ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/8443978

Disentangling Extrinsic from Intrinsic Factors in Disease Dynamics: A Nonlinear Time Series Approach with an Application to Cholera

Article *in* The American Naturalist · July 2004

DOI: 10.1086/420798 · Source: PubMed

CITATIONS
144

READS
151

2 authors:

Katia Koelle
Duke University
105 PUBLICATIONS **2,529** CITATIONS

SEE PROFILE

Mercedes Pascual
University of Chicago
283 PUBLICATIONS **12,599** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Environmental drivers of infectious diseases View project

Influenza Receptor Binding Evolution View project

All content following this page was uploaded by Mercedes Pascual on 15 June 2015.

The user has requested enhancement of the downloaded file.

Disentangling Extrinsic from Intrinsic Factors in Disease Dynamics: A Nonlinear Time Series Approach with an Application to Cholera

Katia Koelle* **and Mercedes Pascual†**

Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan 48109-1048

Submitted March 28, 2003; Accepted December 23, 2003; Electronically published May 18, 2004

Online enhancements: appendix, zip file, literature cited.

ABSTRACT: Alternative explanations for disease and other population cycles typically include extrinsic environmental drivers, such as climate variability, and intrinsic nonlinear dynamics resulting from feedbacks within the system, such as species interactions and density dependence. Because these different factors can interact in nonlinear systems and can give rise to oscillations whose frequencies differ from those of extrinsic drivers, it is difficult to identify their respective contributions from temporal population patterns. In the case of disease, immunity is an important intrinsic factor. However, for many diseases, such as cholera, for which immunity is temporary, the duration and decay pattern of immunity is not well known. We present a nonlinear time series model with two related objectives: the reconstruction of immunity patterns from data on cases and population sizes and the identification of the respective roles of extrinsic and intrinsic factors in the dynamics. Extrinsic factors here include both seasonality and long-term changes or interannual variability in forcing. Results with simulated data show that this semiparametric method successfully recovers the decay of immunity and identifies the origin of interannual variability. An application to historical cholera data indicates that temporary immunity can be long-lasting and decays in approximately 9 yr. Extrinsic forcing of transmissibility is identified to have a strong seasonal component along with a long-term decrease. Furthermore, noise appears to sustain the multiple frequencies in the long-term dynamics. Similar semiparametric models should apply to population data other than for disease.

Keywords: temporary immunity, cholera, disease dynamics, seasonality, extrinsic forcing.

* E-mail: kkoelle@umich.edu.

† Corresponding author; e-mail: pascual@umich.edu.

Am. Nat. 2004. Vol. 163, pp. 901–913. © 2004 by The University of Chicago. 0003-0147/2004/16306-30116$15.00. All rights reserved.

Alternative explanations for population fluctuations unavoidably invoke the well-known and often quoted debate of the 1950s and 1960s between Nicholson and Smith, who stressed the importance of density dependence, and Andrewartha and Birch, who proposed density-independent factors as dominant (Davidson and Andrewartha 1948; Andrewartha and Birch 1954; Nicholson 1954). This debate has an interesting parallel in an even earlier dispute on explanations for disease patterns, with contagionists and localists emphasizing, respectively, for cholera the role of disease transmission versus that of the environment and geography (Pollitzer 1959). Well beyond this simple dichotomy, the recent ecological literature recognizes the complex interplay of intrinsic dynamics with extrinsic drivers in nonlinear systems (e.g., Sinclair 1989; Zimmer 1999; Pascual 2001; Rodó et al. 2002). One central problem has been how to identify the contributions of these different components from the irregular temporal patterns of population time series.

Patterns of immunity are critical to the intrinsic dynamics of disease because past infection levels determine the current number of susceptible individuals and hence future infection levels. Extrinsic factors, such as variability in climate or health policy changes, can also impact transmission rates and therefore the pathogen's rate of spread as well as its overall burden in a population. However, the identification and relative importance of extrinsic and intrinsic factors for many diseases remain problematic and even controversial as demonstrated by the current discussions of the role of environmental, particularly climatic, variables on the dynamics of infectious diseases (Hay et al. 2002; Patz et al. 2002; Rogers et al. 2002). This controversy is evident in a recent review of malaria in which Rogers and colleagues, considering the origin of interannual cycles, write, "Longer-term weather cycles such as ENSO have been invoked recently to 'explain' outbreaks of malaria and other diseases. … None of these analyses

allows an alternative explanation involving intrinsic cycles" (Rogers et al. 2002, p. 714).

There are two main difficulties in addressing these alternative explanations. First, many diseases confer only temporary immunity as opposed to the permanent patterns of measles and other childhood diseases, and the duration of immunity as well as the way in which it decays over time are not well known. Temporary immunity is known to be capable, however, of generating interannual cycles in disease models (Cooke et al. 1977; Hethcote et al. 1989; Girvan et al. 2002). Second, statistical approaches are needed that consider both intrinsic and extrinsic factors in the context of the nonlinear dynamics of disease. One such time series approach was applied recently to cholera (Pascual et al. 2000) and was developed as an extension of the nonlinear time series approaches of Ellner and Turchin (1995). This approach is, however, completely phenomenological and relies on lagged disease levels as surrogates for unknown variables in the system. The problem remains of considering a more mechanistic approach that allows for specific patterns of immunity. A further limitation of the approach is that only extrinsic factors that are stationary can be considered, such as seasonality and interannual fluctuations, whose mean does not vary over time. We introduce here a nonlinear time series model for disease dynamics with two related applications: the estimation of the duration and decay pattern of immunity and the question of whether disease dynamics are generated intrinsically, through immunity patterns, and/or extrinsically, through driven changes in the transmission rate. These changes include long-term trends in addition to seasonality and interannual cycles.

Many theoretical studies have shown that periodic forcing can drive populations and disease models into complicated responses whose characteristic frequencies differ from those present in the forcing (e.g., Schwartz and Smith 1983; Schwartz 1992). Long-term trends in birthrates or vaccination can also trigger qualitative changes in the attractors and in the synchronicity patterns of disease (Rohani et al. 1999; Earn et al. 2000). From studies on systems such as the Dungeness crab (Higgins et al. 1997) and childhood diseases (Rohani et al. 2002), we also know that noise can interact with nonlinear feedbacks to generate surprising patterns that are qualitatively different from those produced by the deterministic skeleton alone. In these studies, the extrinsic factors are usually limited to noise and/or seasonality, while the intrinsic factors are the backbone of the mathematical formulation of the model (Higgins et al. 1997; Rohani et al. 2002). One notable exception by Solow uses a semiparametric approach that allows for extrinsic changes other than seasonality and noise in a fish stock–recruitment model (A. Solow, personal communication).

We rely here on a similar semiparametric approach to develop a time series model for diseases with temporary immunity and unspecified variation in the transmission rate. Nonlinear time series models of this sort allow us to combine mechanistic representations of the processes we know with phenomenological representations of unknown processes (Ellner et al. 1998). To our knowledge, this is the first attempt at using a statistical time series approach to understand retrospectively fluctuations in disease cycles as the result of both intrinsic and extrinsic factors, with the latter not limited to noise and seasonality. We illustrate the approach with an application to historical cholera data for Dhaka in former Bengal, the homeland of the disease, at the beginning of the twentieth century. Rainfall data for Dhaka and its dominant frequencies are also presented to further motivate the analysis.

Cholera is an ideal candidate for the application of our model for several reasons. First, temporary immunity is known to occur, but the duration of immunity and the form in which it decays are unknown, and estimates vary widely (Woodward 1971; Levine et al. 1981; Glass et al. 1982; Clemens et al. 1991; Longini et al. 2002). Second, there is considerable interest in the role played by environmental, primarily climatic, factors in driving the disease (Colwell 1996; Pascual et al. 2000; Rodó et al. 2002). This interest stems in large part from the growing recognition in the past decades that the causative agent of cholera, the bacterium *Vibrio cholerae*, inhabits aquatic environments such as brackish waters and estuaries. Environmental factors influencing the survival and growth of the bacterium in its environmental reservoir, such as temperature and salinity, have the potential to affect disease transmission. In fact, the disease is known to have a strong seasonal component attributed to temperature and monsoonal water level fluctuations because the bacterium *V. cholerae* is transmitted through fecal-oral contamination (Pascual et al. 2002). Furthermore, at the interannual timescales, there is evidence for a role of the El Niño Southern Oscillation (ENSO) in driving the variability of cholera in recent decades in Bangladesh. Interestingly, for Dhaka, this role appears to have intensified in recent times, with weak to undetectable effects for the historical period considered here (Rodó et al. 2002). This observation leaves the interannual variation of the historical period essentially unexplained. Finally, the likelihood of long-term decreases in transmissibility from 1892 to 1940 is high, implying extrinsic forcing of disease dynamics and the resulting nonstationarity of parameters.

Our analysis provides evidence for immunity in cholera that lasts several years. Results also show a long-term trend in the transmission rate with time but no clear evidence for interannual forcing of this parameter. These findings are discussed in the context of the epidemiological evidence for temporary immunity in cholera and in the con-

text of the recent proposal that the link between ENSO and cholera has intensified from historical to recent times (Rodó et al. 2002). Simulations of the fitted model with and without noise reveal an interesting interaction between the noise and the extrinsic feedbacks in the system. Noise appears essential to sustain the interannual cycles of the disease. We discuss the applicability of similar semiparametric approaches to differentiate between extrinsic forcing and intrinsic dynamics in interannual population patterns other than those of infectious diseases.

Methods

To isolate the relative contributions of intrinsic population dynamics from extrinsic forcing, we rely on a simple epidemic model that allows us to formulate both of these factors into a single difference equation. This transmission equation allows for temporary immunity as an intrinsic biological mechanism to be the cause of interannual variability. It also allows for such interannual variability to be caused by extrinsic changes in pathogen transmissibility. The transmission equation is a difference equation of the form

$$I_{t+1} = \beta_t I_t^{\alpha}\left(\frac{S_t}{N_t}\right)^{\gamma}\varepsilon_t, \tag{1}$$

where I is the number of infected individuals, S is the number of susceptible individuals, and N is the population size. The time step corresponds to the average time an individual remains infected (and infectious). An individual who is infected at time t was susceptible at time $t-1$ and will be recovered by time $t+1$. The equation states that the number of individuals infected at time $t+1$ is a function of the transmissibility of the pathogen at time t (β_t), the number of infected individuals at time t (I_t), the fraction of individuals susceptible at time t (S_t/N_t), and the multiplicative noise (ε_t). Similar difference equations describing the process of disease transmission have been used by Finkenstädt and Grenfell (2000).

Pathogen transmissibility over time (β_t) may exhibit cycles of longer-period and long-term trends in addition to seasonal fluctuations. These longer-term changes in transmissibility may result from public health measures altering the transmission probability of the pathogen, from changes in the pathogen's reproductive rate, from changes in contact rates over time, or from climatic variability. We therefore consider that pathogen transmissibility β_t is the product of two separate components: a seasonal one, β_{seas}, and a long-term one, β_{lt}, including both cycles of periods longer than a year and trends. Equation (1) becomes

$$I_{t+1} = \beta_{\text{lt}}\beta_{\text{seas}} I_t^{\alpha}\left(\frac{S_t}{N_t}\right)^{\gamma}\varepsilon_t, \tag{2}$$

where β_{seas} denotes a seasonally varying parameter with n distinct values for the number of time steps that comprise a year (Fine and Clarkson 1982; Finkenstädt and Grenfell 1998, 2000) and β_{lt} includes both cycles of periods longer than a year and trends.

The transmission mode of equation (2) is explicitly frequency dependent in that the number of infecteds at time $t+1$ is proportional to the fraction of susceptible individuals in the population at time t. This differs from the mass-action, or density-dependent, mode of transmission whereby the number of infecteds at time $t+1$ is dependent on the absolute number of susceptible individuals at time t (McCallum et al. 2001). However, our model is able to model disease dynamics that rely on either frequency-dependent or density-dependent transmission (or a combination thereof). The exponents α and γ are mixing parameters included to allow for deviations from the random-mixing assumption (Lui et al. 1987; Finkenstädt and Grenfell 2000).

The model is completed with a second equation for the number of susceptible individuals at time t. This number, S_t, can be expressed as the total population size minus the currently infected and recovered individuals still in the population (Cooke et al. 1977; Y. Xia, J. R. Gog, and B. T. Grenfell, personal communication). Thus,

$$S_t = N_t - \sum_{i=0}^{m} I_{t-i}\kappa_i, \tag{3}$$

where the function κ is the product of immunity and survivorship. The value κ_i gives the proportion of the population that was infected i time steps ago and remains immune and alive in the present and is therefore not part of the susceptible pool. The integer m is the minimum amount of time necessary for the function κ to reach 0 such that any individuals who were infected at least m time steps ago are either completely susceptible again or no longer present in the population.

Given only data on the number of infected individuals (I_t) and population size (N_t) over time, we now seek to recover the seasonal transmission rates (β_{seas}), the long-term transmission rates (β_{lt}), the mixing exponents α and γ, and the immunity function κ. To fit the model, we first log transform equation (2), which gives

$$\log(I_{t+1}) = \log(\beta_{\text{lt}}) + \log(\beta_{\text{seas}}) + \alpha\log(I_t) + \gamma\log\left(\frac{S_t}{N_t}\right) + \log(\varepsilon_t), \tag{4}$$

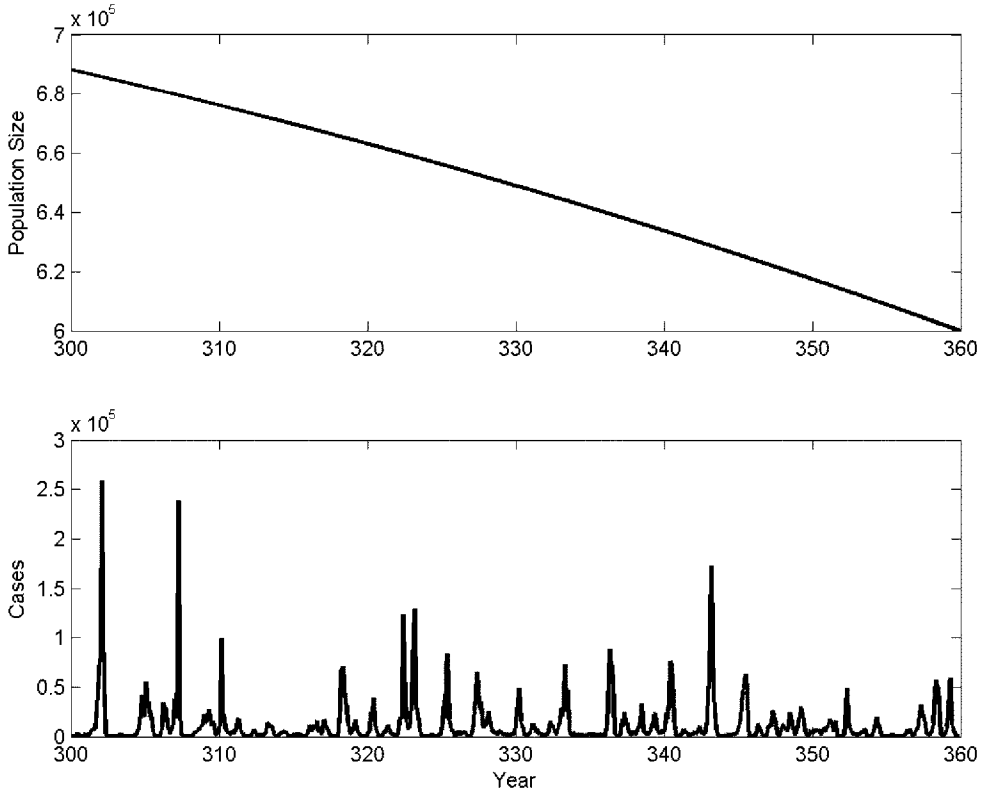

Figure 1: *Top*, population decrease from 688,223 to 600,000 over the 60 yr of simulated time series data. *Bottom*, simulated time series of infecteds over the same period. The model was simulated using multiplicative noise introduced at a level of $\varepsilon_t = \exp(0.6\zeta_t)$, where $\zeta_t \sim N(0, 1)$. The first 300 yr of transient simulated data were removed.

with $\log(\varepsilon_t)$ being Gaussian-distributed noise with mean 0. By initially assuming that the number of susceptibles S_t is close to the total population size N_t, we simplify this equation with a Taylor series expansion of $\log(S_t/N_t)$ to the first approximation. Then,

$$\log\left(\frac{N_t - \sum_{i=0}^{m} I_{t-i}\kappa_i}{N_t}\right) \approx \frac{-\sum_{i=0}^{m} I_{t-i}\kappa_i}{N_t}, \tag{5}$$

where equation (2) was used for the number of susceptibles. From equations (4) and (5), the transmission equation becomes

$$\log(I_{t+1}) \approx \log(\beta_{\text{lt}}) + \log(\beta_{\text{seas}}) + \alpha \log(I_t) - \frac{\gamma}{N_t}\sum_{i=0}^{m} I_{t-i}\kappa_i + \log(\varepsilon_t). \tag{6}$$

Because of the unspecified variation in β_{lt}, a simple parametric regression cannot be applied to fit this model. Instead, a semiparametric method is used, which consists of two main steps (Hastie and Tibshirani 1990). First, all parameters other than β_{lt} are obtained by a regression step using weighted least squares. This regression step is repeated a number of times, with a backfitting algorithm that progressively improves our estimate of the susceptible fraction in the population through the adjustment of the immunity function. The immunity estimates are smoothed with a spline under the constraint of decreasing values. Second, the values of β_{lt}, the parameter that is allowed to vary in an unspecified way with time, are obtained by smoothing the residuals of the regression step. The details of this semiparametric approach are described in the appendix. We also describe a two-step approach to obtain both a confidence set for the immunity kernel and pointwise confidence intervals for the transmissibility values and the mixing exponents. The first step addresses the uncertainty in the shape of the immunity function resulting from the spline fit. The second step assumes that the immunity function is correct, which allows the straightforward computation of pointwise standard error bands for the remaining parameters.

To test the performance of the proposed time series

method, we applied it to simulated data. The top of figure 1 illustrates the decreasing population size used in the simulations, while the bottom of figure 1 shows the time series of infected individuals derived from simulating equations (2) and (3). These two data sets are the only inputs to the analysis. Figure 2 shows the results of the model applied to the simulated data of figure 1. The fitted model has an r^2 value of 0.88. The method is able to recover the immunity function κ and the seasonal and long-term component of transmissibility. Confidence bands for long-term transmissibility confirm the qualitative pattern of interannual variability as well as its dominant period (fig. 2*C*). Finally, confidence intervals for both α and γ confirm the deviation from homogeneous mixing in the transmission process, with both exponents being less than, and different from, 1. While the value of α is accurately recovered, this is not the case for γ. However, from other simulations, we have found that the estimates of all other parameters, including long-term transmissibility, appear only weakly sensitive to the value of γ. Furthermore, the possibility of wider standard error bands for this exponent is discussed in the appendix.

We have assumed transmission to be frequency dependent. However, many diseases exhibit density-dependent transmission or a combination of both types (Roberts and Heesterbeek 1993; De Jong et al. 1995). Two ways in which our frequency-dependent model can be reformulated to explicitly accommodate for density-dependent transmission are detailed in the appendix.

Use of Surrogate Measures for Infected Time Series

Time series of infected individuals are often difficult to obtain directly. Instead, one may have data on hospitalizations due to the disease over time, or mortality estimates, or even severely underreported infected estimates. For any

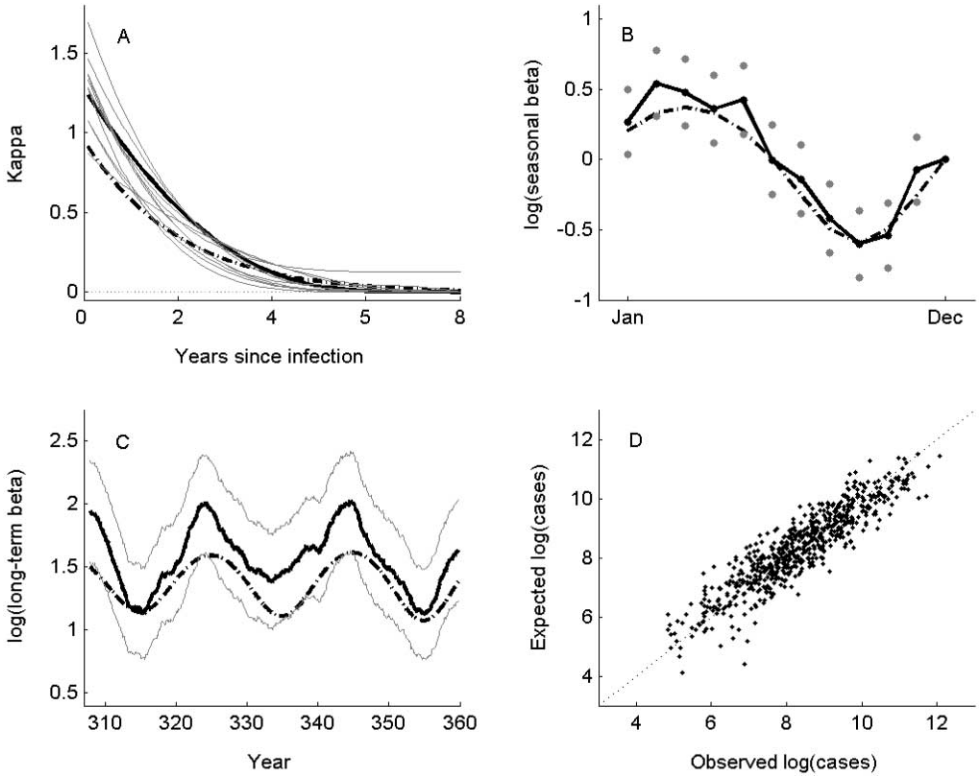

Figure 2: Parameter estimates (*bold black lines*) from the model fit to data from figure 1 alongside actual parameters used to generate the time series (*dashed black lines*). *A*, Immunity function κ. The actual κ is given by the product of survivorship $\text{surv}(i) = 0.95 - 0.0075i$ and immunity $\text{imm}(i) = \exp(-i/30)$. The 10 curves in gray are a subset of the global confidence set for κ created by bootstrapping. *B*, Logarithm of the seasonal transmissibility over the 12 mo comprising an annual cycle, ± 2 times the SE of the point estimates (*gray circles*). Actual seasonal transmissibility values are given by $\beta_{\text{seas}}(t) = 1.0 + 0.45 \sin(2\pi t_{\text{mod12}}/12)$. *C*, Estimated logarithm of the long-term transmissibility, ± 2 SE (*gray bands*), alongside the actual values. *D*, Observed log (cases) versus expected log (cases). The r^2 value for the fit is 0.88. Actual values of $(\alpha, \gamma) = (0.90, 0.95)$. Estimated values of $(\alpha, \gamma) = (0.8688, 0.6535)$, with the ± 2 SE interval range for $\alpha = [0.8262, 0.9113]$ and for $\gamma = [0.5432, 0.7637]$. Results shown use the optimal combination of smoothing bandwidth and penalty weight $(h, \mu) = (28, 10^{10})$, determined by cross-validation.

of these cases, we can still apply the proposed method to disentangle the contributions of extrinsic and intrinsic factors as long as the number of infected individuals is proportional to our surrogate measure. Let

$$I_t = cM_t, \tag{7}$$

where M_t is the surrogate measure of infected individuals and c is a proportionality constant. Substituting equation (7) into equation (1) and using the definition of the number of susceptible individuals from equation (3), we have

$$cM_{t+1} = \beta_t (cM_t)^{\alpha} \left(\frac{N_t - \sum_{i=0}^{m} cM_{t-i}\kappa_i}{N_t} \right)^{\gamma} \varepsilon_t. \tag{8}$$

Rearranging equation (8) into a form similar to that of equation (1), we have

$$M_{t+1} = B_t M_t^{\alpha} \left(\frac{N_t - \sum_{i=0}^{m} M_{t-i}K_i}{N_t} \right)^{\gamma} \varepsilon_t, \tag{9}$$

with $B_t = \beta_t c^{\alpha-1}$ and $K_i = c\kappa_i$. The model can then be used to fit the parameters α, γ, K, and B_t.

Application of the Full Model to Historical Cholera Dynamics

We apply the analysis to the monthly records of cholera mortality for the time period of 1892–1940 in Dhaka (fig. 3*B*). These data were previously extracted from the disease records by the medical examiners for former British India (Sanitary Commissioner's Reports and the Bengal Public Health Reports 1892–1940). The aggregation of infected individuals into monthly intervals is adequate for our purposes because the period of disease communicability is largely over within 3 wk after the acutely infected stage. The mortality data exhibit a clear annual cycle, with both a spring and a fall peak, which is well known in the literature and has been discussed in relation to monsoonal variation (Pascual et al. 2002). In addition to seasonality, there are also multiple interannual cycles present in the time series, as shown by the local wavelet power spectrum (fig. 3*C*). Both an 8-yr and a biennial cycle are evident toward the end of the time series, while interannual cycles of 4 yr are present intermittently throughout the entire time series.

The population size over this time period was reconstructed from census data taken every 10 yr, monthly birth data, and monthly death data. The resulting data show substantial population growth (fig. 3*A*), with a 72% increase over the 49-yr time period.

Besides ENSO, which represents a remote climate forcing, rainfall has been one local environmental covariate of interest in cholera because of its influence on environmental water levels, pathogen environmental concentration, and salinity. Figure 4*A* shows rainfall data for Dhaka extracted from the same historical records as cholera. To further motivate the application of the proposed model, we present the dominant frequencies and their localization in time for the rainfall data resulting from wavelet analysis (fig. 4*B*). Besides pronounced seasonality, weak interannual variability is detected at periods of 2 and 4 yr. There is, however, no interannual variability at the longer period of 8 yr observed in cholera. To go beyond scale matching, application of the model allows us to consider the alternative hypothesis that intrinsic dynamics and its interplay with seasonality play a role in generating these cycles.

The application of the model to the historical disease data yields some interesting results of the duration of temporary immunity and of the roles of extrinsic versus intrinsic factors (fig. 5). The mixing exponent $\alpha = 0.74$ is lower than 1, indicating a deviation from the homogeneous mixing assumption (fig. 5). The value of γ, at 0.97, is close to 1. The immunity function κ declines to 0 over a period of approximately 9 yr (fig. 5*A*). Because this function combines survivorship and temporary immunity, the exact waning period of immunity is expected to last no less than 9 yr. Because the disease data are deaths and not infected numbers, we can interpret the *Y*-axis intercept of this curve according to equation (9), where we can assign the value of 1 to κ_0 and therefore compute c as approximately 72. The interpretation is that approximately one out of every 72 cholera cases (including both symptomatic and asymptomatic cases) died from the disease. Representative curves from the 95% confidence set are typically similar in shape to the estimated function, with a fast initial decline followed by a plateau and long duration of immunity (above 7 yr). The seasonal transmissibility estimated by the model (fig. 5*B*) fits the known pattern for Dhaka; two peaks occur annually, in March and November, resulting in high levels of cholera mortality in April and December. The long-term transmissibility (fig. 5*C*) shows first a steady decline followed by a leveling off around 1922, with no indication of significant interannual variability.

The application to cholera, for which we have mortality instead of incidence data, required the assumption of a constant relationship between incidence and mortality. Although cholera mortality rates are known to have decreased in hospitals over this period, it is highly unlikely that the developing treatments reached the majority of the affected population at the beginning of the twentieth century (M. Bouma, personal communication). In fact, mortality rate estimates for this region, compiled by the World

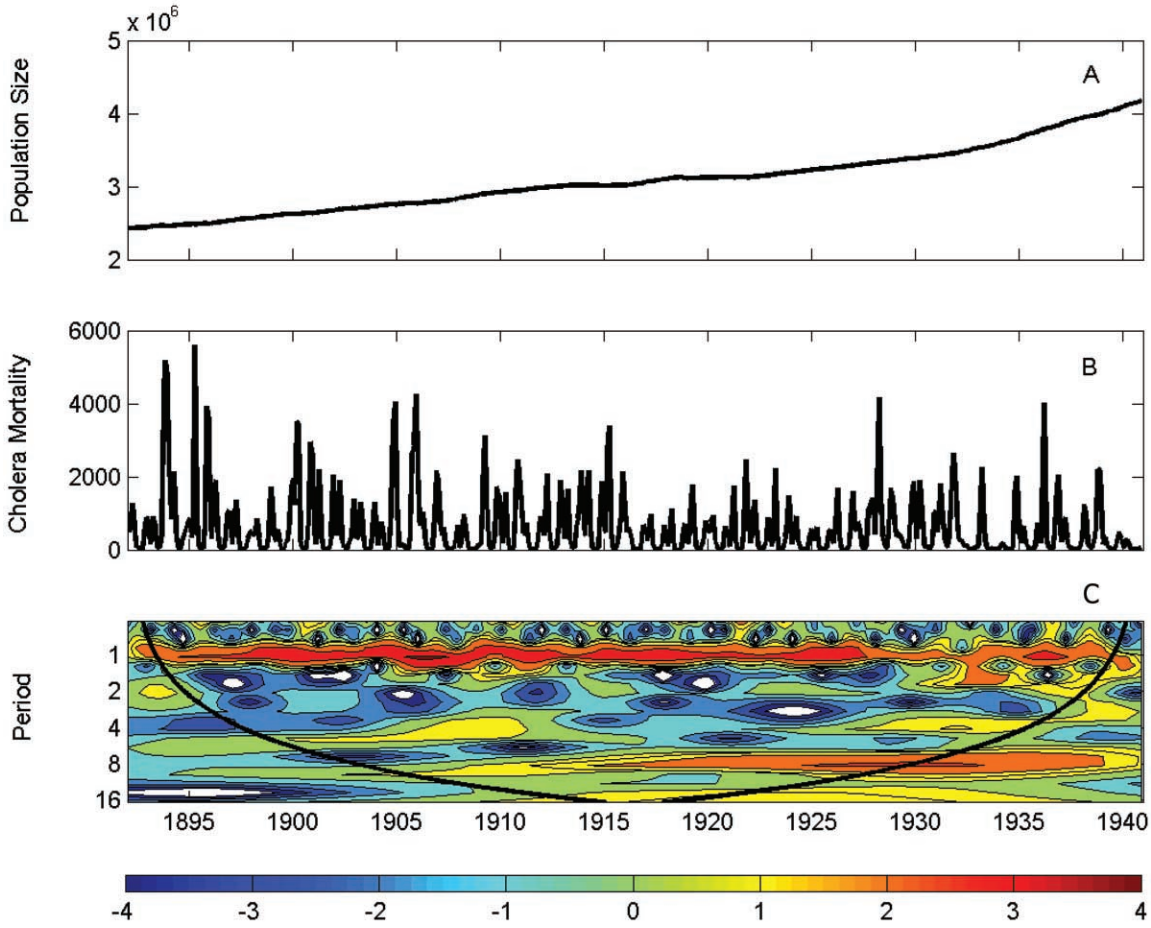

Figure 3: *A*, Population growth in the district of Dhaka over the time period 1892–1940, reconstructed from census, mortality, and birth records. We assumed census and mortality data to be accurate and adjusted the birth reporting rate such that census reports were met. Under this assumption, birth reporting rates between censuses were statistically determined to be 96.8% (1891–1901), 95.41% (1901–1911), 99.33% (1911–1921), 86.45% (1921–1931), and 71.56% (1931–1941). *B*, Monthly cholera mortalities in Dhaka over the same time period. *C*, Wavelet time series analysis for the Dhaka cholera mortality time series. (Wavelets are used to decompose the variance of a time series into different frequencies at different localities in time. Thus, contours of high intensity in the color scale indicate the presence and dominance of a particular period at a given time. For further explanation of wavelet analysis, see Torrence and Compo 1998; Grenfell et al. 2001 for an application to disease data.) Data were log transformed and detrended prior to wavelet analysis. The Morlet wavelet function was used. The logarithm of power is color coded as shown on the bottom bar. Wavelet software was provided by C. Torrence and G. Compo and is available at http://paos.colorado.edu/research/wavelets/.

Health Organization since 1950, indicate that the case fatality rates of symptomatic cholera infections were remarkably constant, at approximately 60% for nearly 2 decades.

The one-step-ahead model predictions, when compared with the actual cholera mortality data, give a good fit, with an r^2 value of 0.82 (fig. 5*D*). However, if we predict the number of cholera deaths further into the future, predictability decreases rapidly in 4 mo, leveling off to an r^2 value of close to 0.63 (not shown).

This rapid loss of predictability can be examined further by simulating the deterministic skeleton from the initial conditions for the entire time span of the data (fig. 6*A*). Given the initial mortality history (1892–1902), the population records, and the fitted parameters, we iterate the deterministic model for 39 yr, 1903–1941. The resulting simulation shows that the interannual variation in mortality is quickly lost through time. In fact, when iterated for extended periods of time with a constant interannual transmissibility and a constant population size, the deterministic skeleton generates a regular annual cycle with no interannual variability.

Interestingly, however, when dynamic noise is added to the simulation by sampling the residuals from the fit, in-

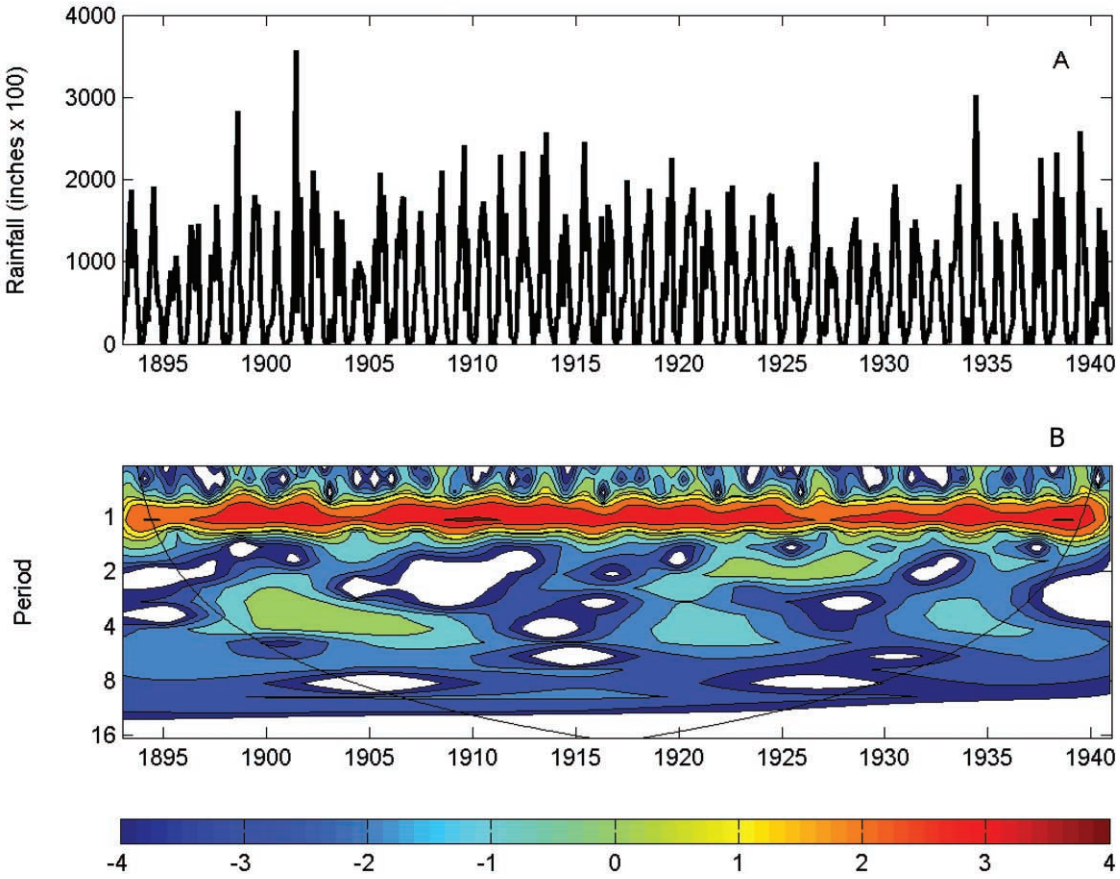

Figure 4: *A*, Monthly rainfall data from the district of Dhaka over the time period 1893–1940. *B*, Wavelet time series analysis for these data. The Morlet wavelet function was again used in the analysis.

terannual variability persists (fig. 6*B*, 6*C*). To determine whether interannual variability would be retained indefinitely, we simulated the mechanistic model with dynamic noise for extended periods of time, again using constant interannual transmission values and constant population sizes. Wavelet analyses of these simulated time series all exhibited sustained interannual variability (not shown). The noise appears to interact with the nonlinear deterministic dynamics generating multiple timescales of interannual variability. This interannual variability clearly differs from that obtained by simply adding measurement noise to the deterministic annual cycles, which does not generate significant interannual cycles.

One pattern observed in the cholera data is the change from an interannual cycle on the order of 4 yr to multiple, coexisting interannual cycles, including a strong 8-yr interannual cycle and evidence for both a biennial cycle and a weaker 4-yr cycle after 1930 (fig. 3*C*). In the stochastic simulations, the exact timing of the interannual frequencies varies with the specific sequence of noise added. However, the transition from one long interannual cycle to multiple coexisting cycles, including a more pronounced biennial cycle, is repeatedly seen in these stochastic simulations and seems to be a general feature that may result from the nonstationarity of population sizes and transmission coefficients (fig. 6*C*).

Discussion

We have presented a nonlinear time series model to identify the respective contributions of extrinsic forcing and intrinsic deterministic feedbacks in infectious disease dynamics. The approach further reconstructs the pattern of decaying immunity from time series data on cases and population sizes. Although recent work has been able to successfully estimate the duration of temporary immunity in seasonally forced disease systems (Y. Xia, J. R. Gog, and B. T. Grenfell, personal communication), our model further allows for interannual fluctuations and long-term changes in transmissibility. Consideration of these factors

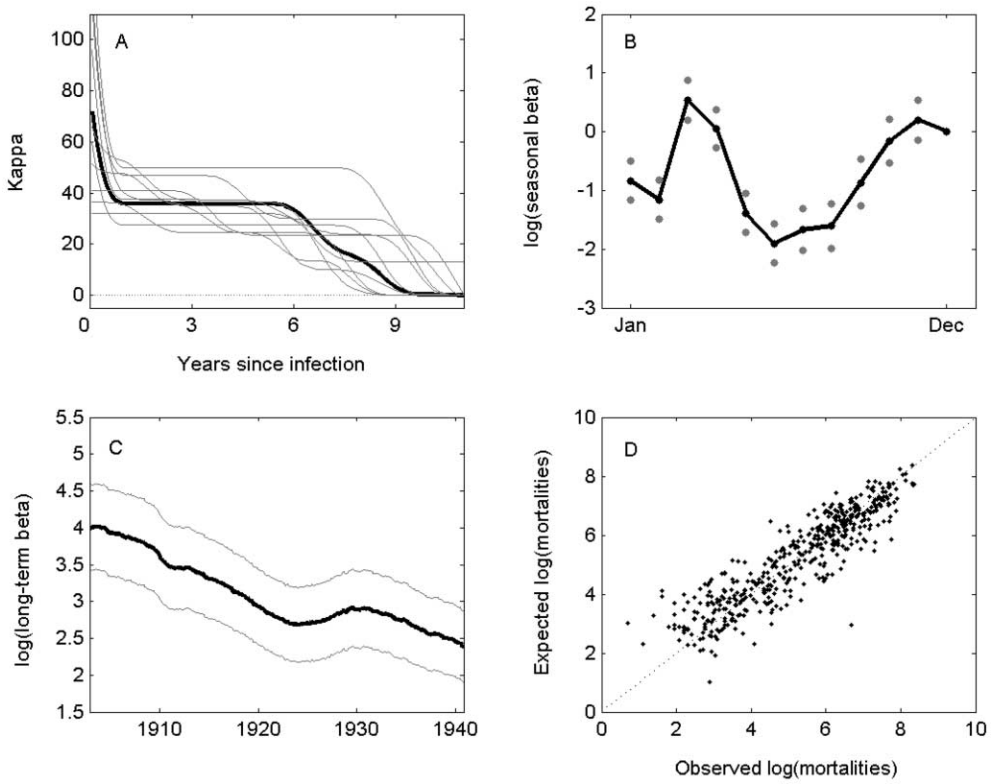

Figure 5: Parameter estimates (*bold black lines*) from the model fit to the Dhaka cholera data from figure 3. *A*, Immunity function κ over 11 yr. The 10 curves in gray are a subset of the global confidence set for κ created by bootstrapping. *B*, Logarithm of the seasonal transmissibility values over an annual cycle (*black*) ± 2 times the SE of the point estimates (*gray circles*). *C*, Logarithm of long-term transmissibility (*black*) ± 2 SE bands (*gray bands*). *D*, Observed log (mortalities) versus expected log (mortalities). The standard deviation of the residuals is 0.72. The r^2 value for the fit is 0.82. Estimated values of $(\alpha, \gamma) = (0.7374, 0.9684)$, with the ± 2 SE interval range for $\alpha = [0.6682, 0.8066]$ and for $\gamma = [0.7874, 1.1495]$. Results shown use the optimal combination of smoothing bandwidth and penalty weight $(h, \mu) = (47, 10^6)$, determined by cross-validation.

in the context of the nonlinear dynamics of disease becomes important to address the role of climate variability and climate change in interannual disease patterns (Pascual et al. 2000; Rogers et al. 2002). A similar semiparametric approach can be developed to address the role of extrinsic drivers in population patterns other than those of infectious diseases.

The application of our model to historical cholera data indicates that interannual cycles in the past can be attributed to the interplay of temporary immunity, seasonality, and noise. Changes in transmissibility over time are present only as a long-term trend and seasonality, with no clear role of extrinsic forcing at interannual timescales. This result is consistent with the recent report of a strong forcing signal by ENSO only in recent decades, which appears weaker or absent in historical times (Rodó et al. 2002). This difference between past and present has been interpreted as evidence for an effect of climate change. We specifically looked here for an ENSO signature in the model's residual noise but found no significant lagged or nonlagged correlation of the residuals with the Southern Oscillation Index, rainfall, sea surface temperature anomalies in the Pacific, or ENSO years derived from historical publications (Quinn et al. 1987). Furthermore, the residuals show no evidence of significant autocorrelation. There is, however, clear seasonal variation in transmissibility, with peaks in spring and late fall, as expected from the seasonality of the disease and from the environmental drivers proposed in the literature, specifically temperature and rainfall (see Pascual et al. 2002 for a review).

Our approach is meant to complement, and not to replace, other correlative or scale matching approaches to determine the role of environmental forcing. By addressing the criticism that no alternative explanation to extrinsic forcing is allowed for interannual cycles (Roger et al. 2002), application of the proposed model can reinforce or weaken

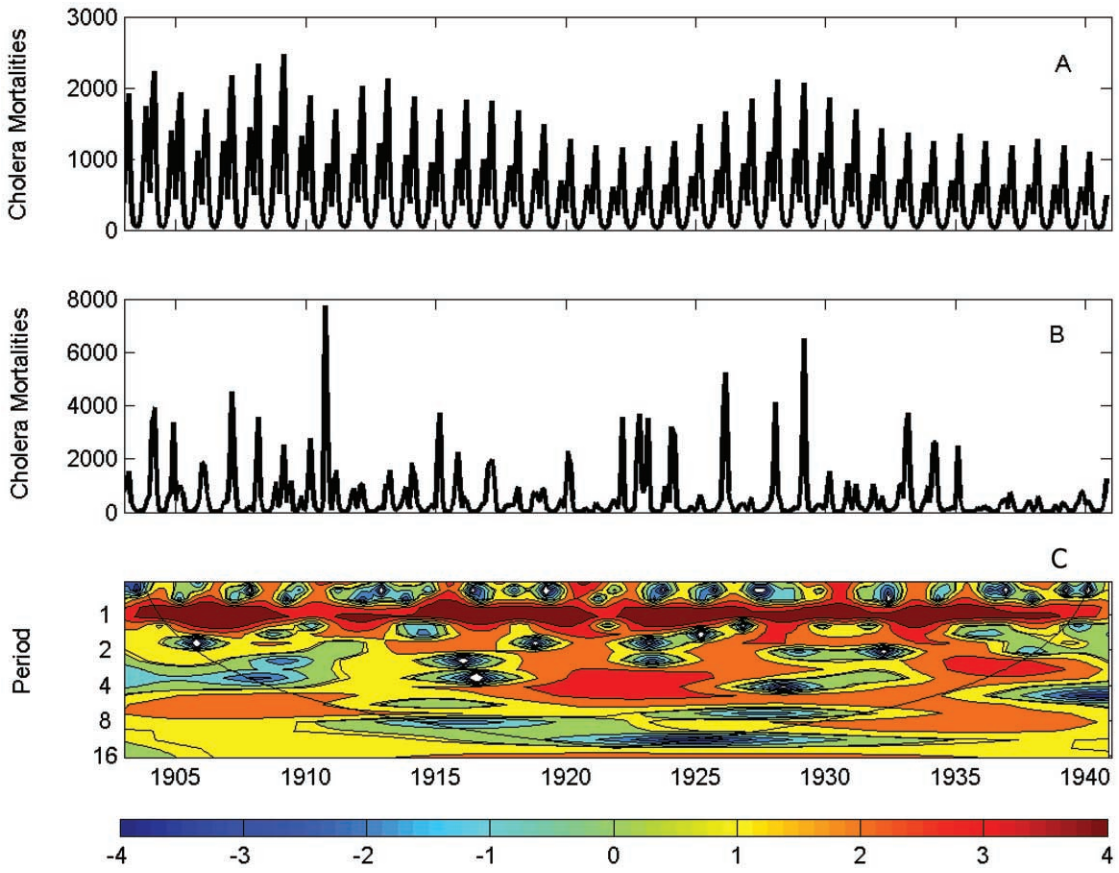

Figure 6: *A*, Deterministic skeleton for the cholera model with the fitted parameters shown in figure 5. *B*, Representative stochastic realization of the mechanistic cholera model with the same fitted parameters. Noise was added dynamically by sampling (with replacement) from the residuals of the model's fit. *C*, Wavelet time series analysis for the noisy deterministic skeleton shown in *B*. Data were log transformed and detrended before wavelet analysis, as for figure 3*C*. The Morlet wavelet function was used.

our conclusions from correlative evidence. Here, our results provide an explanation based on temporary immunity for the long cycles in cholera of approximately 8 yr, which are absent in the rainfall record and remain unexplained by other environmental covariates so far examined, such as ENSO. Temporary immunity and seasonality also appear able to account for the periods of 2 and 4 yr observed in both cholera and rainfall. One assumption of the proposed method is, however, the smooth variation of long-term transmissibility and, therefore, of the response of cholera to environmental variables. Threshold behavior would receive the use of a different type of time series model.

Although it has been known that individuals recently recovered from cholera experience some temporary immunity, quantitative estimates of the duration of immunity have not been previously available. Here, we find that partial immunity lasts no less than 9 yr. The exact degree to which an average individual is immune, however, is unknown because of the model's inability to separate immunity levels from survivorship. Field trials have shown that in Bangladesh, where cholera is endemic, the disease's case rate falls sharply with age, whereas in epidemic areas with no previous cholera occurrences, adults experience a higher incidence rate (Mosley et al. 1968). This pattern suggests that long-term immunity to cholera exists, and additional studies confirm this conjecture. In a 42-mo surveillance program, only seven out of 2,214 individuals were reinfected with cholera, which corresponded to a 61% lower incidence rate for reinfections than for primary infections (Clemens et al. 1991). In another field trial, only three reinfections occurred over a 9-yr period, compared with an expected 29 reinfections in the case of no temporary immunity (Glass et al. 1982). Protection against

classical cholera reinfection lasting at least 36 mo has also been observed in experimental rechallenge studies (Levine et al. 1981). These studies are consistent with our results that indicate a high degree of immunity for the first 5 yr following an infection and then a subsequent waning of immunity over the four subsequent years. Further complications in patterns of immunity arise from the existence of different strains in more recent times. We are currently extending the approach to analyze recent records in which two different biotypes are present.

In our simulations, the interaction of exogenous noise with the seasonally forced deterministic dynamics generates interannual variability with multiple temporal scales instead of the simple annual cycle observed in the absence of noise. It remains to be determined whether the more complicated patterns seen in the presence of noise result from the proximity to a bifurcation point and the related "ghost" of attractors that have lost stability. There are other examples in the ecological literature of the complex interactions that are possible between noise and nonlinear feedbacks (e.g., Higgins et al. 1997; Rohani et al. 1999, 2002). The cholera example shows that irregular cycles with multiple dominant frequencies are another possible outcome.

The model considers that transmission dynamics result primarily from short-term contacts between infected and susceptible individuals. The existence of an environmental aquatic reservoir for *Vibrio cholerae* introduces a possible alternative route of transmission through contaminated water in the environment. However, if the abundance of the pathogen in the reservoir experiences an important feedback from the levels of infection in the population, the treatment of transmission as in other infectious diseases is justified. Furthermore, recent evidence has uncovered heightened pathogen infectivity following passage of the pathogen in the human host (Merrell et al. 2002), which would reinforce the transmission feedback from infected to susceptible individuals.

The application of the semiparametric model to diseases with long-term immunity requires an extensive data set. This requirement is dictated primarily by the duration of immunity itself; data limitations clearly arise when the duration of temporary immunity becomes comparable to, or a significant portion of, data set length. Application of the model is also problematic when the prevalence of the disease in the population is low, such that the fraction of susceptible individuals in the population is close to 1, and when the dynamics are perfectly periodic. Sufficient interannual variability is necessary in order to successfully separate temporary immunity from extrinsic forcing. With simulated data, we have extensively examined conditions leading to the model's failure as well as signatures of this failure. When the approach fails to recover the underlying forcing and/or immunity pattern, the results of the fit are never biologically plausible. In other words, failure of the model is clearly recognizable. Particular signatures include a mixing exponent γ that does not converge and goes to large negative or positive values together with values of κ that converge to 0.

Another technical question is in regards to the convergence of our method, including a backfitting algorithm to a global optimum. Unique solutions have been shown for simpler semiparametric models (Hastie and Tibshirani 1990, p. 118). Although the convergence of our method remains to be analytically demonstrated, extensive simulation results suggest the existence of a global optimum. These results include modification of the initial κ function (obtained in the first step of the backfitting algorithm) by reducing its values by a fixed fraction. In this case, the method produces the same final immunity function up to a point where the initial values are too low and result in the pathology described above, of no convergence for γ together with values of κ approaching 0.

The model presented here uses a semiparametric approach. The parametric part of the model allows for a portion of the dynamics to be defined mechanistically, while the nonparametric part of the model allows for flexibility in long-term or interannual changes in a parameter of interest. Although the usefulness of semiparametric models has been recently underscored in the ecological literature (Ellner et al. 1998), their application to nonlinear dynamics that are driven by long-term or interannual variation has not been recognized (with the exception of a fisheries model; A. Solow, personal communication). For the relatively long time series of disease records, gradual changes in parameter values are inevitable. Many time series are therefore detrended before they are fit with fully parametric models. The semiparametric model is able to detrend such a time series without imposing a specific form for the trend (linear, quadratic, etc.) while allowing for the possible interplay of such a trend with the nonlinear dynamics.

Interannual variability with irregular patterns is common in ecological data. Ecological models have shown innumerable examples of complex responses to forcing, as nonlinearity allows the transfer of variability across temporal scales. While separating and identifying from data the factors responsible for such patterns appear to be daunting tasks, it is exactly the information contained in these highly irregular patterns that can make it possible.

Acknowledgments

We thank B. Grenfell and Y. Xia for sharing their work and ideas on time series models for disease dynamics with temporary immunity and for comments on the manu-

script. We also thank M. Bouma for providing the historical cholera and rainfall records and answering many questions on the data, A. Solow for advice on the semiparametric approach, S. Ellner for suggesting the use of splines with constraints, and two anonymous reviewers for their detailed comments. This work was conducted in part at the National Center for Ecological Analysis and Synthesis (NCEAS; Marine Diseases working group). We are grateful to D. Harvell and other participants of the working group for stimulating and interesting discussions at NCEAS, Papagallos, and Rocks. This research was supported by a National Oceanic and Atmospheric Administration grant (Joint Program on Climate Variability and Human Health, with Electric Power Research Institute/National Science Foundation/Environmental Protection Agency/National Aeronautics and Space Administration) and by a Centennial fellowship from the James S. McDonnell Foundation to M.P.

Literature Cited

Andrewartha, H. G., and L. C. Birch. 1954. The distribution and abundance of animals. University of Chicago Press, Chicago.

Clemens, J. D., F. Van Loon, D. A. Sack, M. R. Rao, F. Ahmed, J. Chakraborty, B. A. Kay, et al. 1991. Biotype as determinant of natural immunising effect of cholera. Lancet 337:883–884.

Colwell, R. 1996. Global climate and infectious disease: the cholera paradigm. Science 274:2025–2031.

Cooke, K. L., D. F. Calef, and E. V. Level. 1977. Stability or chaos in discrete epidemic models. Pages 73–93 *in* V. Lakshmikantham, ed. Nonlinear systems and applications: an international conference. Academic Press, New York.

Davidson, J., and H. G. Andrewartha. 1948. The influence of rainfall, evaporation and atmospheric temperature on fluctuations in the size of a natural population of *Thrips imaginis* (Thysanoptera). Journal of Animal Ecology 17: 200–222.

De Jong, M. C. M., O. Diekmann, and H. Heesterbeek. 1995. How does transmission of infection depend on population size? Pages 84–94 *in* D. Mollison, ed. Epidemic models: their structure and relation to data. Cambridge University Press, Cambridge.

Earn, D. J., P. Rohani, B. M. Bolker, and B. T. Grenfell. 2000. A simple model for complex dynamical transitions in epidemics. Science 287:667–670.

Ellner, S., and P. Turchin. 1995. Chaos in a noisy world: new methods and evidence from time-series analysis. American Naturalist 145:343–375.

Ellner, S., B. A. Bailey, G. V. Bobashev, A. R. Gallant, B. T. Grenfell, and D. W. Nychka. 1998. Noise and nonlinearity in measles epidemics: combining mechanistic and statistical approaches to population modeling. American Naturalist 151:425–440.

Fine, P. E., and J. A. Clarkson. 1982. Measles in England and Wales. I. An analysis of factors underlying seasonal patterns. International Journal of Epidemiology 11:5–14.

Finkenstädt, B. F., and B. T. Grenfell. 1998. Empirical determinants of measles metapopulation dynamics in England and Wales. Proceedings of the Royal Society of London B 265:211–220.

———. 2000. Time series modelling of childhood diseases: a dynamical systems approach. Applied Statistics 49: 187–205.

Girvan, M., D. Callaway, M. Newman, and S. Strogatz. 2002. Simple model of epidemics with pathogen mutation. Physical Review E 65:031915.

Glass, R. I., S. Becker, M. I. Huq, B. J. Stoll, M. U. Khan, M. H. Merson, J. V. Lee, and R. E. Black. 1982. Endemic cholera in rural Bangladesh, 1966–1980. American Journal of Epidemiology 116:959–970.

Grenfell, B. T., O. N. Bjornstad, and J. Kappey. 2001. Travelling waves and spatial hierarchies in measles epidemics. Nature 414:716–723.

Hastie, T. J., and R. J. Tibshirani. 1990. Generalized additive models. Chapman & Hall, London.

Hay, S. I., J. Cox, D. J. Rogers, S. E. Randolph, D. I. Stern, G. D. Shanks, M. F. Myers, and R. W. Snow. 2002. Climate change and the resurgence of malaria in the East African Highlands. Nature 445:905–909.

Hethcote, H. W., M. A. Lewis, and P. van der Driessche. 1989. An epidemiological model with a delay and a nonlinear incidence rate. Journal of Mathematical Biology 27:49–64.

Higgins, K., A. Hastings, J. N. Sarvela, and L. W. Botsford. 1997. Stochastic dynamics and deterministic skeletons: population behavior of Dungeness crab. Science 276: 1431–1435.

Levine, M. M., R. E. Black, M. L. Clements, L. Cisneros, D. R. Nalin, and C. R. Young. 1981. Duration of infection-derived immunity to cholera. Journal of Infectious Diseases 143:818–820.

Longini, I. M., M. Yunus, Y. K. Zaman, A. K. Siddique, R. B. Sack, and A. Nizam. 2002. Epidemic and endemic cholera trends over a 33-year period in Bangladesh. Journal of Infectious Diseases 186:246–251.

Lui, W. M., H. W. Hethcote, and S. A. Levin. 1987. Dynamical behavior of epidemiological models with nonlinear incidence rates. Journal of Mathematical Biology 25:359–380.

McCallum, H., N. Barlow, and J. Hone. 2001. How should pathogen transmission be modelled? Trends in Ecology & Evolution 16:295–300.

Merrell, D. S., S. M. Butler, F. Qadri, N. A. Dolganov, A. Alam, M. B. Cohen, S. B. Calderwood, G. K. Schoolnik, and A. Camilli. 2002. Host-induced epidemic spread of the cholera bacterium. Nature 417:642–645.

Mosley, W. H., A. S. Benenson, and R. Barui. 1968. A serological survey for cholera antibodies in rural east Pakistan. Bulletin of the World Health Organization 38: 327–334.

Nicholson, A. J. 1954. An outline of the dynamics of animal populations. Australian Journal of Zoology 2:9–65.

Pascual, M. 2001. Scales that matter: untangling complexity in ecological systems. Pages 255–286 *in* Carving our destiny: scientific research faces a new millennium. Commemorative volume, James S. McDonnell Centennial Fellowships. Joseph Henry, Washington, D.C.

Pascual, M., X. Rodó, S. P. Ellner, R. Colwell, and M. J. Bouma. 2000. Cholera dynamics and El Niño–Southern Oscillation. Science 289:1766–1769.

Pascual, M., M. J. Bouma, and A. P. Dobson. 2002. Cholera and climate: revisiting the quantitative evidence. Microbes and Infection 4:237–245.

Patz, J., M. Hulme, C. Rosenzweig, T. D. Mitchell, R. A. Goldberg, A. K. Githeko, S. Lele, A. J. McMichael, and D. Le Sueur. 2002. Climate change (communication arising): regional warming and malaria resurgence. Nature 420:627–628.

Pollitzer, R. 1959. Cholera. World Health Organization, Geneva.

Quinn, W. H., V. T. Neal, and S. E. Antunez De Mayolo. 1987. El Niño occurrences over the past four and a half centuries. Journal of Geophysical Research 92:14449–14461.

Roberts, M., and H. Heesterbeek. 1993. Bluff your way in epidemic models. Trends in Microbiology 1:343–348.

Rodó, X., M. Pascual, G. Fuchs, and A. S. G. Faruque. 2002. ENSO and cholera: a nonstationary link related to climate change? Proceedings of the National Academy of Sciences of the USA 99:12901–12906.

Rogers, D. J., S. E. Randolph, R. W. Snow, and S. I. Hay. 2002. Satellite imagery in the study and forecast of malaria. Nature 415:710–715.

Rohani, P., D. J. Earn, and B. T. Grenfell. 1999. Opposite patterns of synchrony in sympatric disease metapopulations. Science 286:968–971.

Rohani, P., M. J. Keeling, and B. T. Grenfell. 2002. The interplay between determinism and stochasticity in childhood diseases. American Naturalist 159:469–481.

Schwartz, I. B. 1992. Small amplitude, long period outbreaks in seasonally driven epidemics. Journal of Mathematical Biology 30:473–491.

Schwartz, I. B., and H. L. Smith. 1983. Infinite subharmonic bifurcations in an SEIR model. Journal of Mathematical Biology 18:233–253.

Sinclair, A. R. E. 1989. The regulation of animal populations. Pages 197–241 *in* J. M. Cherrett, ed. Ecological concepts. Blackwell, Oxford.

Torrence, C., and G. P. Compo. 1998. A practical guide to wavelet analysis. Bulletin of the American Meteorological Society 79:61–78.

Woodward, W. E. 1971. Cholera reinfection in man. Journal of Infectious Diseases 123:61–66.

Zimmer, C. 1999. Complex systems: life after chaos. Science 284:83–86.

Associate Editor: Benjamin M. Bolker

Supplemental material

Statistical procedure for fitting the semiparametric model

The steps described in this supplemental material were implemented as a set of Matlab (version 6.5). The total transmissibility values (β) contain a smooth long-term component (β_{lt}) and a seasonal component (β_{seas}) with *n* distinct values, where *n* is the number of time intervals that comprise a year. Since β takes unique values while its factors β_{lt} and β_{seas} do not, we fix the β_{seas} value of every data point in the n^{th} seasonal time interval at 1. In practice, to allow for seasonal variation in transmissibility, we introduce a vector Δ into equation (6) that specifies a given data point's seasonal occurrence by containing a one in the data point's seasonal time interval and zero otherwise. We let this vector Δ be of length *n-1*, such that $\log(\beta_{seas})_n = 0$. We allow $\log(\beta_{lt})$ to vary smoothly in time in an unspecified fashion by defining it non-parametrically. Equation (2.6) can then be written as:

$$\log(I_{t+1}) = \log(\beta_{lt})_t + \sum_{j=1}^{n-1} \log(\beta_{seas})_j \Delta_{j,t} + \alpha \log(I_t) - \gamma \frac{\sum_{i=0}^{m} I_{t-i}\kappa_i}{N_t} \qquad (2.S1)$$

Semi-parametric equations of this form can be solved explicitly (see Hastie and Tibshirani 1990), by performing a weighted least-squares regression on the parametric part, followed by smoothing the residuals from the regression fit to obtain the non-parametric part. Letting the row vector $X_t = [\Delta_t, \log(I_t), Z_t]$, where $\Delta_t = (\Delta_{1,t}, \Delta_{2,t}, \ldots \Delta_{n-1,t})$ and $Z_t = \left(-\frac{I_t}{N_t}, -\frac{I_{t-1}}{N_t}, \ldots -\frac{I_{t-m}}{N_t}\right)$, we solve for the parametric part of equation (2.S1):

$$\hat{\theta} = (X^T(I-W)X)^{-1} X^T(I-W)Y \qquad (2.S2)$$

where X is a matrix with rows X_t, Y is a column vector with row values $\log(I_{t+1})$, I is the identity matrix of size t x t, and W is a weight matrix of size t x t. We use a truncated

Gaussian kernel with a specified smoothing bandwidth *h* in the construction of the weight matrix W:

$$W_{ij} = \begin{cases} Ce^{-0.5\left(\frac{i-j}{h}\right)^2} & for\ |i-j| <= h \\ 0 & otherwise \end{cases} \qquad (2.S3)$$

with constant *C* chosen for each row such that the sum of the weights in each row equals 1. The resulting regression coefficients are $\hat{\theta} = [\log(\beta_{seas}), \alpha, \gamma\kappa]^T$, where $\log(\beta_{seas}) = (\log(\beta_{seas})_1, \log(\beta_{seas})_2, \ldots, \log(\beta_{seas})_{n-1})$, and $\gamma\kappa = (\gamma\kappa_0, \gamma\kappa_1, \ldots, \gamma\kappa_m)$. We set $\gamma=1$ as a first approximation, which allows us to compute the κ values from the fitted parameters $\gamma\kappa$. These κ values are then fitted with a penalized regression spline (Ruppert and Carroll 1997; Ellner et al. 2002) that is constrained to be monotonically decreasing and positive. The spline used here is a cubic regression spline:

$$\hat{\kappa}_i = f(X_i) + \varepsilon_i \qquad \text{where} \quad i = 0, 1 \ldots m \quad \text{and}$$

$$f(X) = \lambda_0 + \lambda_1 X + \lambda_2 X^2 + \lambda_3 X^3 + \sum_{j=1}^{k} \lambda_{3+j} \max(X - \eta_j, 0)^3 \qquad (2.S4)$$

where $\{\eta_j, j=1, 2, \ldots, k\}$ is the set of knots. We use $k=10$ knots. The fitted spline minimizes the quantity:

$$\sum_i (\hat{\kappa}_i - \kappa_i)^2 + \mu \sum_{j=1}^{10} \lambda_{3+j}^2 \qquad (2.S5)$$

where μ is the specified penalty weight. By requiring the vector of differences between the $\hat{\kappa}$ data points to be negative, the spline values are constrained to be monotonically decreasing. With this constraint in place, the requirement that the last data point be greater than or equal to 0 is sufficient to fulfill the positivity constraint for all other spline values. The problem of fitting the constrained spline is solved by mapping it into a standard quadratic programming problem (Wood 1997; Ellner et al. 2002).

The initial estimated number of susceptibles is then calculated as:

$$\hat{S}_t = N_t - \sum_{i=0}^{m} I_{t-i}\hat{\kappa}_i. \qquad (2.S6)$$

Since our initial Taylor series expansion (equation 2.5 in the main text) was only to the first approximation, and therefore rough, this estimate of the number of susceptibles, $\hat{S}$, deviates from the true number of susceptibles, S. This deviation must be due to errors in the estimated immunity kernel $\hat{\kappa}$. Let ψ be the necessary changes to the $\hat{\kappa}$ values that would bring $\hat{S}$ closer to S:

$$S_t = N_t - \sum_{i=0}^{m} I_{t-i}(\hat{\kappa}_i + \psi_i) = N_t - \sum_{i=0}^{m} I_{t-i}\hat{\kappa}_i - \sum_{i=0}^{m} I_{t-i}\psi_i \qquad (2.S7)$$

We can now use a backfitting algorithm to obtain better parameter estimates and to allow γ to deviate from its assigned value of 1. With this expression for S, equation (2.4) becomes:

$$\log(I_{t+1}) = \log(\beta_{lt})_t + \sum_{j=1}^{n-1} \log(\beta_{seas})_j \Delta_{j,t} + \alpha \log(I_t) + \gamma \log\left(\frac{\hat{S}_t - \sum_{i=0}^{m} I_{t-i}\psi_i}{N_t} \right) \qquad (2.S8)$$

A Taylor series expansion, now around $\frac{\hat{S}_t}{N_t}$, gives:

$$\log(I_{t+1}) \approx \log(\beta_{lt})_t + \sum_{j=1}^{n-1} \log(\beta seas)_j \Delta_{j,t} + \alpha \log(I_t) + \gamma \log\left(\frac{\hat{S}_t}{N_t} \right) - \frac{\gamma \sum_{i=0}^{m} I_{t-i}\psi_i}{\hat{S}_t} \qquad (2.S9)$$

This approximation is better than the one in equation (2.6), since $\left| \sum_{i=0}^{m} I_{t-i}\psi_i \right| << \left| \sum_{i=0}^{m} I_{t-i}\kappa_i \right|$. Equation (2.S9) is then solved explicitly through a weighted

least-squares regression, similar to (2.S2). By redefining the row vectors $X_t = \left[\Delta_t, \log(I_t), \log\left(\frac{\hat{S}_t}{N_t}\right), Z_t \right]$ with $Z_t = \left(-\frac{I_t}{\hat{S}_t}, -\frac{I_{t-1}}{\hat{S}_t}, \ldots -\frac{I_{t-m}}{\hat{S}_t} \right)$, we obtain

$$\hat{\theta} = (X^T (I - W) X)^{-1} X^T (I - W) Y = [\log(\beta_{seas}, \alpha, \gamma, \gamma\psi]^T \qquad (2.S10)$$

where $\gamma\psi = (\gamma\psi_0, \gamma\psi_1, \ldots, \gamma\psi_m)$. By dividing the fitted values $\gamma\Psi$ by the estimated γ, we compute the values Ψ. The improved immunity function estimate κ' can then be written as:

$$\kappa'_i = \hat{\kappa}_i + \psi_i \quad \text{for } i=0, 1\ldots, m. \qquad (2.S11)$$

These κ' values are then fitted with a constrained penalized regression spline as above, yielding the new immunity kernel estimate, $\hat{\kappa}'$. The next iteration begins with replacing $\hat{\kappa}$ with $\hat{\kappa}'$. We can repeat this backfitting algorithm, consisting of steps (2.S6) through (2.S11), until $\hat{\kappa}$ has converged.

The second step of the semi-parametric approach consists of obtaining the $\log(\beta_{lt})$ values from the resulting residuals by smoothing them with the W matrix:

$$\log(\beta_{lt}) = W(Y - X\hat{\theta}) \qquad (2.S12)$$

where X is a matrix with rows $X_t = \left[\Delta_t, \log(I_t), \log\left(\frac{\hat{S}_t}{N_t}\right) \right]$ and $\hat{\theta} = [\log(\beta_{seas}), \alpha, \gamma]^T$. Finally, the unique total transmissibility β values, for all time points *t*, can be calculated by:

$$\beta_t = \exp(\log(\beta_{lt})_t + \sum_{j=1}^{n-1} \log(\beta_{seas})_j \Delta_{j,t}). \qquad (2.S13)$$

Cross-validation: choosing the smoothing bandwidth and the penalty weight

Fitting the semi-parametric model requires the specification of a smoothing bandwidth h (used in the construction of the weight matrix W) and a penalty weight μ (used in the spline fit). To identify the combination of (h, μ) that neither over- nor under-smoothes the data, we use a cross-validation approach. The objective is to minimize the residual sum of squared differences between Y and the estimated cross-validated fit $\hat{Y}^{cv}$:

$$RSS = \sum_t (Y_t - \hat{Y}_t^{cv})^2 . \qquad (2.S14)$$

Our cross-validation estimate $\hat{Y}^{cv}$ differs from our fit $\hat{Y}$ in two respects: (1) The estimates of long-term transmissibility, $\log(\beta_{lt})$, are replaced with cross-validation estimates of long-term transmissibility, $\log(\beta_{lt})^{cv}$, and (2) susceptible estimates, $\hat{S}$, are replaced with cross-validation susceptible estimates, $\hat{S}^{cv}$.

To obtain $\log(\hat{\beta}_{lt})^{cv}$, which depends on h, we smooth the original residuals from the model fit ($Y - X\hat{\theta}$), with a matrix W^{cv} with elements

$$W_{ij}^{cv} = \begin{cases} Ce^{-0.5\left(\frac{i-j}{h}\right)^2} & for \quad 0 < | i - j | <= h \\ 0 & otherwise \end{cases} \qquad (2.S15)$$

with constant C again chosen for each row such that the sum of the weights in each row equals 1. The cross-validation matrix W^{cv} thereby does not include a data point in its own estimation, i.e. $W_{ii}^{cv} = 0$ for all i.

To obtain $\hat{S}^{cv}$, which depends on μ, we first fit a cross-validation spline $\hat{\kappa}^{cv}$ to the last κ', the estimate of the immunity function before the final constrained spline fit, $\hat{\kappa}$. This cross-validated spline is obtained using the same procedure as in (A4), with the exception that $\hat{\kappa}_i$ is computed from coefficients λ_0, λ_1, …, λ_{3+k} that were estimated without using X_i:

$$\hat{\kappa}_i = f^{-i}(X_i) + \varepsilon_i \qquad \text{where} \quad i = 0, 1, \ldots, m \quad \text{and}$$

$$f^{-i}(X^{-i}) = \lambda_0 + \lambda_1(X^{-i}) + \lambda_2(X^{-i})^2 + \lambda_3(X^{-i})^3 + \sum_{j=1}^{k} \lambda_{3+j} \max((X^{-i}) - \eta_j, 0)^3 . \quad (2.S16)$$

For all time points *t*, we then calculate $\hat{S}_t^{cv}$ as:

$$\hat{S}_t^{cv} = N_t - \sum_{i=0}^{m} I_{t-i} \hat{\kappa}_i^{cv} . \quad (2.S17)$$

The cross-validated estimate $\hat{Y}_t^{cv}$ can now be evaluated for all time points *t*:

$$\hat{Y}_t^{cv} = \log(\hat{\beta}_{lt})_t^{cv} + \sum_{j=1}^{n-1} \log(\beta_{seas})_j \Delta_{j,t} + \alpha \log(I_t) + \gamma \log\left(\frac{\hat{S}_t^{cv}}{N_t}\right) \quad (2.S18)$$

with α, γ, and the *n-1*. log(β_{seas}) parameter values taken from the $\hat{Y}$ fit that uses the same (h, μ) combination. The optimal combination of (h, μ) for obtaining $\hat{Y}$ is then the one that minimizes the cross-validated RSS.

Confidence intervals

To quantify the degree of uncertainty in our results, we use an approach consisting of two separate steps. We first use a bootstrapping approach to create a confidence region around the immunity kernel $\hat{\kappa}$. This region quantifies the uncertainty in the shape of the immunity kernel due to the spline fit. For the second step, we then assume that $\hat{\kappa}$ is accurate, which allows us to compute pointwise confidence intervals for the remaining parameters (i.e. transmissibility values and mixing exponents).

To create a confidence region around $\hat{\kappa}$, we start with the last κ', the estimate of the immunity function before the final constrained spline fit, $\hat{\kappa}$. The errors of this fit are $\varepsilon = \kappa' - \hat{\kappa}$. We generate κ'^*, a bootstrap sample of κ', by

$$\kappa'^{*}_{i} = \hat{\kappa}_i + \varepsilon^{*}_{i} \text{ for all } i = 0, 1, \ldots, m \qquad (2.S19)$$

where each ε^{*}_{i} is an error sampled with replacement from ε. We then fit a constrained spline to this κ'^{*}, yielding a bootstrapped $\hat{\kappa}^{*}$. After generating 1000 bootstrapped immunity kernels $\hat{\kappa}^{*}$, we choose to show 10 of them that lie in the 95% confidence set for $\hat{\kappa}$. Which curves lie in the 95% confidence set is established through generating a bootstrap analogue of the approximate pivotal, as explained in (Hastie and Tibshirani, p. 65).

Assuming that $\hat{\kappa}$ in the model's fit is accurate, we can use a method presented in Hastie and Tibshirani (p. 127) to generate standard-error bands for the *n-1* $\log(\beta_{seas})$ values, the mixing exponents α and γ, and $\log(\beta_{lt})$. The explicit solution to the parametric part of the semi-parametric model is given by:

$$\hat{\theta} = (X^T (I - W) X)^{-1} X^T (I - W) Y \qquad (2.S20)$$

where X is a matrix with rows $X_t = \left[\Delta_t, \log(I_t), \log\left(\frac{\hat{S}_t}{N_t} \right) \right]$ and $\hat{\theta} = [\log(\beta_{seas}), \alpha, \gamma]^T$. The estimated variance-covariance matrix of $\hat{\theta}$ is:

$$\sigma^2(\hat{\theta}) = \sigma^2 (X^T (I - W) X)^{-1} \qquad (2.S21)$$

where σ^2 is an estimate of the variance of the errors $r = Y - \hat{Y}$ (see Neter et al. 1983 (pp. 219-220) for weighted least-squares regression and Hastie and Tibshirani 1990 (p. 128) for estimating the semi-parametric model's degrees of freedom for error). The pointwise standard-error bands are obtained using +/- 2 times the square-root of the diagonal of this matrix.

The explicit solution to the nonparametric part of the semi-parametric model is $\log(\beta_{lt}) = W(Y - X\hat{\theta})$. By using equation (A20) for $\hat{\theta}$ and rearranging, we can write:

$$\log(\beta_{lt}) = \left[W(I - X(X^T(I-W)X)^{-1}X^T(I-W))\right]Y = RY \qquad (2.S22)$$

The covariance matrix of $\log(\beta_{lt})$ is then:

$$\mathrm{cov}(\log(\beta_{lt})) = RR^T\sigma^2. \qquad (2.S23)$$

Pointwise standard-error bands are then computed using +/- 2 times the square root of the diagonal elements of this covariance matrix.

This two-step approach may underestimate the magnitudes of the standard-error bands, since the fraction susceptible S/N is taken as given data as opposed to a fitted variable. This assumption may be one reason why, for the simulated dataset, the actual value of γ lies outside the +/- 2 standard-error range of the estimated value of γ.

An alternative approach would generate global confidence intervals by producing bootstrap samples of Y (the dependent variable), using resampled residuals r, and fitting the whole model to these samples. We chose the two-step approach, due to its tractability and ease of use, while being aware of its limitations.

Density-dependent versus frequency-dependent transmission

Consider that transmission is density-dependent, such that the dynamics are best modeled as

$$I_{t+1} = \beta_t I_t^{\alpha} S_t^{\gamma} \varepsilon_t \qquad (2.S24)$$

instead of as equation (2.1).

Multiplication of the left-hand side of the equation by $(N_t/N_t)^{\gamma}$ (obviously equal to 1) and a simple rearrangement of the variables gives

$$I_{t+1} = \beta_t N_t^{\gamma} I_t^{\alpha} \left(\frac{S_t}{N_t}\right)^{\gamma} \varepsilon_t \qquad (2.S25)$$

The first way to allow for density-dependent transmission is therefore to fit the frequency-dependent full model, and then interpret the parameter estimates according to equation (2.S25). The density-dependent β_t can therefore be calculated as:

$$\beta_t(dd) = \frac{\beta_t(fd)}{N_t^{\gamma}} \qquad (2.S26)$$

The absolute number of susceptible individuals (S_t) in the population at any time can also be calculated from the frequency-dependent results by multiplying the current fraction of susceptibles (S_t/N_t) by the current population size N_t. This conversion between frequency-dependent results and the density-dependent interpretation of them is valid as long as the change in N_t over time is smoother than the long-term transmissibility component in β_t.

A second way to allow for density-dependent transmission can be used when changes in N_t are not sufficiently smooth. Using equation (2.S25) as our density-dependent formulation, we can divide both sides by N_t^{γ}, with the initial assumption that γ is equal to 1. The dependent variable is now I_{t+1}/N_t^{γ}. As the backfitting algorithm described in Appendix A is reiterated, the value of γ is able to deviate from its initial value of 1. Therefore, for each subsequent backfit, the dependent variable in the density-dependent case (I_{t+1}/N_t^{γ}) needs to be recalculated. The final results are again density-dependent β_t estimates and the reconstructed susceptible population sizes.

Statistical procedure for fitting the parametric model

The semiparametric model presented above allows for fluctuations in the long-term transmission rate. A simpler model, which we constructed during the development of this semiparametric model, is one that assumes that there are no significant long-term or interannual changes in the transmission rate. In this case, we seek to reconstruct the fraction of susceptible individuals over time (via an immunity kernel), seasonal variation in the transmission rate, and the mixing exponents α and γ. As in the semiparametric model, we use as data inputs only information on population sizes and case dynamics. The parametric model has fewer degrees of freedom than the semiparametric model- an

advantage of the parametric model over the semiparametric model in the absence of interannual or long-term transmission rate changes. Fewer degrees of freedom would generate higher confidence in the shape of the fit immunity kernel. Here we briefly present the method for fitting this parametric model.

The transmission equation without interannual variability in β is:

$$I_{t+1} = \beta_{seas} I_t^{\alpha} \left(\frac{S_t}{N_t} \right)^{\gamma} \varepsilon_t \qquad (2.S27)$$

The expression for the fraction of susceptible individuals in terms of the present population size and the history of infected individuals is still equation 3.

Taking the logarithm of equation 2.S27 and substituting expression 2.5 for $\gamma \log\left(\frac{S_t}{N_t} \right)$, as for the semiparametric model, yields the statistical expression analogous to equation 2.S1 for the parametric model:

$$\log(I_{t+1}) = \sum_{j=1}^{n} \log(\beta_{seas})_j \Delta_{j,t} + \alpha \log(I_t) - \gamma \log\left(\frac{\sum_{i=0}^{m} I_{t-i} \kappa_i}{N_t} \right) \qquad (2.S28)$$

Note that in this case, the vector Δ is of length n, instead of n-1 for the semiparametric model. Defining the row vector $\mathbf{X}_t = [\Delta, \log(I_t), Z_t]$, where $\Delta_t = (\Delta_{1,t}, \Delta_{2,t}, \ldots, \Delta_{n,t})$ and $Z_t = (-I_t/N_t, -I_{t-1}/N_t, \ldots, -I_{t-m}/N_t)$, we solve for equation (2.S28) using a least squares fit:

$$\hat{\theta} = \left[X^T X \right]^{-1} X^T Y \qquad (2.S29)$$

where $\mathbf{X}$ is a matrix with rows $\boldsymbol{X}_t$, $\boldsymbol{Y}$ is a column vector with row values $\log(I_{t+1})$. The resulting regression coefficients are $\hat{\theta} = [\log(\hat{\beta}_{seas}), \alpha, \gamma\kappa]^T$, where $\log(\hat{\beta}_{seas}) = [\log(\hat{\beta}_{seas})_1, \log(\hat{\beta}_{seas})_2, \ldots, \log(\hat{\beta}_{seas})_n]$ and $\gamma\kappa = (\gamma\kappa_0, \gamma\kappa_1, \ldots, \gamma\kappa_m)$. As in the semiparametric approach, we set $\gamma = 1$ as a first approximation, which allows us to compute the κ values

from the fitted parameters γ_k. These κ values are then fitted with a constrained, cubic, penalized regression spline as described above in equations (2.S4) and (2.S5). The backfitting algorithm described for fitting the semiparametric model, consisting of re-iterations of equations (2.S6) through (2.S11) until convergence of the immunity kernel, is now used for fitting the parametric model. The backfitting algorithm used in fitting the parametric model differs slightly from the one used in fitting the semiparametric model: 1) In equations (2.S8) and (2.S9), the parametric model leaves out the long-term transmission rate term $\log(\beta_{lt})_t$. B) In equations (2.S8) and (2.S9), the dummy vector Δ used in the parametric model is of length n instead of n-1. C) In equation (2.S10), a least squares regression is used, as in equation (2.S29), instead of the weighted least squares regression used in the semiparametric model fit. Cross-validation is now used only to objectively choose the penalty weight μ for the immunity kernel spline. Confidence intervals for the parametric model's parameter estimates are generated similarly to the way they are generated for the semiparametric model's parameter estimates: we first obtain bootstrapped immunity kernels and then calculate the standard error bands for the seasonal transmission rates β_{seas} and the mixing exponents α and γ, assuming that the immunity kernel is correct.

References

Andrewartha, H. G., and L. C. Birch. 1954. The distribution and abundance of animals. University of Chicago Press, Chicago.

Clemens, J. D., F. Van Loon, D. A. Sack, M. R. Rao, F. Ahmed, J. Chakraborty, B. A. Kay, et al. 1991. Biotype as determinant of natural immunising effect of cholera. Lancet 337:883–884.

Colwell, R. 1996. Global climate and infectious disease: the cholera paradigm. Science 274:2025–2031.

Cooke, K. L., D. F. Calef, and E. V. Level. 1977. Stability or chaos in discrete epidemic models. Pages 73–93 *in* V. Lakshmikantham, ed. Nonlinear systems and applications: an international conference. Academic Press, New York.

Davidson, J., and H. G. Andrewartha. 1948. The influence of rainfall, evaporation and atmospheric temperature on fluctuations in the size of a natural population of *Thrips imaginis* (Thysanoptera). Journal of Animal Ecology 17:200–222.

De Jong, M. C. M., O. Diekmann, and H. Heesterbeek. 1995. How does transmission of infection depend on population size? Pages 84–94 *in* D. Mollison, ed. Epidemic models: their structure and relation to data. Cambridge University Press, Cambridge.

Earn, D. J., P. Rohani, B. M. Bolker, and B. T. Grenfell. 2000. A simple model for complex dynamical transitions in epidemics. Science 287:667–670.

Ellner, S., and P. Turchin. 1995. Chaos in a noisy world: new methods and evidence from time-series analysis. American Naturalist 145:343–375.

Ellner, S., B. A. Bailey, G. V. Bobashev, A. R. Gallant, B. T. Grenfell, and D. W. Nychka. 1998. Noise and nonlinearity in measles epidemics: combining mechanistic and statistical approaches to population modeling. American Naturalist 151:425–440.

Ellner, S., Y. Seifu, and R. H. Smith. 2002. Fitting population models to time-series data by gradient matching. Ecology 83:2256–2270.

Fine, P. E., and J. A. Clarkson. 1982. Measles in England and Wales. I. An analysis of factors underlying seasonal patterns. International Journal of Epidemiology 11:5–14.

Finkenstädt, B. F., and B. T. Grenfell. 1998. Empirical determinants of measles metapopulation dynamics in England and Wales. Proceedings of the Royal Society of London B 265:211–220.

Finkenstädt, B. F., and B. T. Grenfell. 2000. Time series modelling of childhood diseases: a dynamical systems approach. Applied Statistics 49:187–205.

Girvan, M., D. Callaway, M. Newman, and S. Strogatz. 2002. Simple model of epidemics with pathogen mutation. Physical Review E 65:031915.

Glass, R. I., S. Becker, M. I. Huq, B. J. Stoll, M. U. Khan, M. H. Merson, J. V. Lee, and R. E. Black. 1982. Endemic cholera in rural Bangladesh, 1966–1980. American Journal of Epidemiology 116:959–970.

Grenfell, B. T., O. N. Bjornstad, and J. Kappey. 2001. Travelling waves and spatial hierarchies in measles epidemics. Nature 414:716–723.

Hastie, T. J., and R. J. Tibshirani. 1990. Generalized additive models. Chapman & Hall, London.

Hay, S. I., J. Cox, D. J. Rogers, S. E. Randolph, D. I. Stern, G. D. Shanks, M. F. Myers, and R. W. Snow. 2002. Climate change and the resurgence of malaria in the East African Highlands. Nature 445:905–909.

Hethcote, H. W., M. A. Lewis, and P. van der Driessche. 1989. An epidemiological model with a delay and a nonlinear incidence rate. Journal of Mathematical Biology 27:49–64.

Higgins, K., A. Hastings, J. N. Sarvela, and L. W. Botsford. 1997. Stochastic dynamics and deterministic skeletons: population behavior of Dungeness crab. Science 276:1431–1435.

Levine, M. M., R. E. Black, M. L. Clements, L. Cisneros, D. R. Nalin, and C. R. Young. 1981. Duration of infection-derived immunity to cholera. Journal of Infectious Diseases 143:818–820.

Longini, I. M., M. Yunus, Y. K. Zaman, A. K. Siddique, R. B. Sack, and A. Nizam. 2002. Epidemic and endemic cholera trends over a 33-year period in Bangladesh. Journal of Infectious Diseases 186:246–251.

Lui, W. M., H. W. Hethcote, and S. A. Levin. 1987. Dynamical behavior of epidemiological models with nonlinear incidence rates. Journal of Mathematical Biology 25:359–380.

McCallum, H., N. Barlow, and J. Hone. 2001. How should pathogen transmission be modelled? Trends in Ecology & Evolution 16:295–300.

Merrell, D. S., S. M. Butler, F. Qadri, N. A. Dolganov, A. Alam, M. B. Cohen, S. B. Calderwood, G. K. Schoolnik, and A. Camilli. 2002. Host-induced epidemic spread of the cholera bacterium. Nature 417:642–645.

Mosley, W. H., A. S. Benenson, and R. Barui. 1968. A serological survey for cholera antibodies in rural east Pakistan. Bulletin of the World Health Organization 38:327–334.

Neter, J., M. Wasserman, and H. Kutner. 1983. Applied linear regression models. Irwin, Homewood, Ill.

Nicholson, A. J. 1954. An outline of the dynamics of animal populations. Australian Journal of Zoology 2:9–65.

Pascual, M. 2001. Scales that matter: untangling complexity in ecological systems. Pages 255–286 *in* Carving our destiny: scientific research faces a new millennium. Commemorative volume, James S. McDonnell Centennial Fellowships. Joseph Henry, Washington, D.C.

Pascual, M., X. Rodó, S. P. Ellner, R. Colwell, and M. J. Bouma. 2000. Cholera dynamics and El Niño–Southern Oscillation. Science 289:1766–1769.

Pascual, M., M. J. Bouma, and A. P. Dobson. 2002. Cholera and climate: revisiting the quantitative evidence. Microbes and Infection 4:237–245.

Patz, J., M. Hulme, C. Rosenzweig, T. D. Mitchell, R. A. Goldberg, A. K. Githeko, S. Lele, A. J. McMichael, and D. Le Sueur. 2002. Climate change (communication arising): regional warming and malaria resurgence. Nature 420:627–628.

Pollitzer, R. 1959. Cholera. World Health Organization, Geneva.

Quinn, W. H., V. T. Neal, and S. E. Antunez De Mayolo. 1987. El Niño occurrences over the past four and a half centuries. Journal of Geophysical Research 92:14449–14461.

Roberts, M., and H. Heesterbeek. 1993. Bluff your way in epidemic models. Trends in Microbiology 1:343–348.

Rodó, X., M. Pascual, G. Fuchs, and A. S. G. Faruque. 2002. ENSO and cholera: a nonstationary link related to climate change? Proceedings of the National Academy of Sciences of the USA 99:12901–12906.

Rogers, D. J., S. E. Randolph, R. W. Snow, and S. I. Hay. 2002. Satellite imagery in the study and forecast of malaria. Nature 415:710–715.

Rohani, P., D. J. Earn, and B. T. Grenfell. 1999. Opposite patterns of synchrony in sympatric disease metapopulations. Science 286:968–971.

Rohani, P., M. J. Keeling, and B. T. Grenfell. 2002. The interplay between determinism and stochasticity in childhood diseases. American Naturalist 159:469–481.

Ruppert, D., and R. J. Carroll. 1997. Penalized regression splines. Technical report TR1249. Department of Operations Research and Industrial Engineering, Cornell University, Ithaca, N.Y.

Schwartz, I. B. 1992. Small amplitude, long period outbreaks in seasonally driven epidemics. Journal of Mathematical Biology 30:473–491.

Schwartz, I. B., and H. L. Smith. 1983. Infinite subharmonic bifurcations in an SEIR model. Journal of Mathematical Biology 18:233–253.

Sinclair, A. R. E. 1989. The regulation of animal populations. Pages 197–241 *in* J. M. Cherrett, ed. Ecological concepts. Blackwell, Oxford.

Torrence, C., and G. P. Compo. 1998. A practical guide to wavelet analysis. Bulletin of the American Meteorological Society 79:61–78.

Wood, S. N. 1997. Inverse problems and structured-population dynamics. Pages 555–586 *in* S. Tuljapurkar and H. Caswell, eds. Structured-population models in marine, terrestrial, and freshwater systems. Chapman & Hall, New York.

Woodward, W. E. 1971. Cholera reinfection in man. Journal of Infectious Diseases 123:61–66.

Zimmer, C. 1999. Complex systems: life after chaos. Science 284:83–86.

View publication stats