Distinguishing Error from Chaos in Ecological Time Series [and Discussion]

George Sugihara; Bryan Grenfell; Robert M. May; P. Chesson; H. M. Platt; M. Williamson

Stable URL:

http://links.jstor.org/sici?sici=0962-8436%2819901129%29330%3A1257%3C235%3ADEFCIE%3E2.0.CO%3B2-O

Philosophical Transactions: Biological Sciences is currently published by The Royal Society.

Distinguishing error from chaos in ecological time series

GEORGE SUGIHARA[1], BRYAN GRENFELL[2] AND ROBERT M. MAY[3]

[1] *Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093, U.S.A.*
[2] *Department of Zoology, Cambridge University, Cambridge, U.K.*
[3] *Department of Zoology, Oxford University, Oxford, OX1 3PS, U.K. And Imperial College, Exhibition Road, London SW7 2AZ, U.K.*

SUMMARY

Over the years, there has been much discussion about the relative importance of environmental and biological factors in regulating natural populations. Often it is thought that environmental factors are associated with stochastic fluctuations in population density, and biological ones with deterministic regulation. We revisit these ideas in the light of recent work on chaos and nonlinear systems. We show that completely deterministic regulatory factors can lead to apparently random fluctuations in population density, and we then develop a new method (that can be applied to limited data sets) to make practical distinctions between apparently noisy dynamics produced by low-dimensional chaos and population variation that in fact derives from random (high-dimensional)noise, such as environmental stochasticity or sampling error.

To show its practical use, the method is first applied to models where the dynamics are known. We then apply the method to several sets of real data, including newly analysed data on the incidence of measles in the United Kingdom. Here the additional problems of secular trends and spatial effects are explored. In particular, we find that on a city-by-city scale measles exhibits low-dimensional chaos (as has previously been found for measles in New York City), whereas on a larger, country-wide scale the dynamics appear as a noisy two-year cycle. In addition to shedding light on the basic dynamics of some nonlinear biological systems, this work dramatizes how the scale on which data is collected and analysed can affect the conclusions drawn.

1. INTRODUCTION

The classical debate between the biotic and climatic schools has divided opinion over the relative importance of deterministic versus stochastic forces in controlling ecological populations (Sinclair 1989). This long-standing debate over random versus determined variation has begun to take on new meaning with recent interest in chaos and nonlinear dynamics, and with the ever-increasing demonstrations of the applicability of these ideas to real data. Until recently, one would have viewed a time series such as the one shown in figure 1a and concluded that the ecologically important information here rested in the smooth fitted line. There is, however, a change of view occurring in dynamics, similar to the change that fractals is bringing to geometry and the study of spatial pattern, which suggests that the most interesting things may be found in the irregularities rather than in the smoothed pattern (Lorenz 1969; Takens 1981; Schaffer & Kot 1986; Sugihara & May 1990). Although initially it appears that incorporating such detail into the population dynamics debate may further cloud the problem, we shall argue that the end result is not new difficulty, but the prospect of a new clarity and simplicity.

The paper is divided into four sections. In the first, we discuss the limitations of traditional approaches to analysing deterministic influences on population dynamics. The second section outlines the light which recent advances in nonlinear dynamics theory shed on these problems. In particular, it summarizes a new method for distinguishing noise from low dimensional determinism in ecological time series, based on their internal predictability (Sugihara & May 1990). The third section applies this method to time series of notified case reports for childhood diseases in developed countries. Because they are often relatively long, and reflect comparatively simple host–microparasite population interactions, these series are among the best ecological candidates for applying nonlinear methods (Schaffer & Kot 1986; Schaffer *et al.* 1988). After reviewing previous work in this area, we present a new analysis of the dynamics of measles in English cities, which has significant implications for the question of spatial scaling in ecological systems and the concept of stationarity as defined in traditional time series analysis. The final section draws these conclusions together and suggests lines for future work.

2. TRADITIONAL APPROACHES

The classical approaches to analysing population
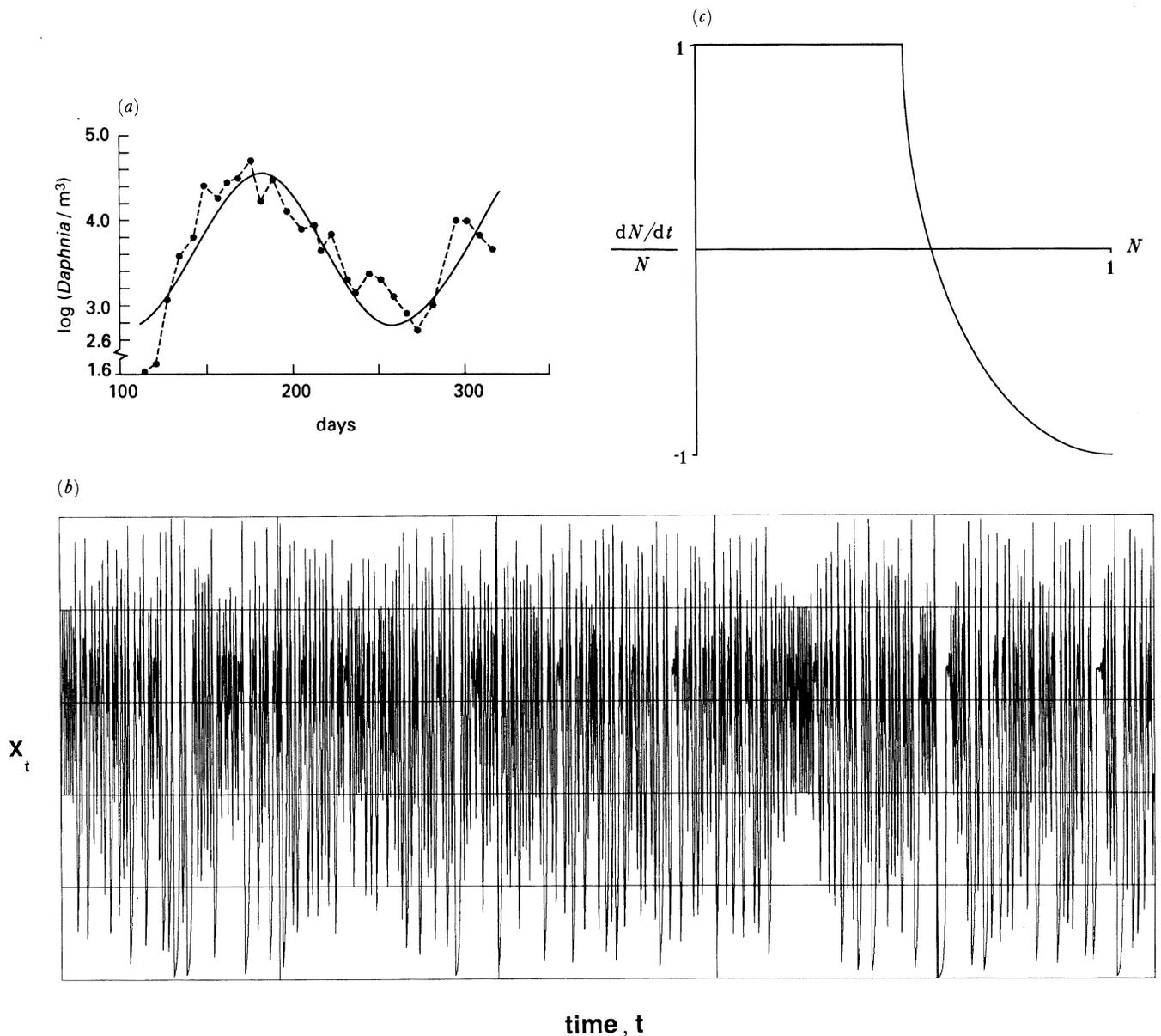
[113]

235

16-2

Figure 1. (a) An example of population dynamics for *Daphnia* and edible algae from Lake Washington (Murdoch & McCaughley 1985). The dashed line joins the observations and the smooth solid line is a sine wave fitted to the seasonal pattern. (b) Time series of 1000 points for the chaotic tent map: $x_{t+1} = 2x_t$ if $x_t < 0.5$; $x_{t+1} = 2 \ (1-x_t)$ if $x_t > 0.5$. These data are in many ways indistinguishable from white noise. (c) Key factor analysis applied to the data in figure 1b. Although the signature appears to be an example of density vague population regulation (no density dependence at low densities with severe control at high densities), it was derived from a low-dimensional chaotic process.

dynamics are key factor analysis of observed data (to distinguish density-independent and nonlinear density-dependent influences) and the exploration of laboratory and mathematical models that reduce the complexity of real systems to a single or a few factors (Sinclair 1989). Although these methods have been highly successful in the main, recent developments in nonlinear theory identify a number of problems, which we summarize as follows.

(a) *Limitations of key factor analysis*

Consider the model time series shown in figure 1b, which appears to be stochastic. If this were a time series for a real population, one might easily conclude that it

represents a population that is being buffeted by frequent random shocks with apparent occasional returns to a quasi-equilibrium. Indeed, as shown in figure 1c, a key factor analysis based on these data would lead one to the opinion that this is a classic example of what some would call 'density-vague control', i.e. no regulation at low densities with control occurring only at high densities. Yet these data do not represent density-vagueness at all, but are an example of simple chaotic dynamics that were generated from the deterministic tent map (Sugihara & May 1990). Thus, in this case, an interpretation of these data as arising from external unpredictibilities would have been incorrect. Conventional approaches would have mis-identified what was in fact simple chaotic dynamics
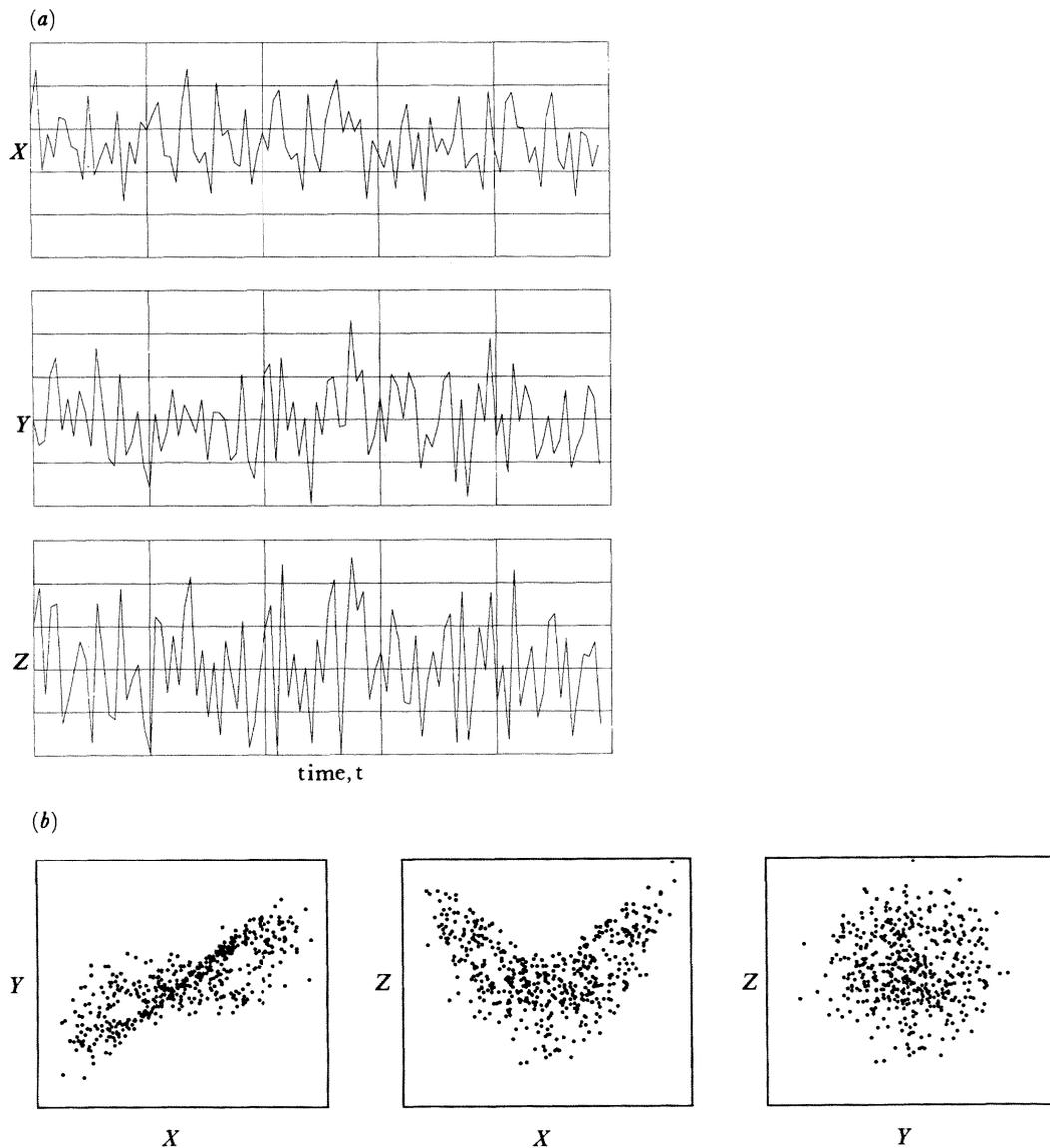
Figure 2. (a) Simultaneous time series for the three variables of the Lorenz system: X, Y and Z as functions of time. Each time-series represents the projection onto one axis of the Lorenz attractor (the so-called butterfly attractor) as it is embedded in three dimensions. (b) The time series in figure 2a were randomly sampled at 500 simultaneous times, and the correlations between all pairs of variables for the Lorenz attractor are shown here. The correlation between X and Y has $r = 0.636$; that between X and Z has $r = 0.001$; and Y and Z have $r = 0.000$. That is, although Y and Z are deterministically coupled they show no correlation.

(i.e., low dimensional chaos) as random variation within a density-vague envelope.

(b) *Limitations of the single factor approach*

To complicate matters further (as discussed here by Godfray & Blythe (1990)), the dynamics of real populations can only be properly understood if they are considered in their ecological context, that is in a way that recognizes the inherent complexity or multidimensionality of the problem. Populations do not exist singly, but are embedded in a dynamic web of other species and environmental forces. While we are consciously willing to acknowledge this fact, it is important to realize how such complexity could blur any relations that one could ever hope to see in a typical single-factor analysis (regressing one explana-

tory variable against another: a mainstay of ecological research) such as the key factor approach.

This can be shown with the following simple example. The three time series shown in figure 2a were generated from the three variables that describe the well known Lorenz equations (Lorenz 1969). The system is completely determined by these three variables. Suppose now that we do not know that these time series came from a Lorenz system. We only know that they are parallel measurements from some system. Typically when one is trying to understand a phenomenon one looks for patterns, in this case correlations between the time series. If we randomly sample these series at simultaneous times and look for pairwise correlations we get the result shown in figure 2b. While there is a significant relation between X and Y, there is no significant linear relation between X and Z

(although there is a significant parabolic relation), and no relation of any kind between Y and Z. Thus even though Y and Z are deterministically coupled they appear completely unrelated. That is to say, lack of correlation between pairs of variables does not imply lack of causation.

We are faced, therefore, with two dilemmas that at first glance are not obviously related: (i) how to distinguish randomness from low dimensional determinism, and (ii) how to explicitly acknowledge higher dimensions. These two problems are not only central to population regulation debates, but are at the heart of nonlinear theory as shall be discussed in the next section.

3. NONLINEAR PERSPECTIVE

Suppose we have perfect knowledge about the internal dynamics of a population, so that without arbitrary disturbance everything is determined and nothing is random. That is, we know all the variables and the functions describing how they are related. The space whose axes consist of each of these dynamically coupled variables is known as the 'state space.' For example, if this is an experimental multispecies system where the only important variables are other species, then the axes of the state space would be the population sizes of each of the coupled species. All population variability and motion would be constrained to some sub-manifold of that space set by the relations (functions) among variables. If, after an arbitrary perturbation, the population returns to it this sub-manifold, it is called an 'attractor' (Takens 1981; Abraham & Shaw 1982; Schaffer & Kot 1985; Godfray & Blythe 1990).

The central concepts here, of state space and attractor, represent the scientific ideal of perfect deterministic knowledge. In practice, however, we do not know the state space. Rather, we might have time series for one or more of the variables. Thus, in practice, the state space and its attractor are a black box, and the time series are observables or outputs. Each of these time series can be viewed essentially as a projection onto one axis of the state space through time (e.g. figure 2a shows the three time series for the Lorenz attractor). That is, each time series is a view in one dimension of a process occurring in higher dimensions. Therefore, in a perfectly deterministic world, much of the complexity or apparent randomness in a time series will arise from a state space having a high number of dimensions, or alternatively (and possibly in conjunction with), an attractor with chaotic dynamics.

Let us focus on the case where we have only one time series. How can we get information about the state space and attractor that produced it? The standard method here is Taken's (1981) technique of using lagged coordinates to embed a time series in higher dimensions (Godfray & Blythe 1990). Although embeddings can be created from the original state-space time series themselves, here we shall consider the worst case where there is only one time series to work with; this is also the case where the lagged coordinate idea is

most valuable. Again, the space created by such an embedding is not the original state-space, but a mock version of it: something we shall call a 'phase space'. As proved by Takens (1981), the phase space retains essential properties of the original state-space including the dimensionality of the attractor it contains. Moreover, as we shall see, a phase space can be used to make forecasts, properties of which provide practical distinctions between low dimensional determinism and noise, even with limited data of the kind encountered in population biology (Sugihara & May 1990).

Thus even though our ultimate goal is to understand the population in state-space (i.e., how many dimensions does it have? are the dynamics low dimensional chaos or simply noisy? etc.), we may only have a time series for one variable of this space (namely for the population itself whose dynamics we are trying to understand). To get around this problem we shall construct a phase space as a surrogate having the same topological characteristics as the original state-space (again, a higher dimensional embedding of the time series by using time-lagged coordinates). Information about the original attractor can then be gained by exploring the properties of the mock attractor in phase space.

(a) Nonlinear forecasting to differentiate noise from chaos: basic ideas

The method outlined here is discussed in fuller detail in Sugihara & May (1990) and is based on theories of short-range prediction proposed in general terms by Lorenz (1969) and others (Tong & Lim 1980; Priestly 1980; Farmer & Sidorowich 1989). The basic idea here (which is classical in prediction) is that, if indeed deterministic laws govern a system, then even if the dynamical behaviour is chaotic, the future may be reckoned from the behaviour of past values that are similar to the present. The key, however, is in knowing the dimensionality within which the past, present and future are embedded.

Suppose (as discussed by Godfray & Blythe) we have properly embedded a time series in an E-dimensional phase space so that each lagged sequence of data points, $Z_t = \{x_t, x_{t-\tau}, \ldots, x_{t-(E-1)\tau}\}$ is a point in this E-dimensional space. Here we usually choose $\tau = 1$, but the results do not appear to be too sensitive to the choice of τ, provided it is not too large. This is like taking an E-pronged fork whose tines are separated by a distance τ, and dragging it sideways along the time series; the vector of time-series values, Z_t, formed by where the tines land at each instant, describes another E-dimensional point, and the set of vectors $\{Z\}$ describes the attractor. In general, $E_{min} \leqslant 2D+1$, where D is the attractor dimension. Each predictee Z_i is now to be regarded as an E-dimensional point, comprising the present value x_i and the $E-1$ previous values each separated by one lag time τ. We now locate all nearby E-dimensional points in the phase space and choose a minimal neighbourhood so that the predictee is contained within the smallest simplex formed from its $E+1$ closest neighbours; a simplex containing $E+1$ vertices is the smallest simplex that can contain an

E-dimensional point as an interior point. When a minimal bounding simplex cannot be found (for example, for boundary points), we use the *E*+1 nearest neighbours. To obtain a prediction, we simply project the domain of the simplex into its range (that is, we keep track of where the other points in the simplex end up at *p* time steps), and compute the weighted centre of mass of the simplex to get the predicted value. In other words, we follow the short term destiny of nearby points in the attractor to see where they end up after *p* time steps. This is a non-parametric method, and it should apply to any stationary or quasi-ergodic process, including chaos. It uses no previous information about the model used to generate the series, only the information in the time series output itself. Unless otherwise stated, we shall construct the phase space from the first half of the time series to make predictions on the second half.

(b) *Nonlinear forecasting to differentiate noise from chaos: examples from models*

Figure 3*a* shows an application of this method to the white-noise time series produced by taking first differences of the tent map series shown in figure 1*b*. It compares predicted against actual results two time steps into the future: a time step where there is no significant correlation between values. Notice again, the phase space constructed from the first half is used to predict the values in the second half. Thus, this time-series, which by standard statistical analysis is un-correlated white noise (unpredictable), in fact becomes strongly predictable when embedded in higher dimensions.

Figure 3*b* shows how predictability, as measured by the standard correlation coefficient, declines as the prediction interval T_p (i.e., how far into the future one projects) increases. Such a decrease in the correlation coefficient with increasing prediction time is the hallmark of chaos (or equivalently, of the presence of a positive lyapunov exponent, with the magnitude of the exponent being related to the rate of decrease of ρ with T_p). This property is noteworthy because it suggests a simple way to distinguish between additive noise and multiplicative chaos: predictions with uncorrelated additive noise will appear to have a fixed amount of error, regardless of how far or close into the future one tries to predict, whereas predictions with multiplicative chaos tend to deteriorate as one tries to forecast further into the future. This can be seen in figure 3*c* where it is shown that the characteristic signature of ρ decreasing with T_p does not arise when the erratic time series is in fact a noisy limit cycle (here additive noise superimposed on a sine wave). With uncorrelated additive noise, such as sampling variation, the error remains constant as the simplex is projected further into the future. In contrast, the dashed line in figure 3*c* represents the correlation coefficient (ρ) against prediction time (T_p) relation for a chaotic sequence generated as the sum of two independent runs of first-differences of the tent map. Although the two time series here both look alike as sample functions of some random process, the characteristic signatures differen-

tiate random noise in one instance from deterministic chaos in the other.

The predictions in figure 3*a–c* are based on an embedding dimension of *E* = 3. These results are, however, sensitive to the choice of *E*. This is shown in figure 3*d* where results are summarized for prediction accuracy (correlation between predicted and observed, here at $T_p = 1$) versus embedding dimension *E*. It may at first sight appear surprising that having potentially more information erodes the accuracy of the predictions, since for large *E* there are more data summarized in each *E*-dimensional point, and a higher dimensional simplex of neighbours for each predictee. Sugihara & May (1990) have suggested that this effect may be caused by contamination of nearby points in the higher dimensional embeddings with points whose earlier coordinates are close but whose recent (and more relevant) coordinates are distant. If this is so, this method may have additional application as a trial and error method of computing an upper bound on the embedding dimension, and thence on the attractor dimension.

So far, we have compared relations between ρ and T_p for chaotic time series with the corresponding relation for additive white noise. More problematic, however, is the comparison with ρ–T_p relations generated by coloured noise spectra where there are significant short-term correlations but no long term ones. As with chaos, such correlated noise can also lead to declining ρ–T_p curves. Although, in the limit, the shallow form of the decline in simple cases may distinguish correlated noise from a chaotic signature (Sugihara & May 1990; Farmer & Sidorowich 1989), in a practical sense, particularly with limited data of the kind available in population biology, such distinctions may be difficult to find. One practical solution to this dilemma, suggested by Sugihara & May (1990), is that coloured noise may tentatively be distinguished from deterministic chaos if in addition to an exponentially declining ρ–T_p curve, the correlation, ρ, between predicted and observed values obtained by nonlinear methods is significantly better than the correlation obtained by the best-fitting autoregressive linear model. That is, if a time series is chaotic it should have both a steeply declining ρ–T_p curve and more predictability under the nonlinear hypothesis (i.e. that it was produced by nonlinear mechanisms), than if one assumed it was produced by linear mechanisms (i.e. by the linear superposition of simple cycles of various period and amplitude).

Perhaps most germane to the biotic/climate issue is the possibility of noise entering multiplicatively as a disturbance to population numbers which is then fed back into the dynamics. An example would be noise entering in the form $X_{t+1} = F(X_t + \Theta_t)$, where *F* is assumed to have stable dynamics. Here, if *F* is nonlinear, a nonlinear predictor may still perform better. However, one can again expect a slower than exponential decline in the ρ–T_p curve. An exponential decline which arises from locally exponentially diverging trajectories may be taken as the operational definition of chaos. A simple way to use nonlinear forecasting to distinguish this possibility is by examin-
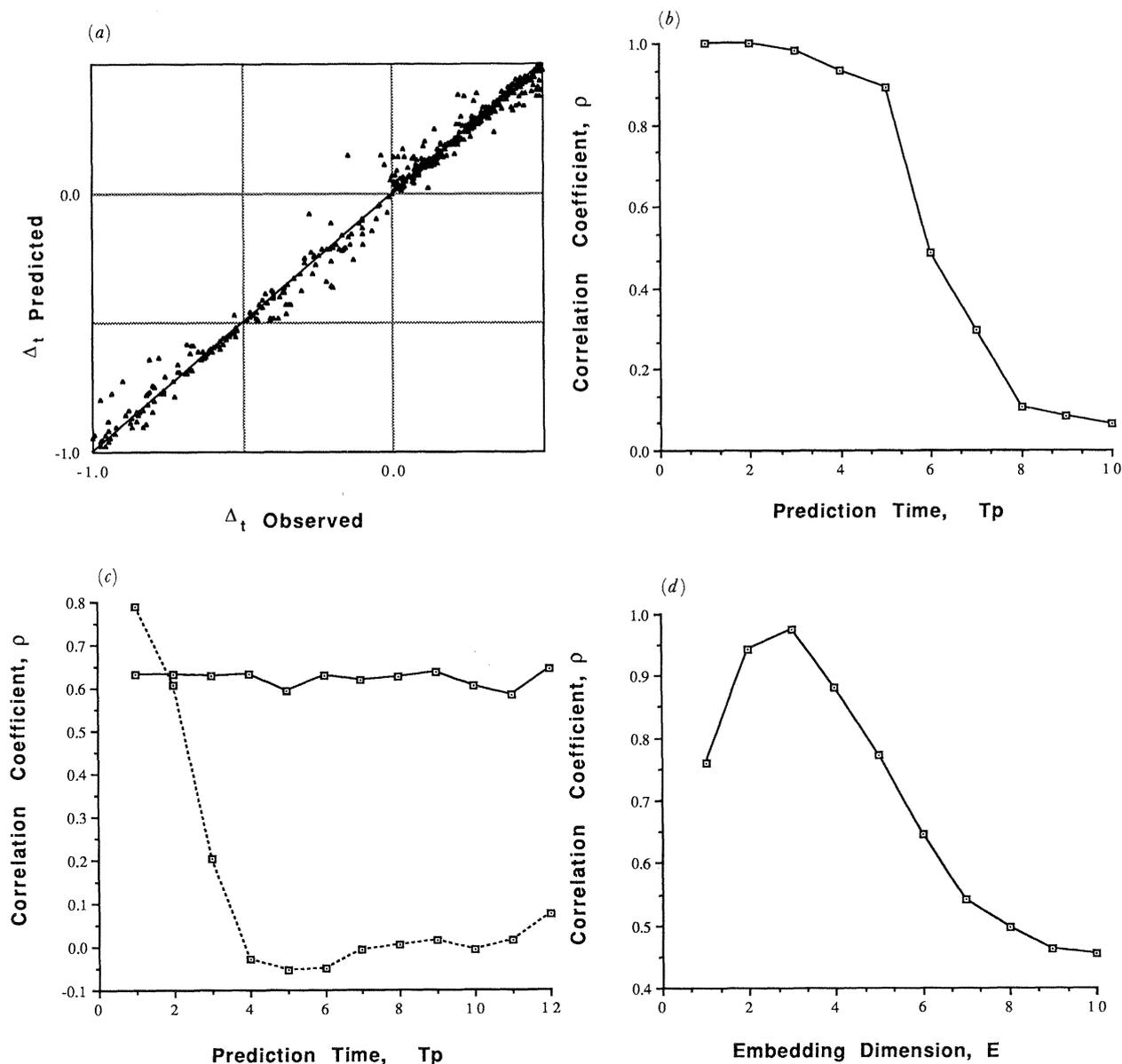
Figure 3. (a) Predicted values two steps into the future ($T_p = 2$) versus observed values for the white-noise time series produced by taking first differences of the tent map series shown in figure 1b (after Sugihara & May 1990). Specifically, the first 500 points in the series were used to generate a library of patterns, which were then used as a basis for making predictions for each of the second 500 points. As described in the text, the predictions were made using a simplex projection method (Sugihara & May 1990) with an embedding dimension and lag time of $E = 3$ and $\tau = 1$, respectively. Here the coefficient of correlation between predicted and actual values is $\rho = 0.997$ ($N = 500$). For comparison, we note that the corresponding correlation coefficient using an autoregressive linear model is $\rho = 0.04$. (b) Predictability measured by the standard correlation coefficient, ρ, as a function of how far into the future the forecast is made, T_p. The exponential decay in predictability with increasing prediction time, T_p, shown here is a characteristic of chaotic dynamics. (c) Additive noise (solid line) versus chaos (dashed line). The solid line shows that the correlation, ρ, between predicted and observed values for the case of additive noise (here white noise superimposed on a sine wave) does not decline as one tries to forecast further into the future. By contrast, the dashed line shows the declining signature characteristic of a chaotic sequence (here the sum of two separate tent map sequences). (d) Correlation coefficient, ρ, between predicted and observed values as a function of embedding dimension, E, for predictions one time step into the future ($T_p = 1$); like figures 3a and b, the figure is based on the time series shown in figure 1b.

ing whether the E-dimensional simplex tends to expand (chaos) or contract (noise fed through stable dynamics) when projected from its domain into its range; this idea will be developed further elsewhere.

4. APPLICATION: THE DYNAMICS OF CHILDHOOD MICROPARASITIC INFECTIONS

Because of their public health importance, the epidemiology of childhood viral diseases such as

measles and chickenpox in developed countries is especially well documented (Nokes & Anderson 1986). In particular, the relatively long time series of case reports accumulated from disease notification programmes in Europe and the U.S.A. provide a mass of information about the characteristically oscillatory dynamics of these infections in human communities (Anderson *et al.* 1984). The dynamic origin of this recurrent epidemic behaviour has been extensively examined, both in terms of mathematical models (Bartlett 1957; Anderson & May 1985; May 1986; Hethcote & Levin 1989), and time-series analysis of epidemiological data (Anderson *et al.* 1984). The data for measles have received considerable attention recently, and have been the focus of a debate as to whether measles dynamics is simply a noisy limit cycle (Schwartz 1985), or low dimensional chaos superimposed on a seasonal cycle (Schaffer & Kot 1985, 1986; Schaffer *et al.* 1989). Much of this controversy has centred on Schaffer's pioneering analyses of case reports for New York, and so we begin by applying the prediction method to these data.

(a) Measles and chickenpox in New York

The methods described above have been used to analyse public health records of monthly changes in the reported incidence of measles and chickenpox in New York City (Sugihara & May 1990). The results are summarized in figure 4. The earlier arguments for chaos, based largely on qualitative judgements as to static properties of the attractor and model simulations (see, for example, Schaffer *et al.* 1989), were supported by the results of the forecasting analysis presented here. However, because prediction is a harder test of E-dimensional determinism than judgements as to the geometry of a putative attractor, we think this analysis constitutes the strongest evidence so far, for the measles attractor. Here we see a steeply declining ρ–T_p curve, with the characteristic signature of a chaotic process. The result is supported by the fact that the nonlinear predictor performs significantly better than the best linear predictor ($P < 0.0005$). An optimal embedding dimension of 5–8 is roughly consistent with our independent estimates of an attractor dimension of 2.5–3.5 by using the Grassberger–Procaccia (1983) algorithm (Schaffer & Kot (1985) also report an estimated attractor dimension of 2.5 for measles). Thus, we believe the apparent irregularity in the measles time series is not due to random effects (environmental shocks or measurement errors), but is generated by low dimensional chaos.

In contrast, the analysis for chickenpox (figure 4b) suggests that complexity here is not due to low dimensional chaos, but to noise (possibly high dimensional chaos), superimposed on a strong annual cycle. These data produce a flat ρ–T_p curve, similar to the additive noise case in figure 3c. Moreover, the best linear predictor was found to perform at least as well as our nonlinear predictor (the correlation between the predicted and observed monthly change in the numbers of chickenpox cases using the linear predictor was $\rho = 0.84$, and for the nonlinear predictor $\rho = 0.82$

where $N = 266$ and $T_p = 1$). Thus, there is no evidence that the irregularity in the chickenpox data is due to anything other than random noise.

As discussed in more detail elsewhere (Sugihara & May 1990), there are biological reasons to explain why measles may exhibit chaotic dynamics (essentially deriving from a roughly two-year 'inter-epidemic period' interacting with annual variations in transmissibility), while chickenpox (where infectiousness can recrudesce at older ages) exhibits only annual periodicities.

(b) Measles in England and Wales

We now extend the analysis to data on changes in the monthly incidence of measles in England and Wales (figure 5a). These and subsequent data were extracted from the Registrar General's Weekly Returns, for the period 1948 (when measles notifications began) to 1967 (just before the onset of mass measles vaccination in 1968 significantly altered the dynamics of the infection; Anderson *et al.* (1984)). As with earlier analyses (Sugihara & May 1990) we begin by transforming the data to first differences, partly to remove such linear trends as may exist and partly to increase the density of points in phase space. Because the time series here are very short ($N = 240$, roughly half the size of the New York series), to maximize the information content in estimating E we allow the library and predictions to span the full time period. However, to avoid circularity between our forecasts and the model, we sequentially exclude points from the library that are in the neighbourhood of each predictee (specifically the $(E-1)\tau$ points preceding and following each forecast). Similar but much noisier results for estimating the embedding dimension were obtained by using the standard protocol of the first half predicting the second half. The standard protocol was used for the ρ–T_p curves, where the pattern, though noisy, appeared to be more robust. The qualitative appearance of these curves was found to be much the same for all choices of library.

As shown in figure 5b, we obtain optimal embeddings at $E = 7$–10, which is similar to the range of values found for measles in New York ($E = 5$–8). However, unlike New York, it appears that the dynamics here are not produced by low dimensional chaos. The ρ–T_p curve (figure 5c) does not decay exponentially as it does for the measles incidence data from New York City, but rather has a flat appearance more reminiscent of the additive noise case we saw for chickenpox (figure 4b). This result is corroborated by the comparison with the optimal linear model. For predicted changes in measles frequency one month in advance ($N = 120$), the best linear autoregressive model gives the result $\rho = 0.797$, which is not significantly different from the correlation obtained with the nonlinear predictor, where $\rho = 0.790$. Thus, unlike measles incidence in New York City, measles in England and Wales appears not to be chaotic.

Thus we are faced with an apparent contradiction: why should measles in New York City be chaotic while the same disease in the United Kingdom is a simple

[119]

(a)

Time in Months, t

(b)

Embedding Dimension, E
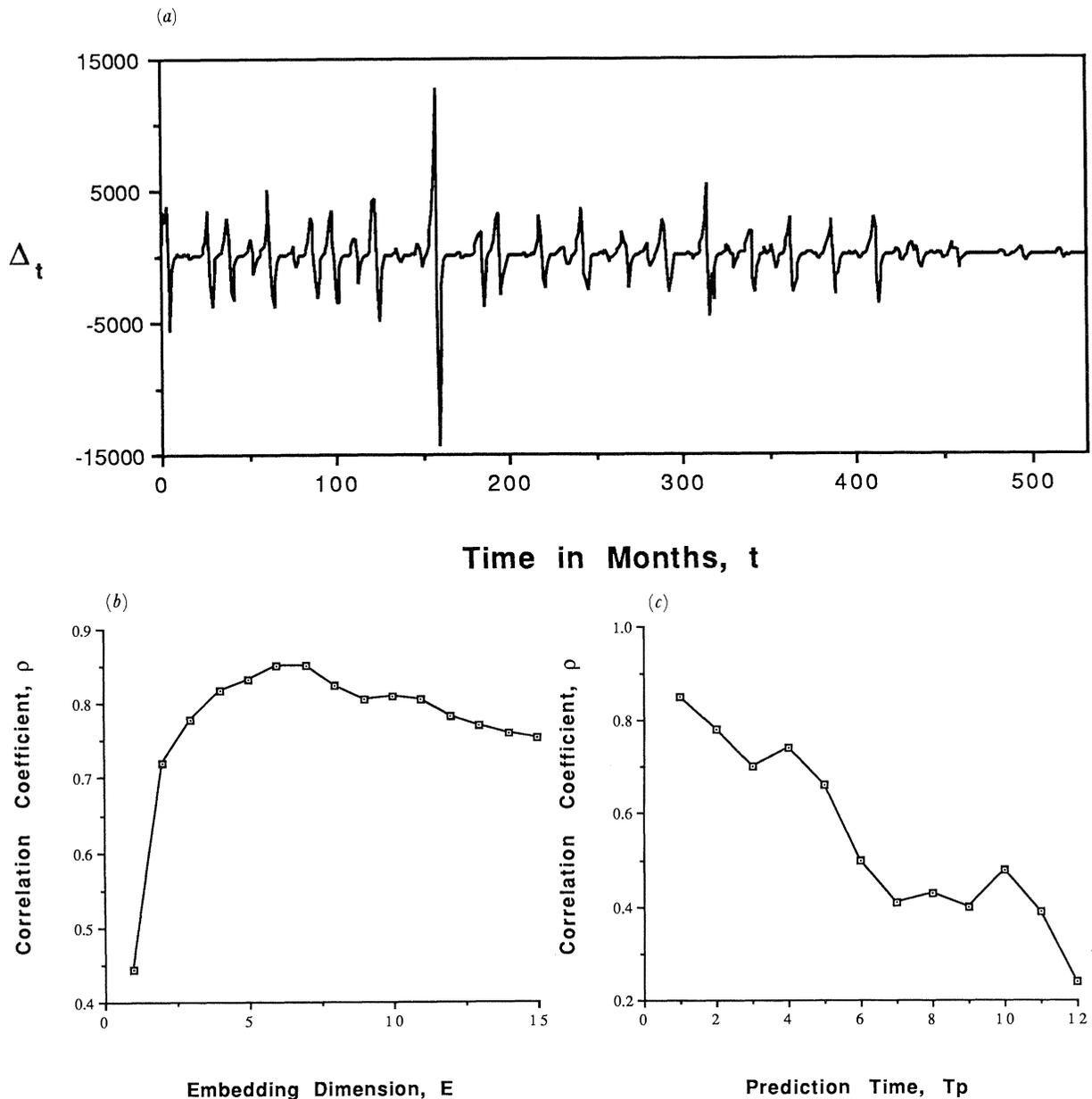
(c)

Prediction Time, Tp

Figure 4. For description see opposite.

biennial cycle with additive noise? Could the contradiction be explained by differences in the population or spatial scales involved? For example, could individual cities in the U.K. be chaotic and nonlinear, but produce an emergent behaviour that appears linear when they are aggregated?

(c) Spatial dynamics of measles

The analysis of the spatial dynamics of measles has a distinguished pedigree in both biomathematics (Bartlett 1957) and spatial geography (Cliff & Haggett 1988). Before exploring the dynamics of measles in English cities, we clarify the dynamical implications of spatial heterogeneity with a simple model.

(i) *Scale dependence in models*

The sensitivity of ecological models to aggregation and scaling has been discussed in a number of different

contexts (see, for example, Cohen 1979; Livdahl & Sugihara 1983; Sugihara *et al*. 1984; Ives & May 1985; Allen & Starr 1985; O'Neill *et al*. 1986; Sugihara *et al*. 1989). Here we test the theoretical possibility of emergent linearity from nonlinear parts with the following simple experiment.

We approximate measles dynamics within a single city as a chaotic logistic map $(X_{t-1} = aX_t(1-X_t))$, superimposed on a sine wave. We then investigate what happens as more of these 'sine + logistic' series are summed. In summing the series, we require the linear part (the sine waves) to be synchronized as the seasonal patterns in cities would be, but we allow the nonlinear parts to be independent to approximate spatial decoupling (or weak coupling by contagion). In effect, this is equivalent to averaging the output from independent logistic maps, and superimposing this net output on a sine wave. Because the dynamics in each city may not be perfectly identical, the logistic maps
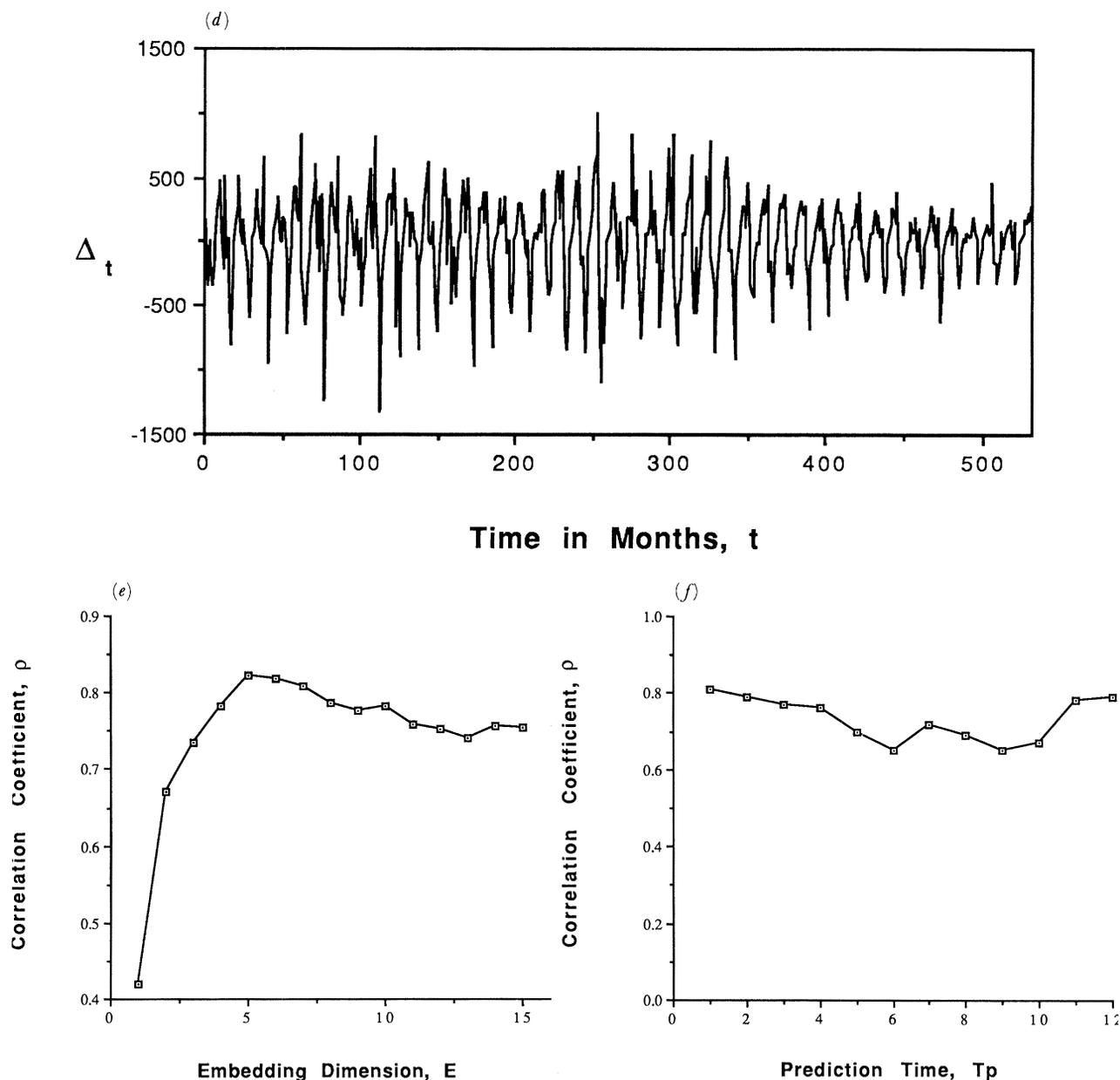
Figure 4. (a) Time series generated by taking first differences, $x_{t+1} - x_t$, of the monthly number of cases of measles reported in New York City between 1928 and 1972 (the first 532 points in the sequence shown here). After 1963, the introduction of immunization against measles had a qualitative effect on the dynamics of infection; this can be seen in the later part of the sequence shown here. (b) By using the methods described earlier, the first part of the measles time series (216 points, from 1928 to 1946) was used to construct a library, which was then used to predict forward from each point in the second part of the series (1946 to 1963). The correlation coefficient, ρ, between predicted and observed results is shown as a function of the embedding dimension, E, for predictions one time-step ahead, $T_p = 1$ (and $\tau = 1$). The figure suggests an optimal embedding dimension of $E \sim 5-7$. (c) Here ρ, between predicted and observed results for measles, is shown as a function of prediction interval T_p (for $E = 6$ and $\tau = 1$). The overall decline in prediction accuracy with increasing time into the future may be taken as indicative of chaotic dynamics, as distinct from uncorrelated noise. Figures 4d, e, f as for figures a, b, c, respectively, except now the data are for monthly case reports of chickenpox in New York City, from 1928 to 1972. Here, all 532 points are used in the analysis. Again figure 4e suggests an optimal embedding dimension $E \sim 5-7$. In marked contrast to figure c, f (calculated on the basis of $E = 5$ and $\tau = 1$) indicates pure additive noise, superimposed on a basic seasonal cycle. For a more detailed discussion of figure 4, see Sugihara & May (1990).

are given some variability by choosing the parameter a for each map uniformly in the interval (2.67, 3.67). However, similar results are obtained with independently initialised logistic maps having identical parameters.

Figure 6a shows how the $\rho-T_p$ signature varies with

increasing aggregation. As more independent logistic maps (cities) are folded into the picture the $\rho-T_p$ signature becomes ever more shallow, giving much the appearance of the linear noise case. This is corroborated by figure 6b where the linear predictor tends to match the nonlinear predictor more closely as more

[121]
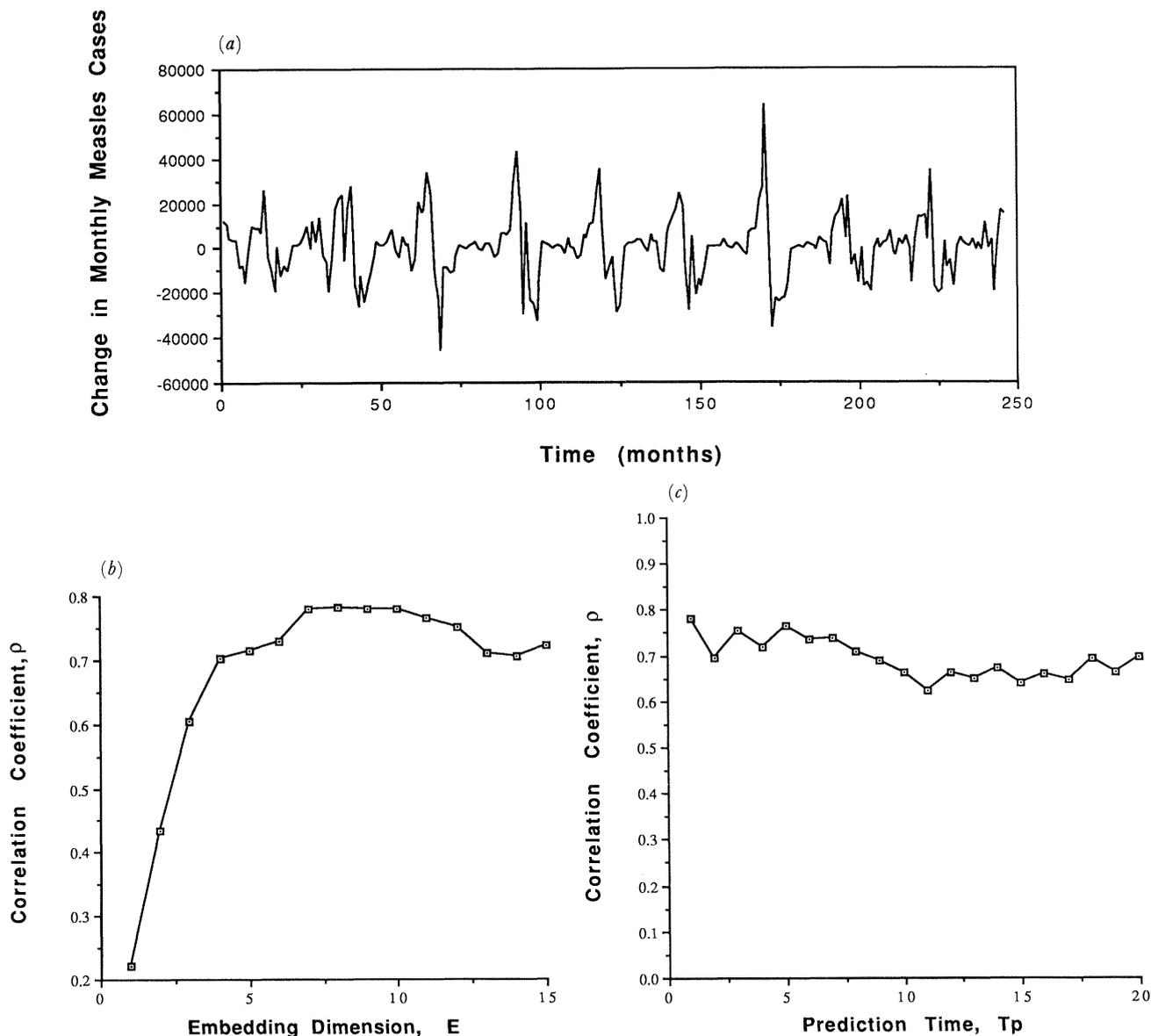
Figure 5 (a) Time-series of first differences in the number of cases of measles reported across England and Wales between 1948 and 1966. For data-source see Anderson *et al.* (1984). (b) Correlation coefficient or predictability, ρ, as a function of embedding dimension, E, for measles from England and Wales. Because of the low number of data points available ($N = 216$), we employed the whole series in forecasting in order to obtain the clearest estimate of optimal embedding (unbiased by the possibility of nonstationarity in the data). However, to ensure independence, the library used for each forecast was constructed to exclude points near the predictee in the time-series. A lag of $\tau = 2$ was used to embed these data. (c) Predictability, ρ, as a function of prediction time, T_p, for measles from England and Wales. Here the optimal parameters obtained above ($\tau = 2$ and $E = 8$) were used in forecasting; the library of patterns from the first half of the data was used to predict the second half of the data. The relatively flat pattern shown here is similar to the additive noise case seen for chickenpox in New York City. This figure shows that the large-scale aggregate behaviour of measles across England and Wales does not appear to be chaotic (in curious contrast to measles in New York City, figure 4a).

logistic series are summed. These trends are understandable in light of the following two facts. First, as more logistic maps are superimposed, the nonlinear signal becomes ever more complicated. High dimensional dynamics, chaotic or otherwise, are regarded as noise. Secondly, as more such series are superimposed, the amplitude of the nonlinear signal should decrease roughly as the square root of the number of independent chaotic logistic maps; this exposes more clearly the linear parts (seasonal sine wave) of the time series which are synchronized. Thus as more in-

dependent chaotic nonlinear series are aggregated, the nonlinear part should begin to resemble noise superimposed on a sine wave.

(ii) *Measles in English cities*

To test the applicability of these ideas to the observed patterns for measles in England and Wales, we have disaggregated the data, focusing on individual cities. The central question here is whether evidence for chaotic behaviour (which is not apparent in the countrywide analysis) emerges on a single-city scale.
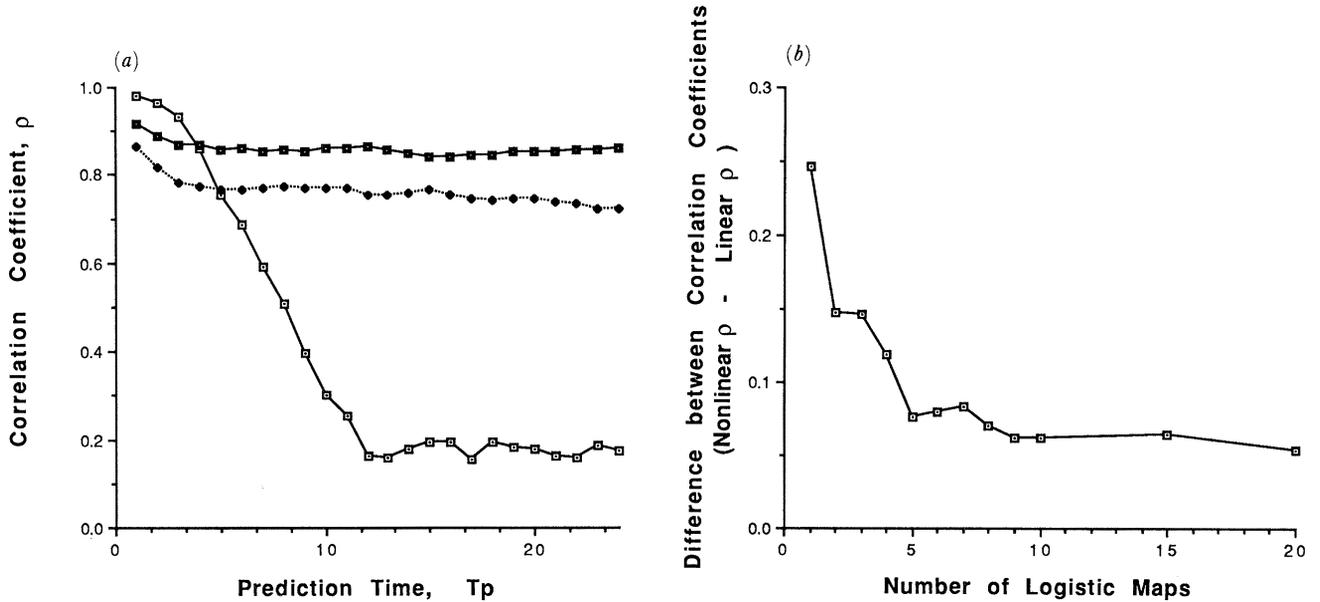
Figure 6. (*a*) Predictability, ρ, as a function of prediction time, T_p, at different levels of aggregation. Here we approximate measles dynamics in a single city as a chaotic logistic map superimposed on a sine curve, and investigate behaviour as more of these (sine + logistic) series are lumped (summed). The solid line with open boxes is for one (sine + logistic) ($\tau = 1$, $E = 3$), the dashed line with closed diamonds is for 10 (sine + logistic) series summed ($\tau = 1$, $E = 7$), and the solid line with solid boxes is for 20 series summed ($\tau = 1$, $E = 9$). The effect of such aggregation on the dynamics is to diminish the nonlinear chaotic portion of the signal, so that the $\rho-T_p$ signature looks increasingly like the additive noise case. (*b*), the difference in predictability, ρ, between optimal linear autoregressive methods versus our nonlinear methods is shown, as a function of the number of (sine + logistic) maps that are lumped together. The maps are as described in figure 6*a*, and here $T_p = 1$ (and $E = 3$, $\tau = 1$). Note that the difference in ρ decreases with increasing aggregation.

Table 1. *Demographic summary for the seven English cities used in the spatial analysis of measles*

distance matrix	distance (road miles)						
	London	Birmingham	Liverpool	Manchester	Sheffield	Bristol	Newcastle
London	—	113	205	189	159	115	281
Birmingham	—	—	94	81	76	88	205
Liverpool	—	—	—	35	74	164	156
Manchester	—	—	—	—	38	164	132
Sheffield	—	—	—	—	—	164	128
Bristol	—	—	—	—	—	—	293
population[a] (thousands)	8282	1096	792	693	514	435	294

[a] Estimated population in 1960; from OPCS (1960).

We shall focus here on a representative sample of seven large English cities: table 1 lists the cities, along with their population sizes and a distance matrix, while figure 7*a* shows the associated measles time-series for the period 1948–67.

Figure 7*b* shows the embedding analyses for each of the seven cities (again, using the full data set ($N = 240$) to compute these correlations). All of the five most populous cities, London, Birmingham, Liverpool, Manchester and Sheffield, had optimal embeddings in a range similar to what was observed for New York ($E = 5$–8; Manchester, however, also had a peak at $E = 4$), and each had a local maximum at $E = 7$. Indeed these results seem to match the embedding results for New York better than those for the pooled data for Britain. On the other hand, the two least populous and

most isolated cities appeared to require higher dimensional embeddings: for Bristol $E = 10$, and for Newcastle $E = 12$. Although one must be cautious not to overinterpret the specific figures obtained here, especially in light of the low number of data points involved ($N = 240$), it is interesting to note that both Bristol and Newcastle fall well below the population threshold believed necessary for the infection to remain endemic (Bartlett 1957). It is possible, therefore, that the higher dimensionality of the embedding here is because of the higher complexity coming from the required coupling to the outside world.

Figure 7*c* shows the $\rho-T_p$ curves for the seven cities. The results here are not flat like the ones obtained with the aggregated data, but rather have a look very similar to the chaotic signature observed in New York

Figure 7. (a) Time-series of first differences in the reported monthly cases of measles in seven major British cities between 1948 and 1966 (arranged in order of city population size).

where predictability falls off steeply with increasing prediction time. Moreover, the results of the comparisons with the optimal linear predictor for each city shown in table 2 (ρ_{linear} against $\rho_{\mathrm{nonlinear}}$ at $T_\mathrm{p} = 1$, with the resulting p-level) firmly support the view that the dynamics are chaotic. Thus, it appears that scale considerations may help to resolve the apparent contradiction between the lumped analysis for measles in England and Wales, and the earlier analysis for New York City measles. In this regard, it is interesting that London, the most populous and geographically the largest of the British cities by almost an order of magnitude, appears to show the most gradual decline in its $\rho-T_\mathrm{p}$ curve.

These results show that the nonlinear dynamical

features that are present in the individual cities of the U.K. are averaged out in the aggregate. Although a linear predictor worked well (at least as well as the nonlinear predictor) at forecasting changes in measles incidence country-wide, the greater success in predicting on a city-scale using a nonlinear predictor suggests that one might expect to produce better forecasts country-wide by combining the results of the component nonlinear predictors.

(d) *Stationarity*

Finally, an issue that seldom appears in simple models but that is important when analysing data from the natural world, concerns the stationarity of the
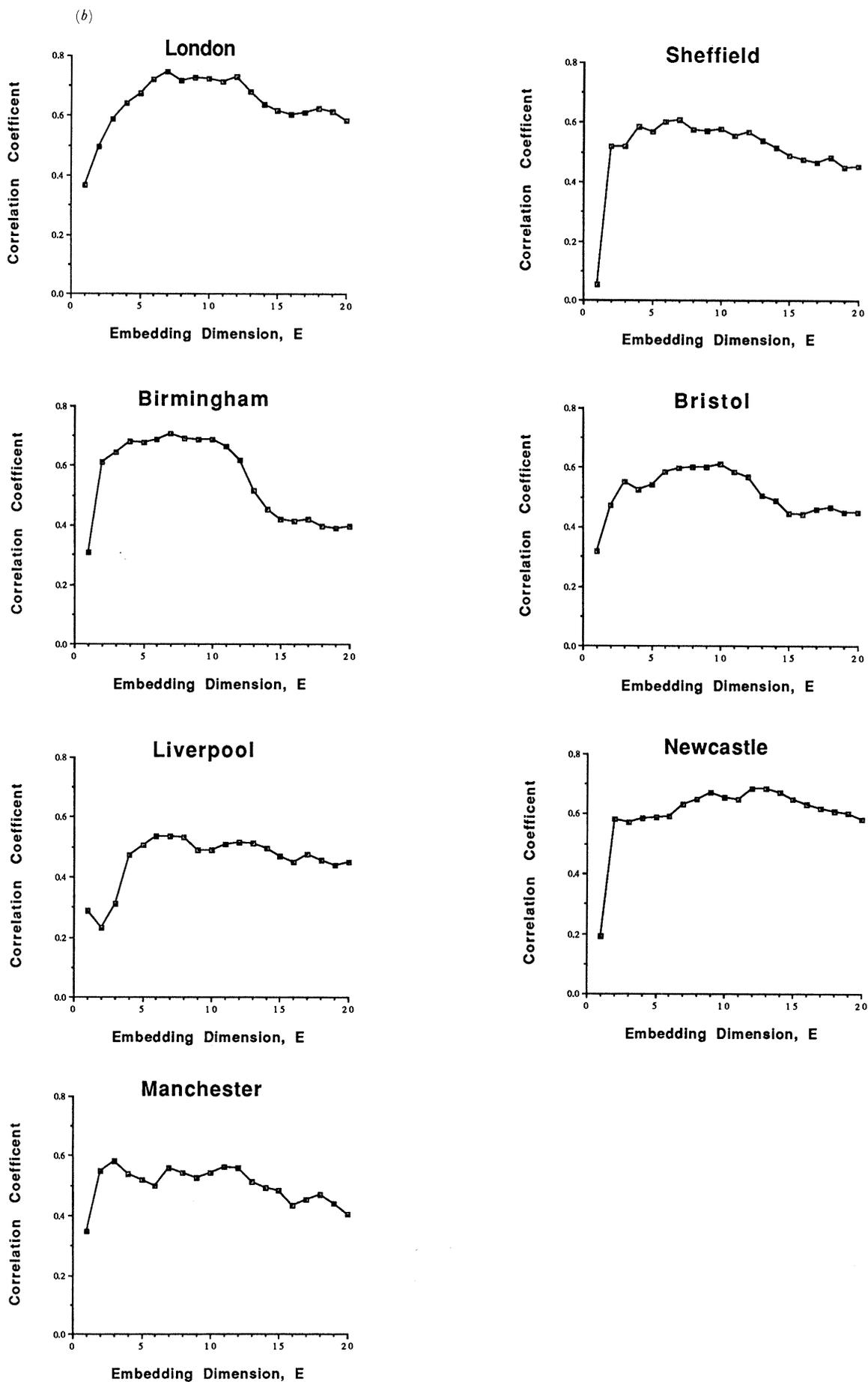
Figure 7. (*b*) Predictability, ρ, as a function of embedding dimension, E, for measles from seven major British cities. Clearest results were obtained by using $\tau = 1$ for Liverpool, Manchester, Sheffield and Newcastle, and $\tau = 2$ for London, Birmingham and Bristol.

[125]
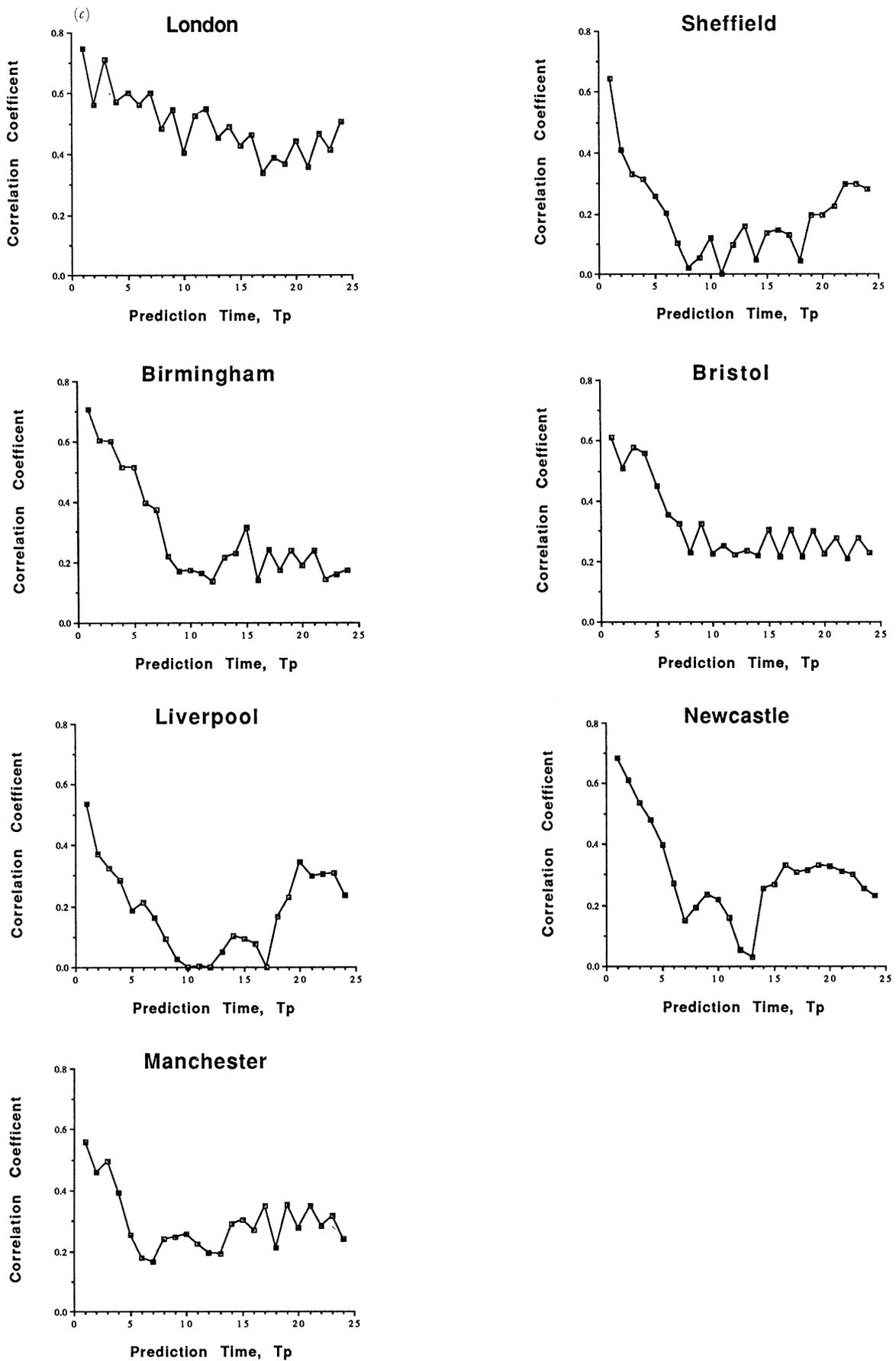
Figure 7.(c) Predictability, ρ, as a function of prediction interval, T_p, for measles from seven major British cities. The following parameters were used: London $\tau = 2$, $E = 7$; Birmingham $\tau = 2$, $E = 7$; Liverpool $\tau = 1$, $E = 7$; Manchester $\tau = 1$, $E = 7$, Sheffield $\tau = 1$, $E = 6$; Bristol $\tau = 2$, $E = 10$; Newcastle $\tau = 1$, $E = 12$. With the possible exception of London, all of the above cities show the characteristic decline in predictability with increasing prediction interval associated with chaotic dynamics, as seen in New York City (figure 4a).

Table 2. *Comparison between linear autoregressive methods and the nonlinear simplex predictor*

City	ρ_{linear}	$\rho_{nonlinear}$	Significance
London	0.63	0.76	$p < 0.001$
Birmingham	0.37	0.71	$p < 0.0005$
Liverpool	0.40	0.48	$p < 0.05$
Manchester	0.38	0.57	$p < 0.0005$
Sheffield	−0.02	0.64	$p < 0.0005$
Bristol	−0.01	0.67	$p \ll 0.0005$
Newcastle	−0.10	0.69	$p \ll 0.0005$

Table 3. *Test of measles data for nonlinear stationarity* ($N = 120$, $T_p = 1$)

	library half→predicted half			
City	$1 \to 1$	$1 \to 2$	$2 \to 1$	$2 \to 2$
London	0.65	0.72	0.73	0.46
Birmingham	0.67	0.71	0.72	0.68
Liverpool	0.51	0.49	0.17	0.53
Manchester	0.48	0.54	0.48	0.27
Sheffield	0.67	0.44	0.47	0.44
Bristol	0.43	0.65	0.70	0.51
Newcastle	0.55	0.69	0.60	0.60

Table 4. *Test of measles data for linear stationarity* ($N = 120$, $T_p = 1$)

	fitted half→predicted half			
City	$1 \to 1$	$1 \to 2$	$2 \to 1$	$2 \to 2$
London	0.64	0.64	0.53	0.64
Birmingham	0.60	0.38	0.45	0.76
Liverpool	0.56	0.43	0.24	0.50
Manchester	0.64	0.38	0.42	0.52
Sheffield	0.60	−0.02	0.28	0.70
Bristol	0.73	−0.07	0.43	0.80
Newcastle	0.60	−0.10	0.35	0.57

process generating the time-series. In all of the ρ–T_p analyses above, the first half of the series was used to construct a library of patterns that was then used to predict the second half of the time series. In the natural world, where parameters can undergo systematic changes over time, past patterns can be of dubious relevance to an altered present or an even more different future. An inspection of several of the time series in figure 7*a*, particularly those for Sheffield, Bristol and Newcastle, suggests that non-stationarity might indeed be a problem here.

One way to gauge whether secular trends might confound the forecasting results given above has been discussed by Sugihara & May (1990). Rather than using the first half of the time series to construct the library, and the second half to compute correlations between predictions and observations, we instead investigate what happens when the library and predicted halves are chosen in all combinations (table 3). That is, we use the first half to predict itself ($1 \to 1$) and then use it to predict the second half ($1 \to 2$), and the second half to predict itself ($2 \to 2$) and then use it to predict the first half ($2 \to 1$). We then compare the correlation coefficients obtained in each of these four cases. If the time-series shows a secular trend, we should find higher correlations when the library and predicted segments span the same time period, and lower ones when a library from one time span (e.g. first half) is used to forecast values from another time span (e.g. the second half of the series).

Table 3 shows the results of such an analysis for each of the English cities in figure 7. Although there is a certain amount of variation in the predictability of each of the reciprocal combinations, there is no systematic trend for higher correlations when the fitted half is used to predict on itself. This is most clearly evidenced in the summary statistics given at the bottom of table 3. Moreover, all reciprocal combinations gave similar ρ–T_p curves. Thus at least from a nonlinear perspective, these time series do not appear to contain secular trends.

On the other hand, when reciprocal pairings between fitted and predicted halves are made using the linear autoregressive approach, they can appear highly nonstationary (table 4). Notice that the linear autoregressive model keys on repeated patterns in one dimension (the time-series itself). Thus the time series for Sheffield, Bristol and Newcastle, for which there was almost no linear predictability between the first half and the second half, appear most clearly non-

stationary to the naked eye. None the less, when these time series are embedded in higher dimensions, the obvious secular trends disappear. The moral that emerges from this is that if a process is truly nonlinear, one needs to be careful in proclaiming nonstationarity based on linear criteria. A stationary process in higher dimensions may only appear to contain secular changes when viewed in one dimension.

5. DISCUSSION

Our preliminary analysis of the spatial dynamics of measles suggests two fruitful avenues for future work. First, the measles data for England and Wales are available on a much finer spatial scale than the crude city-by-city division examined here. In particular, a further subdivision of the London measles data would provide a much more refined test for the 'emergence' of chaos at smaller spatial scales. Secondly, we require more detailed mechanistic models, which allow explicitly for the impact of spatial heterogeneities in transmission on the dynamics of the host–parasite interaction (May 1986; May & Anderson 1984). As shown by the simple ('sine + logistic') spatial model considered above, the analysis of time series simulated from such models can provide important insights into the dynamics of the real system.

Two main points emerge from our paper, the first having to do with dynamical details and the second with general principles. First, growing understanding of deterministically chaotic systems suggests that apparently random time series may in fact be generated by deterministic mechanisms, and that techniques may be available to distinguish such low-

the only alternative to chaotic dynamics is measurement error. However, in nature stochastic factors affect population trajectories, not just their measurement. Thus one should expect populations to fluctuate (to have stochastic dynamics) as a consequence of these stochastic factors, quite apart from any apparent fluctuations that are introduced as an artefact of measurement error. It is quite likely that such stochastic factors will not be additive and will cause effects that are vastly different from the additive measurement error that Professor Sugihara considers.

His technique of distinguishing chaos from additive error will work when these two are the only alternatives, but will be incapable of distinguishing between stochastic dynamics and chaos. I accept Professor Sugihara's point that from some perspectives, high dimensional chaos and stochasticity are the same. It then appears that he intends his technique to distinguish between measurement error and dynamical uncertainty, whatever the cause of the latter, be it low-dimensional chaos or stochasticity.

H. M. PLATT (*The Natural History Museum, London*). As I understand it, Professor Sugihara became interested in the England and Wales measles data because, unlike those of New York, they seemed to be additive noise. However, when he backtracked to individual sets for cities he found the chaos patterns again. How does Professor Sugihara know that those data which he suggests display real additive noise patterns, such as the chicken pox set, are themselves not in fact assemblages of chaos patterns, which he may or may not be able to get at.

M. WILLIAMSON (*University of York, York, U.K.*). Is it not possible to get almost any shape of prediction curve by choosing various models for both dynamical chaos and for systems with measurement error?

G. SUGIHARA. As discussed more fully in the *Nature* paper (Sugihara & May 1990), it seems likely that a specific pattern of autocorrelated noise could be hand-tailored, to mimic any specified relation between ρ (correlation coefficient) and T_p prediction interval), such as that found for the chaotic test map. The converse is surely not true! Chaotic dynamical systems of low dimension will always show a systematic decline in with increasing T_p (at a characteristic prediction interval set by the Lyapunov exponent). We conjecture that, in general, such artificially designed patterns of auto-correlation will typically give flatter ρ–E (embedding dimension) relations than are found for simple time series generated by low-dimensional attractions (see Farmer & Sidorowich 1989; Sugihara & May 1990).

References

Farmer, J. D. & Sidorowich, J. J. 1989 Exploiting chaos to predict the future and reduce noise. In *Evolution, learning and cognition* (ed. Y. C. Lee), pp. 277–304. New York: World Scientific Press.

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

[129]

17-2