Inference for nonlinear dynamical systems, with applications to the ecology of infectious diseases

National Centre for Statistical Ecology

Nov 14, 2007

Ed Ionides

The University of Michigan, Department of Statistics

Acknowledgments

Joint work with Carles Breto (Universidad Carlos III de Madrid) and Aaron King (UM Ecology & Evolutionary Biology).

With advice from Mercedes Pascual (UM Ecology & Evolutionary Biology) and Menno Bouma (London School of Hygiene and Tropical Medicine).

Why quantify biological population dynamics?

- **Conservation**. Mankind is increasingly responsible for managing ecosystems. This requires a quantitative understanding of population behavior.
- **Public health**. Pathogens are also biological populations. Despite successes of vaccination and medical treatment, new diseases are emerging (SARS, HIV/AIDS) and old ones re-emerging due to drug resistant strains (malaria, tuberculosis). Treating the pathogen as part of an ecosystem is one approach to understanding and controlling emergent and re-emergent diseases.
- **Basic scientific interest**.

Six problems of Bjørnstad and Grenfell (Science, 2001)

The following are obstacles for ecological modeling and inference via **nonlinear, mechanistic models**:

1. Combining measurement noise and process noise.
2. Including covariates in mechanistically plausible ways.
3. Continuous time models.
4. Modeling and estimating interactions in coupled systems.
5. Dealing with unobserved variables.
6. Modeling spatial-temporal dynamics.

Wanted:

A framework for modeling and inference allowing consideration of arbitrary nonlinear, partially observed, vector-valued, time series models.

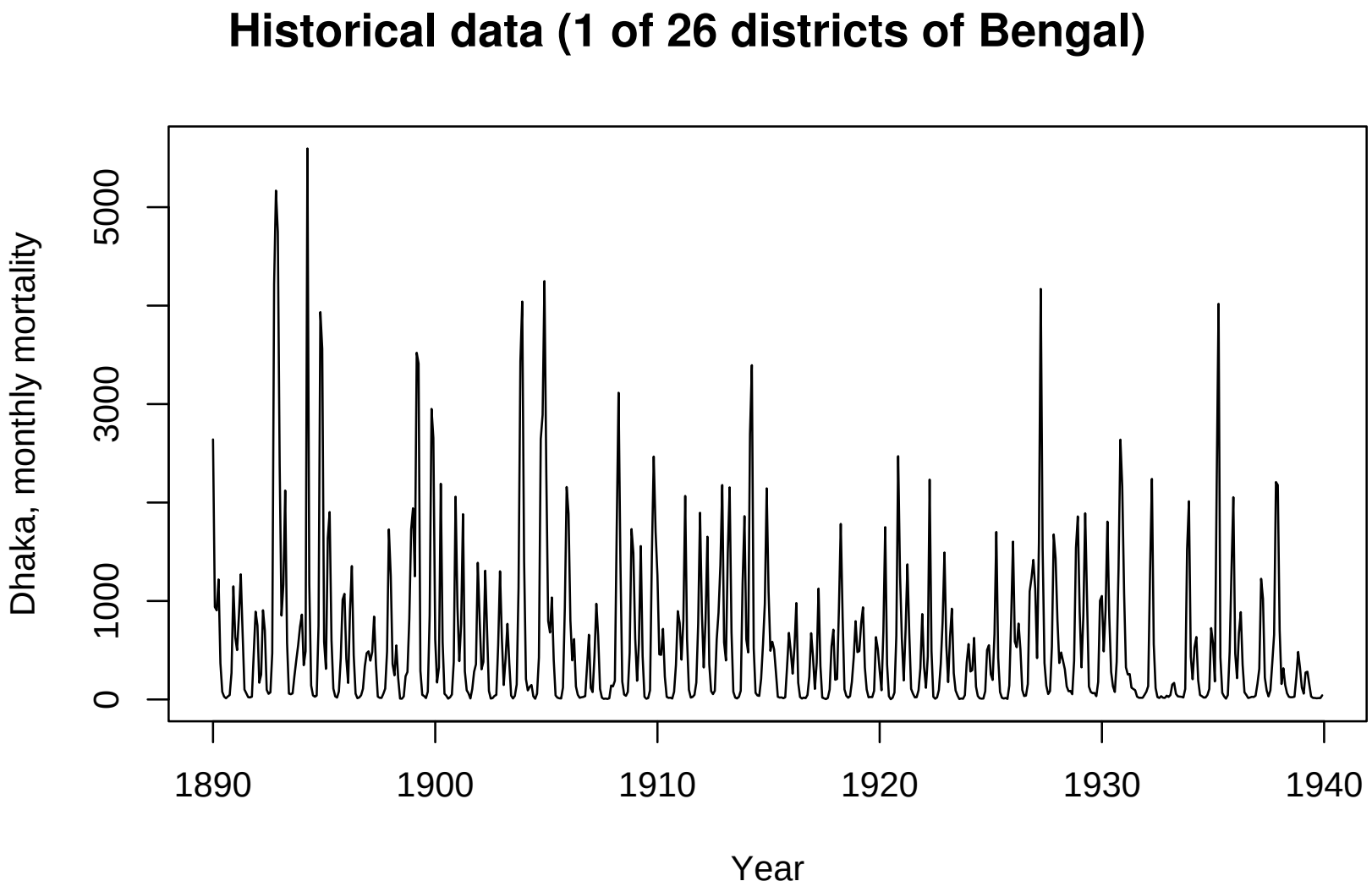
State space models

A state space model is a partially observed Markov process. It consists of an unobserved state process x_t and an observation process y_t which is conditionally independent of the past given x_t.

- x_t models a system (discrete or continuous time, usually with an unkown parameter vector θ).
- y_t models the available observations (discrete time).

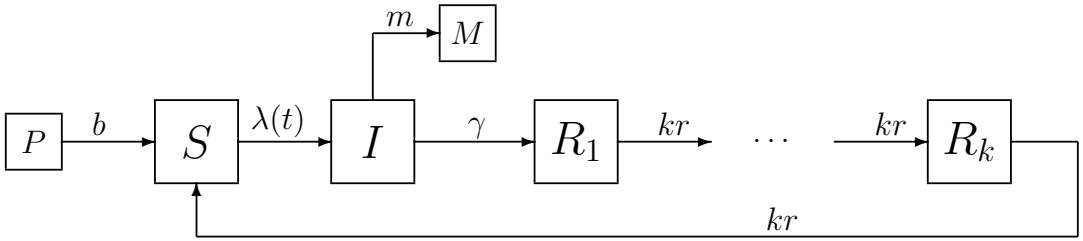
Is this framework appropriate for studying ecological population dynamics? (Thomas et al., 2005; Newman and Lindley, 2006; Buckland et al., 2007)

Example: cholera (a diarrheal disease caused by the bacterium *Vibrio cholerae*)


Historical data (1 of 26 districts of Bengal)
Dhaka, monthly mortality
0
1000
3000
5000
1890
1900
1910
1920
1930
1940
Year

Open questions for cholera epidemiology

- Seasonality!
- Relation to climate drivers.
- A role for bacteriophage?
- forecasting
  - why do some communities have cholera, others not?
- control measures
  - why are tube wells not more effective?



A model for cholera

- Individuals are susceptible S or infected I or in one of k recovery classes R_1 ... R_k
- unobserved state process: $x_t = (S(t), I(t), R_1(t), \ldots, R_k(t))$
- observed process: y_t counts monthly total cholera cases (mortality for historical data).
- population size $P(t)$ is known (a covariate).

Cholera model - nonlinear stochastic differential equations

$$\frac{dS}{dt} = kr\,R_k + \frac{dP(t)}{dt} + \delta\,P(t) - (\lambda(t)+\delta)\,S$$

$$\frac{dI}{dt} = \lambda(t)\,S - (m+\gamma+\delta)\,I(t)$$

$$\frac{dR_1}{dt} = \gamma\,I - (k\,r+\delta)\,R_1$$

$$\vdots$$

$$\frac{dR_k}{dt} = k\,r\,R_{k-1} - (k\,r+\delta)\,R_k$$

Force of infection, with white noise $dW(t)/dt$

$$\lambda(t) = \left(e^{\beta_{trend}\,t}\,\beta_{seas}(t) + \epsilon\frac{dW(t)}{dt}\right)\,\frac{I(t)}{P(t)} + \omega$$

Seasonality

- A principle: when part of the mechanism is not well understood, it must be modeled phenomenologically to make the model fit the data.
- Although we do have data on rainfall, temperature and river discharge, the mechanism driving seasonality is currently unknown.
- We adopt a non-parametric approach:

$$\beta_{seas}(t) = \exp\left\{ \sum_{k=0}^{5} b_k s_k(t) \right\},$$

where $\{s_k(t)\}$ is a periodic cubic B-spline basis.

Why model stochasticity?

- It is present in data.
- Variability (either in the state process or the observation process) is necessary to fit models to data.

Why model in continuous time?

- Eliminates artefacts due to discretization.
- Allows inclusion of covariates measured on various timescales.

Why model with continuous population?

- Can write equations similar to familiar ODE systems.
- Populations here are large (cases are not quite always large, and we have recently developed discrete population, continuous time models with both demographic and environmental stochasticity).

Key idea for our new methodology

- Likelihood based inference for the cholera model is considered difficult (witness: for similar ecological models, others use less efficient criteria, such as least square prediction error and simulated moment matching).
- One exception: Bayesian inference for time-varying parameters becomes a solveable filtering problem. Here, $\theta = \theta_t$ is a random walk with

$$E[\theta_t|\theta_{t-1}] = \theta_{t-1} \qquad \mathrm{Var}(\theta_t|\theta_{t-1}) = \sigma^2$$

- Can the limit $\sigma \to 0$ be used for fixed parameters?

Advantages of likelihood

- **Statistical efficiency**: likelihood methods are typically more efficient (make better use of limited data, giving more precise estimates with smaller standard errors).

- **Asymptotic results**: the 2nd derivative of the log likelihood at its maximum gives computationally convenient standard errors. More computationally intensive methods, such as profile likelihood, are appropriate for investigating uncertainty of key parameters.

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

Previous work on likelihood inference for nonlinear partially observed dynamical systems

- **Bayesian analysis via sequential Monte Carlo**: in practice, this requires approximations whose effect is hard to check (Liu and West, 2001).
- **Direct calculation and maximization of the likelihood surface**: not practically feasible on moderate or large dynamical systems (Hurzeler and Kunsch, 2001).
- **Markov chain Monte Carlo** has been used for Bayesian and frequentist inference–via stochastic Expectation-Maximization (Cappe, 2005). This is not applicable to the cholera model (and other continuous time models) for technical reasons: the non-existence of relative densities between different diffusion measures.

Maximum Likelihood via Iterated Filtering (MIF)

(Ionides, Bretó & King, PNAS, 2006)

Select inital value $\hat{\theta}^{(1)}$ and algorithmic parameters σ_1, c, α and N.

For n in $1, \ldots, N$

(i) set $\sigma = \sigma_1 \alpha^{n-1}$ and initialize $E[\theta_0^{(n)}] = \hat{\theta}^{(n)}$, $\mathrm{Var}(\theta_0^{(n)}) = c\sigma^2$.

(ii) evaluate the filtering means $\hat{\theta}_t^{(n)} = E[\theta_t^{(n)} | y_{1:t}]$ and the prediction variances $V_{t,n} = \mathrm{Var}(\theta_t^{(n)} | y_{1:t-1})$, for $t = 1, \ldots, T$.

(iii) $\hat{\theta}^{(n+1)} = \hat{\theta}^{(n)} + V_{1,n} \sum_{t=1}^{T} V_{t,n}^{-1} (\hat{\theta}_t^{(n)} - \hat{\theta}_{t-1}^{(n)})$

An average of the filtering means, with weights depending on the filtering variances, converges to the maximum of the likelihood (under regularity conditions)

Weighted average interpretation of MIF

- The MIF update can be written as
$\hat{\theta}^{(n+1)} = V_{1,n}\Big(\sum_{t=1}^{T-1}(V_{t,n}^{-1} - V_{t+1,n}^{-1})\hat{\theta}_t^{(n)} + V_{T,n}^{-1}\hat{\theta}_T^{(n)}\Big)$.

- This is a weighted average of the "local" estimates $\{\hat{\theta}_t^{(n)}\}$, since the coefficients sum to 1.

Smoothed likelihood interpretation of MIF

- Adding noise to the parameters is (loosely) equivalent to smoothing the likelihood function.
- Early iterations of MIF try to maximize a heavily smoothed likelihood, helping to avoid getting stuck in local maxima.
- In later iterations, as the noise decreases the function being maximized approaches the true likelihood.

Two analogies

- Like the EM algorithm, MIF is an optimization trick that takes advantage of a special model structure (partially observed Markov processes).
- Like simulated annealing, the stochasticity introduced in MIF results in "thermal fluctuations" which "cool" toward a "freezing point" at a likelihood maximum.

Implementing MIF using Monte Carlo: A brief tutorial

- Let $\{X^F_{t,j}, j = 1, \ldots, J\}$ solve the filtering problem at time t by having (approximately) marginal density $f(x_t|y_{1:t})$.

- **Move particles according to the state process dynamics:**
  Make $X^P_{t+1,j}$ a draw from $f(x_{t+1}|x_t{=}X^F_{t,j})$. Then $\{X^P_{t+1,j}\}$ is a draw from $f(x_{t+1}|y_{1:t})$, solving the prediction problem at time $t+1$.

- **Prune particles according likelihood given data:**
  Make $X^F_{t+1,j}$ a drawn from $\{X^P_{t+1,j}\}$ with probability proportional to $w_j = f(y_t|x_t{=}X^P_{t,j})$. Then $\{X^F_{t+1,j}\}$ solves the filtering problem at $t+1$.

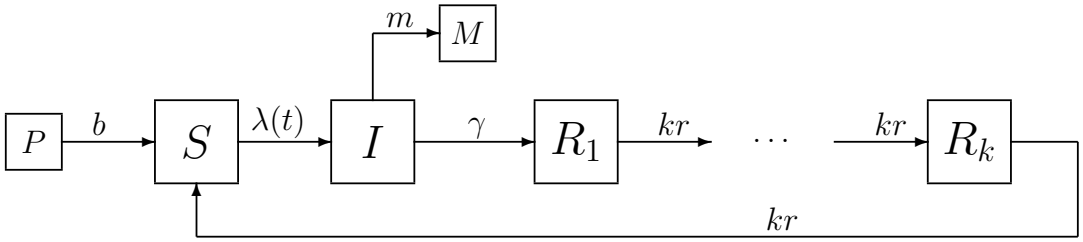
- $E[x_t|y_{1:t}]$ and $\mathrm{Var}(x_t|y_{1:t-1})$ are calculated as the sample mean and variance of $X^F_{t,k}$ and $X^P_{t,k}$ respectively.

Plug and play property of MIF via particle filter

- All MIF needs to know about the state process is how to simulate sample paths. If someone gives you simulation code for a model, you do not need to read it to carry out inference.
- Such methods have been called “equation free” (Kevrekidis et al., 2004).
- I prefer the description “**plug and play**.”
- MCMC or EM approaches are not plug and play.

Implementing MIF for the cholera model

- Here, the Euler method was used to solve the system of stochastic differential equations. This is analogous to numerical solution of ordinary differential equations.
- Algorithmic parameters (initial values and “cooling parameters”) were initialized via heuristic scaling arguments, and adjusted by inspecting convergence diagnostic plots.
- Efficiency of maximization does depend on algorithmic parameters. Once successful (local) maximization has been confirmed, algorithmic parameters play no role in the scientific conclusions.



- $\lambda(t) = \left(e^{\beta_{trend}\, t}\, \beta_{seas}(t) + \epsilon\, dW(t)/dt\right)\ \frac{I(t)}{P(t)} + w$
- ϵ is an *environmental stochasticity* parameter, w is a *reservoir*.
- $\left(e^{\beta_{trend}\, t}\, \beta_{seas}(t) + \epsilon\, dW(t)/dt\right)$ is the transmission rate for *human-to-human* contact.
- Observations: $y_t | M_t \sim \mathcal{N}[M_t, \tau^2 M_t^2]$ with $M_t = \int_{t-1}^{t} mI(t)\, dt$

(a)

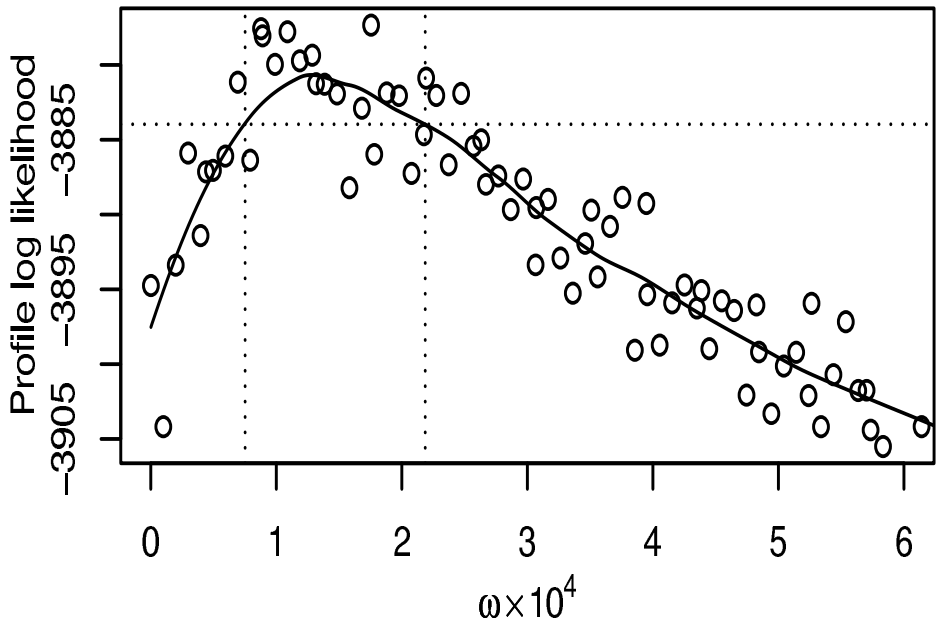
	Estimated Parameters								
	b_0	b_1	b_2	b_3	b_4	b_5	$\omega \times 10^4$	τ	ϵ
θ^*	-0.58	4.73	-5.76	2.37	1.69	2.56	1.76	0.25	0.80
$\hat{\theta}$	-0.50	4.66	-5.58	2.30	1.77	2.47	1.81	0.26	0.78
$SE(\hat{\theta})$	0.13	0.15	0.42	0.14	0.08	0.09	0.26	0.01	0.06

(b)

	Fixed Parameters				
	$1/\gamma$	m	$1/\delta$	$1/r$	k
θ^*	0.75	0.046	600	120	3

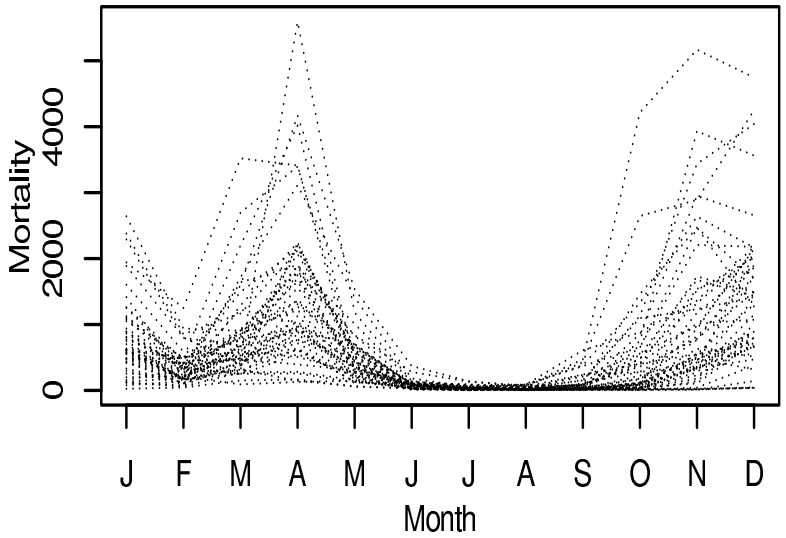
Confidence intervals via profile likelihood

- A profile likelihood gives a 99% C.I. of $[75 \times 10^{-6}, 210 \times 10^{-6}]$ for the environmental reservoir parameter, ω, i.e., 30–80 infections per million inhabitants per month for a population with 38% susceptibility.
- Previous analyses required log linearity, so could not include a reservoir term.



Residual analysis

- Prediction residuals are
$u_t(\hat{\theta}) = (y_t - E[y_t|y_{1:t-1}, \hat{\theta}])/\sqrt{\mathrm{Var}(y_t|y_{1:t-1}, \hat{\theta})}.$
- A correlation of 0.44 ($p = 0.001$) was found between Niño3.4 (an index of El Niño) and September residuals. In September, cholera increased after the monsoon.
- The effect appears only after removing intrinsic system dynamics.



Theorem 1. (Ionides, Bretó & King, 2006)

Suppose $\hat{\theta}_0$, C *and* $y_{1:T}$ *are fixed and define*

$$\begin{aligned} \hat{\theta}_t &= \hat{\theta}_t(\sigma) = E[\theta_t | y_{1:t}] \\ V_t &= V_t(\sigma) = \mathrm{Var}(\theta_t | y_{1:t-1}) \end{aligned}$$

Assuming sufficient regularity conditions for a Taylor series expansion,

$$\lim_{\sigma \to 0} \sum_{t=1}^{T} V_t^{-1} (\hat{\theta}_t - \hat{\theta}_{t-1}) = (\partial/\partial\theta) \log f(y_{1:T} | \theta, \sigma{=}0) \Big|_{\theta=\hat{\theta}_0}$$

The limit of an appropriately weighted average of local filtered parameter estimates is the derivative of the log likelihood. Thus, a fixed point of MIF is (in a limiting sense) a local maximum of the log likelihood.

Theorem 2. (Ionides, Bretó & King, 2006)

Set $\hat{\theta}^{(n+1)} = \hat{\theta}^{(n)} + \sigma_n^2 M(\nabla \ell(\hat{\theta}^{(n)}) + \eta_n)$, where M is a positive definite symmetric matrix. Suppose the following:

1. $\ell(\theta)$ is twice continuously differentiable and uniformly convex.
2. $\lim_n \sigma_n^2 n^{1-\alpha} > 0$ for some $\alpha \in (0,1)$.
3. $\{\eta_n\}$ has $E[\eta_n] = o(1)$, $\mathrm{Var}(\sigma_n^2 \eta_n) = o(1)$, $\mathrm{Cov}(\eta_m, \eta_n) = 0$ for $m \neq n$.

If there is a $\hat{\theta}$ with $\nabla \ell(\hat{\theta}) = 0$ then $\hat{\theta}^{(n)}$ converges in probability to $\hat{\theta}$.

With appropriate assumptions, MIF does converge to a local maximum if "cooled" sufficiently slowly.

Sketch proof of Theorem 1.

- Taylor expansion of $f(\theta_t|y_{1:t})$ about $\hat{\theta}_{t-1}$

$$
\begin{aligned}
f(\theta_t|y_{1:t}) &= \frac{f(y_t|\theta_t, y_{1:t-1}) f(\theta_t|y_{1:t-1})}{\int f(y_t|\theta_t, y_{1:t-1}) f(\theta_t|y_{1:t-1})\, d\theta_t} \\
&= \frac{\{f_t + (\theta_t - \hat{\theta}_{t-1})\nabla f_t + (\theta_t - \hat{\theta}_{t-1})^2 \frac{\nabla^2 f_t}{2} + R_t\} f(\theta_t|y_{1:t-1})}{\int \{f_t + (\theta_t - \hat{\theta}_{t-1})\nabla f_t + (\theta_t - \hat{\theta}_{t-1})^2 \frac{\nabla^2 f_t}{2} + R_t\} f(\theta_t|y_{1:t-1})\, d\theta_t} \\
&= \frac{\{f_t + (\theta_t - \hat{\theta}_{t-1})\nabla f_t + (\theta_t - \hat{\theta}_{t-1})^2 \frac{\nabla^2 f_t}{2} + R_t\} f(\theta_t|y_{1:t-1})}{f_t + (\frac{V_t}{2})\nabla^2 f_t + o(\sigma^2)} \\
&= \Big(1 + (\theta_t - \hat{\theta}_{t-1})\nabla f_t/f_t + (\theta_t - \hat{\theta}_{t-1})^2 \nabla^2 f_t/2f_t + R_t\Big) \times \\
&\quad \big(1 - (V_t/2f_t)\nabla^2 f_t + o(\sigma^2)\big)\, f(\theta_t|y_{1:t-1})
\end{aligned}
$$

Sketch of Proof of Theorem 1. Cont'd

- Difference of expected filtered means of θ_t

$$\begin{aligned}
\hat{\theta}_t - \hat{\theta}_{t-1} &= E[\theta_t - \hat{\theta}_{t-1} | y_{1:t}] \\
&= \int (\theta_t - \hat{\theta}_{t-1}) f(\theta_t | y_{1:t}) \, d\theta_t \\
&= V_t \nabla f_t(\hat{\theta}_{t-1}) / f(\hat{\theta}_{t-1}) + o(\sigma^2) \\
&= V_t \nabla \log f_t(\hat{\theta}_{t-1}) + o(\sigma^2) \\
&= V_t \nabla \log f_t(\hat{\theta}_0, \sigma{=}0) + o(\sigma^2)
\end{aligned}$$

- The result follows by summing up over t

$$\sum_{t=1}^{T} (\hat{\theta}_t - \hat{\theta}_{t-1})/V_t = \sum_{t=1}^{T} \nabla \log f_t(\hat{\theta}_0, \sigma{=}0) + o(1)$$

Conclusions from the cholera case study

1. Statistical methodology

- Likelihood-based inference is possible for rather general partially observed stochastic dynamical systems. Such systems arise in ecology, epidemiology and no doubt elsewhere.
- Iterated filtering allows likelihood maximization (and therefore profile likelihood, likelihood ratio tests) for some models where likelihood-based inference has previously been considered unfeasible.
- Stochastic differential equations provide a flexible class of tractable models. Technical difficulties can often be avoided.

Conclusions from the cholera case study

2. Some continuing scientific work

- Further data analysis leads us to a novel suggestion that short-term "natural immunization" plays a major role in cholera epidemiology. [this is another talk].
- In fact, this idea is not new as a clinical theory, but the population level significance has not been realised.
- "Plug and play" is useful for investigating alternative models.
- Compartment models have limitations, but nevertheless form the basis for current understanding of disease dynamics.
- Current work extends the cholera model to multiple interacting strains.

Markov chains with noise added to the rates

(one-slide summary of recent work)

- Using gamma noise on the rates of a Markov chain, one can enjoy the flexibility of SDE modeling (i.e. specifying a model by mean and variance of infinitesimal increments) for discrete population models.
- Standard continuous time models for discrete populations have only demographic stochasticity—usually inappropriate for data analysis.

Future possibilities

- Spatio-temporal nonlinear modeling requires modifications of state space methods (the curse of dimensionality).
- Is it plausible to develop methods that do not rely on exhaustive Monte Carlo simulation? Maybe not, particularly if one wishes for the flexibility of specifying a model via an algorithm to simulate from it (plug and play).
- Various refinements of iterated filtering are possible: employing computationally efficient filters and fine-tuning the iterated fitering algorithm.

Thank you!

References

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

Buckland, S., Newman, K., Fernndez, C., Thomas, L., and Harwood, J. (2007). Embedding population dynamics models in inference. *Statistical Science*, 22:44–58.

Cappé, O., Moulines, E., and Rydén, T. (2005). *Inference in Hidden Markov Models*. Springer, New York.

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

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

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., Bretó, C., and King, A. A. (2006). Inference for nonlinear dynamical systems. *Proc. Natl. Acad. Sci. USA*, 103:18438–18443.

Kevrekidis, I. G., Gear, C. W., and Hummer, G. (2004). Equation-free: The computer-assisted analysis of complex, multiscale systems. *AIChE J.*, 50:1346–1354.

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

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.

Newman, K. B. and Lindley, S. T. (2006). Accounting for demographic and environmental stochasticity, observation error and parameter uncertainty in fish population dynamic models. *North American Journal of Fisheries Management*, 26:685–701.

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.

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