

Noise and Nonlinearity in Epidemics:  
combining statistical and mechanistic modeling to  
characterize and forecast population dynamics.

S. Ellner<sup>1</sup>, B. Bailey<sup>1</sup>,  
G. Bobashev<sup>1</sup>, A. R. Gallant<sup>2</sup>,  
B. Grenfell<sup>3</sup>, D.W. Nychka<sup>1</sup>

- (1) Biomathematics Program, Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, USA
- (2) Department of Economics, University of North Carolina, Chapel Hill NC, USA.
- (3) Department of Zoology, Cambridge University, Cambridge, UK.

DRAFT March 25, 1996

**Abstract.** Childhood disease epidemics in large cities, notably measles, have been proposed as a well-supported example of deterministic chaos underlying complex population dynamics, but this remains controversial. Methods based on nonlinear time series modeling identify these epidemics as nonlinear with substantial random noise, clustering near the transition between stability and chaos. This conclusion has been challenged on the ground that the time series models have lower forecasting accuracy than mechanistic models with chaotic dynamics. However, the time series models in this comparison were all linear. Here we broaden the comparison to include nonlinear time series models for the "noisy nonlinearity" hypothesis, introducing an intermediate class of "semi-mechanistic" models which incorporate some mechanistic structure while retaining statistical flexibility. All of the nonlinear time series models exhibited higher prediction accuracy than deterministic chaotic models, but the semi-mechanistic model was by far the most accurate. This comparison suggests that for forecasting, control, and other practical applications on populations other than measles, semi-mechanistic modeling may be the most effective approach for characterizing and predicting population dynamics from limited data. Characterization of measles dynamics based on the semi-mechanistic models indicates that the dynamics have an appreciable random component, are near the border between stability and chaos, and vary between local (in state space) stability and chaos.

## INTRODUCTION

Historical data on recurrent epidemics have generated sustained interest among epidemiologists and population biologists interested in the causes, and the consequences for control and prediction, of the complex mix of seasonal and non-seasonal oscillations often observed. Much recent interest has resulted from influential papers by Schaffer, Kot, Olsen and coworkers (reviewed by Kot et al. 1988, Schaffer et al. 1990) contending that the historical record on measles in the developed world provides evidence of chaotic dynamics occurring in a natural biological population. It remains uncertain whether or not any populations of macro-organisms exhibit chaos, either in the wild or in the lab (Hastings et al. 1993). Recent analyses based on nonlinear time-series modeling (Turchin 1993, Ellner and Turchin 1995) and mechanistic population modeling (Hanski et al. 1993, Dennis et al. 1995, Costantino et al. 1995) have identified several likely cases, but these methods need further testing and the data sets on macro-organisms are all short (usually 30-100 data points). Epidemic data series are longer, typically over 400 data points, and their accuracy is probably better due to the importance attached to human disease notifications. Consequently as Tidd et al. (1993; hereafter TOS) observe, "chaos in childhood diseases is thus something of a test case on which hinges a good deal more than the dynamics of some half dozen pathogenic agents". A convincing "case for chaos" in epidemics would increase the credibility of evidence for chaos in other natural populations for which the data are less accurate and sparser.

The initial "case for chaos in childhood epidemics" (reviewed by Schaffer et al. 1990) was based on methods developed in theoretical physics, such as algorithms to calculate fractal dimension. Parallel studies of mechanistic epidemic models -- mainly SEIR models (described below) revealed that certain features of the data were consistent with model output only for parameter values where the models are chaotic. However it is now widely recognized that those methods often give misleading results for data series (such as epidemics) that are short and noisy by the standards of experimental physics (e.g., Stone 1992, Tidd et al. 1993, Hastings et al. 1993, Ellner et al. 1995). Moreover, methods intended for controlled-environment laboratory data are seriously confounded by seasonal and school-year driven variations in transmission rate (Ellner et al. 1995); models with seasonal variation in parameters can generate spurious "fieldmarks" of chaos that mimic closely those found in epidemic data (Ellner 1991, Stone 1992).

Elsewhere (Nychka et al. 1992, Ellner & Turchin 1995, Ellner et al. 1995) we have presented a different approach based on fitting a nonlinear time series model (NTSM)

$$x_{t+T} = f(x_t, x_{t-1}, \dots, x_{t-n}) + e_t. \quad (1)$$

Here  $x_t$  is the log of population size or number of cases at time  $t$ ,  $e_t$  represents exogenous environmental noise affecting the dynamics, and  $f$  is a flexible nonlinear model whose parameters are estimated from the data. The model in (1) is based on attractor reconstruction in time-delay coordinates, generalized to allow for exogenous perturbations. The fitted model can be used to characterize the dynamics for stability vs. chaos (by calculating the Lyapunov exponent  $\lambda$ ), and the level of predictability vs. randomness in the dynamics (quantified by the  $r^2$  of the fitted model). Ellner and Turchin (1995) explain these procedures and their theoretical background; several related measures have been developed by H. Tong and co-workers (see Yao and Tong 1995). Using NTSM, measles epidemics in large developed-world cities were found to have non-negligible noise (Ellner et al. 1995; mean  $r^2 = 0.82 \pm 0.09$  (SD),  $n=12$  cities, for predictions one quarter ahead). The dynamics tended to be weakly stable, with the distribution of estimated Lyapunov exponents clustering near the border between stable and chaotic dynamics (dominant Lyapunov exponent  $\lambda=0$ ).

Proponents of NTSM have argued that SEIR models are a drastic simplification of the system, so SEIR-based data analyses are over-constrained and can be strongly biased. In particular, the omission of environmental variability biases the outcome towards chaos, because chaos is then the only way to generate dynamics that are not almost exactly periodic. TOS criticized NTSM for using strictly phenomenological models that treat the data as just a series of numbers and ignore the considerable biological information that is incorporated in mechanistic models (e.g., transmission from infectives to susceptibles, seasonal trends in contact rates). TOS argue that NTSM is therefore prone to miss departures from nonchaotic null hypotheses, because the omission of information reduces the precision of estimates. To support this claim TOS compared the forecasting accuracy of the phenomenological and mechanistic models, on data that were not used to fit the model (out-of-sample forecasting, as in Casdagli 1989, Sugihara and May 1990). TOS found that finite-population simulations of a deterministic SEIR model had higher prediction accuracy than linear stochastic time series models on measles data from several large cities, and interpreted this result as evidence for underlying determinism in childhood epidemics.

However, the comparison in TOS did not contain any models that represent the hypothesis supported by NTSM, that epidemics are both nonlinear and "noisy" (beyond the inevitable demographic stochasticity due to finite population size). If this "noisy nonlinearity" hypothesis is admitted, the comparison in TOS is seen to have an incomplete experimental design (Table 1). Because TOS only compare linear, stochastic, phenomenological models with nonlinear, deterministic, mechanistic models, it is not possible to determine from their

comparison where the dynamics fall on the attributes of interest: linear vs. nonlinear, deterministic vs. stochastic, SEIR vs. not-SEIR.

Our goals in this paper are twofold. First, we complete the programme initiated by TOS to provide a quantitative basis for choosing between these alternate characterizations of the dynamics, by adding to the comparison models that represent the "noisy nonlinearity" hypothesis. These complete the experimental design of TOS by filling in the lower-right corner of Table 1A. Secpnd, we introduce and evaluate a class of models intermediate between mechanistic and phenomenological, which we call "semi-mechanistic". The semi-mechanistic models incorporate some SEIR-style structure, and use mechanistic information to define *a priori* a meaningful state space, but estimate the form of the contact rate equation from the data. When the dust settles, a semi-mechanistic model is the "winner" in our comparison. This outcome suggests that TOS were correct in claiming that NTSM analyses are under-constrained and imprecise, but it also indicates that SEIR is over-constrained and thus biased. We expect (and are investigating) improvements in forecasting accuracy by using extensions of SEIR with additional mechanistic detail. However any mechanistic model is inevitably incomplete, especially if model complexity is constrained by the need to estimate parameters from limited available data. Spatial structure, age structure, individual heterogeneity in susceptibility, and long-term trends in birth rates are among the factors that a "valid" mechanistic epidemic model would need to include, but there is little hope of obtaining data to estimate all of them accurately.

Data analysis based on a completely specified mechanistic model is tacitly assuming that we know everything there is to know about the dynamics and have the data needed to estimate parameters for all relevant processes; in this ideal situation state we could, and should, use the mechanistic model as the basis for data analyses, forecasting, and control. Data analysis based on a purely phenomenological model assumes that we know nothing whatsoever about the processes underlying the data. Our actual state of knowledge is typically somewhere between these extremes, for epidemics and (we would argue) for virtually all populations and ecosystems. A model representing our knowledge would therefore be partially mechanistic, but would retain flexibility about poorly known aspects of the system. The comparisons here illustrate how quantitative comparisons can be used to decide which mechanistic information should be "hard-wired" into the model, and how the final model can be employed to characterize the dynamics. These methodological aspects are generally relevant to modeling the dynamics of populations and ecosystems, for a range of objectives including forecasting, control, and elucidating mechanisms by comparing the goodness-of-fit of models based on alternative mechanistic hypotheses.

The literature on nonlinear dynamics and forecasting related to this paper is enormous and mostly not written for biologists, so we do not attempt a review. For a selective review see Hastings et al. (1993); recommended entry points into the primary literature include Casdagli and Eubank (1992), Ott et al. (1994), Weigend and Gershenfeld (1994), and Tong (1995).

## METHODS AND MODELS

*Data.* The data are monthly case report totals for measles in 5 large cities (Figure 1). The time span for each data series was chosen to terminate before vaccination (which began in the 1960's) had any evident effect, and to avoid any obvious causes of non-stationarity in the dynamics. The data for London therefore does not include the years of the Second World War, during which many children were evacuated. For US cities those years are included; virtually all cases occur before the age of 15, so the absence of those old enough for military service would not greatly affect disease transmission.

*Prediction accuracy.* To assess prediction accuracy we followed the approach of TOS with some technical modifications that increased accuracy. The forecasting problem is to predict the number of cases in future months, given monthly case report totals  $C_t$  up to the present. We use  $x_t$ =log-transformed monthly case report totals in order to mitigate the "nonuniformity" in measles data (Kendall et al. 1994) that reduces forecasting accuracy. Forecasts forward from time  $t$  are then based on a vector of past values,

$$X_t = (x_t, x_{t-L}, \dots, x_{t-nL}) . \quad (2)$$

$X_t$  is a reconstructed state-vector for the system in which past values act as surrogates for unobserved variables, such as the number of susceptible individuals.

For each time  $t$  and each prediction interval  $T_p$ , each model  $M$  was used to produce a forecast  $\hat{x}_{t+T_p}^M$  of  $x_{t+T_p}$ , which was the conditional mean of  $x_{t+T_p}$  given  $X_t$  in the model. The conditional mean calculations differ among models, and are described below. Predictions are made only for the second half of each data series, which was not used in fitting the models. Prediction accuracy was measured by a quantity which we call the "prediction  $r^2$ ". Prediction  $r^2$  is computed by the usual formula for  $r^2$  in a regression analysis:

$$\begin{aligned} \text{Prediction } r^2 &= 1 - \frac{\text{Mean square of residuals}}{\text{Variance of data}} \\ &= 1 - \left( \frac{\sum_j (x_j - \hat{x}_j^M)^2}{\sum_j (x_j - \bar{x})^2} \right) \end{aligned} \quad (3)$$

The sums in (3) run over the second half of the data series, and  $\bar{x}$  is the mean of  $x$ , over the same set of times. TOS report results for two different measures of prediction accuracy (the correlation coefficient, and the slope of the linear regression, between observed and predicted values) but state that a measure equivalent to prediction- $r^2$  (the scaled error) gave "roughly comparable results" (TOS p. 265). Here we use prediction- $r^2$  because of its familiar interpretation as the proportion of total variance accounted for by the model; the correlation coefficient used by TOS (which is for the linear regression between observed and predicted) does not have that interpretation.

*Mechanistic models.* We consider only the SEIR model, which TOS found to give the highest forecasting accuracy. The SEIR model is a system of nonlinear differential equations that models the changes over time in the fraction of Susceptible, Exposed, Infective, and Recovered individuals in the population (exposed individuals have caught the disease but are asymptomatic and do not transmit the disease to others; infectives are symptomatic and can transmit the disease). The model is

$$\begin{aligned} dS/dt &= m(1-S) - b(t)SI \\ dE/dt &= b(t)SI - (m+a)E \\ dI/dt &= aE - gI \\ R &= 1 - (S + E + I) \end{aligned} \tag{4}$$

where  $m$  is the mortality rate (assumed to be the same for all classes),  $b(t)$  is the contact intensity, and  $1/a$  and  $1/g$  are the mean duration of the exposed and infectious periods. This form of the model assumes that all newborns are susceptible and that recovered individuals have permanent immunity (which are appropriate for measles), and further that population size is constant.  $b(t)$  is assumed to follow a deterministic annual cycle due to seasonality and the school year, transmission being more likely when school is in session. A sine-wave model for  $b(t)$  is often used for convenience, but we used the more realistic form introduced by Kot et al. (1988),  $b(t) = \beta_0(1 + \beta_1\varphi(t))$ ,  $\varphi(t) = 1.5(0.68 + \cos(2\pi t)) / (1.5 + \cos(2\pi t))$ .

Forecasts for the SEIR model were obtained by a refinement of the method in TOS. We generated a 200-year-long series of simulated case totals  $\{C_t^S, t = 0, 1, \dots\}$ , which were then log-transformed and scaled in the same way as the real data, and converted into an "atlas" of simulated state vectors  $\{X_t^S, t = 0, 1, \dots\}$ . (Spot-checks on atlases up to 500 years long indicated that an atlas longer than 200 years yields only minuscule improvement). Predictions forward in time from any state-vector  $X_t$  in the data series were obtained as a weighted average over

trajectories originating at nearby state vectors in the atlas, in which closer vectors are weighted more heavily:

$$\hat{x}_{t+T_p} = \frac{\sum_i w_{i,t} x_{i+T_p}^S}{\sum_i w_{i,t}} \quad (5)$$

Equation (5) is what statisticians call a *kernel* regression model, which provides a consistent estimate of the conditional mean (Härdle 1990, Cheng and Tong 1992). The weights  $w_{i,t}$  are given by  $k(d_{i,t}/h)$ , where  $d_{i,t}$  is the distance from  $X_t$  to the  $i^{\text{th}}$  point in the atlas,  $k$  is a weighting function which falls off with distance, and  $h$  (called the *bandwidth*) is a constant which controls the rate at which more distant points are down-weighted. We used  $k(z) = 1/(1+z^2+0.5z^4+0.3z^6)$ , which falls off roughly as  $\exp(-z^2)$  but is computationally faster. For each prediction time  $T_p$ , the value of  $h$  was chosen by ordinary cross validation on the simulated data (specifically, we used the value of  $h$  which gave the highest prediction  $r^2$  when the atlas for predictions from  $X_j^S$  consisted of all  $X_k^S$  with  $|j-k|>12$  months). Because the optimal  $h$  depends on the length of the atlas, it would be inappropriate to choose  $h$  by cross-validation on the empirical data.

Following TOS we first implemented SEIR as a finite-population Monte Carlo simulation, with a small amount of immigration (as in Kendall et al. 1994) to prevent an unrealistically high number of months with 0 cases. However, we found that substantially more accurate forecasts could be obtained using the differential equations (4) to create the atlas. We used the values of  $m$ ,  $a$ , and  $g$  (which can be estimated directly) and of  $\beta_0$  from TOS. For each data series we computed forecasting accuracy for  $\beta_I=0.20, 0.22, 0.24, 0.26,$  and  $0.28$ , which spans the range of empirical estimates and runs from simple periodic dynamics to chaos. Forecasting accuracy was improved, especially at lower values of  $\beta_I$ , by adding small random variations about the seasonal trend in contact rate intensity,  $b(t) = \beta_0(1 + \beta_1\varphi(t) + \sigma z_t)$  where  $z(t)$  was a first-order Gaussian autoregressive process with unit variance and autocorrelation of 0.5 at time lag of 2 weeks. For each value of  $\beta_I$  the value of  $\sigma$  was increased from 0 in increments of 0.01 until the model dynamics were qualitatively similar to the data (i.e., a mix of 2-year-periodic, 3-year-periodic, and aperiodic dynamics). These values of  $\sigma$  are thus the minimal modification of the deterministic model necessary to generate realistic simulations. The largest value of  $\sigma$  was 0.03, at  $\beta_I=0.20$ ; even small amounts of noise can move the SEIR model from periodic to chaotic dynamics for  $\beta_I$  in the range considered here (Rand and Wilson 1991).

*Time-series models.* We considered two phenomenological time-series models, feedforward neural networks (FNN) and the semi-nonparametric (SNP) model of Gallant and Tauchen (1992). The FNN model is equation (1) with  $T=1$  and

$$f(x_1, x_2, \dots, x_d) = \beta_0 + \sum_{i=1}^k \beta_i G\left(\sum_{j=1}^d \gamma_{ij} x_j + \mu_i\right) \quad (6)$$

where  $G$  is a sigmoid function such as  $G(u) = \exp(u)/(1 + \exp(u))$ . In the neural interpretation, each term in ( )'s represents the stimuli to a single neuron and  $G(\cdot)$  is that neuron's firing rate. However FNN was used here strictly as a statistical model. Given the amount of data available on measles, our past experience and comparative studies favor FNN over competing models for characterizing nonlinear dynamics (McCaffrey et al. 1992, Ellner & Turchin 1995). The parameters  $(\beta_i, \gamma_{ij}, \mu_i)$  were estimated by nonlinear least squares for each given value of  $k$ , and the value of  $k$  was chosen by the Bayes Information Criterion (BIC), using methods described elsewhere (Ellner et al. 1992, Nychka et al. 1992).

The SNP model, developed for financial and macroeconomic time series, approximates the one-step-ahead conditional density using a series expansion in which the leading term is a linear autoregression with Gaussian noise. Departures from the linear model are described by polynomials multiplying the leading term's density and modifying its shape, in which the coefficients are polynomial functions of the current state. A complete description is given elsewhere (Gallant and Tauchen 1992). After some trial-and-error (using only the first half of each data series) the SNP model was constrained to use 24 past monthly values in the linear autoregression but only the 4 most recent values in the polynomials. Methods used to estimate parameter values and select the polynomial order are described elsewhere (Gallant and Tauchen 1992).

The SNP and FNN models were fitted for a prediction time of one month ahead. Predictions farther ahead were obtained by simulating the dynamics and averaging over simulations to obtain the conditional mean. For SNP this involves repeated random draws from the fitted one-step-ahead transition density. For FNN we iterated forward equation (1) with  $T=1$  and random shocks  $e_t$ , generated by random sampling with replacement from the residuals of the one-month-ahead model.

*Semi-mechanistic model.* Our objective in constructing a semi-mechanistic model was to “hard-wire” into the model the qualitative structure of disease transmission without specifying the exact form of the rate equations. The model we used was

$$\begin{aligned}x_{t+1} &= g(S_t, x_t, x_{t-L}, x_{t-2L}, \dots, x_{t-nL}) + e_t \\ S_{t+1} &= S_t - C_{t+1} + r_{t+1}\end{aligned}\tag{7}$$

where  $C_t$  is the (untransformed) total number of cases in month  $t$ ,  $S_t$  is an estimate of the number of susceptibles at the end of month  $t$ , and  $r_t$  is the total net recruitment into the susceptible population in month  $t$ .

The first line in equation (7) represents new cases resulting from contact between susceptibles and infectives, as in equation (4).  $(S_t, x_t, x_{t-L}, x_{t-2L}, \dots, x_{t-nL})$  should be viewed as a reconstructed state vector in which the lagged case totals are surrogates for unobserved variables. The second line in equation (7) is just mass-balance for the number of susceptibles. To be exact it should read  $S_{t+1} = S_t - E_{t+1} + r_{t+1}$ , where  $E_t$  is the number of individuals catching the disease in month  $t$ . However, because the latent period for measles (roughly 1 week) is short compared with the sampling interval (one month), we can assume that  $C_t \doteq E_t$  and thereby avoid using  $E_t$  as an additional state variable. We used the FNN model for  $g$  in equation (7), which we call the “SC\_FNN” model; parameters of  $g$  were estimated by least squares. We also used a kernel model (as described above) for  $g$ , in which the atlas for prediction forward from  $X_j$  consisted of all  $X_k$  with  $|j-k| > 24$  months, bandwidth chosen by ordinary cross validation. The kernel model was used mainly to obtain an estimate of forecasting accuracy based on the entire data set, using ordinary cross validation (as described above). This could not be done with SC\_FNN because of the computational cost of re-fitting the model hundreds of times with different subsets of the data deleted.

Values of  $S_t$  and  $r_t$  were estimated from case report data using methods described elsewhere (Bobashev et al. 1995).  $S_t$  differs from the actual number of susceptibles by an unknown shift of location and scale;  $r_t$  differs from actual recruitment by the same shift of scale, which is approximately equal to the fraction of cases that are reported. The method guarantees that the second line of equation (7) is valid. Predictions were generated by simulation of model (7), as for FNN and SNP. The recruitment rate  $r_t$  was treated as a known covariate, since it is an exogenous factor not predictable from epidemic data, but both  $x_t$  and  $S_t$  were forecast by iterating equation (7) with the fitted  $g$ .

*Seasonal covariates.* In previous studies (Ellner et al. 1995, Ellner and Turchin 1995) the prediction accuracy of FNN models for monthly population data was greatly improved by using time-of-year as a covariate for predictions. This was done by adding a seasonal “clock” to the basic model (equation 1), giving the seasonal FNN model

$$x_{t+T} = f(x_t, x_{t-L}, \dots, x_{t-mL}, \cos(2\pi t / 12), \sin(2\pi t / 12)) + e_t \quad (8)$$

in which  $f$  is given by equation (6), and the seasonal semi-mechanistic model

$$\begin{aligned} x_{t+1} &= g(S_t, x_t, \sin(2\pi t / 12), \cos(2\pi t / 12)) + e_t \\ S_{t+1} &= S_t - C_{t+1} + r_{t+1} \end{aligned} \quad (9)$$

where  $g$  is given by equation (6) or by a nonparametric kernel. Additional lagged values of  $x_t$  (as in equation 7) were also tried in equation (9), but this did not improve forecasting accuracy. Seasonal models (equations 8 and 9) uniformly out-performed their nonseasonal counterparts (equations 6 and 7), so we report results only for the seasonal models.

SEIR-based predictions implicitly include seasonality, because seasonal variation in the contact intensity  $b(t)$  is built into the model. With time delay  $L=1$  the TOS prediction method often bases predictions on neighboring state vectors that come from the wrong time of year (Grenfell et al. 1994), but using larger values of  $L$  (as we do here) avoids this error and the forecasting accuracy is much better. As a result we were unable to find any way of incorporating seasonal covariates that improved SEIR forecasting accuracy. Use of susceptibles actually reduced the average SEIR forecasting accuracy. These results indicate that reconstruction of the SEIR state space from lagged cases really worked: seasonality and susceptibles did not provide any additional information that was useful for prediction.

*Model-free prediction.* As baselines for interpreting prediction accuracies, we determined the prediction  $r^2$  for two methods that make direct use of the data. The first uses only the average seasonal trend: the forecast for time  $k$  is simply the mean of all values from the same month in the first half of the data series. We refer to this as the “seasonal trend” forecast. The second method is the same as that used for the SEIR-based prediction, except that the atlas for each prediction consists of all data outside a 24-month window on either side of the time for which predictions are made. This is similar to using the first half of the data to predict the second, but makes more efficient use of the limited data. We refer to this as “data atlas” forecasting. As a third baseline, note that a prediction  $r^2 < 0$  occurs if model-based

predictions are less effective than simply taking the unconditional mean value as the prediction.

## RESULTS

*Embedding parameters.* The forecasting accuracy of each method depends on the choice of embedding parameters in the state vector  $X_t$  (equation 2), namely the time-delay  $L$ , and the number of lags  $D=m+1$ . TOS used  $L=1$  month and  $D=4$  or  $6$ , but as noted above  $L=4$  greatly improved the forecasting accuracy on measles data for England and Wales (Grenfell et al. 1994). We used data-atlas forecasting accuracy as the criterion for choosing embedding parameters (as in Sugihara and May 1990), on the principle that a "good" embedding for the data should also be "good" for an accurate model. For our data sets,  $L=3$  or  $4$  with  $D=5$  or  $6$  gave much more accurate forecasts (Figure 2). In most of the cities, forecasts with  $L=3$  or  $4$  were better, and forecasts with  $L=1$  were worse, than forecasts based solely on the seasonal trend. The highest average forecasting accuracy (averaging over all values in Figure 2) was obtained at  $L=3, D=6$ , which has been the most popular choice for analysis of measles data in past studies. We therefore report results here using  $L=3, D=6$  for the SEIR, FNN, and data-atlas forecasts; all results were very similar for any of the better embedding parameters ( $L=3$  or  $4, D=5$  or  $6$ ). For the seasonal FNN model, the 6-dimensional state vector consisted of 4 past values plus the 2 sin/cos covariates; otherwise it consisted of 6 past lags.

*Seasonality in the SEIR model.* SEIR prediction accuracy also depends on the level of seasonality ( $\beta_I$ ) used in the simulations generating the atlas (Figure 3). For each of the 5 cities, we report results for the value of  $\beta_I$  that gave the highest average forecasting accuracy for 1-24 months ahead using embedding parameters  $L=3, D=6$ :  $\beta_I=0.20$  for New York City and London,  $\beta_I=0.22$  for Detroit,  $\beta_I=0.24$  for Baltimore and  $\beta_I=0.26$  Milwaukee.

*Prediction accuracies of the models.* SEIR-based forecasts were slightly less accurate overall than forecasts based on phenomenological time series models representing the "noisy nonlinearity" hypothesis, SNP and the seasonal FNN model (Figure 4). SNP forecasts were markedly better than SEIR for London and Milwaukee, and nearly identical to SEIR in prediction accuracy on the other cities. FNN forecasts were markedly better than SEIR for Milwaukee, but were only slightly better on average than SEIR for the other four cities. The semi-mechanistic SC\_FNN model was better overall than SEIR in each of the cities (Figure 5), and had a higher prediction  $r^2$  than SEIR in 89% of the (city  $\times$  prediction interval) combinations examined.

Comparing all four models, the highest average prediction  $r^2$  was achieved by SC\_FNN in three cities (New York, Detroit, Milwaukee), FNN in one city (Baltimore), and SNP in one city (London). SC\_FNN outperformed FNN on average (mean prediction  $r^2=0.61$  for SC\_FNN, 0.53 for FNN) and in 71% of the (city  $\times$  prediction interval) combinations: the semi-mechanistic model “beat” the purely phenomenological model. SNP and SEIR had lower average prediction  $r^2$  than either of the neural-net based models (mean prediction  $r^2=0.52$  for SNP, 0.32 for SEIR).

SEIR-based forecasting was reported to be more successful at “peak-to-peak” prediction, in which the next outbreak maximum is predicted based on the most recent outbreak maximum, with intermediate months ignored (Tidd et al. 1993, Kendall et al. 1994). We therefore compared the peak-to-peak prediction accuracy of the SEIR and SC\_FNN models (by fitting a smoothing spline to the time-series of successive maxima in long simulations of the models). Again SC\_FNN gave consistently more accurate forecasts (Table 2), however (in contrast to SEIR) the peak-to-peak forecasts were less accurate than one-year-ahead forecasts obtained by iterating forward one month at a time (Figure 5).

*Characterizing the dynamics: Unpredictability.* The model fitted to a data series can be used to estimate significant features of the dynamics. Some features of interest for population and epidemic dynamics are the level of predictability vs. unpredictable “noise” in the dynamics, the overall extent of chaos vs. stability, and how both of these vary as a function of the system’s current state.

Estimates of short-term unpredictability ( $1-r^2$ ) for each city are summarized in Table 3. The SEIR model has lower forecasting accuracy and therefore gives slightly higher estimates of unpredictability, but the three models give similar estimates. Moreover, there is good (though imperfect) agreement between the models on the relative unpredictability for the different cities, as indicated by linear correlations ( $\rho$  values in the Note to Table 3).

The interpretation of Table 3 is confounded by the effect of measurement errors in the data. Even a perfectly deterministic system will appear to be somewhat unpredictable if the data are corrupted by random errors, and estimates of unpredictability such as those in Table 3 are consequently biased upwards. Thus Table 3 really gives an estimated *upper bound* on the unpredictability. A *lower bound* is also available, based on the fact that the effect of measurement errors on apparent predictability is roughly independent of the prediction interval (Sugihara and May 1990). Figure 6 illustrates this relationship for our data. When

simulated measurement errors were added to the data, the forecasting profiles  $\hat{r}^2(T_p)$  (estimated prediction  $r^2$  as a function of prediction interval) shifted progressively downward. Thus if we could somehow remove the measurement errors, the result would be an upward shift in the forecasting profile. The magnitude of the upward shift can be at most  $1 - \hat{r}^2(1)$ , because even with perfect data the one-month-ahead forecasting accuracy cannot be above 1. An upper bound on the forecasting accuracy can therefore be estimated by applying this maximum possible upward shift, which yields  $\hat{r}^2(1) - \hat{r}^2(j)$  as a lower bound on the  $j$ -month-ahead unpredictability  $1 - r^2(j)$ .

For our data, the lower bound on unpredictability is apparently quite conservative (Figure 7). Figure 7 was derived from the same simulations as Figure 6. For each data series we applied the upward-shift method to each of the replicates with additional simulated measurement errors (CV=0.1, 0.2 or 0.3); the resulting 30 lower bounds on unpredictability are shown as dashed lines in Figure 7. In all cases the lower bound is well below the unpredictability for the data without any simulated measurement errors.

We thus have upper and lower bounds for the intrinsic unpredictability of the system itself, given by  $1 - \hat{r}^2(j)$  and  $\hat{r}^2(1) - \hat{r}^2(j)$  respectively. These bounds on short-term unpredictability are summarized for each city in Table 4. For predictions 6 months ahead, the average lower bound on unpredictability was 23%, with considerable variation among cities.

To give a standard of reference for these unpredictability estimates, we computed unpredictability by the same method for the data and for SEIR model output (Figure 8), in the finite-population Monte Carlo implementation which incorporates demographic stochasticity (see *Methods: Mechanistic models*). The models exhibited lower unpredictability on average than any of the 5 cities, but for London and New York there was little difference between the models' unpredictability and the lower bound on the data unpredictability.

*Characterizing the dynamics: stability vs. chaos.* Overall stability vs. chaos can be quantified by the dominant Lyapunov exponent  $\lambda$ . The Lyapunov exponent gives the long-term sensitivity to initial conditions, i.e. the rate of growth (or decay) over time of the effect of a small perturbation in the system's state. Thus  $\lambda > 0$  indicates sensitive dependence on initial conditions, which is the classical defining feature of chaotic dynamics. The mathematical definition of  $\lambda$  for systems with random perturbations, and methods for estimating  $\lambda$  in such cases, are discussed by Ellner and Turchin (1995).

Based on the comparison of forecasting accuracy, we used the SC\_FNN model to estimate  $\lambda$ . Parameter values were estimated from the entire data series, retaining the model specification that gave the best out-of-sample prediction  $r^2$ . The estimates of  $\lambda$  are shown in Figure 9 (top panel), with 95% confidence intervals based on asymptotic distribution theory for a likelihood ratio statistic (see Bailey 1996, Bailey et al. 1996 for details). These estimates support the previous conclusion (Ellner et al. 1995) that measles epidemic dynamics are clustered near the transition between stable and chaotic dynamics ( $\lambda \approx 0$ ). They are neither strongly stable, nor strongly chaotic, and for 4 of the 5 cities the 95% confidence interval for  $\lambda$  contains 0.

Short-term stability vs. instability can be characterized by the local Lyapunov exponents (LLE)  $\lambda_m(t)$  (e.g., Abarbanel et al. 1991, 1992, Wolff 1992, Bailey 1995).  $\lambda_m(t)$  is the short-term rate of growth ( $\lambda_m(t) > 0$ ) or decrease ( $\lambda_m(t) < 0$ ) in the effect of a perturbation at time  $t$ , over the time interval from  $t$  to  $(t+m)$ . If the system is deterministic,  $\lambda_m(t)$  depends only on the system state at time  $t$ ; in noisy systems it is a random quantity that depends on the system trajectory between times  $t$  and  $(t+m)$ .

In each of the cities, there was substantial variation in the local Lyapunov exponents over the course of epidemics, including a roughly 50:50 split between positive and negative values (Figure 9, lower 3 panels). The variation in local exponents within each city (as indicated by the gap between the 10th and 90th percentiles of the distribution) is far larger than the variation between cities in the global exponent. Thus in all cases the near-0 global exponent results from alternation between short-term sensitive and insensitive dependence on initial conditions. The local Lyapunov exponents are constant in a linear system, so the large variation seen here is additional evidence of significant nonlinearity in measles dynamics.

## DISCUSSION

The main biological implication of our analyses is that measles epidemics are best described as a mixture of nonlinearity and "noise", with neither component being small enough to disregard. The dynamics are not essentially deterministic, nor are they just random deviations from a seasonal trend. The data exhibit unpredictable variation beyond what can be accounted for by measurement error, or by finite-population effects in an otherwise deterministic SEIR model. The nature of this "noise" around the model predictions is an

important question that our analyses here do not resolve. Epidemics are likely to be perturbed by both environmental variability and demographic stochasticity (Bartlett 1957, Grenfell et al. 1995b). The biological details omitted from our models, such as age structure and spatial heterogeneity, may also contribute to forecasting errors. However the observed levels of predictability are quite high for at least 6 months into the future, and the dynamics are reasonably well described by a simple semi-mechanistic model.

The estimates of the global Lyapunov exponent  $\lambda$ , and the local exponents  $\lambda_m(t)$  (Figure 9) indicate that the epidemic dynamic are clustered very near the border between stable ( $\lambda < 0$ ) and unstable ( $\lambda > 0$ ) dynamics, but there are large fluctuations in the degree of short-term chaos vs. stability (i.e., short term sensitivity vs. insensitivity to initial conditions). These findings imply that interactions between noise and nonlinearity are a significant aspect of the dynamics. During periods of short-term sensitivity to initial conditions ( $\lambda_m(t) > 0$ ) the system is acting as a "noise amplifier", in which the disease transmission dynamics amplify the effect of random "shocks" to the system. During these periods, much of the unpredictability is internally generated, as in a chaotic system. When  $\lambda_m(t) < 0$  there is short-term insensitivity to initial conditions, and unpredictability is generated only by the external shocks, as in a stable system with random perturbations. A characterization of this kind of dynamics as "chaotic" or "nonchaotic" might be correct, but it would be incomplete and misleading.

Our results also have more general implications for the study of dynamics in other ecological and epidemiological systems. Our results support TOS's argument that analyses based on phenomenological time series models omit useful information and therefore sacrifice precision. However, contrary to TOS, our results do not support analyses based on a simple mechanistic model such as SEIR. The main flaw in the TOS study was the use of linear models as the benchmark to evaluate the performance of a candidate mechanistic model. A linear model is not a fair "straw man" because it is not the best of the phenomenological models now in use. We have demonstrated that a nonparametric autoregressive estimator such as a neural net that attempts to track the first conditional moment, or a completely nonparametric time series estimator such as SNP that attempts to track all conditional moments, are the appropriate phenomenological models to use as benchmarks. However the most accurate description of the dynamics, as measured by out-of-sample forecasting accuracy, was obtained

with a semi-mechanistic model that incorporates some reliable biological information but uses the data to estimate the correct form of some terms in the rate equations.

The main difference between the SEIR and semi-mechanistic models is that the former uses the conventional "strong homogeneous mixing" contact rate equation (Anderson and May 1991), in which the infection rate is proportional to the product of the number of susceptibles and the number of infectives. The SEIR and semi-mechanistic models operate at exactly the same levels of detail - citywide monthly case totals - so the difference in performance is not simply a matter of one model being more detailed than the other. The lower predictive power of the SEIR model therefore suggests that the conventional contact rate equation may not be adequate. If so, this might provide an explanation for the persistent discrepancies between the empirical data and SEIR-type models (Bolker and Grenfell 1993). Alternatively, the flexibility of the semi-mechanistic model may be allowing it to account, at least in part, for some of the biological details omitted from our models. The most serious omission is probably age structure, which can have significant effects on the dynamics of epidemic models (Schenzle 1984, Anderson and May 1991, Bolker and Grenfell 1993). It is also not yet clear to what extent the irregular, aperiodic fluctuations observed in the data (Figure 1) are intrinsic to the disease dynamics (Grenfell et al. 1995a), or arise from demographic factors such as secular changes in birth rates (Grenfell et al. 1995b,c). Of the models considered here, only the semi-mechanistic model takes account explicitly of variations in birth rates.

The polar alternatives of fitting simple mechanistic models, and purely descriptive statistical techniques, have been the dominant approaches for modeling and quantifying population fluctuations. While it is recognized that both approaches have limitations, both are widely used. Our results suggest that better forecasts, and more accurate characterizations of observed dynamics, may be possible by an appropriate combination of statistical and mechanistic modeling approaches. What is "appropriate" will be difficult to prescribe in general. Rather, the analyses here and in TOS illustrate how a suite of models can be evaluated objectively to identify the kind, and amount, of mechanistic information that ought to be hard-wired into the model, to obtain the most reliable and accurate predictions. The same kinds of comparisons could provide a basis for dissecting the deterministic and stochastic components to identify underlying mechanisms, and determining the importance of factors omitted from candidate mechanistic models. In principle our approach could be recommended for any long-

term data set on population or ecosystem dynamics, but there are few that approach the duration and accuracy of human disease notifications. It therefore remains an important challenge (May 1992) to invent and validate methods which compensate for limited duration in time by exploiting spatial or other forms of replication.

### Literature Cited

- Abarbanel, H. D. I., R. Brown, and M. B. Kennel. 1991. Variation of Lyapunov exponents on a strange attractor. *Journal of Nonlinear Science* 1: 175-199.
- Abarbanel, H. D. I., R. Brown, and M. B. Kennel. 1992. Local Lyapunov exponents computed from observed data. *Journal of Nonlinear Science* 2: 343-365.
- R. M. Anderson and R. M. May, 1991. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford
- Bailey, B. A. 1996. Local Lyapunov exponents: predictability depends on where you are. In: W. Barnett, A. Kirman, and M. Salmond (eds.) *Nonlinear Dynamics in Economics. Proceedings of the Ninth International Symposium in Economic Theory and Econometrics*. Cambridge University Press (*in press*). {Also available as: Institute of Statistics Mimeo Series #2270, Statistics Department, North Carolina State University, Raleigh NC 27695-8203}.
- Bailey, B.A., S. Ellner, D.W. Nychka. 1996. Chaos with Confidence: asymptotics and applications of local Lyapunov exponents. Submitted to: C. D. Cutler and D. Kaplan (eds.) *Nonlinear Dynamics and Time Series: Building a Bridge Between the Natural and Statistical Sciences*. American Mathematical Society, Providence, RI.
- Bartlett, M.S. 1957. Measles periodicity and community size. *Journal of the Royal Statistical Society A* 120:48-70.
- Bobashev, G., S. Ellner, D.W. Nychka, and B.T. Grenfell. 1996 Reconstructing susceptible and recruitment dynamics from measles epidemic data. Submitted to *Mathematical Biosciences*.
- B. M. Bolker and B. T. Grenfell. 1993. Chaos and biological complexity in measles dynamics. *Proceedings of the Royal Society of London B* 251:75-81.
- Casdagli, M. 1989. Nonlinear prediction of chaotic time series. *Physica D* 35: 335-356.
- Casdagli, M. 1992b. A dynamical systems approach to modeling input-output systems. pp. 265-281 in: M. Casdagli and S. Eubank (eds.) *Nonlinear Modeling and Forecasting*. Addison-Wesley, NY.
- Casdagli, M. and S. Eubank (eds.) 1992. *Nonlinear Modeling and Forecasting*. Addison-Wesley, NY.
- Cheng, B. and H. Tong. 1992. On consistent nonparametric order determination and chaos. *Journal of the Royal Statistical Society B* 54: 427-450.
- Costantino, R.F., J.M. Cushing, and R.A. Desharnais. 1995. Experimentally induced transitions in the dynamics behavior of insect populations. *Nature* 375: 227-229.
- Deissler, R. J. and J.D. Farmer. 1992. Deterministic noise amplifiers. *Physica D* 55: 155-165.

- Dennis, B., R.A. Desharnais, J.M. Cushing, and R.F. Costantino. 1995. Nonlinear demographic dynamics: mathematical models, statistical methods, and biological experiments. *Ecological Monographs* 65: 261-281.
- S. Ellner, 1991. Detecting low-dimensional chaos in population dynamics data: a critical review. pp. 63-90 in: J. Logan and F. Hain (eds.), *Chaos and Insect Ecology*. VAES, VPI & SU, Charlottesville, VA.
- Ellner, S., D. W. Nychka, and A. R. Gallant. 1992. LENNS, a program to estimate the dominant Lyapunov exponent of noisy nonlinear systems from time series data. Institute of Statistics Mimeo Series #2235, Statistics Department, North Carolina State University, Raleigh NC 27695-8203.
- S. Ellner, A. R. Gallant and J. Theiler. 1995. Detecting nonlinearity and chaos in epidemic data, in *Epidemic Models: Their Structure and Relation to Data*, D. Molson, ed., Proceedings of NATO ARW on Epidemic Models, Cambridge Univ. Press, Cambridge (1995).
- S. Ellner and P. Turchin, 1995. Chaos in a noisy world: new methods and evidence from time series analysis. *American Naturalist* 145: 343-375.
- Gallant, A.R. and G. Tauchen. 1992 A nonparametric approach to nonlinear time series analysis: estimation and simulation. pp. 77-92 in: D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt, and M.S. Taqqu (eds.) *New Directions in Time Series Analysis, part II*. Springer-Verlag, NY.
- Grenfell, B., B.M. Bolker. and A. Kleczkowski. (1995a). Seasonality and extinction in chaotic metapopulations. *Proc. Roy. Soc Lond. B* 259: 97-103.
- Grenfell, B., B.M. Bolker. and A. Kleczkowski. (1995b). Seasonality, demography and the dynamics of measles in developed countries. pp. 248-268 in: D. Mollison (ed). *Epidemic Models: Their Structure and Relation to Data*. Cambridge University Press, Cambridge.
- B.T. Grenfell, A.Kleczkovski, S. Ellner and B.M. Bolker. 1995c. Non-linear forecasting and chaos in ecology and epidemiology: measles as a case study. pp 345-371 in: H. Tong (ed.) *Chaos and Forecasting: Proceeding of the Royal Society Discussion Meeting*. World Scientific, Singapore.
- Härdle, W. 1990. *Applied Nonparametric Regression*. Cambridge University Press, Cambridge.
- Hansski, I., P. Turchin, E. Korpimäki, and H. Henttonen. 1993. Population oscillations of boreal rodents: regulation by mustelid predators leads to chaos. *Nature* 364: 232-235.
- A. Hastings, C.L. Hom, S. Ellner, P. Turchin, and H.C.J. Godfray, 1993. Chaos in ecology: is Mother Nature a strange attractor? *Annual Reviews of Ecology and Systematics* 24: 1-33.
- Kendall, B.E., W.M. Schaffer, L.F. Olsen, C.W. Tidd, and B.L. Jorgensen. 1994. Using chaos to understand biological dynamics. pp. 184-203 in: J. Grasman and G. van Straten (eds.) *Predictability and Nonlinear Modeling in Natural Sciences and Economics*. Kluwer Academic, Dordrecht.

- May, R.M. 1992. Discussion on the meeting on chaos. *Proceedings of the Royal Statistical Society B* 54: 451-452.
- D. McCaffrey, S. Ellner, A.R. Gallant, and D. Nychka, 1992. Estimating the Lyapunov exponent of a chaotic system with nonparametric regression. *J. Amer. Stat. Assoc.* 87: 682-695.
- D. Nychka, S. Ellner, A.R. Gallant, & D. McCaffrey. 1992. Finding chaos in noisy systems (with discussion). *J. Royal Stat. Soc. Series B* 54, 399-426.
- L. F. Olsen, G. L. Truty and W. M. Schaffer Oscillations and chaos in epidemics: a nonlinear dynamic study of six childhood diseases in Copenhagen, Denmark. *Theor. Popul. Biol.* 33:344-370. (1988).
- Ott, E. 1993. *Chaos in Dynamical Systems*. Cambridge University Press, Cambridge.
- Ott, E., T. Sauer, and J. Yorke (eds.). 1994. *Coping with Chaos: Analysis of Chaotic Data and the Exploitation of Chaotic Systems*. John Wiley and Sons, New York
- D. A. Rand and H. Wilson, 1991. Chaotic stochasticity: a ubiquitous source of unpredictability in epidemics. *Proc. R. Soc. Lond. Biol.* 246:179-184.
- W. M. Schaffer, L. M. Olsen, G. L. Truty and S. L. Fulmer, The case for chaos in childhood epidemics. In: *The Ubiquity of Chaos*, S. Krasner, ed., Washington, D.C., AAAS, 1990, pp. 138-166.
- Schenzle, E. 1984. An age-structured model of pre- and post-vaccination measles transmission. *IMA Journal on Mathematics Applied to Medicine and Biology* 1: 169-191.
- Stone, L. 1992. Coloured noise or low-dimensional chaos? *Proceedings of the Royal Society of London Series B* 250, 77-81.
- Sugihara, G. and R. M. May, Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, 344:734-741 (1990)
- C. W. Tidd, L. F. Olsen and W. M. Schaffer, The case or chaos in childhood epidemics. II. Predicting historical epidemics from mathematical models, *Proc. R. Soc. Lond. B* 254:257-273 (1993).
- Turchin, P. 1993. Chaos and stability in rodent population dynamics: evidence from nonlinear time-series analysis. *Oikos* 68: 167-172.
- Weigend, A.S. and N.A. Gershenfeld (eds.) 1994. *Time Series Prediction: Forecasting the Future and Understanding the Past*.
- Wolff, R. C. W. 1992. Local Lyapunov exponents: looking closely at chaos. *Journal of the Royal Statistical Society Series B* 54: 353-371.
- Yao, Q.W. and H. Tong. 1995. On prediction and chaos in stochastic systems. pp. 57-86 in: H. Tong (ed.) *Chaos and Forecasting: Proceedings of the Royal Society Discussion Meeting*. World Scientific, Singapore.

Table 1

## A. The experimental design in Tidd et al. (1993).

	<u>Mechanistic</u>	<u>Phenomenological</u>
<u>Linear</u>	-	<i>M1, M2, M3</i>
<u>Nonlinear</u>	<i>SEIR, RAS</i>	-

## B. The experimental design in this paper.

	<u>Mechanistic</u>	<u>Semi-mechanistic</u>	<u>Phenomenological</u>
<u>Nonlinear</u>	SEIR	<i>SC_FNN</i>	<i>FNN, SNP</i>

**Note.** - (A) The prediction models compared by Tidd et al. (1993) were either nonlinear mechanistic models (SEIR, RAS) in which the only noise is demographic stochasticity due to finite population size, or linear phenomenological models driven by high levels of environmental noise (M1, M2, M3). SEIR and RAS are described in the text. M1 is a linear Gaussian autoregressive process, M2 is nonlinear transformation of a linear Gaussian autoregressive process, and M3 is a linear Gaussian autoregressive process superimposed on an annual periodic trend. Thus M1 is a special case of M2, which is a special case of M3. (B) In this paper we compare SEIR with models that allow both nonlinearity and substantial levels of environmental noise (FNN, SNP, SC\_FNN models described in the text)..

Table 2. Prediction  $r^2$  for peak-to-peak prediction by the SEIR and semi-mechanistic neural net (SC\_FNN) models for measles epidemics.

	New York	Baltimore	Detroit	Milwaukee	London
SC_FNN	0.51	0.43	0.47	-0.11	-0.05
SEIR	0.43	0.17	0.19	-0.76	-0.09

**Table 3.** Short-term unpredictability estimates (1-prediction  $r^2$ ) by SEIR, SNP, and semi-mechanistic neural net (SC\_FNN) models for measles epidemics, for prediction intervals 1 and 3 months.

	SEIR		SNP		SC_FNN	
	1 month	3 months	1 month	3 months	1 month	3 months
New York	0.12	0.19	0.06	0.22	0.04	0.13
Baltimore	0.22	0.46	0.14	0.55	0.12	0.31
Detroit	0.18	0.36	0.09	0.31	0.08	0.24
Milwaukee	0.30	0.66	0.15	0.53	0.11	0.37
London	0.17	0.30	0.07	0.20	0.09	0.22
Mean	0.20	0.39	0.10	0.36	0.09	0.25
Std. Dev.	0.07	0.18	0.04	0.15	0.03	0.09

Note: Table entries are the mean square prediction error expressed as a fraction of the overall variance of the data. This value can be interpreted as the fraction of the data variance that is not accounted for by the model. Correlations ( $\rho$ ) between the unpredictability estimates from the three models are significant for both one month ahead and three month ahead forecasts (One month:  $\rho(\text{SNP}, \text{FNN})=0.85$ ,  $P=.03$ ;  $\rho(\text{SNP}, \text{SEIR})=0.92$ ,  $P=.01$ ;  $\rho(\text{FNN}, \text{SEIR})=0.81$ ,  $P=.04$ . Three months:  $\rho(\text{SNP}, \text{FNN})=0.88$ ,  $P=.05$ ;  $\rho(\text{SNP}, \text{SEIR})=0.86$ ,  $P=.04$ ;  $\rho(\text{FNN}, \text{SEIR})=0.98$ ,  $P=.009$ .  $P$ -values are for randomization tests against the null hypothesis of independence).

Table 4. Short-term unpredictability (1-prediction  $r^2$ ) estimated by kernel-based semi-mechanistic model with bandwidth chosen by ordinary cross validation. "Upper bound" is the unpredictability from applying the model to the empirical data, corresponding to the estimates in Table 2. "Lower bound" is the under-estimate obtained by the vertical shift method, which over-corrects for effects of measurement error: the curve of prediction  $r^2$  vs. prediction interval is shifted vertically upward so that the value for one month ahead is  $r^2=1.0$ .

	Upper bound			Lower bound	
	1 month ahead	3 months ahead	6 months ahead	3 months ahead	6 months ahead
New York	.07	.13	.20	.06	.13
Baltimore	.10	.23	.33	.13	.23
Detroit	.08	.20	.27	.11	.18
Milwaukee	.14	.36	.55	.23	.41
London	.08	.18	.27	.10	.19
Mean	.09	.22	.32	.13	.23
Std. Dev.	.03	.09	.13	.06	.11

**Note:** Table entries are the mean square prediction error expressed as a fraction of the overall variance of the data. This value can be interpreted as the fraction of the data variance that is not accounted for by the model forecasts.

## FIGURE LEGENDS

**Figure 1.** Data used to evaluate forecasting accuracy, measles monthly case report totals for New York City, Baltimore, Detroit, Milwaukee, and London. New York City data from London and Yorke (1973), London data from the Registrar General's Weekly Report; Baltimore, Detroit, and Milwaukee compiled and provided by William M. Schaffer.

**Figure 2.** Comparison of data-atlas prediction accuracy for different embedding parameters. Predictions for the second half of each data series were made by a kernel regression model fitted to the first half of the data series (see text for details).

**Figure 3.** Comparison of SEIR prediction accuracy for different levels of the seasonality parameter  $\beta_j$ . The results shown are for embedding parameters  $L=3$ ,  $D=6$ , chosen on the basis of data-atlas forecasting accuracy.

**Figure 4.** Comparison of prediction accuracy between SEIR, SNP, and FNN model-based forecasts 1 to 24 months ahead.

**Figure 5.** Comparison of prediction accuracy between the SEIR model and the SC\_FNN semi-mechanistic model.

**Figure 6.** Forecasting accuracy using the semi-mechanistic model (equation 9) with nonparametric kernel estimate of  $g$ , for complete case report time series (solid curve) and for the same data with simulated measurement errors (dotted curves). Measurement errors were Gaussian, with zero mean and standard deviation proportional to the data value; the proportionality constant is the coefficient of variation (CV), which was 0.1, 0.2, or 0.3. Dotted curves show the mean over 10 replicates for each value of CV.

**Figure 7.** Lower bounds on unpredictability derived from the simulations in Figure 8. The upward-shift method (described in the text) was applied to the forecasting profiles with simulated additional measurement errors (10 replicate trials each at  $CV=0.1$ ,  $0.2$ , and  $0.3$ ). This procedure resulted in 30 different estimated lower bounds for the unpredictability in the original data. The lower bounds are plotted as dashed curves; the solid curve is the unpredictability  $1-r^2$  in the original data, corresponding to the solid curve in Figure 6.

**Figure 8.** (a) Comparison of unpredictability between the data, and output of the SEIR model in finite-population Monte Carlo simulations. The unpredictabilities plotted for the data are the lower bound from the upward-shift method, which overcorrects for the effects of measurement errors. The model output are "measured" exactly and therefore we did not do any adjustment of the estimated unpredictability. The solid curves are the average of  $1-r^2$  over 25 successive segments of length 450 months each, from two long simulation run at populations of 1 million and 5 million. Standard errors are below 0.01 for all plotted values. (b) Typical segments of SEIR model output: monthly case totals expressed as percentage of the total population (1 million or 5 million).

**Figure 9.** Estimates of global and local Lyapunov exponents for each of the cities. Estimated values are indicated by the solid box, and the error bars show 95% confidence intervals obtained by the method of Bailey et al. (1996). The top panel shows the estimated global Lyapunov exponent  $\lambda$ , with the dashed line at 0 indicating the transition between stable ( $\lambda < 0$ ) and chaotic ( $\lambda > 0$ ) dynamics. The three other panels refer to the local Lyapunov exponents  $\lambda_m(t)$  for  $m=12$  months ahead, which give the short-term rate of growth ( $\lambda_m(t) > 0$ ) or decrease ( $\lambda_m(t) < 0$ ) in the effect of a perturbation at time  $t$ . The panels show confidence intervals for the fraction of  $\lambda_m(t)$  values which are positive, and for the 10th and 90th percentiles of the distribution of  $\lambda_m(t)$  over the course of the data series.

Fig 1

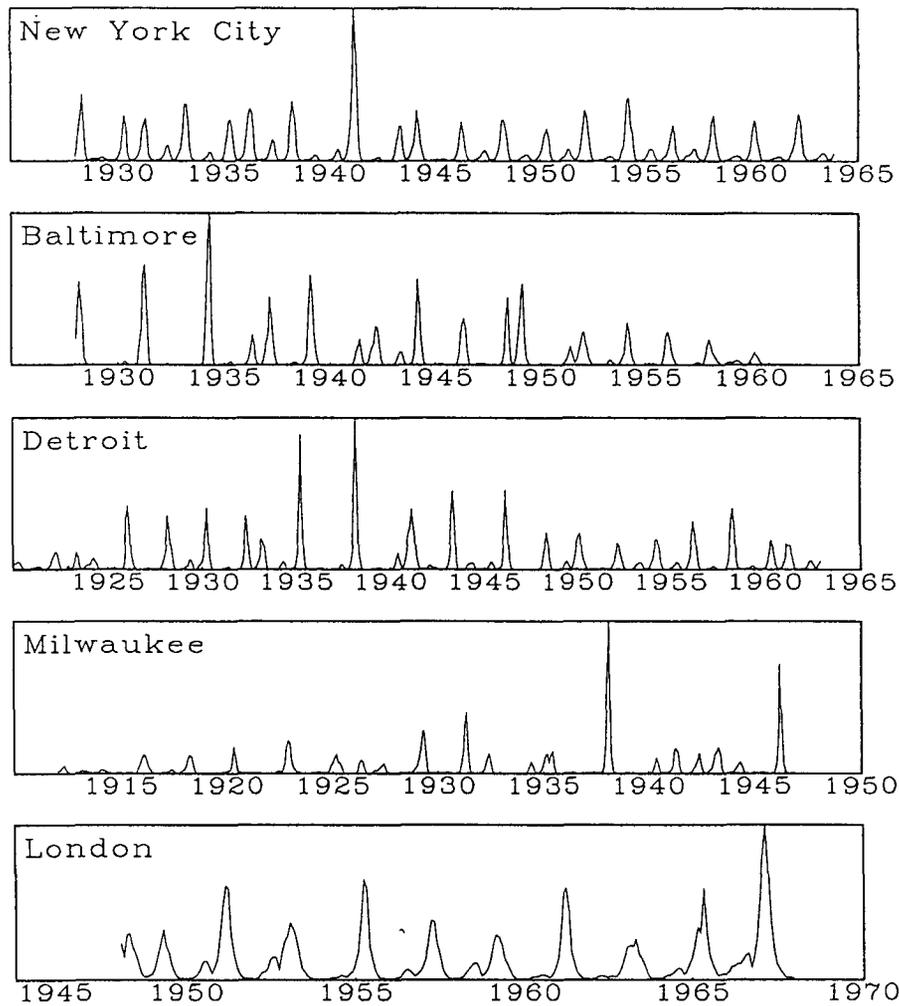
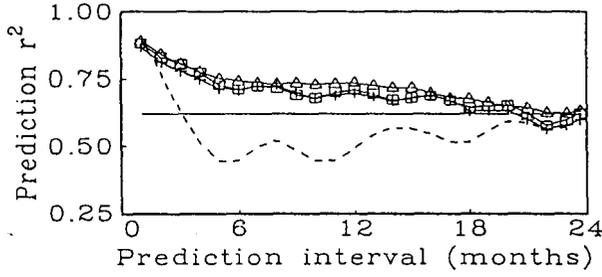
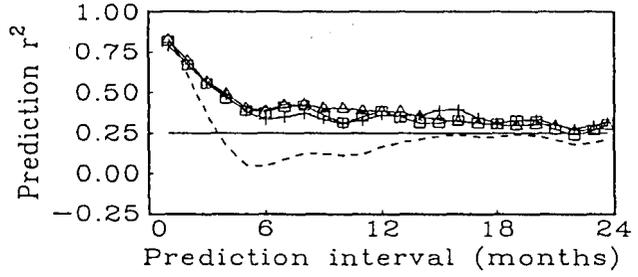


Fig 2

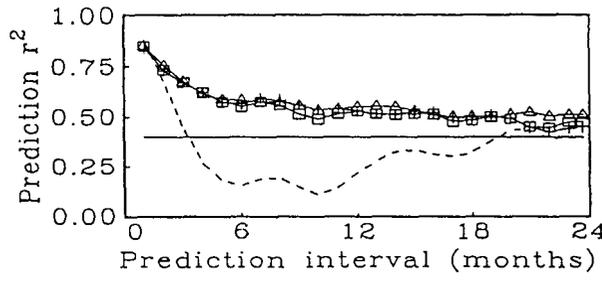
New York City



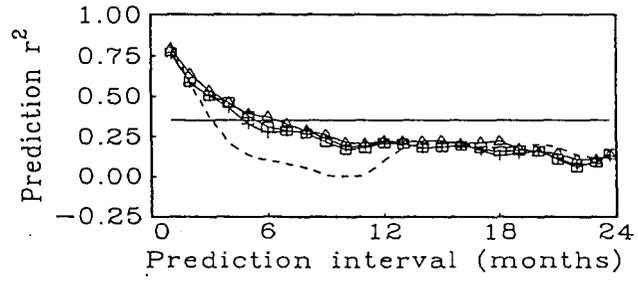
Baltimore



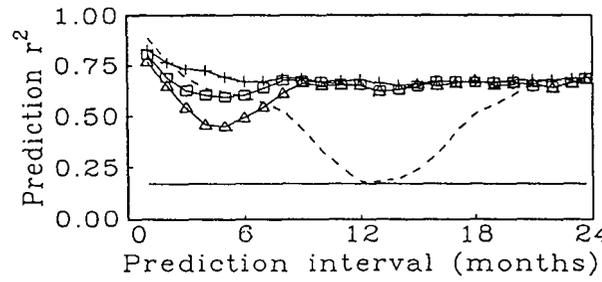
Detroit



Milwaukee



London



- △ L=3, D=6
- L=4, D=5
- + L=4, D=6
- L=1, D=4
- Annual Trend

Fig 3

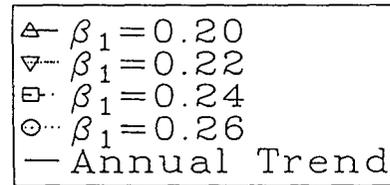
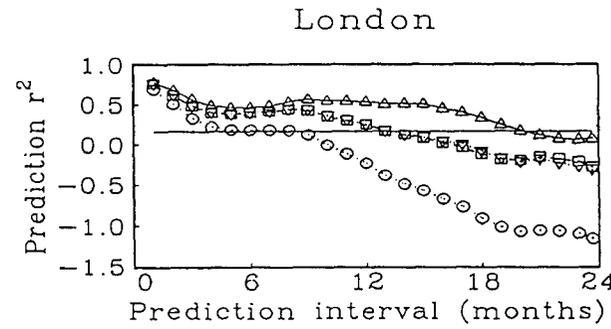
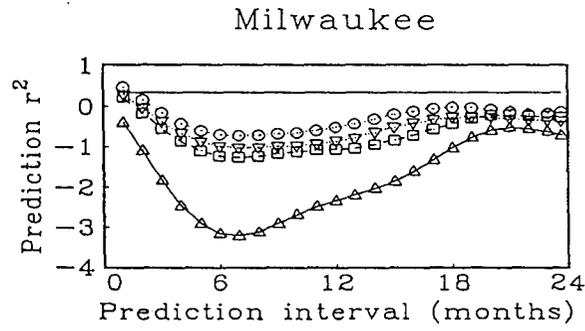
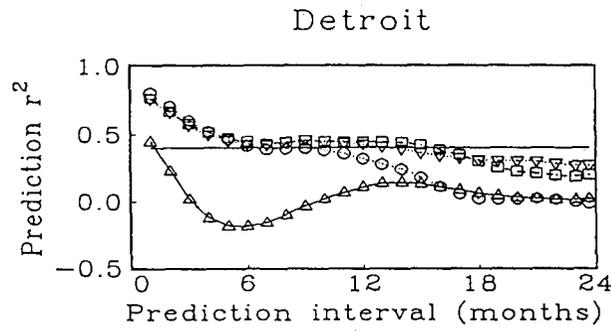
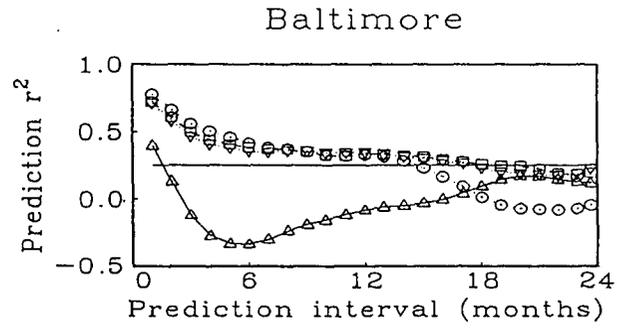
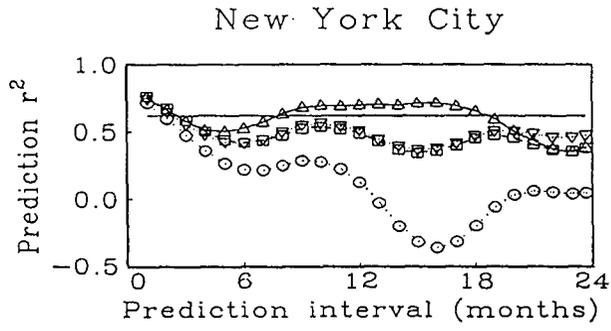
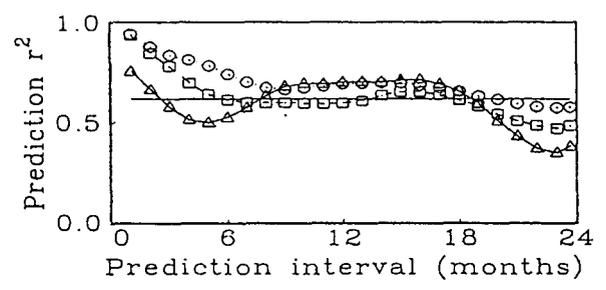
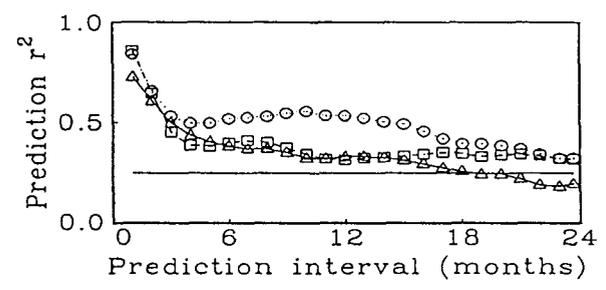


Fig 4

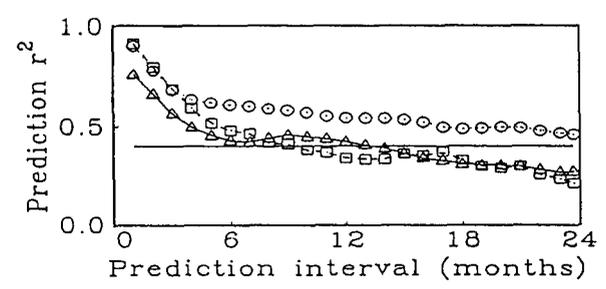
New York City



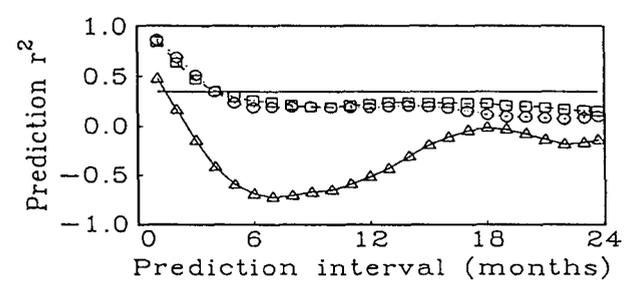
Baltimore



Detroit



Milwaukee



London

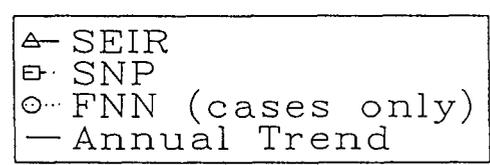
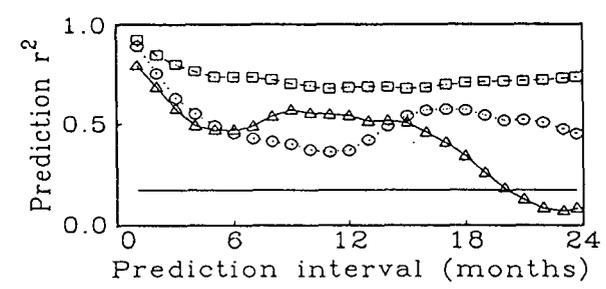
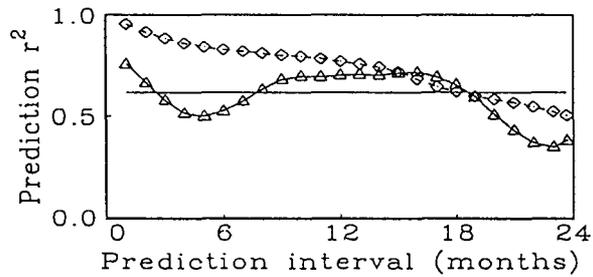
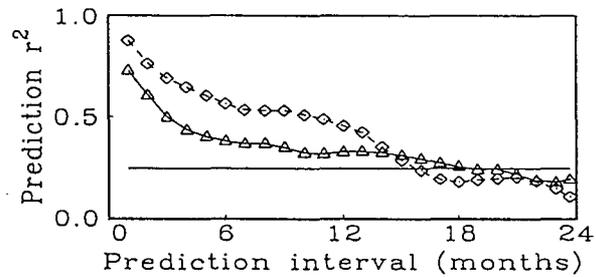


Fig 5

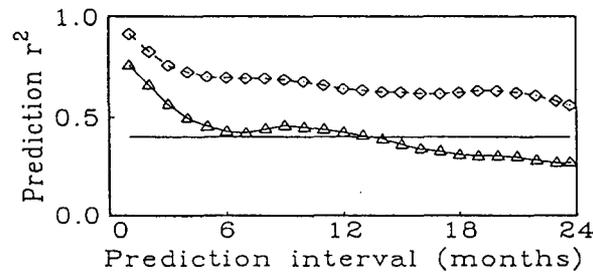
New York City



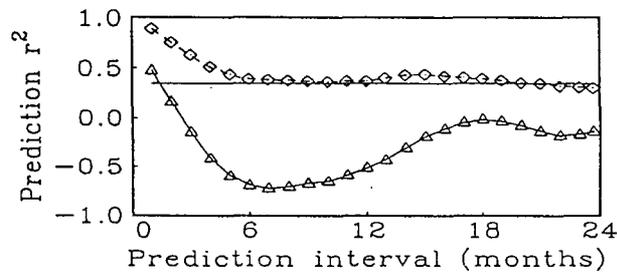
Baltimore



Detroit



Milwaukee



London

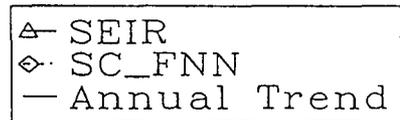
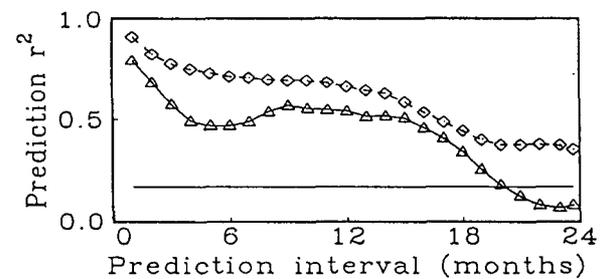


Fig 6

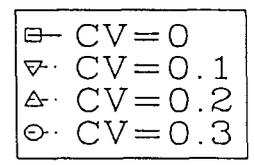
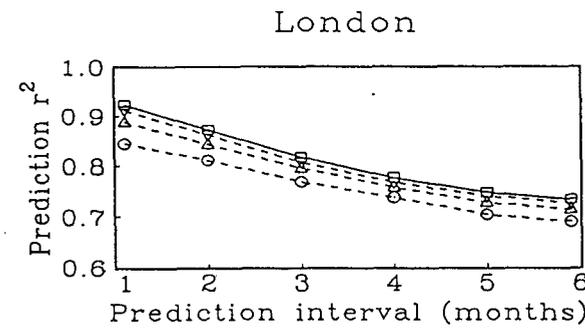
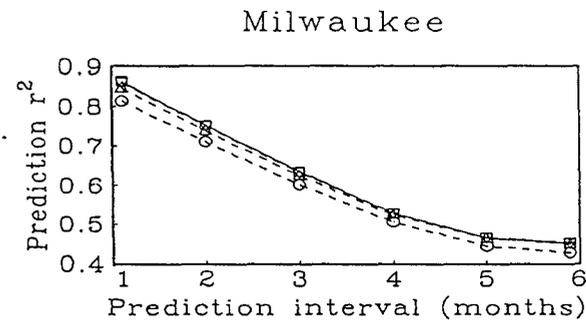
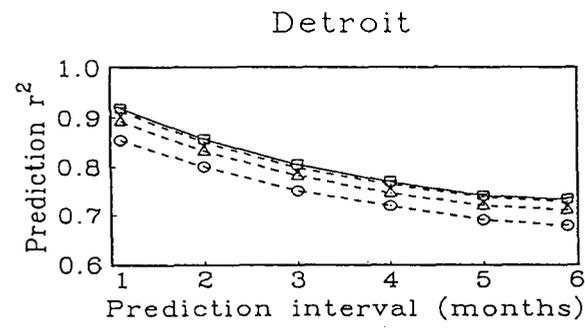
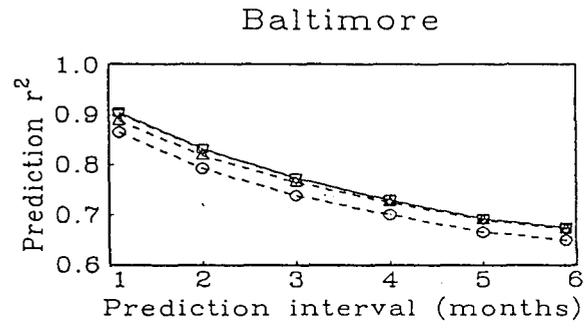
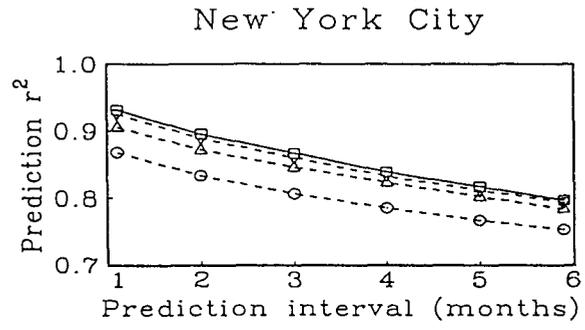


Fig 7

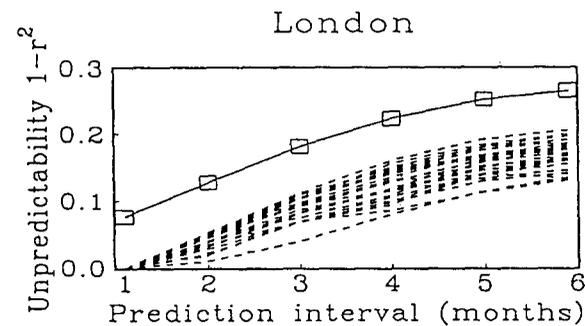
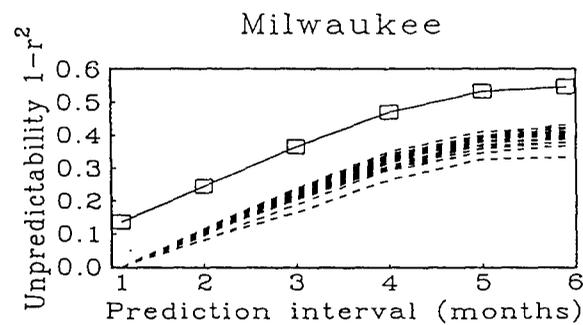
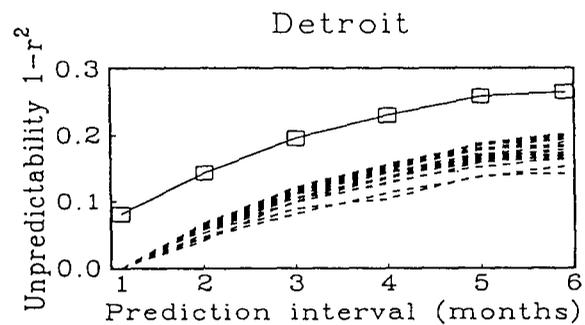
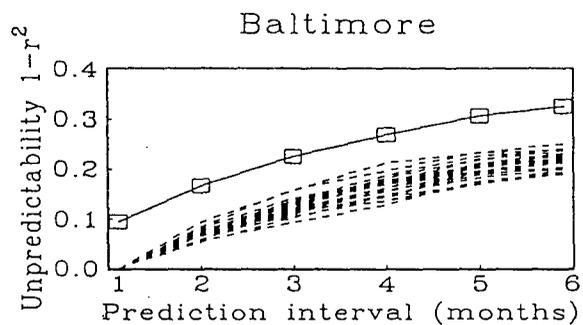
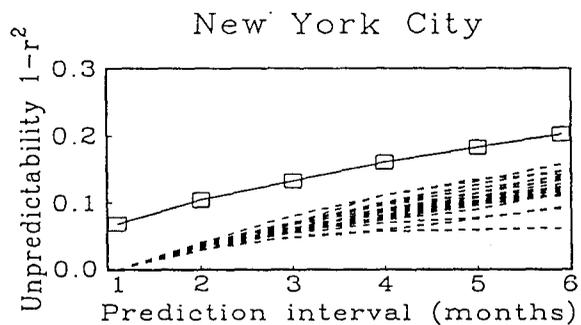
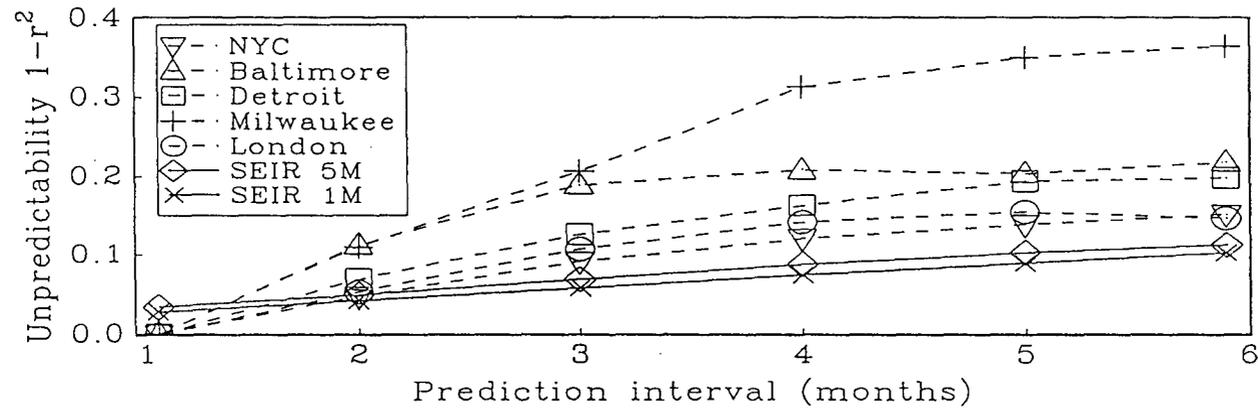


Fig 8

Unpredictability: Data vs. MC-SEIR



MC-SEIR simulations

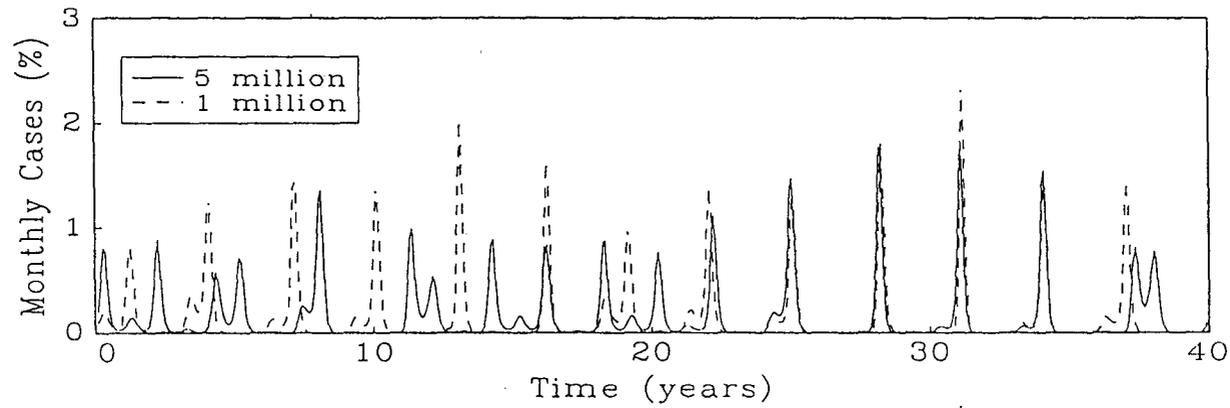


Fig 9

