

# Detecting nonlinearity and chaos in epidemic data

Stephen Ellner<sup>1</sup>, A. Ronald Gallant<sup>2</sup>, and James Theiler<sup>3,4</sup>

<sup>1</sup>*Biomathematics Graduate Program, Department of Statistics,  
North Carolina State University, Raleigh NC 27695-8203, USA;*

<sup>2</sup>*Department of Economics, University of North Carolina, Chapel Hill NC 27533-3305, USA;*

<sup>3</sup>*Santa Fe Institute, 1660 Old Pecos Trail, Santa Fe NM 87501 USA;*

<sup>4</sup>*Center for Nonlinear Studies & Theoretical Division,  
Los Alamos National Laboratory, Los Alamos, NM 87545 USA.*

## 1. Introduction

Historical data on recurrent epidemics have been central to the debate about the prevalence of chaos in biological population dynamics. Credit for this interest in epidemics goes to Schaffer and Kot (1985, 1986), who first recognized that the abundance and accuracy of disease incidence data opened the door to applying a range of methods for detecting chaos that had been devised in the early 1980's. Using attractor reconstruction, estimates of dynamical invariants, and comparisons between data and simulation of SEIR models, the "case for chaos in childhood epidemics" was made through a series of influential papers beginning in the mid 1980's (reviewed by Schaffer et al. 1990). The proposition that the precise timing and magnitude of epidemic outbreaks are deterministic but chaotic is appealing, since it raises the hope of finding determinism and simplicity beneath the apparently stochastic and complicated surface of the data.

However the initial enthusiasm for methods of detecting chaos in data has been followed by critical re-evaluations of their limitations. Early hopes of a "one size fits all" algorithm to diagnose chaos vs. noise in any data set have given way to a recognition that a variety of methods must be used, and interpretation of results must take into account the limitations of each method and the imperfections of the data (e.g., Theiler 1990).

Our goals here are twofold. First, we present an overview of methods for detecting nonlinearity and chaos in epidemic data. We identify features of epidemic data that create problems for the older, better known methods of detecting chaos (Section 2), and we then review some newer methods for detecting nonlinearity and chaos that are suited to epidemic data, and have a more solid statistical basis (Sections 3-5). Our emphasis is on the essential ideas of each method, referring the reader to the relevant literature for

the technical details which we omit. Second, we begin a re-evaluation of the claims for nonlinear dynamics and chaos in epidemics, by applying each of the newer methods to a collection of data sets on measles, mumps, rubella, and chicken pox. These results are new, and publication elsewhere is not planned, so a fairly complete description is given.

When we ask “are epidemics nonlinear?”, we are not questioning the existence of global nonlinearities in epidemic dynamics, such as nonlinear transmission rates. Our question is whether the data’s deviations from an annual cycle (for example, the biennial or triennial cycles that are often observed) are adequately described by a linear, noise-driven stochastic process, or whether a nonlinear description is mandated by the data.

Our conclusion is that evidence for chaos is generally lacking, but at least for measles we can reject the hypothesis of linear noise superimposed on an annual cycle. Thus nonlinearity in the dynamics, and its interactions with stochastic perturbations, are manifested in the data and should be taken into account when interpreting or attempting to predict fluctuations in the number of cases. In particular, our results suggest that short-term noise amplification (Deissler and Farmer 1992) and “transient chaos” are likely to be common.

## **2. Noise, seasonality, and the hunt for chaos**

The task of detecting nonlinearity or chaos in epidemics is complicated by two unavoidable features of the data: dynamic noise and seasonality. The literature on detecting chaos mostly ignores these features (apart from lip service), so many “consumers” of the literature are unaware of their immense effects on methods for detecting chaos. Those effects are the subject of this section.

### DYNAMIC NOISE

The prevalent attitude in the chaos literature is that any stochasticity is an undesirable corruption of the data. This attitude is reasonable for random measurement errors – accurate data is indeed better than inaccurate data – and physicists have devoted considerable effort to methods for reducing measurement errors. However epidemic dynamics also are affected by “dynamic noise” – external, unpredictable perturbations (e.g., fluctuations in weather, teacher strikes, etc.) that affect disease transmission and consequently are an intrinsic part of the dynamics.

Here we take the view, following Eckmann & Ruelle (1985), that the defining feature of chaos is bounded fluctuations with sensitive dependence on initial conditions. This

definition of chaos applies equally to completely deterministic systems and to systems with dynamic noise. Formally, suppose that the data are generated by a stationary ergodic process of the form

$$(1) \quad X_{t+1} = F(X_t, E_t)$$

where  $X_t \in \mathbb{R}^d$  and  $E_t$  is a sequence of *iid* random variables. The system's sensitivity to small changes in initial conditions is quantified by the dominant Lyapunov exponent  $\lambda$ , given by

$$(2) \quad \lambda = \lim_{m \rightarrow \infty} \frac{1}{m} \log \|DF(X_m, E_m)DF(X_{m-1}, E_{m-1}) \cdots DF(X_1, E_1)\| ,$$

where  $DF(\bullet, E)$  is the Jacobian matrix of  $F(\bullet, E)$ .  $\lambda$  is well-defined and constant with probability 1 under some mild regularity conditions (Kifer 1986). Thus  $\lambda$  is a specific number, rather than a random variable, even for systems with dynamic noise. Note that the Jacobians in (2) only involve derivatives with respect to the state ( $X$ ). Thus for noisy systems the Lyapunov exponent characterizes the exponential divergence of two trajectories with slightly perturbed initial conditions, but subject to the same random shocks ( $E$ ).

Dynamic noise can move systems into or out of chaos (Crutchfield et al. 1982); in particular, the stability of seasonally forced SEIR models is very sensitive to small random fluctuations in the contact rate (Rand and Wilson 1991). Removing dynamic noise by “noise reduction” techniques is not desirable: we want to characterize the real dynamics, which are noisy due to random forcing. Most methods for detecting chaos or nonlinearity in data, even methods that are robust against (or explicitly designed to handle) measurement errors, have serious problems with dynamic noise. Methods in this category include:

*Fractal dimension.* Estimates of fractal dimension (see Theiler 1990 for a review) are seriously degraded by dynamic noise much smaller than the system's range of fluctuations, even though much higher levels of measurement error can be dealt with (R. Smith 1992ab). This reflects a fundamental difference between the effects of measurement errors and dynamic noise. With measurement errors, we are viewing a low-dimensional attractor through fogged-up glasses; with dynamic noise the attractor is infinite dimensional.

*Lyapunov exponents by the Wolf et al. (1985) method.* This method quantifies the sensitive dependence on initial conditions by finding segments of the time series that come close together in phase space, and monitoring their subsequent divergence. Because

divergence due to dynamic noise is confounded with divergence due to sensitive dependence on initial conditions, dynamic noise generates “false positives” in the hunt for chaos (Sayers 1990).

*Nonlinear prediction (Sugihara & May 1990).* The method of Sugihara and May (1990) distinguishes between measurement error and deterministic chaos by comparing the accuracy of short-term and long-term out of sample forecasts. In a chaotic system, long-term forecasts are less accurate due to sensitive dependence on initial conditions. However dynamic noise also decreases long-term forecast accuracy, so distinguishing between chaos and dynamic noise by this method is generally not possible (Ellner 1991, Stone 1992).

## SEASONALITY

Seasonality should not create any problems for methods of detecting chaos in data, because any periodically forced system can be re-expressed as an equivalent autonomous system by adding a state variable to serve as a clock. This frees us (in theory) to behave as if our data come from an autonomous system. In practice, however, data analyses can be confounded by strong seasonal forcing:

*Attractor reconstruction.* The early claims of evidence for chaos in epidemics was based on the now-classic method of attractor reconstruction in time delay co-ordinates (Packard et al. 1980, Takens 1981, Sauer et al. 1991). However, Ellner (1991) showed that the fieldmarks of low-dimensional chaos which had been observed in measles data – graphically reconstructed attractors, Poincare sections, and Poincare maps – were also observed in a seasonally forced nonchaotic stochastic model which is really infinite-dimensional.

*Lyapunov exponents.* In the Wolf et al. (1985) method, and the modified Wolf method proposed by Rand and Taylor (this volume), data segments nearby in phase space can correspond to different times of year. Subsequent trajectories will diverge simply because they are following different “clocks”, creating a positive bias in estimates of  $\lambda$ . The spurious neighbors also affect our method based on time series modeling (Nychka et al. 1992, McCaffrey et al. 1992); this invalidates Ellner’s (1991) conclusion that measles exhibits weak chaos. A fix-up for the method and updated conclusions are described below.

Some influential figures are now arguing, based on the problems with the older methods, that the program of “detecting chaos” is doomed to failure by the need for

massive amounts of very accurate data. We disagree, so long as the standards of “success” are those of field biology, where imperfect and limited data are the norm, rather than those of laboratory physics. A limited data set may not allow us reject a null hypothesis that could be rejected with additional data, but with methods grounded in experimental statistics we can still say that a given data set does or does not provide evidence for a given hypothesis, and attach statistical measures of confidence to our conclusions. A chance of error is unavoidable, so overall conclusions often must emerge from a series of studies with different limitations, rather than from a single decisive experiment (Hastings et al. 1993).

### 3. Surrogate data

We now turn to some more promising methods for epidemic data. Surrogate data methods provide a Monte-Carlo approach for testing whether data are consistent with a (possibly transformed) linear autoregressive model with Gaussian dynamic noise (Theiler et al. 1992, and references therein). The basic procedure is to simulate “surrogate” data sets which have the same power spectrum as the real data, and compare the values of a test statistic on the real and simulated data. One method for generating surrogates is to Fourier transform the data, randomize the phases in the complex Fourier coefficients while preserving the amplitudes, and inverse Fourier transform to obtain a surrogate data set. The surrogate data have the same discrete power spectrum and therefore the same (circular) linear autocorrelations as the real data, but any couplings between modes due to nonlinear structure in the data have been obliterated. Repeat as often as desired, using an apt test statistic, and you have a statistical hypothesis test of

**$H_0$ : the data arise from a static transform of a Gaussian linear autoregressive process.**

An important strength of the surrogate data method is that any computable measure of nonlinearity can be used as the test statistic. Even if the numerical value of the measure on any single data set may be inaccurate (e.g., biased due to dynamic noise), differences between the real and surrogate data still can provide evidence of nonlinearity which is no less reliable than any other statistical test of a null hypothesis.

Of course it is not quite that simple. If the real data aren't Gaussian they should be made Gaussian by a transformation; care is needed when computing the power spectrum; and it is not clear how to generate good surrogates for data with strong spectral peaks. See Theiler et al. (1992, 1993) for the details. To avoid false negatives, the test statistic must key into some difference between linear and nonlinear dynamics: a statistic that can be computed from the linear autocorrelations is useless because it will have exactly the

same value on the real and surrogate data. It is also helpful if the test statistic measures a physical or intuitively identifiable quantity. Detecting nonlinearity is just the first step; ultimately, one wants to be able to characterize it.

For epidemic data the null hypothesis given above is clearly false due to seasonality. We therefore examined the more interesting null hypothesis

$$(3) \quad H_0: \text{data} = \text{seasonal trend} + \text{transform of a Gaussian linear AR process.}$$

In many cases the data appear to have a biennial or triennial cycle. This is clearly a departure from an annual cycle, but it is not necessarily a *nonlinear* departure, because a linear filter acting on white noise can produce spectral peaks at biennial or triennial periods. If the data exhibit a biennial or triennial cycle, failure to reject (3) would indicate that the multi-year cycle can be described as a linear response by the system to external forcing; while rejection of (3) would imply that the system's response is nonlinear.

To test (3) we subtracted off the seasonal trend (estimated by averaging over years in the data), normalized the deviations from the trend to have seasonally constant variance, and generated surrogates for the normalized deviations. We used several test statistics:

1. The Ramsay & Rothman "time-reversal" statistic

$$\rho_{i,j}(m) = \text{Sample average of } (x_t^i x_{t+m}^j - x_t^i x_{t-m}^j) ,$$

for  $i \neq j$  (Rothman 1990, Ramsay and Rothman 1991). The distribution of a linear process with independent Gaussian innovations is unchanged by time reversal, so excessively large values of  $|\rho_{i,j}|$  signal a departure from  $H_0$ . We calculated  $|\rho_{1,2}(m)|$  for  $m = 1$  through 16 quarters and used the maximum and median of the 16 values as our test statistics.

2. Statistics related to the correlation integral  $C(r)$ , which is the fraction of reconstructed data vectors whose distance apart is less than or equal to  $r$ . The statistics we used were two percentiles of the distance distribution,  $r_{.01}$  and  $r_{.99}$ , defined by  $C(r_p) = p$ , and a crude estimate of the correlation dimension  $D_2$  (Grassberger and Procaccia 1983):

$$\hat{D}_2 = \frac{\log C(r_1) - \log C(r_2)}{\log(r_1) - \log(r_2)} ,$$

using  $r_{.01}$  and  $r_{.99}$  as  $r_1$  and  $r_2$ . This formula for  $\hat{D}_2$  does not give a very accurate estimate of fractal dimension. However, as noted above, such inaccuracies do not affect the validity of surrogate data methods because the conclusions are based on differences between real and surrogate data sets. We used reconstructed state vectors of dimension 8 (i.e, each state

vector consisted of 8 consecutive quarterly case counts), so that these statistics would be looking for “long-range” structures not captured by the linear autocorrelations.

3. “Prediction” accuracy backwards in time (suggested by Robert May following our talk). For nonlinear maps with stretching and folding, the folds make it hard to tell where you came from even if you can predict where you’re going. For example in the logistic map, given  $x_t$  you can predict  $x_{t+1}$  exactly but there are two possibilities for  $x_{t-1}$  and no way to identify the correct choice. Our test statistics were the “prediction” accuracy 1 year into the past for kernel time series models using 2, 3, and 4 future values, with the kernel bandwidth chosen by ordinary cross validation. For these statistics only the seasonal trend was not removed from the data, because trend removal could obscure a simple nonlinear relationship.

The results (Table 1) give consistent, and occasionally very strong, evidence for nonlinearity in measles. Of the 12 measles series analyzed, 10 were significantly nonlinear at the .05 level for at least one of the test statistics. This conclusion is modest relative to other claims which have been made about measles, but it rests on solid statistical foundations and should be difficult to dispute. The pattern is reversed in the other diseases: only 3 of the 10 data sets had a significant nonlinearity at the .05 level.

We chose two of the cases where nonlinearity was detected with  $P < 0.01$ , and plotted the value of the summary statistic for both the original and the surrogate data sets. As Fig. 1 shows, the differences are not only statistically significant, but are numerically substantial as well. The value of  $r_{0.01}$  for detrended Copenhagen measles is roughly 20% smaller than the average value for the surrogate time series; and the crudely estimated dimension  $D_2$  for detrended New York City measles is less than half of the average value for the surrogates.

On the other hand, we remark that no one statistic consistently identifies nonlinearity in all of the measles time series. So we cannot say that measles epidemics in general exhibit low dimension, or high backward predictability. The data provide convincing evidence nonlinearities are present in the underlying process, and also are manifested in the observed dynamics. However, the tests in this section do little to characterize the nature of that nonlinearity.

Our findings are in line with Casdagli’s (1992) results for NYC measles, based on his exploratory method for detecting nonlinear dynamics. Casdagli’s (1992) method involves comparing the short term prediction accuracy of a series of models that range from linear

to strongly nonlinear (specifically, locally affine models based on different numbers of nearest neighbors). A substantial improvement in forecasting accuracy by nonlinear vs. linear models is taken as indicating nonlinear dynamics. For NYC measles, Casdagli (1992) found that nonlinear models could achieve a 25% reduction in RMS forecasting error compared with the linear model, which was interpreted as evidence for nonlinear (though not necessarily chaotic) dynamics.

#### 4. Lyapunov exponents via time series modeling

One rough characterization of nonlinear dynamics is whether the dynamics are chaotic or stable, as indicated by the value of the Lyapunov exponent  $\lambda$  (defined above). In this section we describe methods for estimating  $\lambda$  from time series data, and present estimates of  $\lambda$  for epidemic data.

Our approach is to estimate  $\lambda$  by first estimating the nonlinear map generating the data. This allows us to account explicitly for dynamic noise and to estimate its magnitude, and to estimate  $\lambda$  in a way that is not positively biased by dynamic noise. The first step is reconstruction in time delay coordinates (Sauer et al. 1991, Casdagli 1992b), so in practice the procedure amounts to fitting a nonlinear autoregressive model

$$(4) \quad x_{t+T} = f(x_{t-L}, x_{t-2L}, \dots, x_{t-dL}) + e_t .$$

and using derivatives of the estimated map to compute an estimate of  $\lambda$ . In equation (4),  $L$  is called the “time delay” and  $T$  is the prediction time.

Equation (2) is positively biased for finite  $m$ , and our simulation results suggest that a better estimate is

$$(2') \quad \lambda = \frac{1}{m} \log \|DF(X_m, E_m)DF(X_{m-1}, E_{m-1}) \cdots DF(X_1, E_1)\bar{v}\| ,$$

where  $\bar{v} = (1, 0, 0, \dots, 0)'$ . McCaffrey et al. (1992) give supporting statistical theory, Nychka et al. (1992) discuss practical implementation on short, noisy data series, and Ellner et al. (1991) discuss convergence rates.

Once again, it is not *quite* that simple. Some families of prediction models work much better than others. With short, possibly noisy data sets we have achieved the best overall performance from the “feedforward neural net” (FNN) model. The FNN model decomposes an arbitrary function into a sum of sigmoids,

$$f(x_1, x_2, \dots, x_d) = \beta_0 + \sum_{i=1}^k \beta_i G(\mu_i + \sum_{j=1}^d \gamma_{ij} x_j),$$

where  $G$  is a univariate sigmoid function such as the logistic  $e^u/(1+e^u)$ . FORTRAN source code and a user's manual for our implementation are available by anonymous ftp at `lyapunov.ucsd.edu` in `/pub/ncsu`. Thin-plate splines and similar extensions of polynomial models are also effective for low-dimensional fitting and are much faster to compute, but the number of parameters increases too rapidly for use in higher dimensions (Ellner and Turchin 1993).

Also, precautions must be taken both against overfitting and against underfitting. To guard against overfitting, Nychka et al. (1992) recommend that the model complexity (e.g., the value of  $k$  the FNN model) be chosen by the GCV (generalized cross validation) criterion with the number of model parameters inflated by a factor of 2; this tactic appeared to drastically reduce the chance of overfitting without introducing much bias. If the data are highly autocorrelated, it is likely that GCV will select a linear model that makes accurate short-term predictions but ignores the long-term dynamics. To guard against this, Ellner and Turchin (1993) recommend choosing  $L$  to be the smallest lag such that the autocorrelation between  $x_t$  and  $x_{t-L}$  is below 0.5, and using  $T = L$ . Here we simply used quarterly total case reports, which did not have a tight autocorrelation between successive values.

Seasonality also requires special treatment. When model (4) is fitted to data with a strong seasonal trend, one of the lagged variables usually winds up serving as a surrogate "clock". The estimate of  $\lambda$  then includes derivatives with respect to time, but it shouldn't: resetting the clock is not a perturbation of the system's state. To remove the need for a surrogate clock, we explicitly added a real clock to the model:

$$x_t = f(x_{t-1}, x_{t-2}, \dots, x_{t-d}, \sin(t/K), \cos(t/K)) + e_t$$

where  $K$  is the number of data points per year. The effect of including the clock is as expected (Table 2): the estimated  $\lambda$  drops, and fewer past values are needed to make predictions.

The results on epidemic data (Table 3) are again quite consistent: the dynamics are identified as stable rather than chaotic. In fact there appears to be a mode at or just below the transition to chaos ( $\lambda = 0$ ) in the distribution of Lyapunov exponents (Figure 2). The location of the mode is probably influenced by the weak bias towards underfitting in the procedures used here (Nychka et al. 1992). In simulation trials on low-dimensional models (Ellner and Turchin 1993), the bias towards underfitting was too small to alter the qualitative conclusion from Table 3, that epidemics tend to be neither strongly stable nor

strongly chaotic.

Contrary to our urgings that the hunt for chaos should be pursued in a statistical framework, we have not provided standard errors for the estimates in Table 3. Our feeling is that given the current state of the art, it would be easy to compute a standard error but hard to say exactly what that number means, and we wish to avoid overstating the (statistical) significance of our conclusions. The reliability of our conclusions is best indicated by the consistency across multiple data sets, which we leave for the reader to evaluate.

### 5. Efficient Generalized Method of Moments (GMM)

If enough is known about the system of interest, we may prefer to fit a mechanistic model rather than a purely descriptive time series model. A mechanistic model may be overly (or incorrectly) constrained and therefore unable to really match the dynamics, but mechanistic models have the advantage that time series data can be supplemented with information from other sources. For example, the duration of the infectious period can be hard-wired into an SEIR epidemic model.

Fitting of mechanistic models is frequently complicated by the unavailability of the likelihood in a closed or easily computed form. A popular alternative is to use a “method of moments”: choose parameters so that model output matches some features of the data. The features may be genuine moments (mean, variance, autocorrelations, etc.), or any other functions of a simulated trajectory (period of a limit cycle, fractal dimension, etc.). This leads to fitting criteria such as

$$\text{Minimize } \sum_i C_i \{ M_i(\rho) - M_i(\text{data}) \}^2$$

where  $M_i$  are the features,  $\rho$  is the parameter vector of the mechanistic model, and  $C_i$  are positive weights. However it is not clear which “moments”  $M_i$  to use, and how they should be weighted, to get the most accurate estimates of  $\rho$ .

Gallant and Tauchen (1992) have proved that with appropriate  $M_i$  and weighting, and some smoothness and identifiability conditions, GMM is asymptotically equivalent to maximum likelihood estimation of the mechanistic model’s parameters. The  $M_i$  are obtained by choosing a statistical model  $f(x_{t+1} | x_t, \theta)$  for the transition probabilities governing the time series, such that the parameter vector  $\theta$  of the statistical model is easy to estimate by maximum likelihood. For example,  $f$  may be a nonparametric regression

model with appropriate error structure. The generalized moments to be matched as well as possible are then

$$(5) \quad M_i(\rho) = E_\rho \left\{ \frac{\partial}{\partial \theta_i} \ln f(x_{i+1}|x_i, \tilde{\theta}) \right\},$$

where  $\tilde{\theta}$  is the maximum likelihood estimate of  $\theta$  from the empirical data, and  $E_\rho\{\}$  is expectation with respect to the distribution of  $(x_{i+1}, x_i)$  in the mechanistic model with parameter vector  $\rho$ .

$E_\rho\{\}$  in (5) is computed by simulating the model (i.e., by Monte Carlo integration). If  $E_\rho$  is replaced by the empirical distribution of  $(x_{i+1}, x_i)$ , then the expression in (5) is exactly 0 by the first order condition for maximizing the likelihood. Thus a good mechanistic model should give small values of  $M_i(\rho)$ . The right weighting is a quadratic form  $M^T \tilde{\Gamma}^{-1} M$  where  $\tilde{\Gamma}$  is an estimated information matrix; see Gallant and Tauchen (1992) for precise statement of the results, extension to more general settings, and proofs.

The advantage of GMM is that the statistical model  $f$  doesn't have to be "right", i.e. it doesn't need to duplicate exactly the transition probability of the process generating the data. It just needs to be sufficiently general, or well enough adapted to the application, so that it discriminates parameters of the mechanistic model, i.e.,  $M_i(\rho) = 0$  for all  $i$  if and only if  $\rho = \rho_0$ , where  $\rho_0$  is the true value of  $\rho$ .

For an epidemiological application of this method, we estimated contact rate parameters for a deterministic SEIR model, using the monthly measles case reports series from New York City 1928-1963. To mitigate excessive fadeouts in the model we added a small exchange of individuals (at rate  $\delta$ ) with an "outside world" having fixed levels of the disease:

$$\frac{dS}{dt} = m(1 - S) - \beta(t)SI + \delta(S_0 - S)$$

$$\frac{dE}{dt} = b(t)SI - (m+a)E + \delta(E_0 - E)$$

$$\frac{dI}{dt} = aE - (m+g)I + \delta(I_0 - I)$$

This model is a bit simplistic for measles (see Grenfell, Kleczkowski and Bolker, this volume), and the migration terms are admittedly an *ad hoc* way of bringing the model more closely in line with the data. Our justification is simply that this is our first exploratory attempt at using GMM for fitting a mechanistic epidemic model. As we gain experience and hone the implementation, it should become possible to deal with more realistic models.

We assumed that the contact rate  $\beta(t)$  has the form

$$\beta(t) = b_0 (1 + \sigma e_t) + b_1 \phi(t)$$

where  $\phi(t)$  is the seasonal forcing function proposed by Kot et al. (1988),

$$\phi(t) = 1.5 \left\{ \frac{0.68 + \cos(2\pi t)}{1.5 + \cos(2\pi t)} \right\} - .4$$

(we added the “-.4” so that the average of  $\phi$  over the year would be zero), and  $\sigma e_t$  are autocorrelated random fluctuations with mean 0, variance  $\sigma^2$ , and autocorrelation 0.95 between values 1 month apart. We estimated the values of  $b_0$  and  $b_1$ , assuming that all other parameters of the model were known. We took  $\sigma = .05$  to represent small year-to-year fluctuations in contact intensity; results for  $\sigma = .01$  (not presented) were essentially the same. The statistical model was a neural net with 5 lags and 3 units, as in the best-fit neural net model for monthly NYC measles data (Table 2).

Simulated measurement errors were added to the output from the SEIR model; this was necessary to produce a reasonable match between the power spectra of the simulated and real data at higher frequencies (1-2 months). The simulated errors were lognormal with coefficient of variation based on the estimate that 1/8 of all cases are reported (B. Grenfell, *pers. comm*), and assuming that cases are reported or not in independent clusters of size 2 representing a pair of cases in a family. The clustering assumption increases the variance of the simulated measurement errors, and a cluster size of 2 or 3 was necessary to produce the observed amount of power at high frequencies.

The results are encouraging for the method, but somewhat discouraging for fitting SEIR models from time series data alone. The encouraging result is that our automatic procedure produced a value of the relative forcing intensity  $b_1/b_0$ , that is in line with generally accepted estimates. A contour plot of the GMM fitting criterion (Figure 3a) has a steep “valley” of better fits (smaller values of the criterion) roughly along the line  $b_1 = 0.2b_0$ , and a univariate plot of optimal GMM vs.  $b_1/b_0$  has a well-defined minimum (Figure 3b). The discouraging results are first that, as can be seen in Fig. 3a the terrain along the valley floor is rather flat, so the absolute values of  $b_0$  and  $b_1$  are less well identified. Second, the entire terrain is rough (Figure 4). It is not clear how much of the roughness is due to Monte Carlo error (finite sample size in computing  $E_\rho\{\}$ ), vs. intrinsic roughness of the exact surface. If the latter is dominant, then standard asymptotic methods based on Taylor series approximations will not be available for setting confidence regions or for hypothesis testing based on GMM.

## 6. Conclusions

We would like to close by speculating on the implications of our findings. Our “surrogate data” results indicate that nonlinear departures from annual periodicity are a consistent feature of the measles data, but less common in the other diseases examined. The statistics  $r_{.01}$  and  $r_{.001}$ , based on the correlation integral, were especially powerful at picking out nonlinearity. The property detected by these statistics (as used here, with state vectors corresponding to 2 years of data), is that 2-year-long stretches of data are more similar to each other than would be expected strictly from the linear autocorrelations. Thus nonlinear modeling, and nonlinear forecasting, should be an improvement over linear prediction methods.

According to our Lyapunov exponent estimates, chaos ( $\lambda > 0$ ) appears to be very rare or absent. However, measles is often identified as being near the transition to chaos, with a mode in the distribution of exponents near 0. The same qualitative result was obtained in a survey of natural and laboratory animal populations (Ellner and Turchin 1993). In such cases the dynamics can easily vary between periods of stable behavior, and periods of chaos-like behavior (i.e., finite-time sensitive dependence on initial conditions: see Deissler and Farmer 1992). One way to quantify this type of behavior is by computing local (finite-time) Lyapunov exponents  $\lambda_m$ , defined by equation (2) with a finite value of  $m$  (Abarbanel et al. 1991, 1992; Wolff 1992 and references therein). Figure 5 shows a plot of  $\lambda_m$  over time for the Copenhagen measles series, for  $m=1$  or 2 years; because  $\lambda$  is near 0 there are frequent transitions between sensitive and insensitive short-term dependence on initial conditions. For this type of dynamics, a precise estimate of  $\lambda$  may be less useful than a rough characterization of the pattern of fluctuations in local exponents (e.g., their variance, autocorrelation, frequency of sign changes).

Methods are still evolving rapidly, so our results and conclusions are hardly the last word on nonlinearity and chaos in epidemics. One promising direction, encouraged by the feasibility of GMM model fitting, is to hybridize between mechanistic and statistical modeling. We expect that models that are mechanistic insofar as possible, but rely on state-space reconstruction and nonparametrics where ignorance forces that upon us, have the potential to provide more reliable characterizations of the dynamics, and more reliable prediction methods.

## LITERATURE CITED

Abarbanel, H. D. I., R. Brown, and M. B. Kennel. 1991. Variation of Lyapunov exponents on a strange attractor. *Journal of Nonlinear Science* 1: 175-199.

Abarbanel, H. D. I., R. Brown, and M. B. Kennel. 1992. Local Lyapunov exponents computed from observed data. *Journal of Nonlinear Science* 2: 343-365.

Casdagli, M. 1992a. Chaos and deterministic versus stochastic and non-linear modelling. *Journal of the Royal Statistical Society B*54, 303-328.

Casdagli, M. 1992b. A dynamical systems approach to modeling input-output systems. pp. 265-281 in: M. Casdagli and S. Eubank (eds.) *Nonlinear Modeling and Forecasting*. SFI Studies in the Sciences of Complexity Proc. Vol. XII. Addison-Wesley, NY.

Crutchfield, J.P., J.D. Farmer, and B.A. Huberman. 1982. Fluctuations and simple chaotic dynamics. *Physics Reports* 92, 45-82.

Deissler, R. J. and J.D. Farmer. 1992. Deterministic noise amplifiers. *Physica D* 55: 155-165.

Eckmann, J.-P. and D. Ruelle. 1985. Ergodic theory of chaos and strange attractors. *Reviews of Modern Physics* 57: 617-656.

Ellner S. 1991. Detecting low-dimensional chaos in population dynamics data: a critical review. pp. 63-90 in: J.A. Logan and F.P. Hain (eds.) *Chaos and Insect Ecology*. VPI&SU, Blacksburg VA.

Ellner, S., A.R. Gallant, D. McCaffrey, and D. Nychka. 1991. Convergence rates and data requirements for Jacobian-based estimates of Lyapunov exponents from data. *Physics Letters A* 153, 357-363.

Ellner, S., D. W. Nychka, and A. R. Gallant. 1992. LENNS, a program to estimate the dominant Lyapunov exponent of noisy nonlinear systems from time series data. Institute of Statistics Mimeo Series #2235, Statistics Department, North Carolina State University, Raleigh NC 27695-8203.

Ellner, S. and P. Turchin. 1993. Chaos in a 'noisy' world: new methods and evidence from time series analysis. Preprint.

Gallant, A.R. and G. Tauchen. 1992. Which moments to match? Working Paper, Department of Economics, Duke University, Durham NC.

Grassberger, P. and I. Procaccia. 1983. Measuring the strangeness of strange attractors. *Physica D*9, 189-208.

Hastings, A., C. L. Hom, S. Ellner, P. Turchin, H. J. C. Godfray. 1993. Chaos in ecology: is mother nature a strange attractor? *Annual Review of Ecology and Systematics* (in press).

Kifer, Y. 1986. *Ergodic Theory of Random Transformations*. Birkhäuser, Boston, Massachusetts.

Kot, M., W. M. Schaffer, G.L. Truty, D.J. Graser, and L.F. Olsen. Changing criteria for

imposing order. *Ecological Modelling* 43: 75-110.

McCaffrey, D., S. Ellner, D.W. Nychka, and A.R. Gallant. 1992. Estimating the Lyapunov exponent of a chaotic system with nonlinear regression. *Journal of the American Statistical Association* 87: 682-695.

Nychka, D.W., S. Ellner, D. McCaffrey, and A.R. Gallant. 1992. Finding chaos in noisy systems. *Journal of the Royal Statistical Society Series B* 54: 399-426.

Packard, N.H., J.P. Crutchfield, J.D. Farmer, and R.S. Shaw. 1980. Geometry from a time series. *Physical Review Letters* 45: 712-716.

Ramsay, J.B. and P. Rothman. 1992. Time irreversibility and stationary time series: estimators and test statistics. Preprint.

Rand, D.A. and H.B. Wilson. 1991. Chaotic stochasticity: a ubiquitous source of unpredictability in epidemics. *Proceeding of the Royal Society of London B* 246: 179-184.

Rothman, P. 1990. Characterization of the Time Irreversibility of Economic Time Series. Ph.D. Dissertation, Department of Economics, New York University.

Sauer, T., J. A. Yorke, and M. Casdagli. 1991. Embedology. *Journal of Statistical Physics* 65: 579-616.

Sayers, C. 1990. Chaos and the business cycle. pp. 115-125 in: S. Krasner (ed.), *The Ubiquity of Chaos*. AAAS, Washington DC.

Schaffer W. M. and M. Kot. 1985. Nearly one dimensional dynamics in an epidemic. *Journal of Theoretical Biology* 112, 403-427.

Schaffer W. M., and M. Kot. 1986. Chaos in ecological systems: the coals that Newcastle forgot. *Trends in Ecology and Evolution* 1: 58-63.

Schaffer WM, Olsen LF, Truty GL, Fulmer SL. 1990. The case for chaos in childhood epidemics. pp. 138-166 in: S. Krasner (ed.) *The Ubiquity of Chaos*. AAAS, Washington DC.

Smith, R. L. 1992a. Estimating dimension in noisy chaotic time series. *Journal of the Royal Statistical Society Series B* 54: 329-351.

Smith, R. L. 1992b. Relation between statistics and chaos. *Statistical Science* 7:109-113.

Stone, L. 1992. Coloured noise or low-dimensional chaos? *Proceedings of the Royal Society of London Series B* 250, 77-81.

Sugihara, G. and R.M. May. 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature* 344, 734-741.

Takens, F. 1981. Detecting strange attractors in turbulence. pp. 366-381 in: *Dynamical Systems and Turbulence*, Warwick 1980, Lecture Notes in Mathematics vol. 898 (Eds. Rand, D. and Young, L.-S.) Berlin: Springer.

Theiler, J. 1990. Estimating fractal dimension. *Journal of the Optical Society of America A* 7: 1055-1073.

Theiler, J., S. Eubank, A. Longtin, B. Galdrikian, J.D. Farmer. 1992. Testing for nonlinearity in time series data: the method of surrogate data. *Physica D* 58: 77-94.

Theiler, J., P. S. Linsay and D. M. Rubin. 1993. Detecting nonlinearity in data with long coherence times. in: A.S. Weigend and N.A. Gershenfeld eds. *Predicting the future and understanding the Past*. SFI Studies in the Sciences of Complexity vol XVII. Addison-Wesley (to appear).

Wolf, A., J.B. Swift, H.L. Swinney, and J.A. Vastano. 1985. Determining Lyapunov exponents from a time series. *Physica* 16D: 285-315.

Wolff, R.C.W. 1992. Local Lyapunov exponents: looking closely at chaos. *Journal of the Royal Statistical Society Series B* 54: 353-371



Table 2. Estimated Lyapunov exponents by neural net time series models for measles monthly data. All models used  $L=3$ , with the numbers of lags ( $d$ ) and units ( $k$ ) in the model chosen by the GCV criterion as described in the text. Nonseasonal models only use lagged values of the time series; seasonal models include  $\sin(2\pi j/12)$  and  $\cos(2\pi j/12)$  as covariates ( $j = \text{time}$  in months).

	<u>SEASONAL</u>			<u>NON-SEASONAL</u>		
	#lags	#units	$\lambda$	#lags	#units	$\lambda$
Baltimore	5	4	-0.11	8	7	+0.09
NYC	5	3	-0.08	6	6	+0.02
Detroit	6	5	-0.05	6	6	+0.025
Copenhagen	5	6	-0.01	8	6	+0.06

Table 3. Estimated Lyapunov exponent  $\lambda$  for quarterly case reports using seasonal neural net model. All models used  $L=1$ , with the number of lags ( $d$ ) and units ( $k$ ) chosen by the GCV criterion as described in the text.

	#lags	#units	$\lambda$	$r^2$	df
<b>MEASLES</b>					
NYC	3	2	-0.67	0.93	123
Baltimore	4	2	-0.07	0.83	109
Detroit	6	2	-0.08	0.85	145
Milwaukee	2	2	-7.78	0.77	103
Copenhagen	2	3	-0.06	0.87	135
London	2	1	-0.23	0.67	51
Bristol	3	1	-0.13	0.77	50
Liverpool	2	1	-1.56	0.72	51
Manchester	2	2	-0.24	0.90	45
Newcastle	2	1	-3.61	0.71	51
Birmingham	2	2	-0.16	0.92	45
Sheffield	5	1	-1.93	0.84	48
<b>MUMPS</b>					
NYC	5	2	+0.01	0.94	119
Milwaukee	2	2	-0.39	0.74	153
Copenhagen	2	3	-0.24	0.86	135
<b>RUBELLA</b>					
St.Louis	2	2	-0.27	0.76	61
Copenhagen	2	1	-0.87	0.71	99
<b>CHICKENPOX</b>					
NYC	6	2	-0.14	0.95	117
Detroit	2	1	-0.33	0.78	61
St.Louis	1	1	-1.46	0.86	68
Copenhagen	1	1	-0.61	0.81	107
Milwaukee	2	2	-0.37	0.86	129

### FIGURE LEGENDS

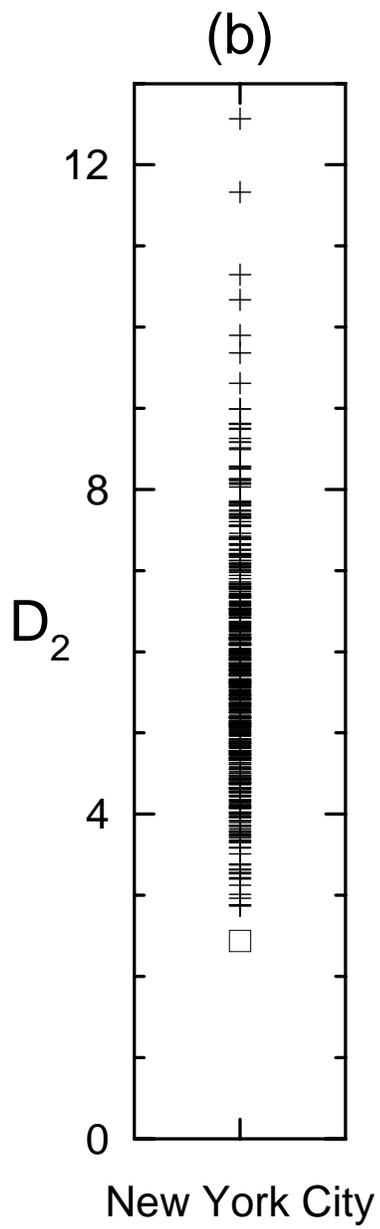
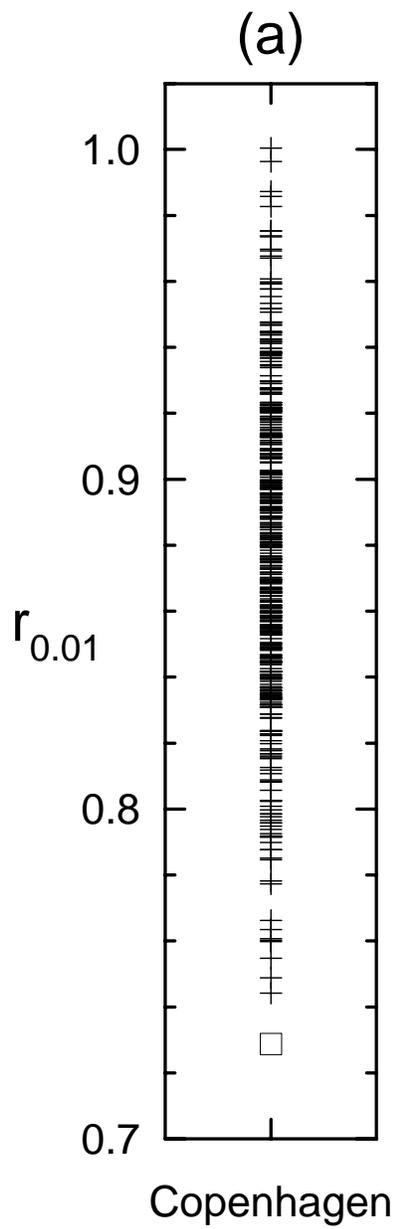
Figure 1. (a) The statistic  $r_{.01}$  is shown for the Copenhagen measles data ( $\square$ ), and for 500 surrogate time series (+). The value is significantly smaller for the actual data than for the surrogates. (b) The estimated correlation dimension  $D_2$  is shown for New York City measles data ( $\square$ ), and for 500 surrogate time series (+). Again, the actual data exhibits a much smaller dimension than is seen in the surrogate time series.

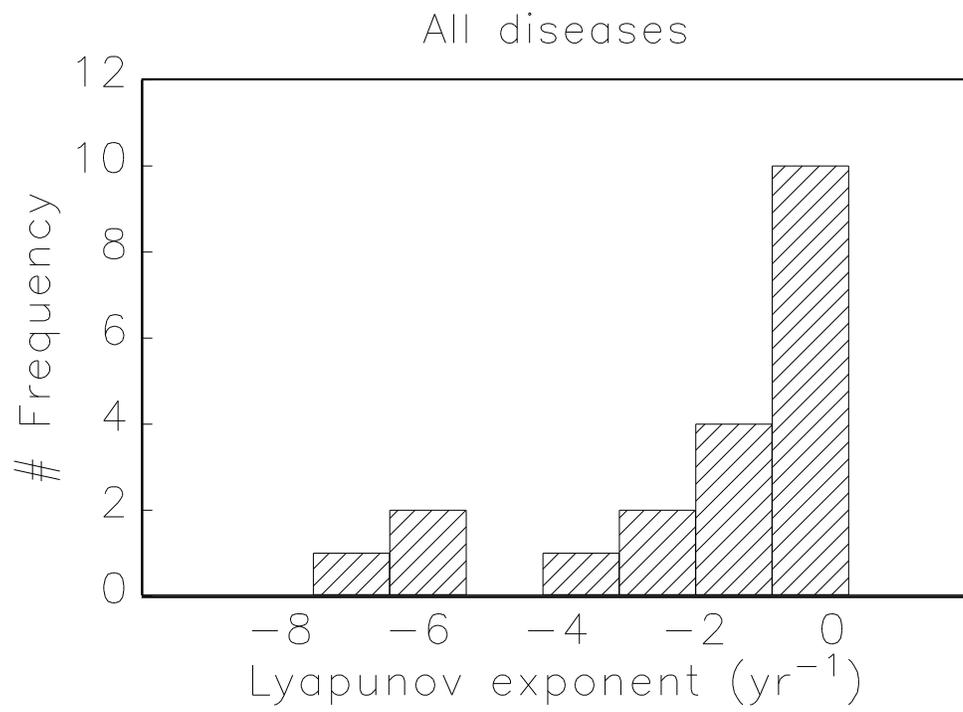
Figure 2. Histogram of estimated Lyapunov exponents for quarterly disease case reports. Values in Table 3 were multiplied by 4 to express exponents in units  $\text{year}^{-1}$ .

Figure 3. (a) Contour plot of the GMM objective function; smaller values correspond to better fits between model and data. Contours based on values computed at a regular  $31 \times 31$  grid over the range of values shown for  $b_0$  and  $b_1$ , with a simulation of 5000 months duration for each parameter combination. (b) Plot of minimum GMM objective function as a function of the relative intensity of seasonal forcing  $b_1/b_0$ . For both (a) and (b) the following parameter values were used in the SEIR model:  $m = 0.02$ ,  $a = 55$ ,  $g = 60$ ,  $S_0 = 0.05$ ,  $E_0 = I_0 = 0.001$ ,  $\delta = .01$ .

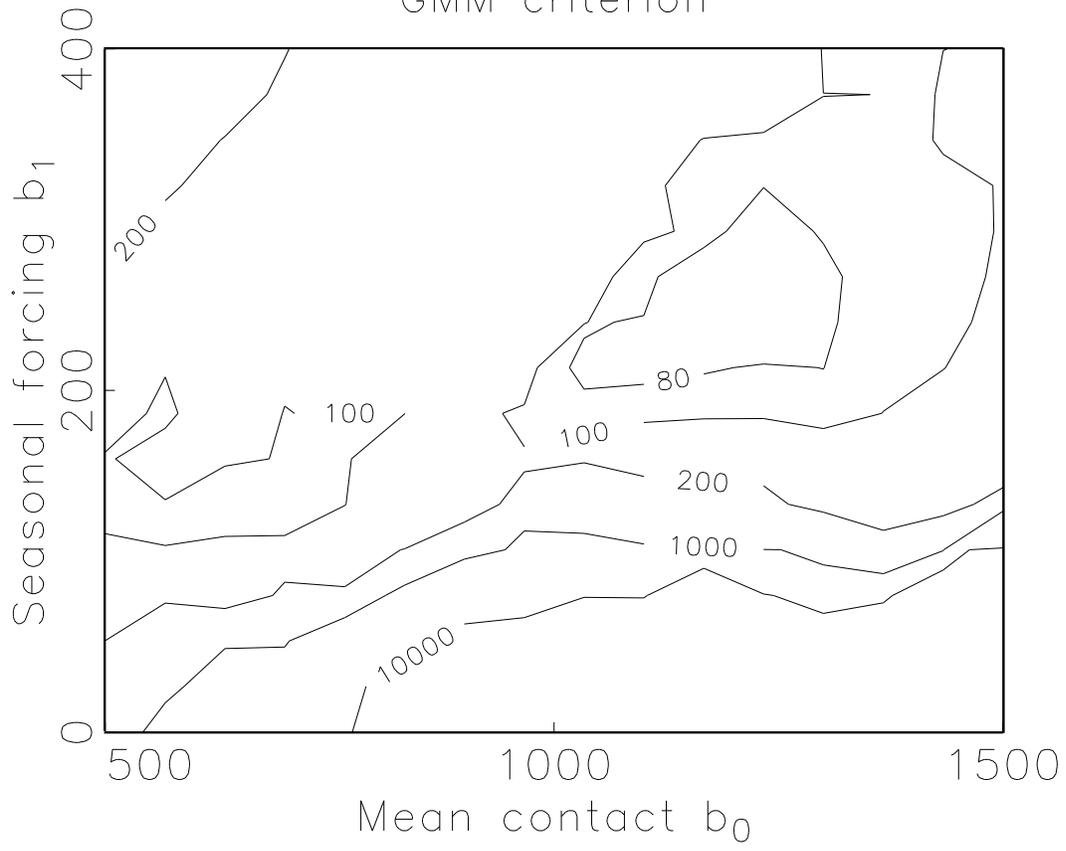
Figure 4. Plot of GMM objective function (computed as in Figure 3) at a grid of values near the best-fit parameter values.

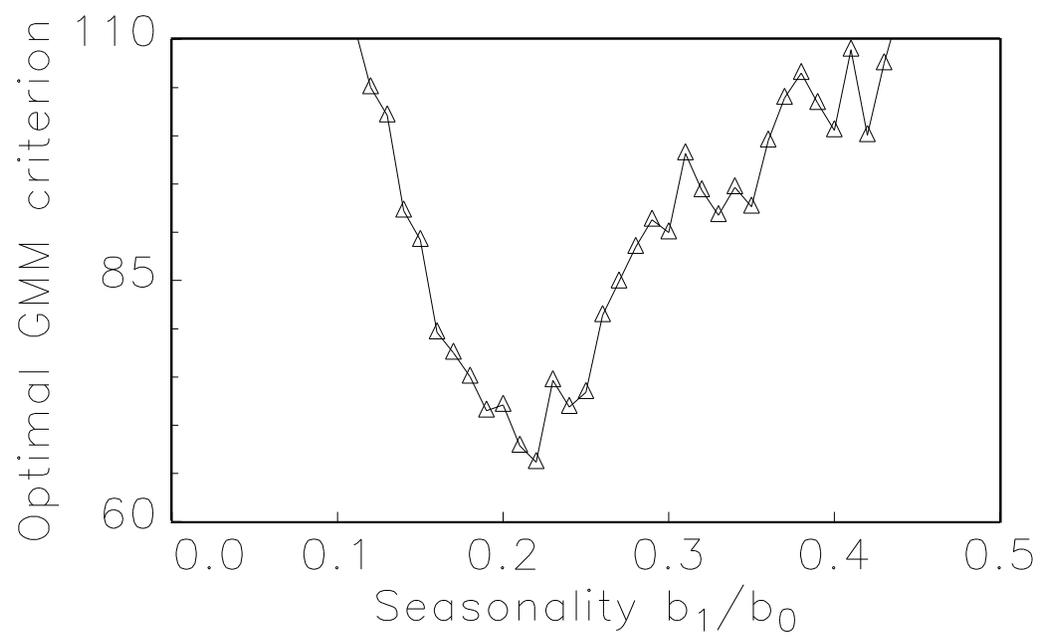
Figure 5. Finite-time “local” Lyapunov exponents for Copenhagen measles, data based on the best-fit seasonal neural net model for quarterly data.

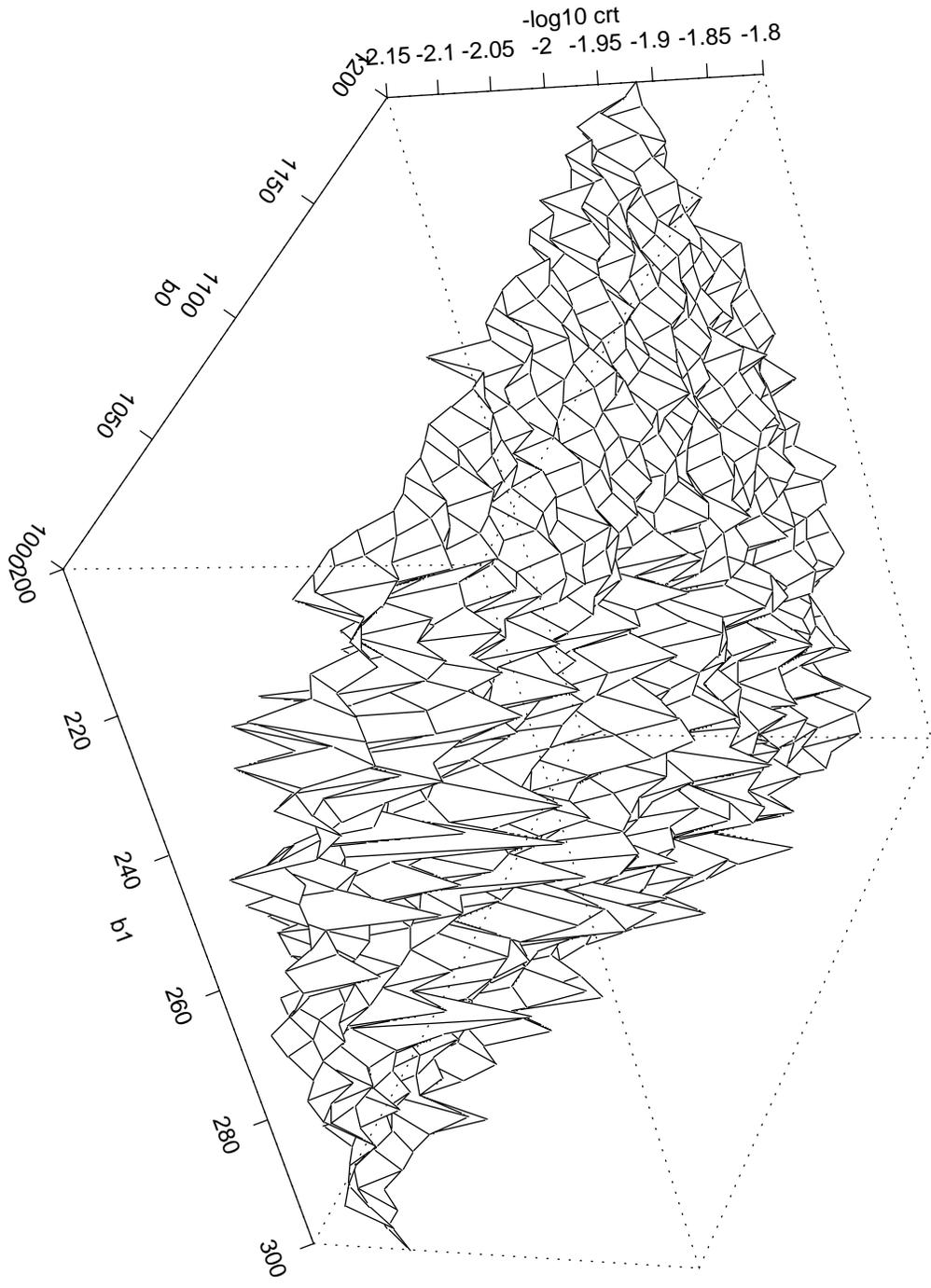




GMM criterion







Copenhagen measles

