8 Modeling Disease Dynamics: Cholera As a Case Study

EDWARD L. IONIDES [1], CARLES BRETÓ [1] AND AARON A. KING [2,3]

[1] Department of Statistics

University of Michigan, Ann Arbor, U.S.A.

[2] Department of Ecology and Evolutionary Biology

University of Michigan, Ann Arbor, U.S.A.

[3] Department of Mathematics

University of Michigan, Ann Arbor, U.S.A.

8.1 INTRODUCTION

Disease dynamics are modeled at a population level in order to create a conceptual framework to think about the spread and prevention of disease, to make forecasts and policy decisions, and to ask and answer scientific questions concerning disease mechanisms such as discovering relevant covariates. Population models draw on scientific understanding of component processes, such as immunity, duration of infection, and mechanisms of transmission, and investigate how this understanding

relates to population-level phenomena. There are several compelling reasons to consider disease processes at this population scale:

1. Anthropogenic change in land use, climate and biodiversity has many potentially large public health impacts (Aguirre et al., 2002). Predicting the future effects of changes to a complex system is difficult. Retrospective studies of the relationship between climate and disease prevalence over space (Greene et al., 2006) and over time (Rodó et al., 2002) can aid predictions and inform policy decisions (Kovats and Bouma, 2002). A major challenge in retrospective studies is to disentangle the extrinsic effects of climate or other environmental drivers from the intrinsic disease dynamics (Koelle and Pascual, 2004).

2. The effectiveness of medical treatment and vaccination strategies for certain infectious diseases, such as malaria and cholera, is limited by drug resistance, genetic shift, and poor medical infrastructure in affected regions. This leads to an emphasis of controlling the disease by behavioral and environmental interventions. An ability to model the disease dynamics can be used to forecast the danger of a major epidemic (Thomson et al., 2006), a step toward implementing effective interventions.

3. Emerging infectious diseases pose a significant public health threat. Many important emerging infectious diseases are zoonotic, i.e. endemic animal diseases which cross over to humans. Examples include HIV/AIDS from chimpanzee and sooty mangabey (Hahn et al., 2000), SARS from bats (Li et al., 2005) and avian flu (Longini et al., 2005). Epidemics are best prevented by early containment of outbreaks. Containment strategies may be evaluated

using population models (Longini et al., 2005). Alternatively, one can attempt to monitor and control the disease in the animal population to reduce contact between humans and infected animals. This can be assisted by employing population models to gain an understanding of the dynamics of the disease in the animal population.

Since the pioneering work of Ross (1916) and Kermack and McKendrick (1927), mathematical modeling has been a mainstay of epidemiological theory. It has also long been recognized that disease models arising in epidemiology are closely related to population models arising in ecology (Bartlett, 1960): the population dynamics of an infectious disease arise from the interaction of host and pathogen species in the context of their environment. This chapter explores some new developments in statistical inference for nonlinear dynamical systems from time-series data, using cholera in Bangladesh as a case study.

8.2 DATA ANALYSIS VIA POPULATION MODELS

A mainstay of population modeling is the compartment model, where the population is divided into groups which can be considered to be homogeneous. The classical SIR compartment model (Kermack and McKendrick, 1927; Bartlett, 1960) groups N_t individuals as susceptible (S_t), infected (I_t), and recovered or removed (R_t). Exposed classes, age-structured classes and geographically structured classes are just some of many possible extensions. Population models may use continuous or discrete time, take continuous or discrete values, and be stochastic or deterministic. Real world processes are continuous time, discrete valued, and stochastic. Stochasticity arises

from demographic noise (variability due to uncertainty of individual outcomes, such as the number of contacts made with an infected individual) and from environmental noise (such as variability due to weather, or economic events affecting the whole population). To a first approximation, demographic stochasticity has variance linear in population size and environmental stochasticity has variance quadratic in population size, though more subtle distinctions can be made (Engen et al., 1998). Models must also choose to be mechanistic or phenomenological, really a continuous scale trade-off between incorporating scientific understanding and aiming for a simple description of relationships observed in data (Ellner et al., 1998). Developing techniques that draw on understanding of population dynamics, while also permitting statistical inference about unknown model parameters and exploration of relevant covariates, is a topic of current research interest (Bjørnstad and Grenfell, 2001).

Data are often aggregated over time and space, such as weekly or monthly counts per region. This has led to the use of discrete-time models for data analysis. Finkenstädt and Grenfell (2000) and Koelle and Pascual (2004) represent the state of the art for data analysis via discrete-time mechanistic modeling, using a Taylor series to generate a log-linear model with unobserved variables reconstructed via back-fitting. There are several reasons to prefer continuous-time models:

1. For discrete-time models, the sampling frequency affects the models available and the interpretation of the resulting parameters. The underlying continuous-time processes are most naturally modeled in continuous time.

2. Continuous-time modeling facilitates the inclusion of covariates measured at various frequencies.

3. Continuous-time disease models have been studied much more extensively from the mathematical point of view than their discrete-time counterparts (Bailey, 1975; Anderson and May, 1991; Hethcote, 2000; Diekmann and Heesterbeek, 2000). This focus represents both that continuous-time models more accurately reflect the real properties of the systems and that such models are relatively easy to analyze. Most data analysis, on the other hand, has made use of discrete-time formulations, which can be fitted to discretely-sampled data in a relatively straightforward fashion. However, the dynamics of discrete-time nonlinear systems are frequently at odds with those of their continuous-time analogues (May, 1976; Glass et al., 2003), a fact which can complicate the interpretation of the parameters of discrete-time models.

Strategies appropriate for fitting continuous time models to discretely observed data include atlas methods (Turchin, 2003), gradient matching (Ellner et al., 2002), and approaches based on nonlinear forecasting (Kendall et al., 2005). Likelihood based analysis (frequentist or Bayesian) has largely been overlooked because finding the likelihood involves the difficult task of integrating out unobserved variables. Maximum likelihood estimates (MLEs) have some considerable advantages:

1. Statistical efficiency: The MLE is typically efficient (makes good use of limited data).

2. Transformation invariance: for example, estimates do not depend on whether the model is written using a log or natural scale.

3. Asymptotic results: the 2nd derivative of the log likelihood at its maximum can be used to give approximate standard errors. This means that simulations to understand variability in estimates are not usually necessary.

4. Model selection: likelihoods are comparable between different models for the same data. In particular, a χ^2 approximation is often appropriate: if p parameters are added to a model and the increase in the log likelihood is large compared to a $(1/2)\chi^2_p$ random variable then the fit is a statistically significant improvement.

Bayesian analysis is also attractive, since previous research may be available to provide an informed prior. Bayesian methods have been used for population models (Thomas et al., 2005; Clark and Bjørnstad, 2004). For this chapter we consider MLE methods, but the computational issue of integrating out unobserved variables arises in a similar way with Bayesian methods.

Evaluation of the likelihood and determination of the conditional distribution of unobserved variables given data are computationally approachable in a broad class of time series models known as state space models (SSMs). SSMs have been proposed as a unifying framework for ecological modeling (Thomas et al., 2005). Likelihood based inference has been shown to outperform other more ad-hoc statistical model fitting criteria for population models incorporating process noise and observation error (de Valpine and Hastings, 2002). The linear, Gaussian SSM (Kalman, 1960) became fundamental to engineering, for signal processing and control theory (Anderson and Moore, 1979), and found applications in economics (Harvey, 1989). Early attempts to handle nonlinear SSMs were plagued by the lack of computational ability to evaluate

the likelihood, so inference resorted to ad-hoc methods (Anderson and Moore, 1979). Brillinger et al. (1980) provides an early ecological application of nonlinear SSMs.

The development of Monte Carlo methods for nonlinear SSMs, combined with increases in computational capability, has made likelihood based inference feasible for increasingly general nonlinear SSMs. This gives the modeler considerable freedom to write down an appropriate model without undue concern for inferential feasibility. There are two main approaches to Monte Carlo inference for SSMs: Sequential Monte Carlo (Gordon et al., 1993; Doucet et al., 2001; Arulampalam et al., 2002) and Markov Chain Monte Carlo (Shephard and Pitt, 1997). This chapter focuses on Sequential Monte Carlo (SMC), which is more widely used for SSMs and simpler to implement. A careful comparison between SMC and Markov Chain Monte Carlo is still, to the authors' knowledge, an open problem.

8.3 SEQUENTIAL MONTE CARLO

An SSM is a partially observed Markov process. The unobserved Markov process, x_t, called the *state process*, takes values in a *state space*, $\mathcal{X}$. The *observation process* y_t takes values in an *observation space*, $\mathcal{Y}$, and y_t is assumed to be conditionally independent of the past given x_t. Here, we take $\mathcal{X}$ to be $\Re^{d_x}$ and $\mathcal{Y}$ to be $\Re^{d_y}$. There is also a vector of unknown parameters $\theta \in \Re^{d_\theta}$. We suppose that observations take place at discrete times, $t = 1, \ldots, T$. We further suppose that all required densities exist, and we adopt a convention that $f(\cdot \mid \cdot)$ is a generic density which is then specified by its arguments. We write concatenated observations as $y_{1:t} = (y_1, \ldots, y_t)$. For the case $t = 0$, $y_{1:0}$ is defined to be an empty vector. The properties

of a state space model are

$$f_\theta(x_t|x_{1:t-1}, y_{1:t-1}) = f_\theta(x_t|x_{t-1}) \quad (8.1)$$

$$f_\theta(y_t|x_{1:t}, y_{1:t-1}) = f_\theta(y_t|x_t) \quad (8.2)$$

The dependence on θ will be written explicitly only when necessary for clarity. In principle, the assumed Markov structure in (8.1) and (8.2) allows the likelihood, $f_\theta(y_{1:T})$, to be found recursively via the identities

$$f(x_t|y_{1:t-1}) = \int f(x_{t-1}|y_{1:t-1}) f(x_t|x_{t-1})\, dx_{t-1}, \quad (8.3)$$

$$f(x_t|y_{1:t}) = \frac{f(x_t|y_{1:t-1}) f(y_t|x_t)}{\int f(x_t|y_{1,t-1}) f(y_t|x_t)\, dx_t}. \quad (8.4)$$

$$f(y_t|y_{1:t-1}) = \int f(y_t|x_t) f(x_t|y_{1:t-1})\, dx_t \quad (8.5)$$

$$f(y_{1:T}) = \prod_{t=1}^{T} f(y_t|y_{1:t-1}) \quad (8.6)$$

In practice, this requires solving potentially challenging integrals. Following Kitagawa (1987), de Valpine and Hastings (2002) showed how these integrals could be solved numerically for relatively simple population models. For more complex models, one may employ an SMC method such as Algorithm 1.

Algorithm 1. Sequential Monte Carlo (SMC).

Initialize: Let $\{X^F_{0,j}, j = 1, \ldots, J\}$ be a sample draw from $f(x_0)$. These J realizations are commonly termed "particles". Each particle will give rise to a trajectory through the state space with distribution $f(x_t|y_{1:t})$.

FOR $t = 1$ to T

- **Move particles according to unconditional state process:**

  Make $X^P_{t,j}$ a draw from $f(x_t|x_{t-1}{=}X^F_{t-1,j})$. Then $\{X^P_{t,j}\}$ has approximate marginal distribution $f(x_t|y_{1:t-1})$. $\{X^P_{t,j}\}$ is said to solve the prediction problem at time t.

- **Calculate conditional likelihood of new observation:**

  Estimate $f(y_t|y_{1:t-1})$ by $(1/J)\sum_{j=1}^{J} f(y_t|x_t = X^P_{t,j})$.

- **Prune particles according likelihood given data:**

  Generate $X^F_{t,j}$ by resampling from $\{X^P_{t,j}\}$ with probability proportional to $w_j = f(y_t|x_t{=}X^P_{t,j})$ using Algorithm 2 (below). $\{X^F_{t,j}\}$ has approximate marginal distribution $f(x_t|y_{1:t})$. $\{X^F_{t,j}\}$ is said to solve the filtering problem at time t.

END FOR

Calculate log likelihood: $\log f(y_{1:T}) = \sum_{t=1}^{T} \log f(y_t|y_{1:t-1})$.

Algorithm 2. Systematic Resampling.

Input: J "particles" $\{X^P_{t,j}, j = 1, \ldots, J\}$ with weights $\{w_j = f(y_t|x_t{=}X^P_{t,j})\}$

Calculate cumulative sum of normalized weights:

FOR $j = 1$ to J set $c_j = \big(\sum_{k=1}^{j} w_k\big)/\big(\sum_{k=1}^{J} w_k\big)$

Resample cumulative sum at intervals of $1/J$:

Set $i = 1$ and $u \sim U[0, 1]$

FOR $j = 1$ to J

- WHILE $(j - u)/J > c_i$ set $i = i + 1$
- Set $X^F_{t,j} = X^P_{t,i}$. This resampling generates a tree structure, where $X^F_{t,j}$ is said to descend from $X^F_{t-1,i}$.

END FOR

Output: J "particles" $\{X^F_{t,j}, j = 1, \ldots, J\}$

The reader is referred to Arulampalam et al. (2002); Doucet et al. (2001); Liu (2001) for extensive discussions of Algorithm 1 and 2, with many possible variations. There are many ways that Algorithm 1 can be fine-tuned to be more computationally efficient. A more critical issue, in the authors' opinion, is how to use the output of Algorithm 1 for effective inference. Although Algorithm 1 is widely applicable for calculating the likelihood at a fixed value of θ, complications arise for both Bayesian and MLE methods, which must compare likelihoods for different values of θ.

Bayesian inference might appear straightforward: simply add θ to the state space. The initial particles are then drawn from $f(x_0, \theta)$ and the particle filter will then

produce a sample from $f(\theta|y_{1:T})$. Each particle at time t has exactly the same value of θ as its ancestor at time $t-1$, and the prior distribution on θ is updated via the SMC algorithm giving particles with successful values of θ more descendants. The catch is that the SMC algorithm degenerates when there is no variability in the θ component of the state process after $t = 0$. Heuristically, the particles in SMC evolve by natural selection according to their plausibility given the data. Particles whose θ component are fixed over time are analogous to natural selection without mutation, which produces only limited scope for evolution. One solution to this is to allow the parameter to vary slowly with time by adding noise (Kitagawa, 1998). If this modification to the model is considered unacceptable, Liu and West (2001) showed how to add noise to the parameters but balance this by simultaneously contracting the parameter distribution toward its mean. The method of Liu and West (2001) has been applied to ecological models by Thomas et al. (2005); Newman and Lindley (2006).

The difficulty for finding the MLE is that the likelihood is calculated with Monte Carlo error. One useful tool for optimizing functions calculated via Monte Carlo is the method of common random numbers (Spall, 2003, Section 14.4), which involves fixing the seed of the random number generator. This method requires synchronization of the Monte Carlo randomness, which is not directly applicable to SMC techniques. General stochastic optimization methods of the Robbins-Monro type (Robbins and Monro, 1951; Kiefer and Wolfowitz, 1952; Spall, 2003) are not applicable for problems where there are many unknown parameters and each function evaluation is a considerable computational expense. The elegant method of Hürzeler and Künsch (2001) for calculating local likelihood surfaces is also not readily ap-

plicable to relatively difficult problems—it is more computationally intensive than standard SMC methods such as Algorithm 1. Ionides et al. (2006) showed how to find the MLE by taking a limit where the noise, added in a similar way to Kitagawa (1998), shrinks to zero. This novel method is described in Algorithm 3 and is applied to a cholera population model in Section 8.4.

Algorithm 3. MLE via iterated filtering.

Initialize: Select $\theta^{\mathrm{hi}} > \theta^{(1)} > \theta^{\mathrm{lo}}$ to be vectors giving a plausible initial value and range for the parameters. Select scalars $0 < \alpha < 1$, C and N.

FOR $n = 1$ to N

- **Apply SMC** (Algorithm 1) with θ included in the state space as a time-varying parameter, evolving as

$$\begin{aligned} \theta_0 &\sim N_{d_\theta}(\theta^{(n)}, C\Sigma_n) \\ \theta_t|\theta_{t-1} &\sim N_{d_\theta}(\theta_{t-1}, \Sigma_n) \quad \text{for} \quad t = 2, \ldots, T, \end{aligned}$$

  where the covariance matrix Σ_n is defined by $[\Sigma_n]_{ii}^{1/2} = [(\theta^{\mathrm{hi}} - \theta^{\mathrm{lo}})/2\sqrt{T}]_i\, \alpha^{n-1}$ and $[\Sigma_n]_{ij} = 0$ for $i \neq j$. Each "particle" is now a pair, e.g. $(X_{t,j}^F, \theta_{t,j}^F)$.

- **Calculate updated estimate:**

$$\begin{aligned} \hat{\theta}_t &= (1/J)\sum_{j=1}^{J} \theta_{t,j}^F \quad \text{for } 1 \le t \le T \\ V_1 &= (C+1)\Sigma_n \\ V_{t+1} &= \Big(\sum_{j=1}^{J} (\theta_{t,j}^F - \hat{\theta}_t)(\theta_{t,j}^F - \hat{\theta}_t)^\tau\Big)/(J-1) + \Sigma_n \\ &\qquad \text{for } 1 \le t \le T-1 \\ \theta^{(n+1)} &= V_1\Big(\sum_{t=1}^{T-1} (V_t^{-1} - V_{t+1}^{-1})\,\hat{\theta}_t + V_T^{-1}\hat{\theta}_T\Big) \end{aligned}$$

END FOR

The MLE is estimated as $\hat{\theta} = \theta^{(N+1)}$

Algorithm 3 is appropriate when information about parameters arrives steadily throughout a time-series. Heuristically, it gains computational efficiency because the parameter estimate is being constantly updated throughout each iteration. Each iteration would correspond to one evaluation of the likelihood for a general purpose optimization algorithm. In Section 8.4, $N = 20$ or $N = 30$ iterations are sufficient to optimize a stochastic function of 13 variables, without availability of analytic derivatives. This computational efficiency is critical when each iteration takes around 30 minutes to compute.

In certain situations, such as estimating the initial value vector x_0, information about a parameter does not arrive steadily throughout a time-series. In this case, Algorithm 3 is not effective. If $\{x_t\}$ is stationary then x_0 can be drawn from the stationary distribution. If $\{x_t\}$ is not stationary, one can either pick some more arbitrary distribution for x_0 or treat x_0 as an unknown parameter (in the frequentist sense). We choose to do the latter, and estimate x_0 by maximum likelihood simultaneously with θ by applying Algorithm 4, which has a similar theoretical justification to Algorithm 3 (Ionides et al., 2006). The value of T_0 in Algorithm 4 should be as small as possible such that $y_{T_0+1:T}$ contains negligible additional information about x_0, beyond that contained in $y_{1:T_0}$. This compromise is known as fixed lag smoothing (Anderson and Moore, 1979).

Algorithm 4. MLE via iterated filtering, for initial values.

Initialize: Select $x_0^{\rm hi} > x_0^{(1)} > x_0^{\rm lo}$ to be vectors giving a plausible initial value and range for the initial values. Select scalars $0 < \alpha < 1$, T_0 and N.

FOR $n = 1$ to N

- **Apply SMC** (Algorithm 1) with $X_{0,j}^F \sim N_{d_x}(x_0^{(n)}, \Phi_n)$ where $[\Phi_n]_{ii}^{1/2} = [(x_0^{\rm hi} - x_0^{\rm lo})/2]_i \, \alpha^{n-1}$ and $[\Phi_n]_{ik} = 0$ for $i \neq k$. For each particle $X_{t,j}^F$, track label of the corresponding initial value, denoted $a(t,j)$. In the terminology of Algorithm 2, $X_{t,j}^F$ descends from $X_{0,a(t,j)}^F$.
- **Calculate updated estimate:** $x_0^{(n+1)} = (1/J) \sum_{j=1}^J X_{0,a(T_0,j)}^F$

END FOR

The MLE is estimated as $\hat{x}_0 = x_0^{(N+1)}$

Algorithms 3 and 4 are different variations on the same theme of using limiting Bayesian posterior distributions to find maximum likelihood estimates. Both algorithms can be combined, so that one filtering iteration updates estimates of all estimated parameters, including initial value parameters.

8.4 MODELING CHOLERA

Cholera is a diarrheal disease endemic to the Ganges delta region (Sack et al., 2004). Global pandemics have occurred throughout recent history. The current (7th) pandemic started in 1960 and has seen the O1 serogroup become established in various locations throughout south Asia, Africa and South America. Cholera is caused by virulent strains of *Vibrio cholerae,* a bacterium that can live and grow in

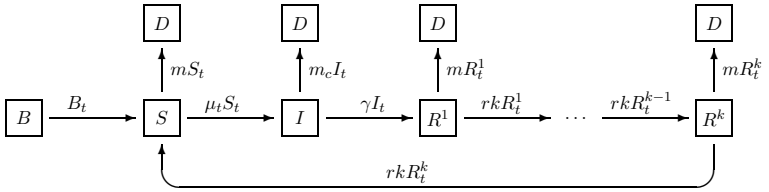

Fig. 8.1 Compartment model for cholera. Each individual is in S (susceptible), I (infected) or one of the classes R^j (recovered). Transitions to B and D model birth and death respectively. The arrows show possible transitions, with superscripts showing transition rates.

brackish, warm water. Human-to-human transmission can be direct, through contact with stool from infected individuals, or indirect, via the environment. There is not a clear distinction between these two paths: we separate them by supposing that the increase in force of infection depending on the number of infected individuals is due to human-to-human transmission. The environmental reservoir is taken to be responsible for the background force of infection (extrapolating to a situation with no infected humans). A compartment model describing the basic features of disease transmission is shown diagrammatically in Figure 8.1. Formally, the diagram in Figure 8.1 corresponds to a set of equations,

$$\begin{aligned}
dS_t &= dN_t^{BS} - dN_t^{SI} - dN_t^{SD} + dN_t^{R^k S} \\
dI_t &= dN_t^{SI} - dN_t^{IR^1} - dN_t^{ID} \\
dR_t^1 &= dN_t^{IR^1} - dN_t^{R^1 R^2} - dN_t^{R^1 D} \\
&\vdots \\
dR_t^k &= dN_t^{R^{k-1} R^k} - dN_t^{R^k S} - dN_t^{R^k D}
\end{aligned}$$

Here, time is measured in months; S_t is the number of individuals in class S (susceptible) and the infinitesimal dS_t is defined such that $S_t = S_0 + \int_0^t dS_u$. For example, N_t^{SI} corresponds to the total number of individuals who have passed from S to I by time t. The k recovered classes allow for flexibility in modeling the time from infection to loss of immunity, at which point an individual becomes newly susceptible. This temporary immunity, with a duration of 3–10 years, is believed to be a key feature of the population dynamics of cholera. We use a model developed by Ionides et al. (2006):

$$\begin{aligned} dN_t^{SI} &= \mu_t S_t\, dt + \psi_t S_t\, dW_t \\ \mu_t &= \beta_t I_t / P_t + w \\ \psi_t &= e I_t / P_t \end{aligned} \tag{8.7}$$

Here, population size P_t is interpolated from available census data, and is presumed to be accurately known; seasonal transmissibility is modeled as $\log(\beta_t) = \sum_{j=0}^{5} b_j s_j(t)$ where $\{s_j(t), j = 0, \ldots, 5\}$ is a periodic cubic B-spline basis; e is an *environmental stochasticity* parameter, modeling noise on the environmental scale (with infinitesimal variance proportional to S_t); w corresponds to a non-human *reservoir* of disease; $\beta_t I_t / P_t$ is *human-to-human* infection; $1/\gamma$ gives mean time to recovery; $1/r$ is the mean time to loss of immunity following recovery, with k giving the shape of this distribution; m and m_c are the death rates among uninfected and infected individuals respectively. The remaining transition equations were modeled

deterministically:

$$\begin{array}{rclrcl} dN_t^{IR^1} &=& \gamma I_t\,dt; & dN_t^{R^{j-1}R^j} &=& rkR_t^{j-1}\,dt; \\ dN_t^{R^kS} &=& rkR_t^k\,dt; & dN_t^{SD} &=& mS_t\,dt; \\ dN_t^{ID} &=& m_cI_t\,dt; & dN_t^{R^jD} &=& mR_t^j\,dt; \\ dN_t^{BS} &=& dP_t + dN_t^{SD} + dN_t^{ID} + \sum_{j=1}^k dN_t^{R^jD} & & & \end{array} \tag{8.8}$$

Defining $C_t = N_t^{ID} - N_{t-1}^{ID} = \int_{t-1}^t dN_s^{ID}$, the number of cholera mortalities between monthly observation times, the data on observed mortality were modeled conditional on C_t as $y_t \sim \mathcal{N}[\rho C_t, \rho(1-\rho)C_t + \tau^2\rho^2C_t^2]$ with reporting rate ρ. The variance component $\rho(1-\rho)C_t$ models demographic stochasticity via binomial sampling variation. Environmental stochasticity is modeled via $\tau^2\rho^2C_t^2$, which dominates demographic variability for large C_t and is found to be appropriate when fitting (8.7) and (8.8) to data. The dominance of environmental stochasticity has been assumed implicitly in previous analyses of similar data, by modeling additive noise of variance τ^2 in $\log(\rho C_t)$ (Finkenstädt and Grenfell, 2000; Koelle and Pascual, 2004). Demographic variability is non-negligible when C_t is small, and can be included in our framework without adding any additional parameters.

Continuous-state population models, such as the model given by (8.7) and (8.8), are more convenient for data analysis than discrete-state population models. Theoretical results and simulation studies of population models often resort to demographic (Poisson) variability, using the rates in Figure 8.1 to define a continuous-time Markov chain. Apart from the inherent appropriateness of discrete populations, the Markov chain approach has the advantage that no extra parameters, beyond the rates, are needed to describe the stochasticity. However, demographic stochasticity alone is not always sufficient to describe observed variations in data; for cholera, demographic stochasticity is entirely inadequate. If extra variability has to be introduced,

stochastic differential equations (SDEs) provide a simple way to do this. SDEs are a natural extension to the ordinary differential equation (ODE) systems already used for describing population dynamics. Other examples of the use of SDEs to provide a framework for modeling and data analysis include Kendall (1974); Brillinger and Stewart (1998); Brillinger et al. (2002); Ionides et al. (2004). There are several misconceptions about SDEs that explain why they are not currently more widely used for modeling. These are listed below, with refutations:

1. The theory of SDEs is inaccessible and obscure. However, numerical solution of SDEs is now well established (Kloeden and Platen, 1999; Higham, 2001). This allows development and exploration of models that would be hard to investigate analytically. In particular, application of the inference methodology in Algorithms 3 and 4 for the model in (8.7) and (8.8) requires only numerical solution of the system of SDEs.

2. There may be little reason to think that Gaussian white noise is a plausible stochastic driver for the system under investigation. Supplying random coefficients to an ODE or Markov chain adds lower frequency "colored noise." However, most practical time series models, such as the ARMA framework (Shumway and Stoffer, 2000), use white noise as the basic building block. This noise is often modeled as Gaussian, for convenience, and the data may sometimes be transformed to increase the plausibility of this assumption. Solutions to SDEs driven by Gaussian white noise include almost all non-Gaussian continuous time, continuous sample path Markov processes (Øksendal, 1998).

Smooth, low frequency noise can be modeled by adding white noise to a derivative of the process of interest.

3. Much discussion has occurred in theoretical modeling literature concerning different possible interpretations of an SDE. The two most popular interpretations are the Itô and Stratonovich solutions (Øksendal, 1998). The distinction, involving the exact way the SDE is solved as a limit of finite sums, should have little scientific relevance. Meaningful scientific conclusions should not depend on the choice of interpretation of SDE (Ionides et al., 2004). Numerical solution is most straightforward for the Itô solution, so that is the one adopted here.

8.4.1 Fitting structural models to cholera data

Maximizing a nonconvex function of more than a few variables is seldom routine, especially when the function is evaluated by Monte Carlo methods. Algorithm 3 provides a way to leverage the special structure of an SSM for optimization, but diagnostic checks are necessary before one has confidence in the results. Beyond the standard approach of trying various initial values ($\theta^{(1)}$, θ^{lo} and θ^{hi}) one should assess the choice of the two variables α and C for Algorithm 3. If α is too small, the rapid decrease in step size in Algorithm 3 may leave the algorithm stranded, unable to reach the maximum. This is analogous to excessively rapid cooling in simulated annealing (Spall, 2003). If α is too large, insufficient cooling will occur within a reasonable computation time. These issues can be diagnosed by plotting $\theta^{(n)}$ against n for several values of α and $\theta^{(1)}$, looking for consistent convergence. C

is a dimensionless constant controlling the initial dispersion of the parameter values, relative to their random perturbations through time. If C is too small, the algorithm converges slowly. If C is too large, the algorithm is less stable and converges erratically. This can be assessed by the same type of convergence plot used for α, or by the observation that a good choice of C is one which makes V_t fairly stable as a function of t.

The likelihood surface near the convergence point, $\hat{\theta}$, can be further examined by "sliced likelihood" plots. Setting $\lambda(\theta) = \log f_\theta(y_{1:T})$, the sliced likelihood for θ_i plots $\lambda(\hat{\theta} + c\delta_i)$ against $\hat{\theta}_i + c$, where δ_i is a vector of zeros with a one in the i^{th} position. If $\hat{\theta}$ is at (or near) the maximum of each sliced likelihood plot then $\hat{\theta}$ is (approximately) a local maximum of $\lambda(\theta)$. Computing sliced likelihoods requires moderate computational effort, linear in the dimension of θ. A smoothed fit (as suggested by Ionides, 2005) is made to the sliced log likelihood, because $\lambda(\hat{\theta} + c\delta_i)$ is calculated with a Monte Carlo error. Figure 8.2 shows a convergence and sliced likelihood plot for a simulation study, presented in Ionides et al. (2006), using the cholera model in (8.7) and (8.8). The deviation between the MLE and the true parameter value is due to the finite length (50 years) of the simulated dataset. In some generality, the MLE for state space models is consistent and asymptotically normally distributed (Jensen and Petersen, 1999).

Sliced likelihoods can be used to generate standard errors, since calculating $\lambda(\hat{\theta} + c\delta_i)$ involves finding $\log f_{\hat{\theta}+c\delta_i}(y_t|y_{1:t-1})$. Regressing $\log f_{\hat{\theta}+c\delta_i}(y_t|y_{1:t-1})$ on c gives an estimate of $(\partial/\partial\theta_i) \log f_{\hat{\theta}}(y_t|y_{1:t-1})$, giving rise to an estimate of the

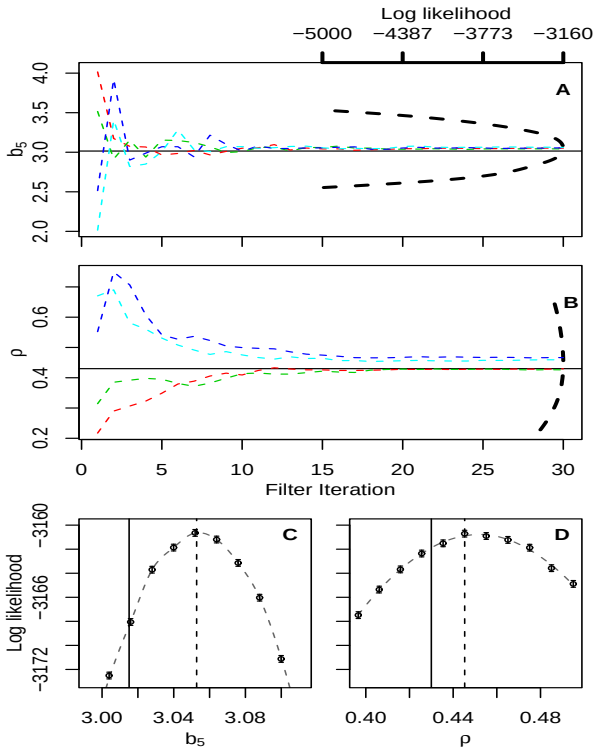

Fig. 8.2 (A, B) Examples of convergence plots for a simulation from (8.7) and (8.8) with four different starting points, validating the convergence of Algorithms 3 with $\alpha = 0.9$ and $C = 20$. The dotted parabolic line corresponds to a sliced likelihood through $\hat{\theta}$. (C, D) Corresponding closeups of the sliced likelihood. The dashed vertical line is at $\hat{\theta}$ and the solid vertical line is at the true value of θ. The simulation was carried out with $\rho = 0.43$; $e = 0.289$; $b_0 = -1.48$; $b_1 = 2.42$; $b_2 = 0.02$; $b_3 = -0.98$; $b_4 = 0.02$; $b_5 = 3.02$; $\tau^2 = 0.02$; $w = 2.5 \times 10^{-6}$; $m_c = 1.19$; $\gamma = 1$; $k = 4$; $1/r = 120$; $1/m = 600$. The last four of these parameters were treated as known, and the remaining parameters were estimated, using Algorithm 3 with $J = 9000$.

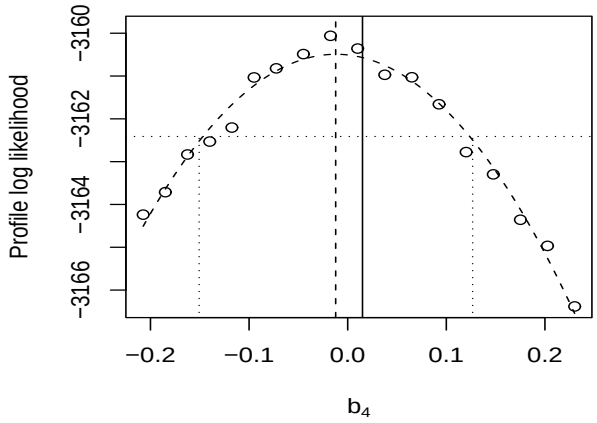

Fig. 8.3 Profile log likelihood $\lambda_{(p)}(b_4))$ for the August seasonal parameter. The log likelihood was maximized over all parameters excluding b_4 (circles) and was then smoothed (dashed line) using non-parametric regression (Ionides, 2005; Cleveland et al., 1993). The dotted lines show the construction of an approximate 95% confidence interval, given by $\{b_4 : 2[\lambda_{(p)}(\hat{b}_4) - \lambda_{(p)}(b_4)] < \chi^2_{0.95}(1)\}$ where $\chi^2_{0.95}(1)$ is the 0.95 quantile of a χ^2 random variable with one degree of freedom and $\hat{b}_4 = \text{argmax}\ \lambda_{(p)}(b_4)$.

observed Fisher information

$$[\hat{\mathcal{I}}_F]_{ij} = \sum_{t=1}^{T} (\partial/\partial\theta_i) \log f_\theta(y_t|y_{1:t-1})(\partial/\partial\theta_j) \log f_\theta(y_t|y_{1:t-1}) \qquad (8.9)$$

where the derivatives are evaluated at $\theta = \hat{\theta}$. This leads to a corresponding estimate $\hat{\mathcal{I}}_F^{-1}$ for the covariance matrix of $\hat{\theta}$.

A superior way to find confidence intervals is via a profile likelihood (Barndorff-Nielsen and Cox, 1994). If θ is partitioned into two components ζ and η then the profile log likelihood of η is defined (Barndorff-Nielsen and Cox, 1994) by $\lambda_{(p)}(\eta) = \sup_\zeta \lambda(\zeta, \eta)$. The optimization required for the profile likelihood can be

carried out using Algorithm 3. Calculating the profile likelihood for each parameter therefore requires approximately N times the computational effort of the sliced likelihood (typically, N is between 20 and 30). The optimization also introduces additional Monte Carlo variability over a simple likelihood evaluation. Figure 8.3 shows the profile likelihood for a parameter of the model in (8.7) and (8.8). This parameter was selected because the profile likelihood confidence interval constructed in Figure 8.3, of width 0.27, was considerably different from the approximation using (8.9), of width 0.10. This rather large discrepancy arose because the quadratic approximation in (8.9) is overly optimistic when some nonlinear combination of the parameters is poorly estimable. The extra computation required to calculate a profile likelihood is evidently worthwhile for a parameter of particular interest. The quadratic approximation can be calculated more routinely, to get a general idea of the scale of uncertainty.

The model in (8.7) and (8.8) was fitted to historical data for Dhaka, Bangladesh (Bouma and Pascual, 2001; Rodó et al., 2002; Koelle and Pascual, 2004), shown in Figure 8.4A. Our resulting estimate of the seasonal transmissibility β_t is shown in Figure 8.4B. Observed mortality is seen to have two seasonal peaks which appear later than the peaks in transmissibility. The winter dip in mortality has been ascribed to reduced environmental viability of *V. cholerae* in colder temperatures. The early January local minimum in transmissibility is consistent with the early January minimum in mean temperature in Dhaka. The summer dip in mortality has been ascribed to dilution of *V. cholerae* due to monsoon rainfall. The monsoon season in Dhaka is May to September, with greatest average rainfall in July. Fitting (8.7) and (8.8),

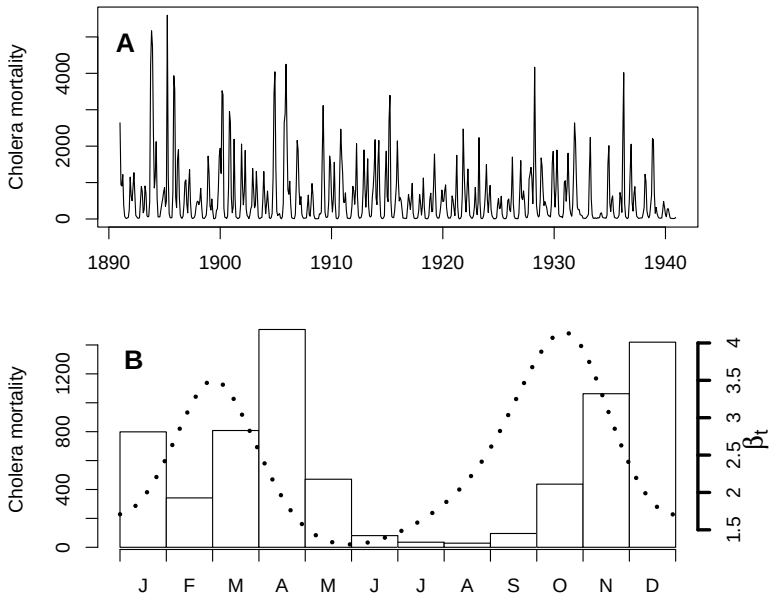

Fig. 8.4 (A) Cholera mortality for Dhaka, Bangladesh, from 1891 to 1940. (B) Monthly averages of Dhaka cholera mortality (boxes) and the seasonal transmissibility β_t (dotted line) from fitting (8.7) and (8.8).

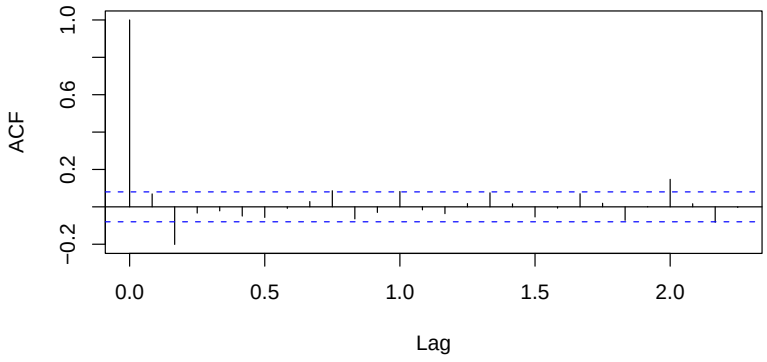

Fig. 8.5 Sample autocorrelation function for the standardized residuals when fitting (8.7) and (8.8) to the data in Figure 8.4.

the transmissibility is seen to decrease too soon to be explained fully by rainfall. Snow-melt from the Himalayas is one candidate to explain this discrepancy.

Investigating residuals is a routine diagnostic check in time series and regression analysis. The most basic residuals to consider for SSMs are the standardized prediction residuals,

$$u_t(\hat{\theta}) = [\mathrm{Var}_{\hat{\theta}}(y_t|y_{1:t-1})]^{-1/2}\big(y_t - E_{\hat{\theta}}(y_t|y_{1:t-1})\big),$$

though there are other possibilities (Ionides et al., 2006; Durbin and Koopman, 2001). Checking whether the residuals are approximately uncorrelated is a way to test the goodness of fit of the model. Residuals also have an important role in the search for covariates. Inasmuch as the model successfully captures the intrinsic dynamics of the disease, the residuals are left with the system noise plus signal from the extrinsic variables, such as climate. From this point of view, features that the intrinsic model

cannot capture are as important as those it can! A more flexible model might fit the data better, but only by explaining variation that in fact has some extrinsic origin. The next step after identifying covariates is to include them in the model. This is not necessarily an easy task — even explaining seasonality can be a challenge (Pascual and Dobson, 2005). For example, both rainfall and drought can initiate cholera epidemics. The low frequency component of residuals from a time series model fit to cholera data has been found to match various plausible environmental drivers, such as rainfall, river discharge and El Niño indices (Koelle et al., 2005). Fitting the model of (8.7) and (8.8) results in less than perfectly white residuals (see Figure 8.5). The residuals nevertheless give evidence of increased cholera infection in Dhaka after the monsoon during El Niño conditions (Ionides et al., 2006), and this association is not evident from the original time series. How best to include environmental covariates in a mechanistic model is a topic for future investigation. However, the methodology in Section 8.3 both provides a tool to identify covariates and a flexible framework for including them in a mechanistic way.

8.5 CONCLUSION

Six key areas requiring further development for time series analysis of population data were identified by Bjørnstad and Grenfell (2001). They may be summarized as follows: *(i)* including measurement error in mechanistic models; *(ii)* mechanistic modeling of environmental forcing; *(iii)* ecologically realistic continuous time models; *(iv)* reconstructing unobserved variables; *(v)* identifying interactions; and *(vi)* spatio-temporal modeling. The cholera modeling example in Section 8.4 demon-

strates that the SSM approach in Section 8.3 can be used to address *(i)-(iv)*. In addition, likelihood based model comparison then provides an approach to *(v)*. In principle, one can write down a spatial-temporal SSM to address *(vi)*. In practice, the dimension of the state space typically scales linearly with the number of spatial locations considered, and high dimensional state spaces increase the numerical burden on the SMC method. For large spatial-temporal problems, such as data assimilation in atmospheric and oceanographic science, SMC is not feasible. Related techniques have been developed for data assimilation (Evensen and van Leeuwen, 1996; Houtekamer and Mitchell, 2001), employing an ensemble of numerical solutions of a spatial-temporal model to approximate the conditional distribution given data. Alternatively, spatial-temporal variability can be incorporated through random effect models (Wikle, 2003; Wikle et al., 1998; Berliner et al., 2000). More progress is necessary before SMC techniques can be routinely applied to spatial-temporal data. However, SMC provides an effective and flexible tool for partially observed stochastic nonlinear dynamical systems of moderate dimension, allowing freedom to develop models based on scientific principles rather than on methodological constraints.

8.6 ACKNOWLEDGMENTS

The authors acknowledge many helpful discussions with Mercedes Pascual and Menno Bouma; the latter provided the cholera data for Figure 8.4. The authors were supported by National Science Foundation grant #0430120.

Bibliography

Aguirre, A. A., Ostfeld, R. S., Tabor, G. M., House, C., and Pearl, M. C., editors (2002). *Conservation Medicine*. Oxford University Press, New York.

Anderson, B. D. and Moore, J. B. (1979). *Optimal Filtering*. Prentice-Hall, New Jersey.

Anderson, R. M. and May, R. M. (1991). *Infectious Diseases of Humans*. Oxford University Press, Oxford.

Arulampalam, M. S., Maskell, S., Gordon, N., and Clapp, T. (2002). A tutorial on particle filters for online nonlinear, non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50:174 – 188.

Bailey, N. T. J. (1975). *The mathematical theory of infectious diseases and its application*. Charles Griffin, London, 2nd edition.

Barndorff-Nielsen, O. E. and Cox, D. R. (1994). *Inference and Asymptotics*. Chapman and Hall, London.

Bartlett, M. S. (1960). *Stochastic Population Models in Ecology and Epidemiology*. Wiley, New York.

Berliner, L. M., Wikle, C. K., and Cressie, N. (2000). Long-lead prediction of Pacific SSTs via Bayesian dynamic modeling. *Journal of Climate*, 13:3953–3968.

Bjørnstad, O. N. and Grenfell, B. T. (2001). Noisy clockwork: Time series analysis of population fluctuations in animals. *Science*, 293:638–643.

Bouma, M. J. and Pascual, M. (2001). Seasonal and interannual cycles of endemic cholera in Bengal 1891–1940 in relation to climate and geography. *Hydrobiologia*, 460:147–156.

Brillinger, D. R., Guckenheimer, J., Guttorp, P., and Oster, G. (1980). Empirical modelling of population time series: The case of age and density dependent rates. In Oster, G., editor, *Some Questions in Mathematical Biology*, pages 65–90. American Mathematical Society, Providence.

Brillinger, D. R., Preisler, H. K., Ager, A. A., Kie, J. G., and Stewart, B. S. (2002). Employing stochastic differential equations to model wildlife motion. *Bull. Brazilian Mathematical Soc.*, 33:385–408.

Brillinger, D. R. and Stewart, B. S. (1998). Elephant-seal movements: Modelling migration. *Canadian J. Statist.*, 26:431–443.

Clark, J. S. and Bjørnstad, O. N. (2004). Population time series: Process variability, observation errors, missing values, lags, and hidden states. *Ecology*, 85:3140–3150.

Cleveland, W. S., Grosse, E., and Shyu, W. M. (1993). Local regression models. In Chambers, J. M. and Hastie, T. J., editors, *Statistical models in S*. Chapman and Hall, London.

de Valpine, P. and Hastings, A. (2002). Fitting population models incorporating process noise and observation error. *Ecological Monographs*, 72:57–76.

Diekmann, O. and Heesterbeek, J. A. P. (2000). *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. J. Wiley and Sons, Chichester.

Doucet, A., de Freitas, N., and Gordon, N. J., editors (2001). *Sequential Monte Carlo Methods in Practice*. Springer, New York.

Durbin, J. and Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford.

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

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

Engen, S., Bakke, O., and Islam, A. (1998). Demographic and environmental stochasticity: Concepts and definitions. *Biometrics*, 54:840–846.

Evensen, G. and van Leeuwen, P. J. (1996). Assimilation of geostat altimeter data for the Agulhas Current using the ensemble Kalman filter with a quasigeostrophic model. *Monthly Weather Review*, 124:58–96.

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

Glass, K., Xia, Y., and Grenfell, B. T. (2003). Interpreting time-series analyses for continuous-time biological models-measles as a case study. *Journal of Theoretical Biology*, 223:19–25.

Gordon, N., Salmond, D. J., and Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings–F*, 140(2):107–113.

Greene, S. K., Ionides, E. L., and Wilson, M. L. (2006). Patterns of influenza-associated mortality among US elderly by geographic region and virus subtype, 1968–1998. *American Journal of Epidemiology*, pre-published on-line.

Hahn, B. H., Shaw, G. M., de Cock, K. M., and Sharp, P. M. (2000). AIDS as a zoonosis: scientific and public health implications. *Science*, 287:607–614.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Hethcote, H. W. (2000). The mathematics of infectious diseases. *SIAM Rev.*, 42:599–653.

Higham, D. J. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Rev.*, 43:525–546.

Houtekamer, P. L. and Mitchell, H. L. (2001). Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 129:123–137.

Hürzeler, M. and Künsch, H. R. (2001). Approximating and maximising the likelihood for a general state-space model. In Doucet, A., de Freitas, N., and Gordon, N. J., editors, *Sequential Monte Carlo Methods in Practice*, pages 159–175. Springer, New York.

Ionides, E. L. (2005). Maximum smoothed likelihood estimation. *Statistica Sinica*, 15:1003–1014.

Ionides, E. L., Bretó, C., and King, A. A. (2006). Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences, USA, 103:18438–18443*.

Ionides, E. L., Fang, K. S., Isseroff, R. R., and Oster, G. F. (2004). Stochastic models for cell motion and taxis. *Journal of Mathematical Biology*, 48:23–37.

Jensen, J. L. and Petersen, N. V. (1999). Asymptotic normality of the maximum likelihood estimator in state space models. *Annals of Statistics*, 27:514–535.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82:35–45.

Kendall, B. E., Ellner, S. P., McCauley, E., Wood, S. N., Briggs, C. J., Murdoch, W. M., and Turchin, P. (2005). Population cycles in the pine looper moth: Dynamical tests of mechanistic hypotheses. *Ecological Monographs*, 75(2):259–276.

Kendall, D. G. (1974). Pole-seeking Brownian motion and bird navigation. *Journal of the Royal Statistical Society, Ser. B*, 36:365–417.

Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A*, 115:700–721.

Kiefer, J. and Wolfowitz, J. (1952). Stochastic estimation of the maximum of a regression function. *Annals of Mathematical Statistics*, 23:462–466.

Kitagawa, G. (1987). Non-Gaussian state-space modelling of non-stationary time series. *Journal of the American Statistical Association*, 82:1032–1063.

Kitagawa, G. (1998). A self-organising state-space model. *Journal of the American Statistical Association*, 93:1203–1215.

Kloeden, P. E. and Platen, E. (1999). *Numerical Soluion of Stochastic Differential Equations*. Springer, New York, 3rd edition.

Koelle, K. and Pascual, M. (2004). Disentangling extrinsic from intrinsic factors in disease dynamics: A nonlinear time series approach with an application to cholera. *American Naturalist*, 163:901–913.

Koelle, K., Rodó, X., Pascual, M., Yunus, M., and Mostafa, G. (2005). Refractory periods and climate forcing in cholera dynamics. *Nature*, 436:696–700.

Kovats, R. S. and Bouma, M. (2002). Retrospective studies: Analogue approaches to describing climate variability and health. In Martens, P. and McMichael, A. J., editors, *Environmental Change, Climate and Health*, pages 144–171. Cambridge University Press, Cambridge.

Li, W., Shi, Z., Yu, M., Ren, W., et al. (2005). Bats are natural reservoirs of SARS-like coronaviruses. *Science*, 310:676–679.

Liu, J. and West, M. (2001). Combining parameter and state estimation in simulation-based filtering. In Doucet, A., de Freitas, N., and Gordon, N. J., editors, *Sequential Monte Carlo Methods in Practice*, pages 197–224. Springer, New York.

Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*. Springer, New York.

Longini, I. M., Nizam, A., Xu, S., Hanshaoworakul, K. W., Cummings, D. A. T., and Halloran, M. E. (2005). Containing pandemic influenza at the source. *Science*, 309:1083–1087.

May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature*, 261:459–467.

Newman, K. B. and Lindley, S. T. (2006). Accounting for demographic and environmental stochasticity, observation error and parameter uncertainty in fish population dynamic models. *To appear in North American Journal of Fisheries Management*.

Øksendal, B. (1998). *Stochastic Differential Equations*. Springer, New York, 5th edition.

Pascual, M. and Dobson, A. (2005). Seasonal patterns of infectious diseases. *PLoS Medicine*, 2:18–20.

Robbins, H. and Monro, S. (1951). A stochastic approximation method. *Annals of Mathematical Statistics*, 22:400–407.

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

Ross, R. (1916). An application of the theory of probabilities to the study of a priori pathometry. part i. *Proceedings of the Royal Society of London, Series A*, 92(638):204–230.

Sack, D. A., Sack, R. B., Nair, G. B., and Siddique, A. K. (2004). Cholera. *Lancet*, 363:223–233.

Shephard, N. and Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, 84:653–667.

Shumway, R. H. and Stoffer, D. S. (2000). *Time Series Analysis and Its Applications*. Springer, New York.

Spall, J. C. (2003). *Introduction to Stochastic Search and Optimization*. Wiley, Hoboken.

Thomas, L., Buckland, S. T., Newman, K. B., and Harwood, J. (2005). A unified framework for modelling wildlife population dynamics. *Aust. N.Z. J. Stat.*, 47:19–34.

Thomson, M. C., Doblas-Reyes, F. J., Mason, S. J., Hagedorn, S. J., Phindela, T., Morse, A. P., and Palmer, T. N. (2006). Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. *Nature*, 439:576–579.

Turchin, P. (2003). *Complex Population Dynamics – A Theoretical/Empirical Synthesis*. Princeton Univ. Press.

Wikle, C. K. (2003). Hierarchical models in environmental science. *International Statistical Review*, 71:181–199.

Wikle, C. K., Berliner, L. M., and Cressie, N. (1998). Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics*, 5:117–154.