*$  is called reachable if, for any  $y$  and any neighborhood  $O$  of  $x^*$ ,  $\sum_n P^n(y, O) > 0$ ). Therefore  $W$  is  $\psi$ -irreducible for  $\psi$  the distribution of  $W_1$  conditional on  $W_0 = W^*$  (Meyn and Tweedie [1993], Prop 6.2.1). This implies that  $W$  cannot have more than one ergodic set, since chains restricted to distinct ergodic sets cannot have the same irreducibility measure.

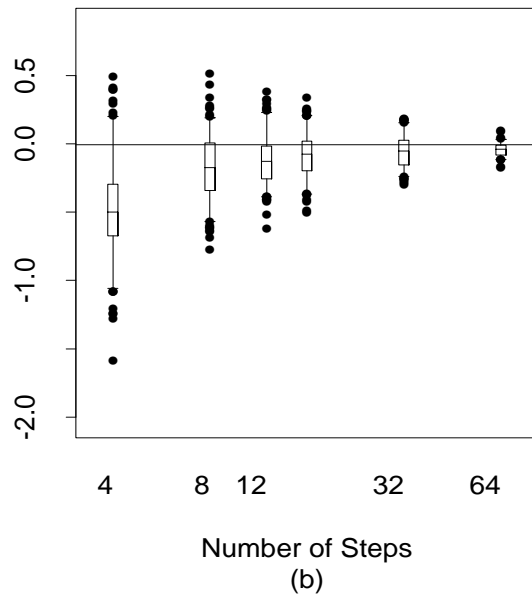
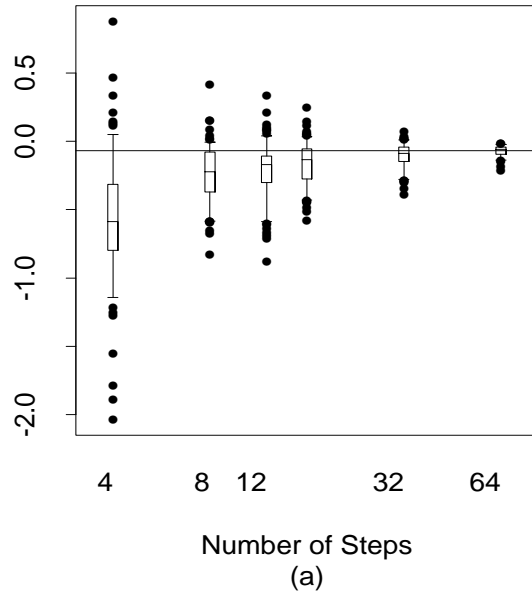
Suppose first that a point  $X^*$  exists such that  $J^* = DF(X^*)$  has a strictly dominant eigenvalue. Let  $v$  be a corresponding eigenvector normalized to unit length, and  $W^* = (X^*, U^*)$  where  $U^* = \{J^*v, -J^*v\}$ . To show that  $W^*$  is reachable, choose any neighborhood  $O$  of  $W^*$  and an  $\varepsilon > 0$  such that  $O$  contains the ball of radius  $\varepsilon$  centered at  $W^*$ . Let  $W_0 = (X_0, U_0)$  be any initial point,  $U_0 = \{u_0, -u_0\}$ . It is a generic property of matrices  $A$  that  $Au_0$  does not lie in  $M$ , the subspace spanned by the non-dominant (generalized) eigenvectors of  $J^*$ . The support of  $(X_m, U_m)$  therefore contains a point  $(\tilde{X}, \tilde{U})$  such that the vectors in  $\tilde{U}$  do not lie in  $M$ . The support of the distribution of  $(X_{m+1}, X_{m+2}, \dots)$  contains  $(X^*, X^*, X^*, \dots)$  by our assumption on the error distribution, and along this sample path we have  $(X, U) \rightarrow W^*$ . Following this sample path long enough there exists a time  $\tau$  such that  $(X_\tau, U_\tau)$  is within  $\varepsilon/2$  of  $W^*$ . Sufficiently nearby sample paths will have the property that  $(X_\tau, U_\tau)$  is within  $\varepsilon$  of  $W^*$ , and these trajectories have positive probability by our assumption on the error distribution, implying that  $W^*$  is reachable. If no such  $X^*$  exists, then following sample paths that repeatedly cycle through  $\Sigma$  (rather than staying at  $X^*$ ) shows that  $(X_1^*, U_1^*)$  is reachable, where  $v$  is an eigenvector of  $J_k^*$  normalized to unit length and  $U_1^* = \{DF(X_1^*)v, -DF(X_1^*)v\}$ .



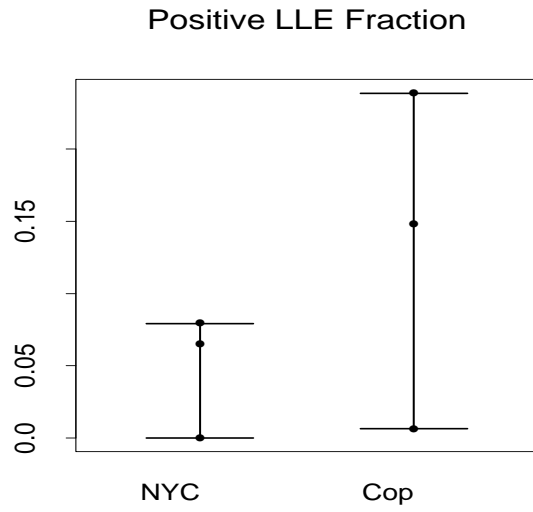
**Figure 1** Histogram of LLEs of the Ikeda map. Bottom right panel shows the interquartile range of the distributions. The dashed line has slope  $-1/2$  corresponding to the expected scaling of the LLEs with increasing number of steps.



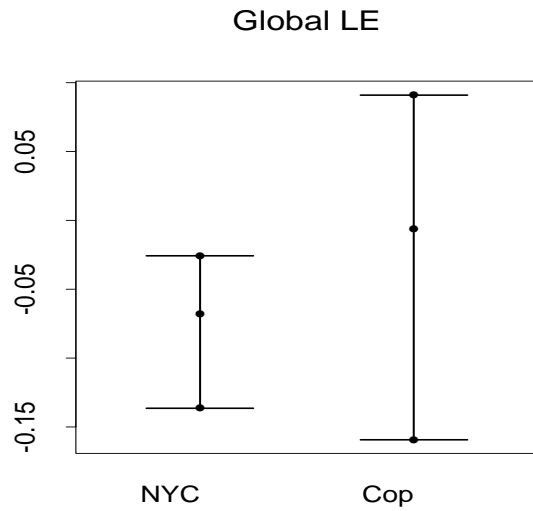
**Figure 2** (a) Coverage probability plot of CI for 10th percentile of LLE5s for noisy Ikeda map. (b) Coverage probability plot of CI for 90th percentile of LLE5s for noisy Ikeda map. (c) Coverage probability plot of CI for global LE for noisy Ikeda map. (d) 400 points from a typical realization of the Ikeda map with noise uniform on  $(-0.1,0.1)$ .



**Figure 3** (a) Distribution of LLEs for New York City measles quarterly cases.  
(b) Distribution of LLEs for Copenhagen measles quarterly cases.

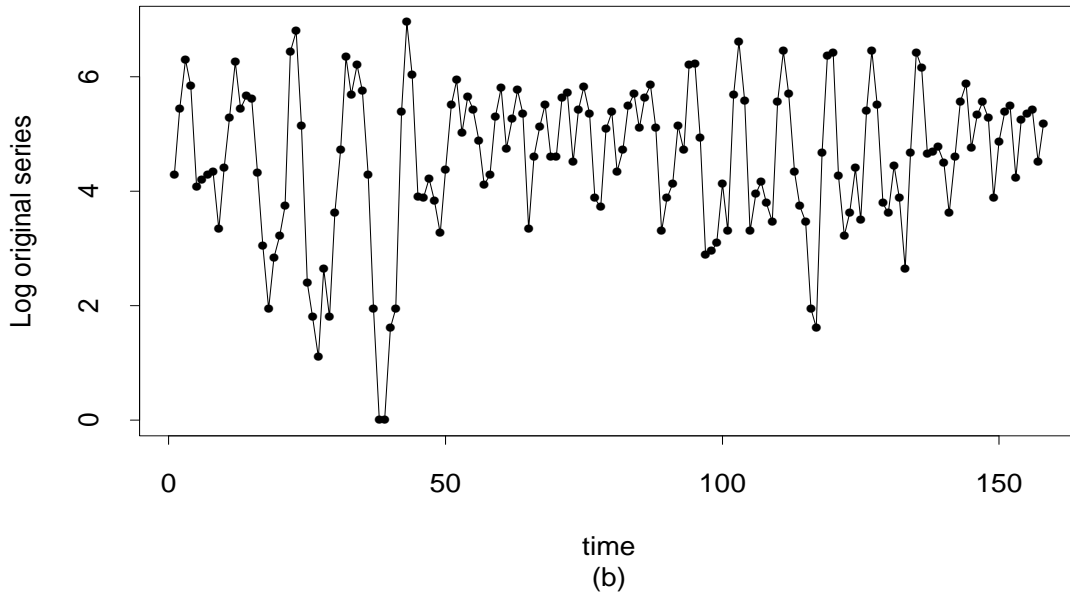
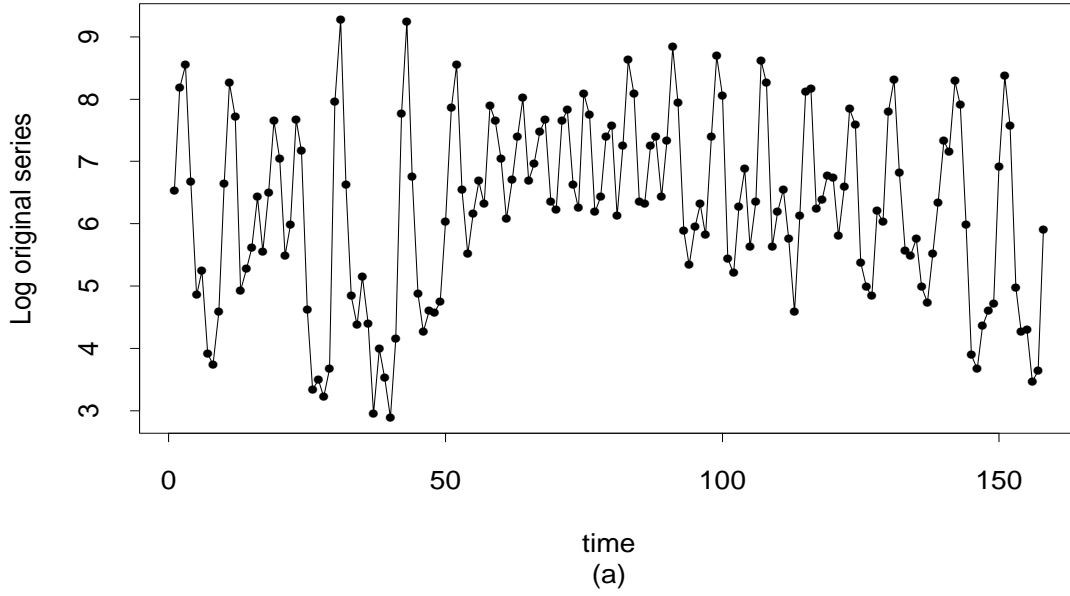


(a)

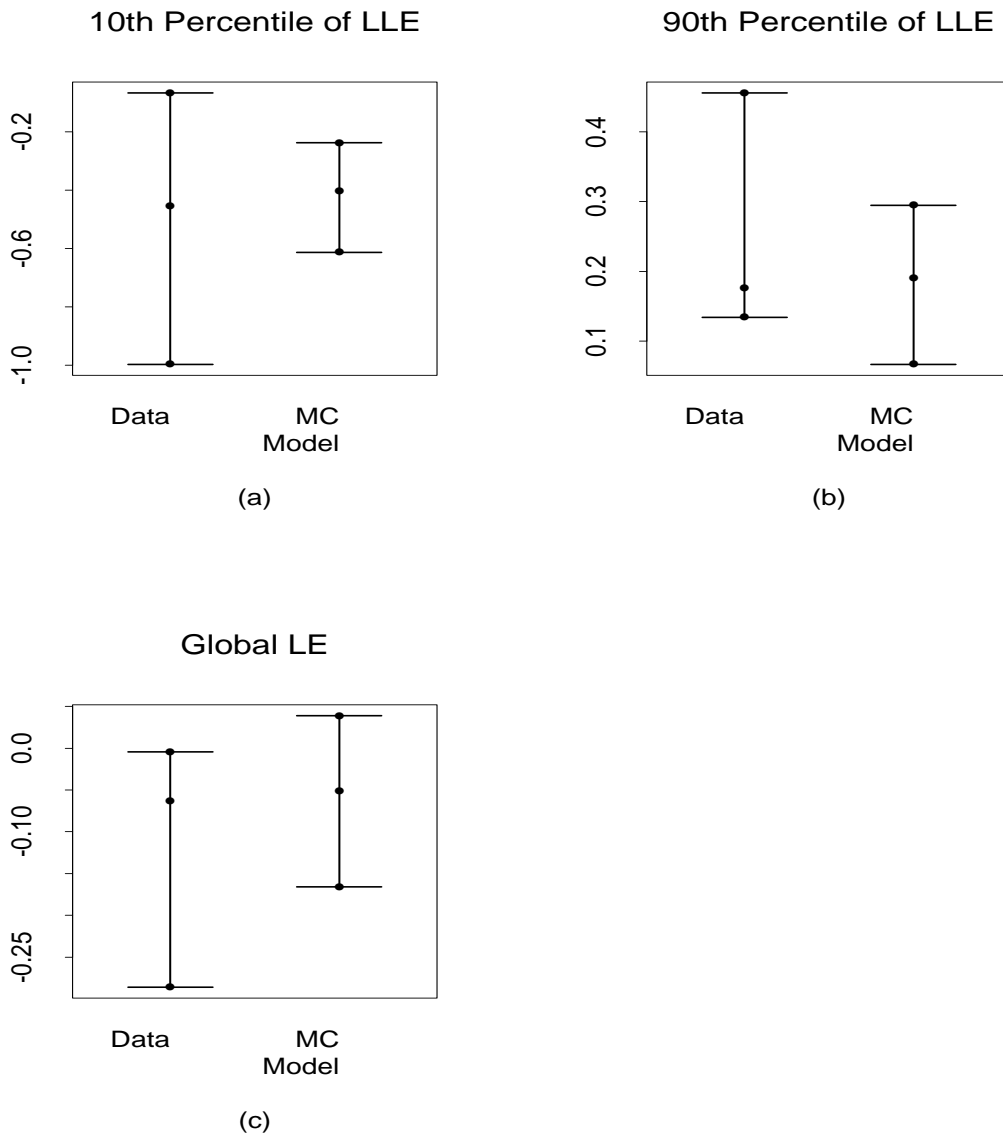


(b)

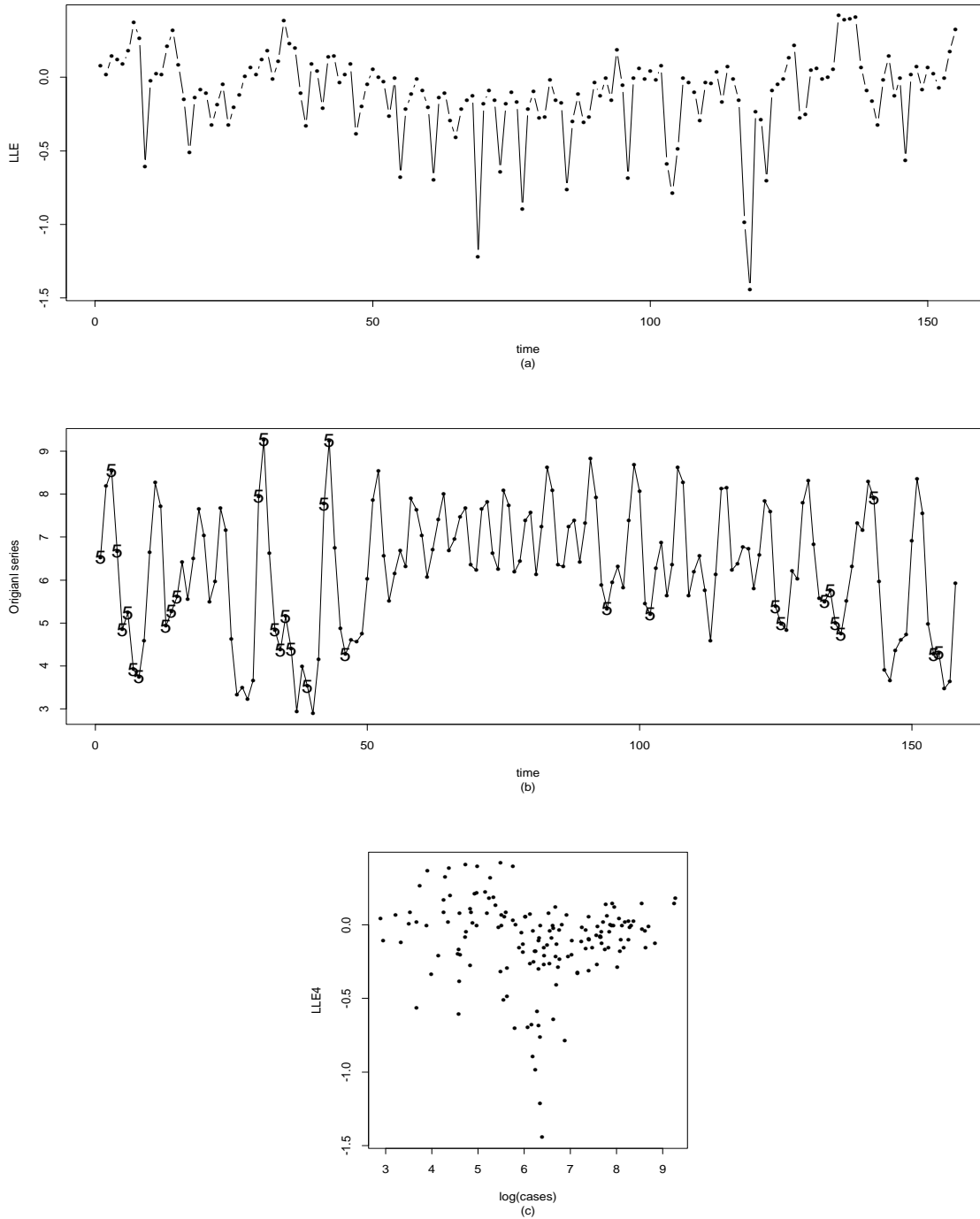
**Figure 4** (a) Positive fraction of LLE4s and corresponding 95% confidence intervals for New York City and Copenhagen measles quarterly cases. (b) Global Lyapunov exponents and corresponding 95% confidence intervals for New York City and Copenhagen measles quarterly cases.



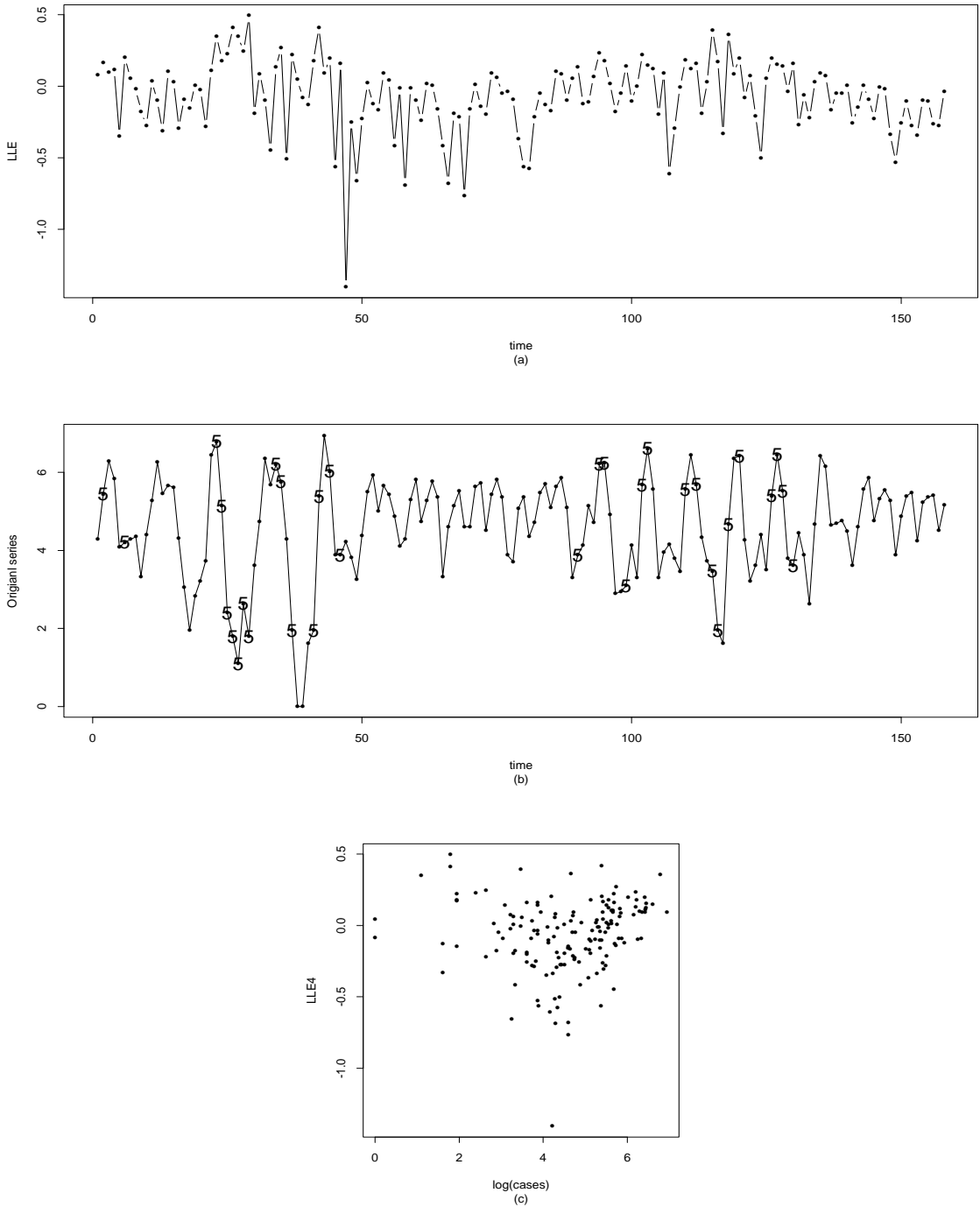
**Figure 5** (a) Copenhagen measles quarterly cases. (b) Monte Carlo age-structured simulation model for measles quarterly cases.



**Figure 6** (a) 10th percentile of LLE4s and corresponding 95% confidence intervals for Copenhagen measles quarterly cases and Monte Carlo model output, (b) 90th percentile of LLE4s and corresponding 95% confidence intervals for Copenhagen measles quarterly cases and Monte Carlo model output, (c) Global Lyapunov exponents and corresponding 95% confidence intervals for Copenhagen measles quarterly cases and Monte Carlo model output.



**Figure 7** (a) 4-step ahead local Lyapunov exponents (LLE4s) for the Copenhagen measles quarterly cases. (b) Series coded to identify the times at which LLE4 was in top 20% of the distribution of LLE4s. (c) Log(cases) vs LLE4s for Copenhagen measles quarterly cases.



**Figure 8** (a) 4-step ahead local Lyapunov exponents (LLE4s) for the Monte Carlo age-structured simulation model of measles quarterly cases. (b) Series coded to identify the times at which LLE4 was in top 20% of the distribution of LLE4s. (c) Log(cases) vs LLE4s for Monte Carlo age-structured simulation model of measles quarterly cases.