Reprinted from Nature, Vol. 344, No. 6268, pp. 734–741, 19th April, 1990

Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series

George Sugihara* **& Robert M. May†**

* Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093, USA
† Department of Zoology, Oxford University, Oxford, OX1 3PS, UK

An approach is presented for making short-term predictions about the trajectories of chaotic dynamical systems. The method is applied to data on measles, chickenpox, and marine phytoplankton populations, to show how apparent noise associated with deterministic chaos can be distinguished from sampling error and other sources of externally induced environmental noise.

Two sources of uncertainty in forecasting the motion of natural dynamical systems, such as the annual densities of plant or animal populations, are the errors and fluctuations associated with making measurements (for example, sampling errors in estimating sizes, or fluctuations associated with unpredictable environmental changes from year to year), and the complexity of the dynamics themselves (where deterministic dynamics can easily lead to chaotic trajectories).

Here we combine some new ideas with previously developed techniques[1-7,16,24-26], to make short-term predictions that are

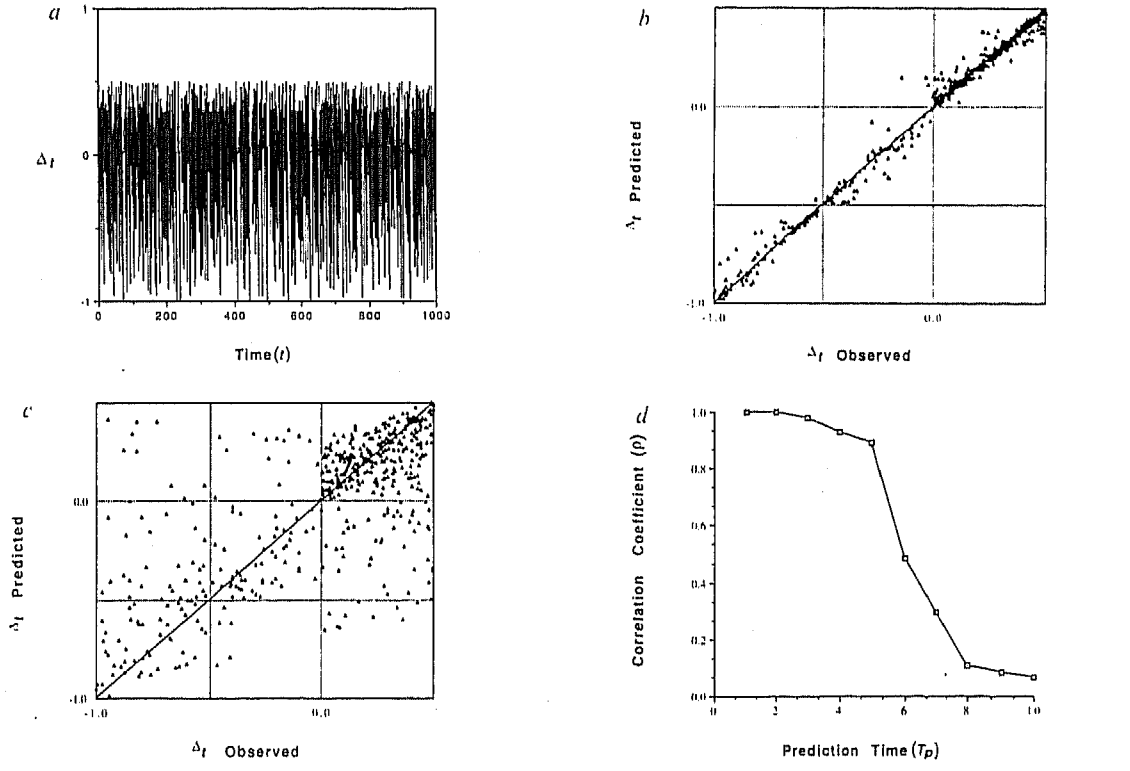

FIG. 1 *a*, Time series of 1,000 points (which in many ways is indistinguishable from white noise) generated by taking first-differences, $\Delta_t = x_{t+1} - x_t$, of the tent map: $x_{t+1} = 2x_t$ for $0.5 > x_t > 0$; $x_{t+1} = 2 - 2x_t$ for $1 > x_t > 0.5$. *b*, Predicted values two steps into the future ($T_p = 2$) versus observed values for the tent delta time series depicted in *a*. 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, and in this figure the embedding dimension and lag time are $E = 3$ 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 obtained using the first half of the series to predict the second half with an autoregressive linear model (where the predictions are based on the weighted average of three linear maps, one for each of the three different τ-values that give the best results in such a linear scheme) is $\rho = 0.04$. *c*, Exactly as for Fig. *b*, except here the predictions are five time steps into the future ($T_p = 5$). The correlation coefficient between predicted and actual values is now $\rho = 0.89$ ($N = 500$). *d*, Summary of the trend between *b* and *c*, by showing ρ between predicted and observed values in the second half (second 500 points) of the time series of *a*, as a function of T_p. As in *b* and *c*, the simplex projection method here uses $E = 3$ and $\tau = 1$. That prediction accuracy (as measured by the coefficient of correlation between predicted and observed values) falls as predictions extend further into the future is a characteristic signature of a chaotic attractor.

based on a library of past patterns in a time series[1]. By comparing the predicted and actual trajectories, we can make tentative distinctions between dynamical chaos and measurement error: for a chaotic time series the accuracy of the nonlinear forecast falls off with increasing prediction-time interval (at a rate which gives an estimate of the Lyapunov exponent[3]), whereas for uncorrelated noise, the forecasting accuracy is roughly independent of prediction interval. For a relatively short time series, distinguishing between autocorrelated noise and chaos is more difficult; we suggest a way of distinguishing such 'coloured' noise from chaos in our scheme, but questions remain, at least for time series of finite length.

The method also provides an estimate of the number of dimensions, or 'active variables', of the attractor underlying a time series that is identified as chaotic. Unlike many current approaches to this problem (for example, that of Grassberger and Procaccia[5]), our method does not require a large number of data points, but seems to be useful when the observed time series has relatively few points (as is the case in essentially all ecological and epidemiological data sets).

Forecasting for a chaotic time series

Below, we outline the method and show how it works by applying it to a chaotic time series generated artificially from the deterministic 'tent map'. We then apply it to actual data on measles and chickenpox in human populations (which have been previously analysed using different techniques[9-13]) and on diatom populations. We conclude that the method may be capable of distinguishing chaos from measurement error even in such relatively short runs of real data.

As an example of the difficulties in short-range forecasting, we consider the chaotic time series shown in Fig. 1*a*. This time series was generated from the first-difference transformation $(x_{t+1} - x_t)$ on the deterministic tent map or triangular 'return map' (described in detail in the legend to Fig. 1*a*). Here and elsewhere is this report, we first-difference the data partly to give greater density in phase space to such chaotic attractors as may exist, and partly to clarify nonlinearities by reducing the effects of any short-term linear autocorrelations. It should be noted, however, that both in our artificial examples and in our later analysis of real data, we obtain essentially the same results if we work with the raw time series (without first-differencing). With the exception of a slight negative correlation between immediately adjacent values, the sequence in Fig. 1*a* is uncorrelated, and is in many ways indistinguishable from white noise: the null hypothesis of a flat Fourier spectrum cannot be rejected using Bartlett's Kolmogorov-Smirnov test, with $P = 0.85$. Because nonadjacent values in the time series are completely uncorrelated, standard statistical methods (that is, linear autoregression) cannot be used to generate predictions two or more steps into the future that are significantly better than the mean value (that is, zero) for the series.

Figure 1*b* and *c* show the results of local forecasting with the above data. The basic idea here, as outlined below, is that if deterministic laws govern the system, then, even if the dynamical behaviour is chaotic, the future may to some extent be predicted from the behaviour of past values that are similar to those of the present.

Specifically, we first choose an 'embedding dimension', E, and then use lagged coordinates to represent each lagged sequence of data points $\{x_t, x_{t-\tau}, x_{t-2\tau}, \ldots, x_{t-(E-1)\tau}\}$ as a point in this E-dimensional space; for this example we choose $\tau = 1$, but the results do not seem to be very sensitive to the value of τ, provided that it is not large[14,15]. For our original time series, shown in Fig. 1*a*, each sequence for which we wish to make a prediction—each 'predictee'—is now to be regarded as an E-dimensional point, comprising the present value and the $E-1$ previous values each separated by one lag time τ. We now locate all nearby E-dimensional points in the state space, and choose a minimal neighbourhood defined to be such that the predictee is contained within the smallest simplex (the simplex with minimum diameter) formed from its $E+1$ closest neighbours; a simplex containing $E+1$ vertices (neighbours) is the smallest simplex that can contain an E-dimensional point as an interior point (for points on the boundary, we use a lower-dimensional simplex of nearest neighbours). The prediction is now obtained by projecting the domain of the simplex into its range, that is by keeping track of where the points in the simplex end up after p time steps. To obtain the predicted value, we compute where the original predictee has moved within the range of this simplex,

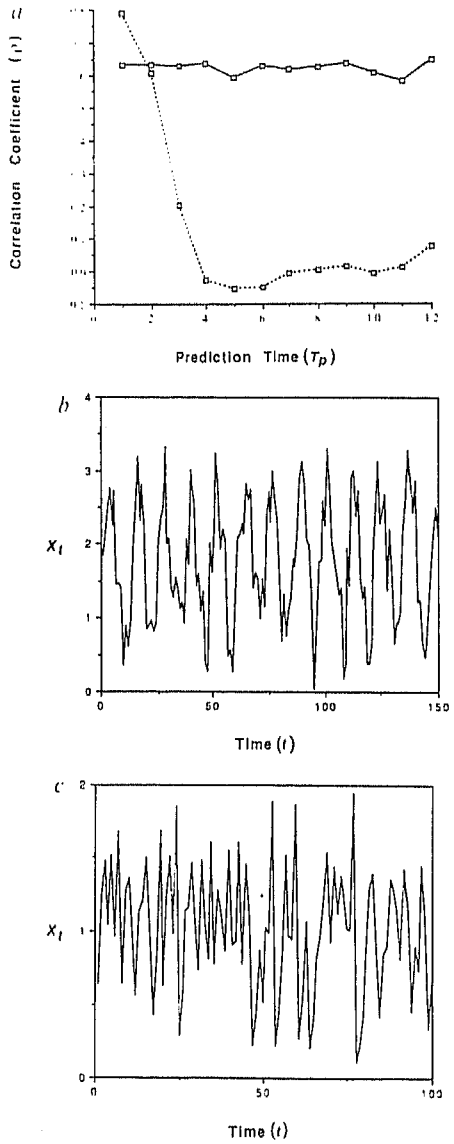

FIG. 2 *a*, Solid line shows ρ between predicted and observed values for the second half of the time series defined in *b* (which is, in fact, a sine wave with additive noise) as a function of T_p. As discussed in the text, the accuracy of the prediction, as measured by ρ, shows no systematic dependence on T_p. By contrast, the time series shown in *c* (which is the sum of two separate tent map series) does show the decrease in ρ with increasing T_p, as illustrated by the dashed line, that is characteristic of a chaotic sequence. Both curves are based on the simplex methods described in the text, with $E = 3$ and $\tau = 1$. *b*, First 150 points in the time series generated by taking discrete points on a sine wave with unit amplitude ($x_t = \sin(0.5t)$), and adding a random variable chosen (independently at each step) uniformly from the interval $[-0.5, 0.5]$. That is, the series is generated as a 'sine wave − 50% noise'. *c*, Time series illustrated here is generated by adding together two independent tent map sequences.

giving exponential weight to its original distances from the relevant neighbours. This is a nonparametric method, which uses no prior information about the model used to generate the time series, only the information in the output itself. It should apply to any stationary or quasi-ergodic dynamic process, including chaos. This method is a simpler variant of several more complicated techniques explored recently by Farmer and Sidorowich[3] and by Casdagli[7].

Figure 1*b* compares predicted with actual results, two time steps into the future. Figure 1*c* makes the same comparison, but at five time steps into the future. There is obviously more scatter in Fig. 1*c* than in Fig. 1*b*. Figure 1*d* quantifies how error increases as we predict further into the future in this example, by plotting the conventional statistical coefficient of correlation, ρ, between predicted and observed values as a function of the prediction-time interval, T_p (or the number of time steps into the future, p). Such decrease in the correlation coefficient with increasing prediction time is a characteristic feature of chaos (or equivalently, of the presence of a positive Lyapunov exponent, with the magnitude of the exponent related to the rate of decrease of ρ with T_p). This property is noteworthy, because it indicates a simple way to differentiate additive noise from deterministic chaos: predictions with additive noise that is uncorrelated (in the first-differences) will seem to have a fixed amount of error, regardless of how far, or close, into the future one tries to project, whereas predictions with deterministic chaos will tend to deteriorate as one tries to forecast further into the future. Farmer and Sidorowich[3,16] have derived asymptotic results (for very long time series, $N \gg 1$) that describe how this error typically propagates, over time, in simple chaotic systems. The standard correlation coefficient is one of several alternative measures of the agreement between predicted and observed values; results essentially identical to those recorded in Figs 1-6 can be obtained with other measures (such as the mean squared difference between predicted and observed values as a ratio to the mean squared error).

Forecasting with uncorrelated noise

Figure 2*a* (solid line) shows that, indeed, this signature of ρ decreasing with T_p does not arise when the erratic time series is in fact a noisy limit cycle. Here we have uncorrelated additive noise superimposed on a sine wave (Fig. 2*b*). Such uncorrelated noise is reckoned to be characteristic of sampling variation. Here the error remains constant as the simplex is projected further into the future; past sequences of roulette-wheel numbers that are similar to present ones tell as much or little about the next spin as the next hundredth spin. By contrast, the dashed line in Fig. 2*a* represents ρ as a function of T_p, for a chaotic sequence generated as the sum of two independent runs of tent map; that is, for the time series illustrated in Fig. 2*c*. Although the two time series in Fig. 2*b* and Fig. 2*c* can both look like the sample functions of some random process, the characteristic signatures in Fig. 2*a* distinguish the deterministic chaos in Fig. 2*c* from the additive noise in Fig. 2*b*.

The embedding dimension

The predictions in Figs 1 and 2 are based on an embedding dimension of $E=3$. The results are, however, sensitive to the choice of E. Figure 3*a* compares predicted and actual results for the tent map two time steps ahead ($T_p=2$), as in Fig. 1*b*, except that now $E=10$ (versus $E=3$ in Fig. 1*b*). Clearly the predictions are less accurate with this higher embedding dimension. More generally, Fig. 3*b* shows ρ between predicted and actual results one time step into the future ($T_p=1$) as a function of E, for two different choices of the lag time ($\tau=1$ and $\tau=2$). It may seem surprising that having potentially more information—more data summarized in each E-dimensional point, and a higher-dimensional simplex of neighbours of the predictee—reduces the accuracy of the predictions; in this respect, these results differ from results reported by Farmer and Sidorowich for parametric forecasting involving linear interpolation to construct local polynomial maps[3]. We think this effect is 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, our method may have additional applications as a trial-and-error method of computing an upper-bound on the embedding dimension, and thence on the dimensionality of the attractor (see also refs 2, 6, 7).

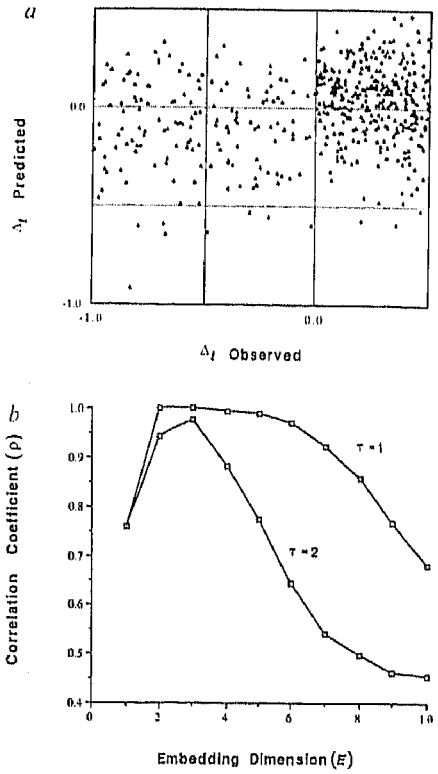

FIG. 3 *a*, Similar to Fig. 1*b*, this figure shows predictions one time step into the future ($T_p=1$) versus observed values, for the second 500 points in the tent map time series of Fig 1*a*, with the difference that here we used an embedding dimension $E=10$ (in contrast to $E=3$ in Fig. 1*b*; the lag time remains unchanged at $\tau=1$). As discussed in the text, the accuracy of the prediction deteriorates as E gets too large $\rho=0.25$, $N=500$). *b*, Correlation coefficient between predicted and observed results, ρ, is shown as a function of E for predictions one time step into the future ($T_p=1$). The relationship is shown for $\tau=1$ and $\tau=2$. The figure indicates how such empirical studies of the relation between ρ and E may be used to assess the optimal E.

Problems and other approaches

We have applied these ideas to a variety of other 'toy models', including the quadratic map along with other first-order difference equations and time series obtained by taking points at discrete time intervals from continuous chaotic systems such as those of the Lorenz and Rossler models (in which the chaotic orbits are generated by three coupled, nonlinear differential equations). The results for ρ as a function of T_p are in all cases very similar to those shown in Fig. 1*d*. Even in more complicated cases, such as those involving the superposition of different chaotic maps, we observe a decline in ρ versus T_p; here, however, the signature can show a step pattern, with each step corresponding to the dominant Lyapunov exponent for each map.

So far, we have compared relationships between ρ and T_p for chaotic time series with the corresponding relations for white noise. More problematic, however, is the comparison with ρ–T_p relationships generated by coloured noise spectra, in which there are significant short-term autocorrelations, although not long-

term ones. Such autocorrelated noise can clearly lead to correlations, ρ, between predicted and observed values that decrease as T_p lengthens. Indeed, it seems likely that a specific pattern of autocorrelations could be hand-tailored, to mimic any given relationship between ρ and T_p (such as that shown in Fig. 1*d*) obtained from a finite time series. We conjecture, however, that such an artifically designed pattern of autocorrelation would in general give a flatter ρ-versus-E relationship than those of simple chaotic time series corresponding to low-dimensional attractors (for example, see Fig. 3*b*). Working from the scaling relations for error versus T_p in chaotic systems[3,16], Farmer (personal communication) has indeed suggested that asymptotically (for very large N), the ρ-T_p relationships generated by autocorrelated noise may characteristically scale differently from those generated by deterministic chaos. Although we have no solution to this central problem—which ultimately may not have any general solution, at least for time series of the sizes found in population biology—we suggest that an observed time series may tentatively be regarded as deterministically chaotic if, in addition to a decaying ρ-T_p signature, the correlation, ρ, between predicted and observed values obtained by our methods is significantly better than the corresponding correlation coefficient obtained by the best-fitting autoregressive linear predictor (see also, ref. 16). For the tent map, as detailed in the legend to Fig. 1, *b* and *c*, our nonlinear method gives ρ values significantly better than those from autoregressive linear models (composed of the weighted average of the three best linear maps).

Most previous work applying nonlinear theory to experimental data begins with some estimate of the dimension of the underlying attractor[2-7]. The usual procedure (for exceptions, see refs 2, 6, 7, 25) is to construct a state-space embedding for the time series, and then to calculate the dimension of the putative attractor using some variant of the Grassberger-Procaccia algorithm[8]. A correlation integral is calculated that is essentially the number of points in E space separated by a distance less than l, and the power-law behaviour of this correlation integral (l^{ν}) is then used to estimate the dimension, D, of the attractor ($D \geq \nu$). This dimension is presumed to give a measure of the effective number of degrees of freedom or 'active modes' of the system. An upper bound on a minimal embedding dimension (which can be exceeded when the axes of the embed-

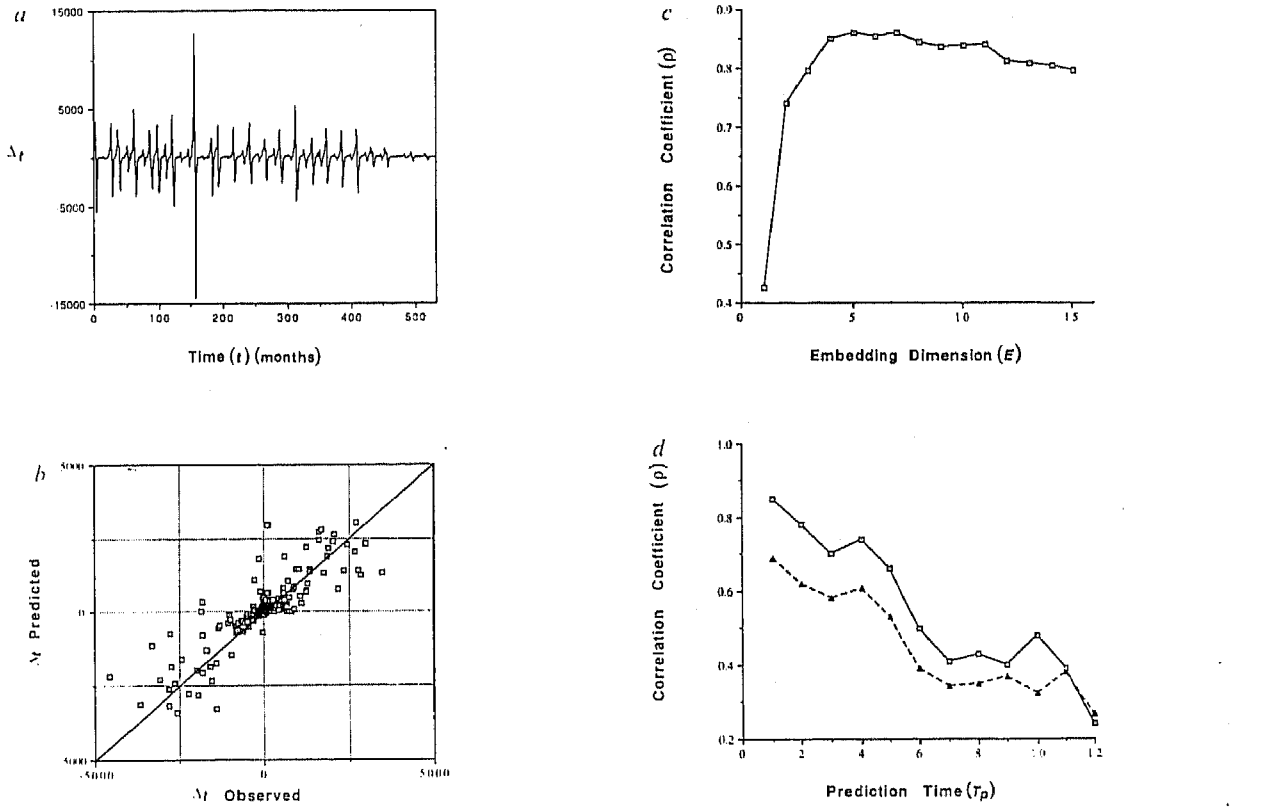

FIG. 4 *a*, Time series generated by taking first differences, $x_{t+1} - x_t$, of the monthly number of cases of measles reported in New York City between 1928 and 1972 (the first 532 points in the sequence shown here). After 1963, the introduction of immunization against measles had a qualitative effect on the dynamics of infection; this can be seen in the later part of the sequence illustrated here. *b*, Using the methods described earlier, the first part of the series in Fig. 4*a* (216 points, 1928 to 1946)) was used to construct a library of past patterns, which were then used as a basis for predicting forward from each point in the second part of the series, from 1946 to 1963. Predicted and observed values are shown here for predictions one time step into the future ($T_p = 1$), using $E = 6$ and $\tau = 1$. The correlation coefficient between predicted and observed values is $\rho = 0.85$ ($P < 10^{-5}$ for $N = 216$). For comparison, the corresponding prediction based on an autoregressive linear model (composed of five optimal linear maps, compare Fig. 1*b*) gives $\rho = 0.72$ (which is significantly different from $\rho = 0.85$ at the $P < 0.0005$ level). *c*, As in Fig. 3*b*, ρ between predicted and observed results, is shown as a function of E (for $T_p = 1$ and $\tau = 1$). This figure indicates an optimal embedding dimension of $E \sim 5$–7, corresponding to a chaotic attractor with dimension 2–3. *d*, Here ρ, between predicted and observed results for measles, is shown as a function of T_p (for $E = 6$ and $\tau = 1$). For the points connected by the solid lines, the predictions are for the second half of the time series (based on a library of patterns compiled from the first half). For the points connected by the dashed lines, the forecasts and the library of patterns span the same time period (the first half of the data). The similarity between solid and dashed curves indicates that secular trends in underlying parameters do not introduce significant complications here. The overall decline in prediction accuracy with increasing time into the future is, as discussed in the text, a signature of chaotic dynamics as distinct from uncorrelated additive noise.

ding are not truly orthogonal) is $E_{min} < 2D+1$, where D is the attractor dimension[8,15]. The scaling regions used to estimate power laws by these methods are typically small and, as a consequence, such calculations of dimension involve only a small fraction of the points in the series (that is, they involve only a small subset of pairs of points in the state space). In other words, the standard methods discard much of the information in a time series, which, because many natural time series are of limited size, can be a serious problem. Furthermore, the Grassberger–Procaccia and related methods are somewhat more qualitative, requiring subjective judgement about whether there is an attractor of given dimensions. Prediction methods, by contrast, have the advantage that standard statistical criteria can be used to evaluate the significance of the correlation between predicted and observed values. As Farmer and Sidorowich[3,16], and Casdagli[7], have also emphasized, prediction methods should provide a more stringent test of underlying determinism in situations of given complexity. Prediction is, after all, the *sine qua non* of determinism.

Time series from the natural world

Measles. For reported cases of measles in New York City, there is a monthly time series extending from 1928 (ref. 17). After 1963, immunization began to alter the intrinsic dynamics of this system, and so we use only the data from 1928 to 1963 ($N = 432$). These particular data have received a lot of attention recently, and they are the focus of a controversy about whether the dynamics reflect a noisy limit cycle[9] or low-dimensional chaos superimposed on a seasonal cycle[10-13]. In particular, the data have been carefully studied by Schaffer and others[13-16,27], who have tested for low-dimensional chaos using a variety of methods, including the Grassberger–Procaccia algorithm[13], estimation of Lyapunov exponents[13], reconstruction of Poincare return maps[10,11], and model simulations[12,13]. Although it is not claimed that any of these tests are individually conclusive, together they support the hypothesis that the measles data are described by a two- to three-dimensional chaotic attractor.

Figure 4*a* shows the time series obtained by taking first differences, $X_{t+1} - X_t$, of these data. As discussed above, the first difference was taken to 'whiten' the series (that is, reduce autocorrelation) and to diminish any signals associated with simple cycles (a possibility raised by proponents of the additive noise hypothesis[9]). We then generated our predictions by using the first half of the series (216 points) to construct an ensemble of points in an E-dimensional state space, that is, to construct a library of past patterns. The resulting information was then used to predict the remaining 216 values in the series, along the lines described above, for each chosen value of E. Figure 4*b*, for example, compares predicted and observed results, one time step into the future ($T_p = 1$ month), with $E = 6$. Figure 4*c* shows ρ between predicted and observed results as a function of E for $T_p = 1$. Taking the optimal embedding dimension to be that yielding the highest correlation coefficient (or least error) between prediction and observation in one time step, it is seen from Fig. 4*c* that $E \simeq 5$–7. This accords with previous estimates[10-13] made using various other methods, and is consistent with the finding of an attractor with dimension $D = 2$–3.

The points joined by the solid lines in Fig. 4*d* show ρ as a function of T_p (for $E = 6$). Prediction error seems to propagate in a manner consistent with chaotic dynamics. This result, in

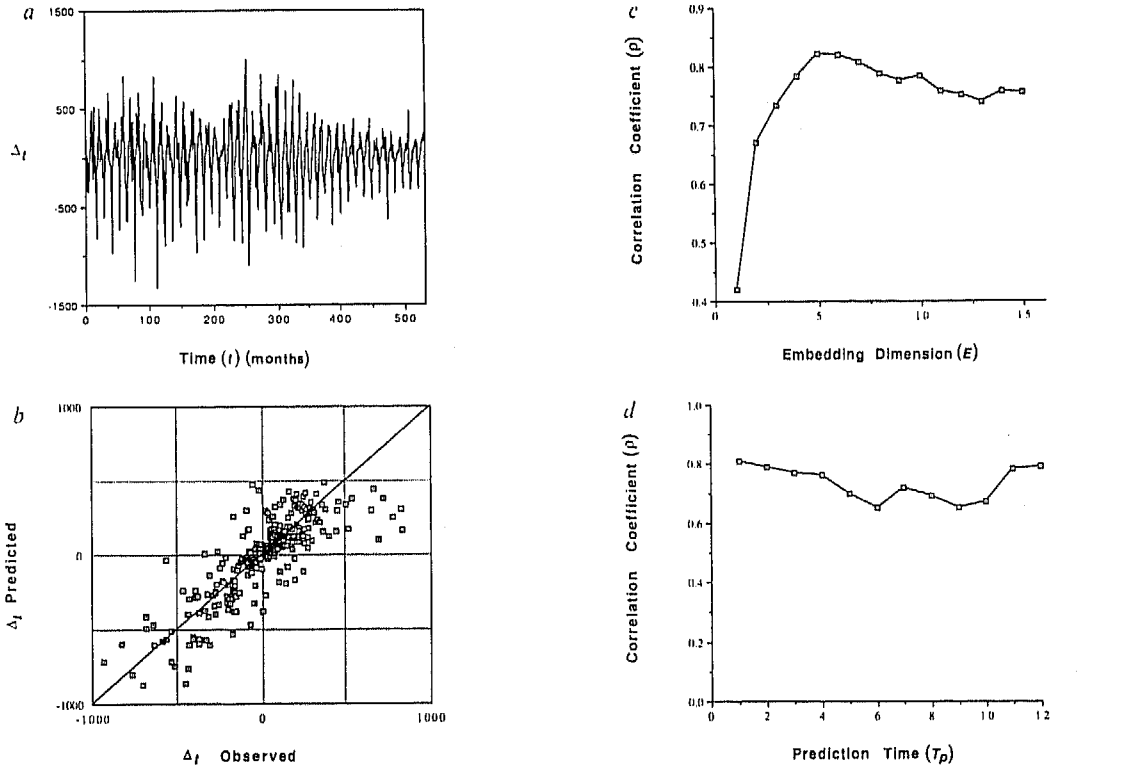

FIG. 5 *a*, As for Fig. 4*a*, except the time series comes from taking first-differences of the monthly numbers of reported cases of chickenpox in New York City from 1928 to 1972. *b*, As in Fig. 4*b*, predicted and observed numbers of cases of chickenpox are compared, the predictions being one time step into the future, $T_p = 1$ (here, $E = 5$ and $\tau = 1$). The correlation coefficient between predicted and observed values is $\rho = 0.82$; an autoregressive linear model alternatively gives predictions which have $\rho = 0.84$. In contrast to Fig. 4*b* for measles, here there is no significant difference between our prediction technique and standard linear autoregressive methods. *c*, Correlation coefficient between predicted and observed results for chickenpox, ρ, shown as a function of E for predictions one time step into the future ($T_p = 1$ and $\tau = 1$). *d*, Compare with Fig. 4*d*; ρ, between predicted and observed values, as a function of T_p (with $E = 5$ and $\tau = 1$) is shown. Here the lack of dependence of ρ on T_p, which is in marked contrast with the pattern for measles in Fig. 4*d*, indicates pure additive noise (superimposed on a basic seasonal cycle).

combination with the significantly better performance ($P < 0.0005$) of our nonlinear predictor as compared with an optimal linear autoregressive model (see legend to Fig. 4*b*) agrees with the conclusion that the noisy dynamics shown in Fig. 4*a* are, in fact, deterministic chaos[10-14].

For data from the natural world, as distinct from artificial models, physical or biological parameters, or both, can undergo systematic changes over time. In this event, libraries of past patterns can be of dubious relevance to an altered present and even-more-different future. In a different context, there is the example of how secular trends in environmental variables can complicate an analysis of patterns of fluctuation in the abundance of bird species[18,19]. We can gauge the extent to which secular trends might confound our forecasting methods in the following way. Rather than using the first half of the time series to compile the library of patterns, and the second half to compute correlations between predictions and observations, we instead investigate the case in which the library and forecasts span the same time period. Therefore we focus our predictions in the first half of the series, from which the library was drawn. To avoid redundancy, however, between our forecasts and the model, we sequentially exclude points from the library that are in the neighbourhood of each predictee (specifically, the $E\tau$ points preceding and following each forecast). The points connected by the dashed lines in Fig. 4*d* show the ρ versus T_p relationship that results from treating the measles data in this way (again with $E = 6$). The fairly close agreement between these results (for which the library of patterns and the forecasts span the same time period) and those of the simpler previous analysis (the solid line in Fig. 4*d*) indicates that within these time frames, secular trends in underlying parameters are not qualitatively important.

Chickenpox. Figure 5*a–d* repeat the process just described for measles, but now for monthly records of cases of chickenpox in New York City from 1949 to 1972 (ref. 20). Figure 5*a* shows the time series of differences, $X_{t+1} - X_t$. The 532 points in Fig. 5*a* are divided into two halves, with the first half used to construct the library, on which predictions are made for the second 266 points. These predictions are compared with the actual data points, as shown for predictions one time step ahead ($T_p = 1$ month) in Fig. 5*b*. In Fig. 5*b*, $E = 5$; Fig. 5*c* shows that an optimum value of E, in the sense just defined, is about 5 to 6. By contrast with Fig. 4*d* for measles, ρ between predicted and observed results for chickenpox shows no dependence on T_p: one does as well at predicting the incidence next year as next month. Moreover, the optimal linear autoregressive model performs as well as our nonlinear predictor. We take this to indicate that chickenpox has a strong annual cycle (as does

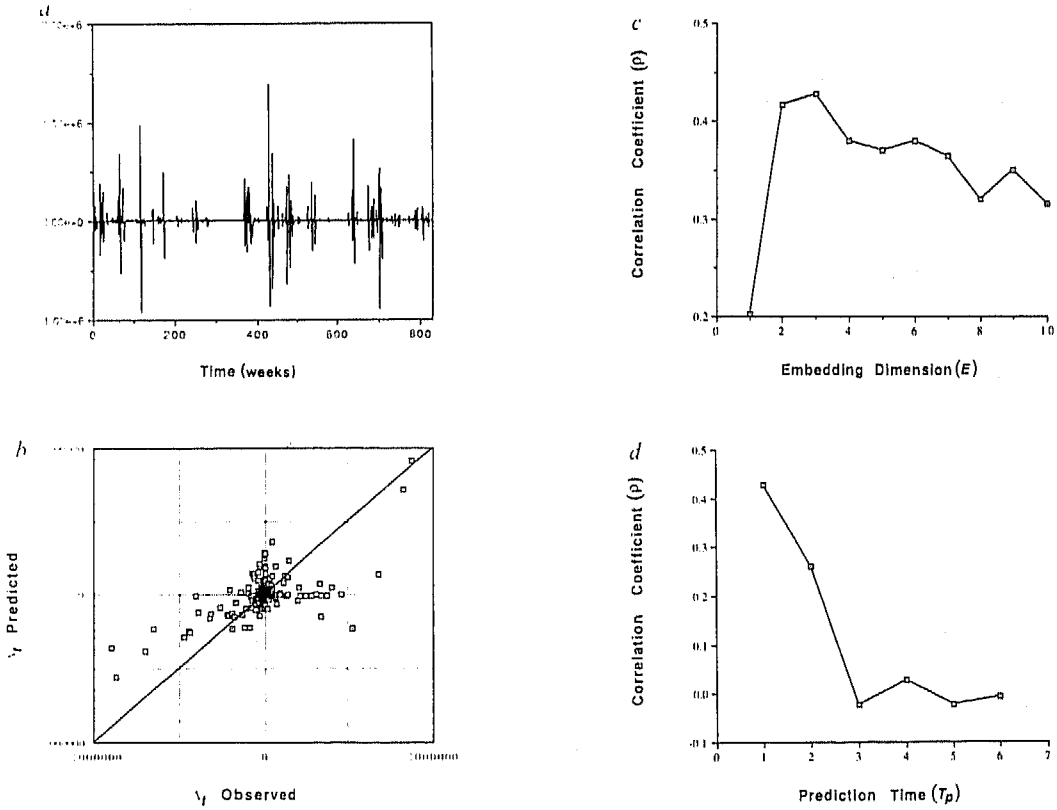

FIG. 6 *a*, Time series of first differences, $x_{t+1} - x_t$, of the weekly numbers of diatoms in seawater samples taken at Scripps Pier, San Diego, from 1929 to 1939 ($N = 830$). *b*, Using the first half of the time series in *a* to construct a library of patterns, we use the simplex projection methods described in the text to predict one week ($T_p = 1$) into the future from each point in the second half of the series ($N = 415$); here $E = 3$, and $\tau = 1$. The correlation coefficient between predicted and observed values is $\rho = 0.42$ ($P < 10^{-4}$ for $N = 415$); the best autoregressive linear predictions (composed of three optimal linear maps) give $\rho = 0.13$, which is significantly less than the nonlinear result ($P < 0.0005$). *c*, As in Figs 4*c* and 5*c*, ρ between predicted and observed values is shown as a function of the choice of E for predictions two time steps into the future ($T_p = 2$ and $\tau = 1$). This figure indicates an optimal E of about 3, consistent with an attractor of dimension about 2. *d*, As in Figs 4*d* and 5*d*, ρ is shown as a function of the T_p (for $E = 3$ and $\tau = 1$). Here the correlation coefficient decreases with increasing prediction interval, in the manner characteristic of chaotic dynamics generated by a some low-dimensional attractor. That ρ is about 50% at best, however, indicates that roughly half the variance in the time series comes from additive noise. The dynamics of this system therefore seem to be intermediate between those of measles (for which Fig. 4*d* indicates deterministically chaotic dynamics) and chickenpox (for which Fig. 5*d* indicates purely additive noise superimposed on a seasonal cycle).

measles), with the fluctuations being additive noise (in contrast to measles, for which the fluctuations derive mainly from the dynamics).

The contrast between measles and chickenpox can be explained on biological grounds[21]. Measles has a fairly high 'basic reproductive rate' ($R_0 = 10$–20), and, after a brief interval of infectiousness, recovered individuals are immune and uninfectious for life; these conditions tend to produce long-lasting 'interepidemic' oscillations, with a period of about 2 years, even in the simplest models[22]. This, in combination with seasonal patterns, makes it plausible that measles has complex dynamics. Chickenpox is less 'highly reproductive' (with R_0 values of about 8 to 10), and may recrudesce as shingles in later life; this makes for an infection less prone to show periodicities other than basic seasonal ones associated with schools opening and closing, and therefore indicates seasonal cycles with additive noise. Furthermore, reporting was compulsory for measles but not for chickenpox over the time period in question, which itself would be likely to make sampling error greater for chickenpox. Whatever the underlying biological explanation, the patterns in Figs 4*d* and 5*d* differ in much the same way as those illustrated in Fig. 2*a* for the artificially generated time series of Fig. 2, *a* and *b*.

Marine plankton. A time series is provided by Allen's weekly record of marine planktonic diatoms gathered at Scripps Pier, San Diego, between 1920 and 1939 ($N = 830$). With the exception of the work of Tont[23], this collection of information has been little analysed, and not at all in the light of contemporary notions about nonlinear dynamics. The data comprise weekly totals of the numbers of individuals of all diatom species, tallied in daily seawater samples collected over ~20 years. As for our analysis of the measles and chickenpox data above, we do not 'smooth' the diatom data in any of the usual ways, although we take first-differences for reasons stated earlier. The resulting time series is shown in Fig. 6*a*.

The results of using the first half of the diatom series to predict the second half are shown in the usual way in Fig. 6*b*. Figure 6*c* shows ρ between predicted and observed results looking one time step ahead ($T_p = 1$), as a function of E. The optimum embedding dimension seems to be about 3. This value for E is consistent with our independent analysis of the data using the Grassberger-Procaccia algorithm, which indicates that $D \simeq 2$. Figure 6*d* shows ρ as a function of T_p (for $E = 3$). The consistent decay in predictive power as one extrapolates further into the future is consistent with the dynamics of the diatom population being partly governed by a chaotic attractor. This view is supported by the significantly better fit of the nonlinear predictor as compared with the optimal linear autoregressive model ($P < 0.0005$). We note, however, that deterministic chaos at best accounts for about 50% of the variance, with the rest presumably deriving from additive noise; the relatively low dimension of the attractor for diatoms compared with measles makes it plausible that the noisier fit of predicted weekly fluctuations in diatoms, versus the predicted monthly fluctuations in measles, reflects a much higher sampling variance for diatoms than for reported measles cases.

Conclusion

The forecasting technique discussed here is phenomenological in that it attempts to assess the qualitative character of a system's dynamics—and to make short-range predictions based on that understanding—without attempting to provide an understanding of the biological or physical mechanisms that ultimately govern the behaviour of the system. This often contrasts strongly with the laboratory and field-experiment approaches that are used to elucidate detailed mechanisms by, for example, many population biologists. The approach outlined here splits the time series into two parts, and makes inferences about the dynamical nature of the system by examining the way in which ρ (the correlation coefficient between predicted and observed results for the second part of the series) varies with prediction interval, T_p, and embedding dimension, E; given the low densities of most time series in population biology, we share Ruelle's[28] lack of confidence in a direct assessment of the dimension of any putative attractor by Grassberger–Procaccia or other algorithms. Our approach works with artificially generated time series (for which we know the actual dynamics, and the underlying mechanisms, by definition), and it seems to give sensible answers with the observed time series for measles, chickenpox and diatoms (deterministic chaos in one case, seasonal cycles with additive noise in another, and a mixture of chaos and additive noise in the third). We hope to see the approach applied to other examples of noisy time series in population biology, and in other disciplines in which time series are typically sparse. □

Received 12 July 1989; accepted 19 February 1990.

1. Lorenz, E. N. *J. atmos. Sci.* **26**, 636–646 (1969).
2. Tong, H. & Lim, K. S. *Jl. R. statist. Soc.* **42**, 245–292 (1980).
3. Farmer, J. D. & Sidorowich, J. J. *Phys. Rev. Lett.* **62**, 845–848 (1987).
4. Priestly, M. B. *J. Time Series Analysis* **1**, 47–71 (1980).
5. Eckman, J. P. & Ruell, D. *Rev. mod. Phys.* **57**, 617–619 (1985).
6. Crutchfield, J. P. & MacNamara, B. S. *Complex Systems* **1**, 417–452 (1987).
7. Casdagli, M. *Physica D*, **35**, 335–356 (1989).
8. Grassberger, P. & Procaccia, I. *Phys. Rev. Lett.* **50**, 346–369 (1983).
9. Schwartz, I. *J. math. Biol.* **21**, 347–361 (1985).
10. Schaffer, W. M. & Kot, M. *J. theor. Biol.* **112**, 403–407 (1985).
11. Schaffer, W. M. & Kot, M. in *Chaos: An Introduction* (ed. Holden, A. V.) (Princeton Univ. Press, 1986).
12. Schaffer, W. M., Ellner, S. & Kot, M. *J. math. Biol.* **24**, 479–523 (1986).
13. Schaffer, W. M., Olsen, L. F., Truty, G. L., Fulmer, S. L. & Graser, D. J. in *From Chemical to Biological Organization* (eds Markus, M., Muller, S. C. & Nicolis, G.) (Springer-Verlag, New York, 1988).
14. Yule, G. U. *Phil. Trans. R. Soc.* **A226**, 267–278 (1927).
15. Takens, F. in *Dynamical Systems and Turbulence* (Springer-Verlag, Berlin, 1981).
16. Farmer, J. D. & Sidorowich, J. J. in *Evolution, Learning and Cognition* (ed. Lee, Y. C.) (World Scientific, New York, 1989).
17. London, W. P. & Yorke, J. A. *Am. J. Epidem.* **98**, 453 (1973).
18. Pimm, S. L. & Redfearn, A. *Nature* **334**, 613–614 (1988).
19. Lawton, J. H. *Nature* **334**, 563 (1988).
20. Helsenstein, U. *Statist. Med.* **5**, 37–47 (1986).
21. Anderson, R. M. & May, R. M. *Nature* **318**, 323–329 (1985).
22. Anderson, R. M., Grenfell, B. T. & May, R. M. *J. Hyg.* **93**, 587–608 (1984).
23. Tont, S. A. *J. mar. Res.* **39**, 191–201 (1981).
24. Varosi, F., Grebogi, C. & Yorke, J. A. *Phys. Lett.* **A124**, 59–64 (1987).
25. Abarbanel, H. D., Kadtke, J. B. & Brown, R. *Phys. Rev.* **B41**, 1782–1807 (1990).
26. Mees, A. I. Research Report No. 8 (Dept Mathematics, University of Western Australia, 1989).
27. Drepper, F. R. in *Erodynamics* (eds Wolff, W., Soeder, C. J. & Drepper, F. R.) (Springer-Verlag, New York, 1988).
28. Ruelle, D. *Proc. R. Soc. A* (in the press).

ACKNOWLEDGEMENTS. We thank Henry Abarbanel, Martin Casdagli, Sir David Cox, Doyne Farmer, Arnold Mandell, John Sidorowich and Bill Schaffer for helpful comments, John McGowan and Sargun Tont for directing us to Allen's phytoplankton records, Alan Trombla for computing assistance, and the United States NSF and the Royal Society for supporting this research.