# Infectious disease dynamics: a statistical perspective

**Epidemiology of Infectious Diseases Seminar**

**Harvard University**

**May 11, 2009**

**Ed Ionides**

The University of Michigan, Department of Statistics

## Collaborators

Anindya Bhadra (University of Michigan, Statistics)

Menno Bouma (London School of Hygiene and Tropical Medicine)

Carles Bretó (Universidad Carlos III de Madrid, Statistics)

Daihai He (McMaster University, Mathematics & Statistics)

Aaron King (Univ. of Michigan, Ecology & Evolutionary Biology)

Karina Laneri (Univ. of Michigan, Ecology & Evolutionary Biology)

Mercedes Pascual (Univ. of Michigan, Ecology & Evolutionary Biology)

## Infectious disease transmission: the public health challenge

### Why do we seek to quantify and understand disease dynamics?

- Prevention and control of emerging infectious diseases (SARS, HIV/AIDS, H5N1 “bird flu” influenza, H1N1 “swine flu” influenza)
- Understanding the development and spread of drug resistant strains (malaria, tuberculosis, MRSA)

## Disease dynamics: epidemiology or ecology, or both?

- Environmental host/pathogen dynamics are close to predator/prey relationships which are a central topic of ecology.
- Analysis of diseases as ecosystems complements more traditional epidemiology (risk factors etc).
- Ecologists typically seek to avoid extinctions, whereas epidemiologists typically seek the reverse.

## Infectious disease transmission: the statistical challenge

- **Time series data of sufficient quantity and quality to support investigations of disease dynamics are increasingly available.**

- **Statistical methods are required for partially observed, nonstationary, nonlinear, vector-valued stochastic dynamic systems.**

## Population models are typically partially observed Markov processes (alias: state space models)

$\mathbf{x_t}$ is an unobserved vector valued stochastic process (discrete or continuous time).

- It is assumed to be Markovian, i.e., the past and future evolution of the system are independent given its current state.
- This is reasonable if all important dynamic processes are modeled as part of the system.

$\mathbf{y_t}$ is a vector of the available observations (discrete time). These are assumed to be conditionally independent given $\mathbf{x_t}$ (a standard measurement model, which can be relaxed).

$\boldsymbol{\theta}$ is a vector of unknown parameters.

## Plug-and-play inference for partially observed Markov processes

- Statistical methods for pomps are **plug-and-play** if they require simulation from the dynamic model but not explicit likelihood ratios.
- Bayesian plug-and-play:
  1. Artificial parameter evolution (Liu and West, 2001)
  2. Approximate Bayesian computation ("sequential Monte Carlo without likelihoods," Sisson et al, *PNAS*, 2007)
- Non-Bayesian plug-and-play:
  3. Simulation-based prediction rules (Kendall et al, *Ecology*, 1999)
  4. Maximum likelihood via iterated filtering (Ionides et al, *PNAS*, 2006)

**Plug-and-play is a VERY USEFUL PROPERTY for investigating scientific models.**

## Plug-and-play in other settings

- **Optimization**. Methods requiring only evaluation of the objective function to be optimized are sometimes called **gradient-free**. This is the same concept as plug-and-play: the code to evaluate the objective function can be *plugged into* the optimizer.

- **Complex systems**. Methods to study the behavior of large simulation models that only employ the underlying code as a "black box" to generate simulations have been called **equation-free** (Kevrekidis et al., 2003, 2004).

  - This is the same concept as plug-and-play, but we prefer our label!

  - A typical goal is to determine the relationship between macroscopic phenomena (e.g. phase transitions) and microscopic properties (e.g. molecular interactions).

## The cost of plug-and-play

- Approximate Bayesian methods and simulated moment methods lead to a loss of statistical efficiency.
- In contrast, iterated filtering enables (almost) exact likelihood-based inference.
- Improvements in numerical efficiency may be possible when analytic properties are available (at the expense of plug-and-play). But many interesting dynamic models are analytically intractable—for example, it is standard to investigate systems of ordinary differential equations numerically.

## Summary of plug-and-play inference via iterated filtering

- **Filtering** is the extensively-studied problem of calculating the conditional distribution of the unobserved state vector $x_t$ given the observations up to that time, $y_1, y_2, \ldots, y_t$.
- **Iterated filtering** is a recently developed algorithm which uses a sequence of solutions to the filtering problem to maximize the likelihood function over unknown model parameters.
  (Ionides, Bretó & King. *PNAS*, 2006)
- If the filter is plug-and-play (e.g. using standard sequential Monte Carlo methods) this is inherited by iterated filtering.

## Key idea of iterated filtering

- Bayesian inference for time-varying parameters becomes a solveable filtering problem. Set $\theta = \theta_t$ to be a random walk with

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

- The limit $\sigma \to 0$ can be used to maximize the likelihood for fixed parameters.

**Theorem 1. (Ionides, Bretó & King, *PNAS*, 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.**

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

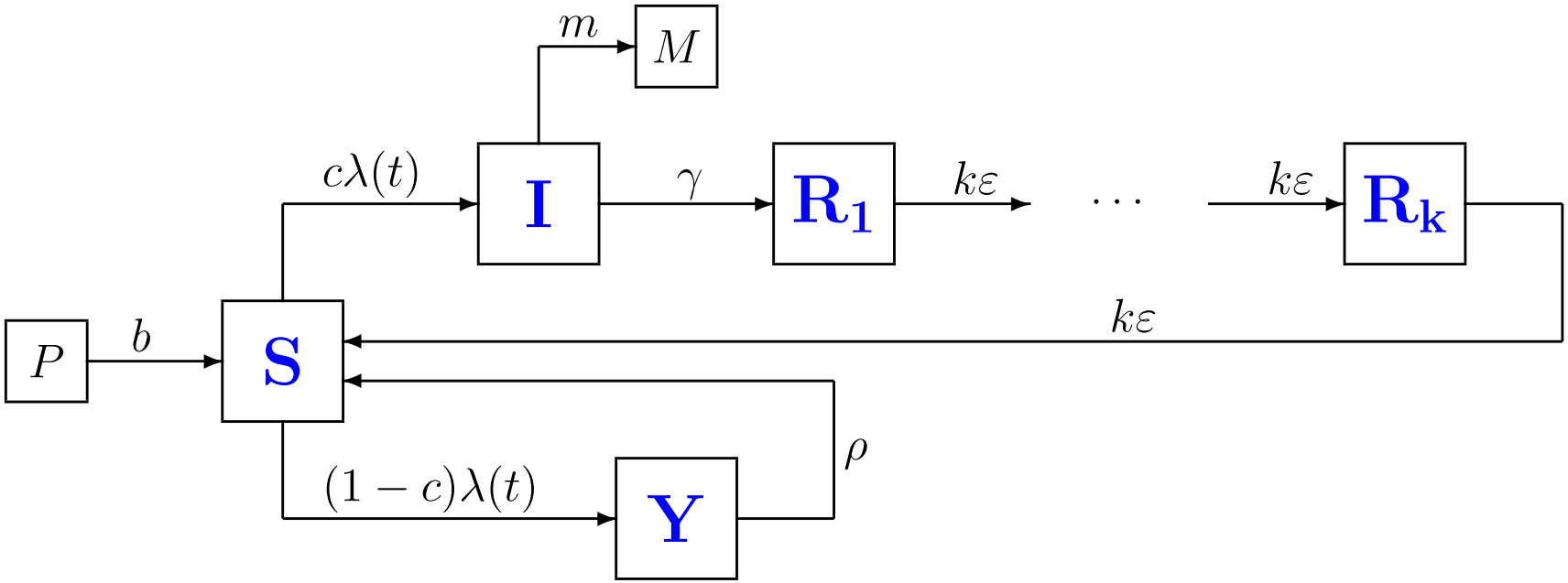
Set $\hat{\theta}^{(n+1)} = \hat{\theta}^{(n)} + \sigma_n^2 M(\nabla \ell(\hat{\theta}^{(n)}) + \eta_n)$, where $M$ is a positive definite symmetric matrix. Suppose the following:

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, iterated filtering does converge to a local maximum if "cooled" sufficiently slowly.**

# Example: cholera (bacterial diarrhea caused by *Vibrio cholerae*)


m
M
cλ(t)
I
γ
R1
kε
...
kε
Rk
kε
P
b
S
ρ
(1 − c)λ(t)
Y

| parameter | symbol |
|---|---|
| force of infection | $\lambda(t)$ |
| probability of severe infection | $c$ |
| recovery rate | $\gamma$ |
| disease death rate | $m$ |
| mean long-term immune period | $1/\varepsilon$ |
| CV of long-term immune period | $1/\sqrt{k}$ |
| mean short-term immune period | $1/\rho$ |

## Cholera model: nonlinear SDE driven by Gaussian noise $\xi(t)$

$$\frac{d}{dt}S(t) = k\epsilon\, R_k + \rho\, Y + \frac{d}{dt}P(t) + \delta\, P(t) - (\lambda(t) + \delta)\, S$$

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

$$\frac{d}{dt}Y(t) = (1-c)\,\lambda(t)\, S - (\rho + \delta)\, Y$$

$$\frac{d}{dt}R_1(t) = \gamma\, I - (k\,\epsilon + \delta)\, R_1$$

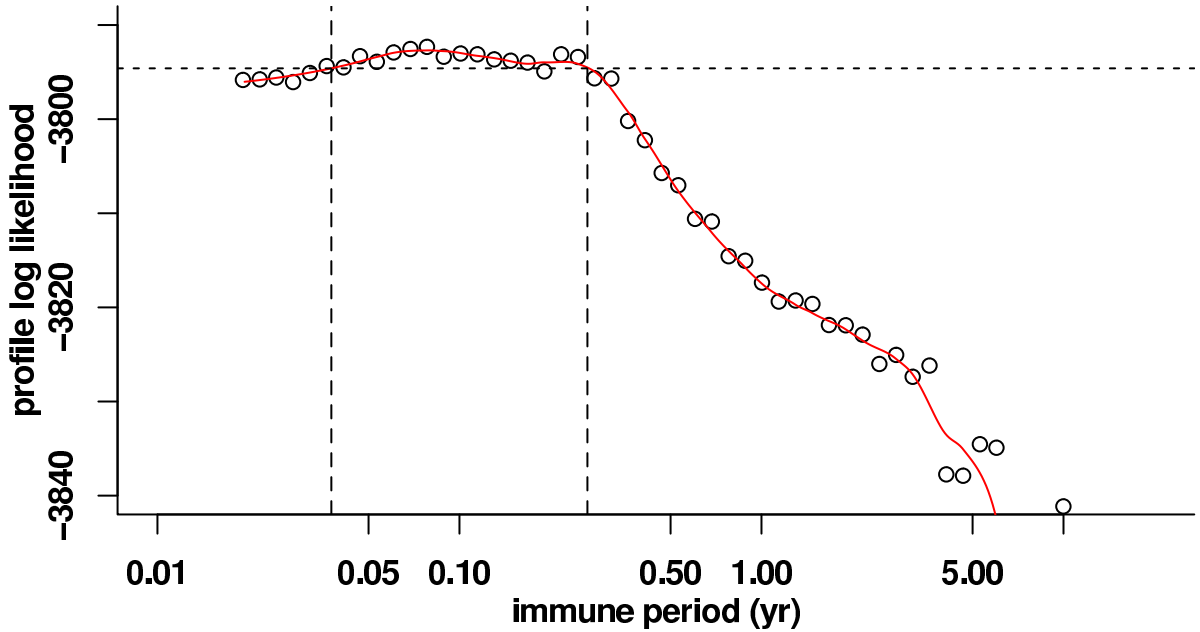
$$\vdots$$

$$\frac{d}{dt}R_k(t) = k\,\epsilon\, R_{k-1} - (k\,\epsilon + \delta)\, R_k$$

## Stochastic force of infection, with periodic cubic spline $\beta_{seas}(t)$:

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

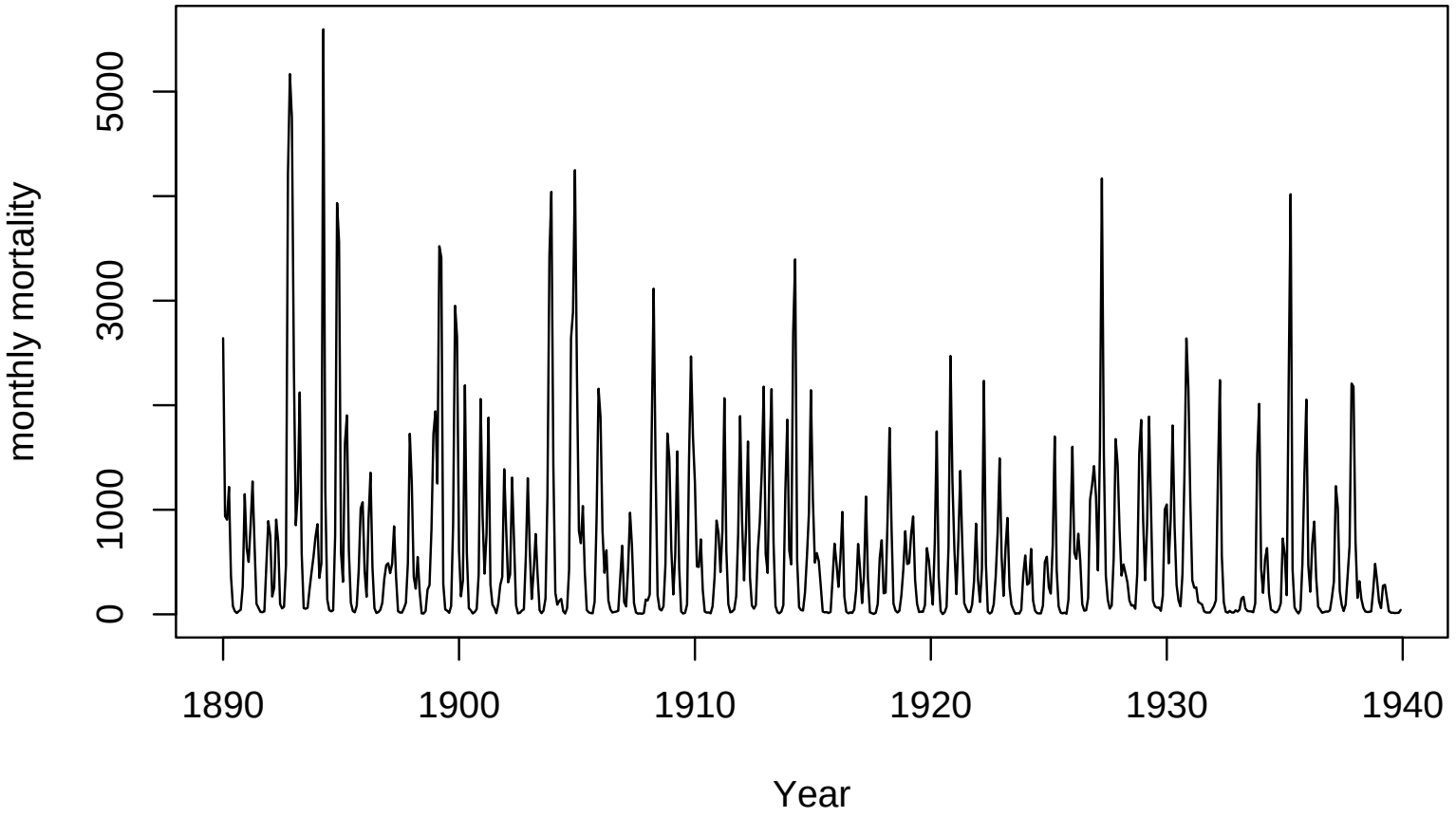
# Duration of immunity



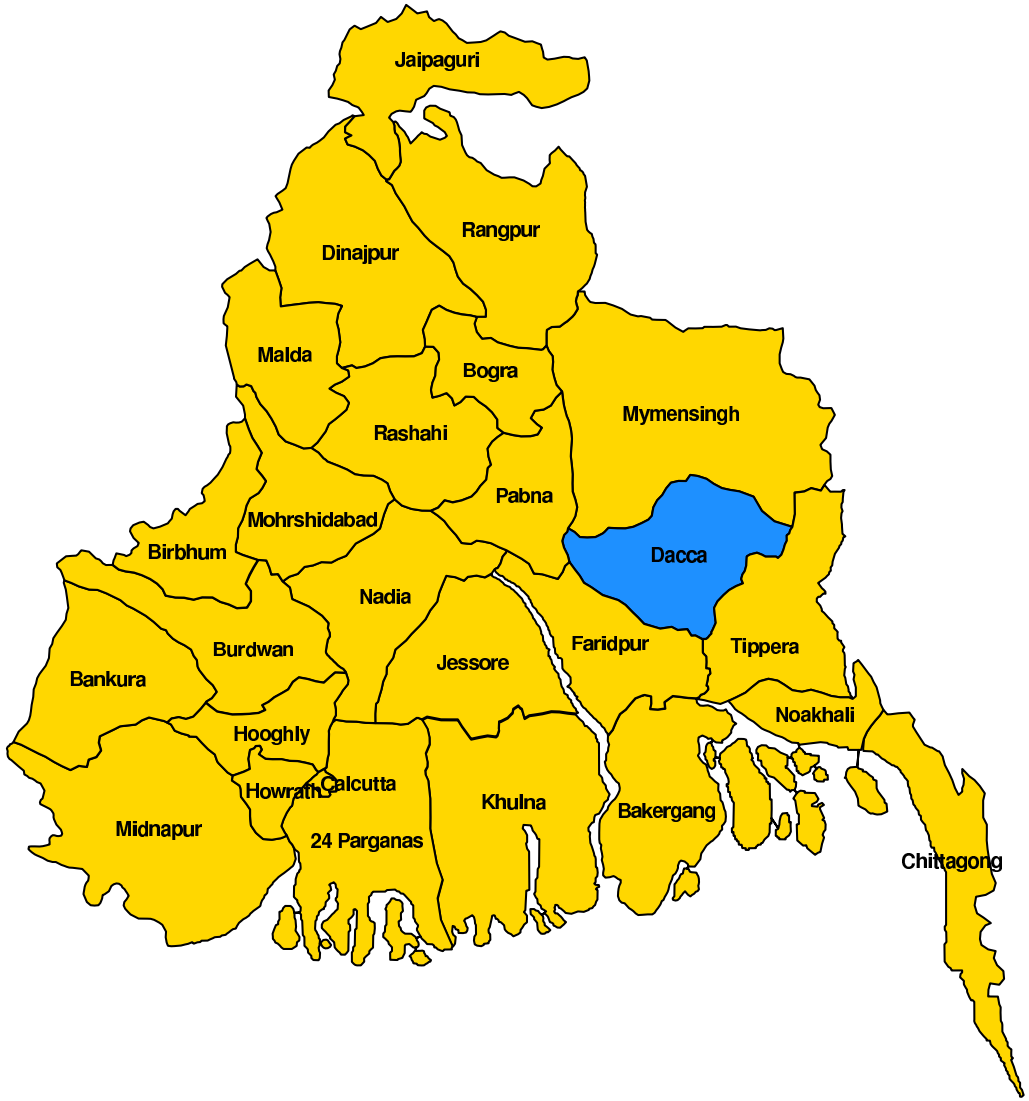

- Profile likelihood of immune period for historical time series data in Dacca, Bangladesh. Other districts give similar results.
- **Conclusion: asymptomatic infections have short-term immunity with epidemiological consequences**

  (King, Ionides, Pascual & Bouma, *Nature*, 2008).

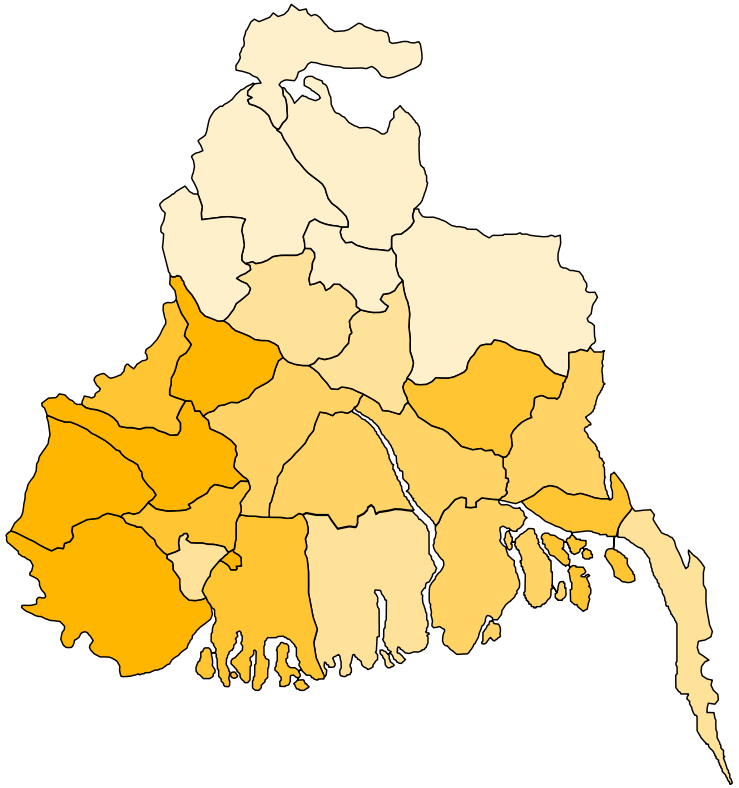
# Monthly reported cholera mortality for Dacca, 1890–1940



## Historic Bengal (now Bangladesh and the Indian state of West Bengal)



## Parameter estimates vary smoothly across space



**Example: estimated effect of environmental cholera (larger environmental reservoirs are shown as darker orange).**

## Stochastic differential equations (SDEs) vs. Markov chains

- SDEs are a simple way to add stochasticity to widely used ordinary differential equation models for population dynamics.
- When some species have low abundance (e.g. fade-outs and re-introductions of diseases within a population) discreteness can become important.
- Discrete population Markov process models are called Markov chains
  Time may be continuous or discrete: for disease transmission, continuous time is usually appropriate.

## Over-dispersion in Markov chain models of populations

- **Remarkably, in the vast literatures on continuous-time individual-based Markov chains for population dynamics (e.g. applied to ecology and chemical reactions) no-one has previously proposed models capable of over-dispersion.**
  - It turns out that the usual assumption that no events occur simultaneously creates fundamental limitations in the statistical properties of the resulting class of models.
- Over-dispersion is the rule, not the exception, in data.
- Perhaps this discrepancy went un-noticed before statistical techniques became available to fit these models to data.

## Implicit models for plug-and-play inference

- Adding "white noise" to the transition rates of existing Markov chain population models would be a way to introduce an infinitesimal variance parameter, by analogy with the theory of SDEs.
- **We do this by defining our model as a limit of discrete-time models. We introduce the term *implicit* for such a model.**
  This is backwards to the usual approach of checking that a numerical scheme (i.e. a discretization) converges to the desired model.
- Implicit models are convenient for numerical solution, by definition, and therefore fit in well with plug-and-play methodology.
- Details are in Bretó, He, Ionides & King "Time series analysis via mechanistic models," *Annals of Applied Statistics*, 2009.

## A general mechanistic model

- A compartment model groups a population into $c$ compartments.
- $X(t) = (X_1(t), \ldots, X_c(t))$ can be written in terms of the flows $N_{ij}(t)$ from $i$ to $j$, via a **conservation of mass** identity:

$$X_i(t) = X_i(0) + \sum_{j \neq i} N_{ji}(t) - \sum_{j \neq i} N_{ij}(t).$$

- Each **flow** $N_{ij}$ is associated with a **rate** function $\mu_{ij} = \mu_{ij}(t, X(t))$.
- Here, $X_i(t)$ is non-negative integer valued. $X(t)$ models a population divided into $c$ groups; $\mu_{ij}$ is the rate at which each individual in compartment $i$ moves to $j$.
- This makes the compartment model closed. Immigration, birth and death can be included via source and sink compartments.

## Noise for rates of Markov chain compartment models

- For each rate $\mu_{ij}$ between pairs of compartments $(i, j)$ we specify an **integrated noise process** $\Gamma_{ij}(t)$ giving rise to a **noise process** $\xi_{ij}(t) = \frac{d}{dt}\Gamma_{ij}(t)$.

## Properties of the integrated noise processes $\Gamma_{ij}(t)$

**P1** Independent increments.

**P2** Stationary increments.

**P3** Non-negative increments. Therefore, $\xi_{ij}(t) = \frac{d}{dt}\Gamma_{ij}(t)$ is **non-negative white noise**.

**P4** Unbiased multiplicative noise: $E[\Gamma_{ij}(t)] = t$.

**P5** Partial independence: $\Gamma_{ij}$ independent of $\Gamma_{ik}$ for $j \neq k$.

**P6** Full independence: all noise processes independent.

**P7** **Gamma noise**: marginally, $\Gamma_{ij}(t)$ is a gamma process.

## A general over-dispersed compartment model

This is an implicit model: we use an Euler approximation to define the process:

1. Divide the interval $[0, T]$ into $N$ intervals of width $\delta = T/N$
2. Set initial value $X(0)$
3. FOR $n = 0$ to $N - 1$
4. Generate noise increments $\{\Delta\Gamma_{ij} = \Gamma_{ij}(n\delta + \delta) - \Gamma_{ij}(n\delta)\}$
5. Generate process increments $(\Delta N_{i1}, \ldots, \Delta N_{i,i-1}, \Delta N_{i,i+1}, \Delta N_{ic}, R_i)$
   $\sim$ Multinomial$\big(X_i(n\delta), p_{i1}, \ldots, p_{i,i-1}, p_{i,i+1}, \ldots, p_{ic}, 1-\sum_{k\neq i} p_{ik}\big)$
   where $p_{ij} = p_{ij}(\{\mu_{ij}(n\delta, X(n\delta))\}, \{\Delta\Gamma_{ij}\})$ is given in (1) below
6. Set $X_i(n\delta + \delta) = R_i + \sum_{j\neq i} \Delta N_{ji}$
7. END FOR

The limiting Markov chain is specified follows:

$$P[\Delta N_{ij} = n_{ij}, \text{ for all } 1 \le i \le c, 1 \le j \le c, i \ne j \mid X(t) = (x_1, \ldots, x_c)]$$

$$= E\left[\prod_{i=1}^{c}\left\{\binom{x_i}{n_{i1} \ \ldots \ n_{ii-1} \ n_{ii+1} \ \ldots \ n_{ic} \ \ r_i}(1-\textstyle\sum_{k\ne i} p_{ik})^{r_i}\displaystyle\prod_{j\ne i} p_{ij}^{n_{ij}}\right\}\right] + o(\delta)$$

where $r_i = x_i - \sum_{k\ne i} n_{ik}$, $\binom{n}{n_1 \ \ldots \ n_c}$ is a multinomial coefficient and

$$\begin{aligned} p_{ij} &= p_{ij}(\{\mu_{ij}(t,x)\}, \{\Delta\Gamma_{ij}(t)\}) \\ &= (1-\exp\{-\sum_k \mu_{ik}\Delta\Gamma_{ik}\})\mu_{ij}\Delta\Gamma_{ij}/\sum_k \mu_{ik}\Delta\Gamma_{ik}, \end{aligned} \quad (1)$$

with $\mu_{ij} = \mu_{ij}(t,x)$.

**Theorem 1 (Breto, He, Ionides & King, 2008)**. Supposing assumptions (P1–P5) about the noise process, this limit does indeed specify a Markov chain.

**Proof**. An explicit construction involving exponential transition clocks for each individual, based on the method of Sellke (1983). Such methods are standard for networks of interacting Poisson processes (i.e., our compartment model with no noise). Care is required here due to the introduction of noise.

**Theorem 1, formal statement**. Suppose (P1–P5) and that $\mu_{ij}(t,x)$ is uniformly continuous as a function of $t$. Let $C(\zeta,0)$ be the compartment containing individual $\zeta$ at time $t=0$. Set $\tau_{\zeta,0}=0$, and generate independent $\mathrm{Exponential}(1)$ random variables $M_{\zeta,0,j}$ for each $\zeta$ and $j \neq C(\zeta,0)$. For $m \geq 1$, recursively set

$$\tau_{\zeta,m,j} = \inf\left\{t : \int_{\tau_{\zeta,m-1}}^{t} \mu_{C(\zeta,m-1),j}(s,X(s))\, d\Gamma_{C(\zeta,m-1),j}(s) > M_{\zeta,m-1,j}\right\}.$$

At time $\tau_{\zeta,m} = \min_j \tau_{\zeta,m,j}$, set $C(\zeta,m) = \arg\min_j \tau_{\zeta,m,j}$ and for each $j \neq C(\zeta,m)$ generate an independent $\mathrm{Exponential}(1)$ random variable $M_{\zeta,m,j}$. The increments

$dN_{ij}(t) = \sum_{\zeta,m} \mathbb{I}\{C(\zeta,m-1)=i, C(\zeta,m)=j, \tau_{\zeta,m}=t\}$

specify a Markov chain $X(t)$ whose infinitesimal transition probabilities are given by the limit of the numerical algorithm as $\delta \to 0$.

**Theorem 2 (Breto, He, Ionides & King, 2008)**. For the case of independent gamma noise, an analytic formula is available for the infinitesimal probabilities and infinitesimal moments of this chain.

**Theorem 2, formal statement**. Supposing (P1–P7), the infinitesimal transition probabilities are

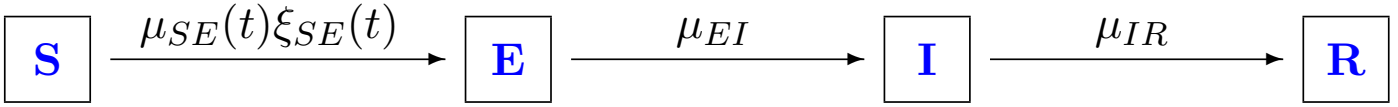
$$P[\Delta N_{ij} = n_{ij}, \text{ for all } i \neq j \mid X(t) = (x_1, \ldots, x_c)]$$
$$= \prod_i \prod_{j \neq i} \pi(n_{ij}, x_i, \mu_{ij}, \sigma_{ij}) + o(\delta)$$

where

$$\pi(n, x, \mu, \sigma) = 1_{\{n=0\}} + \delta \binom{x}{n} \sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k+1} \sigma^{-2} \ln\left(1 + \sigma^2 \mu (x-k)\right)$$

## Measles: an exhaustively studied system

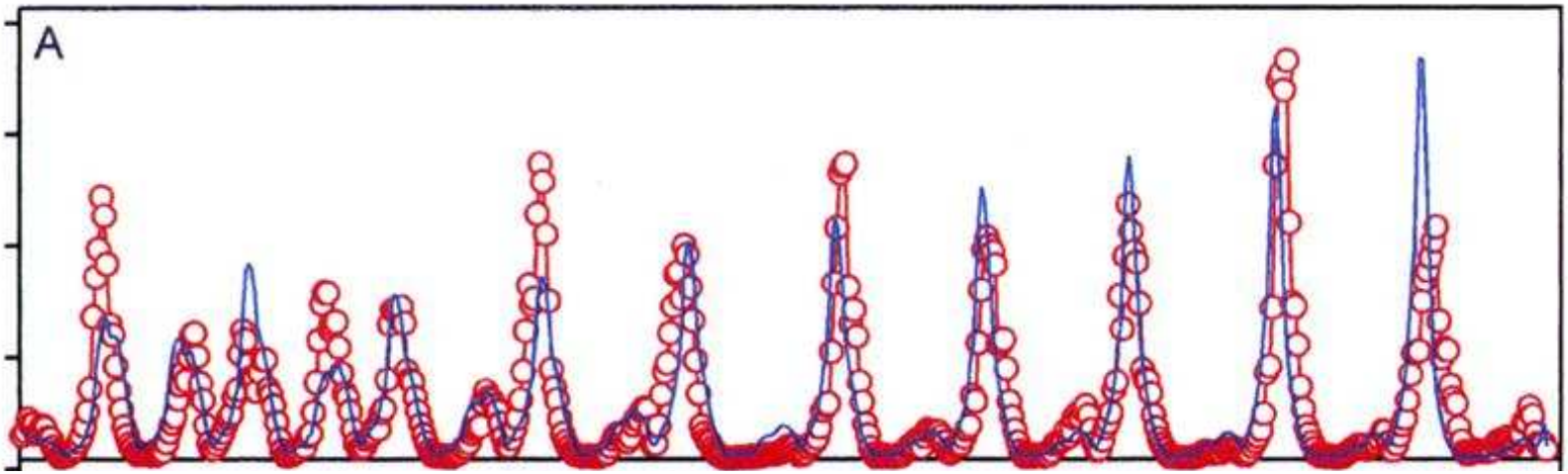
- Measles is simple: direct infection of susceptibles by infecteds; characteristic symptoms leading to accurate clinical diagnosis; life-long immunity following infection.




  Susceptible $\longrightarrow$ Exposed (latent) $\longrightarrow$ Infected $\longrightarrow$ Recovered,
  with noise intensity $\sigma_{SE}$ on the force of infection.

- Measles is still a substantial health issue in sub-Saharan Africa.

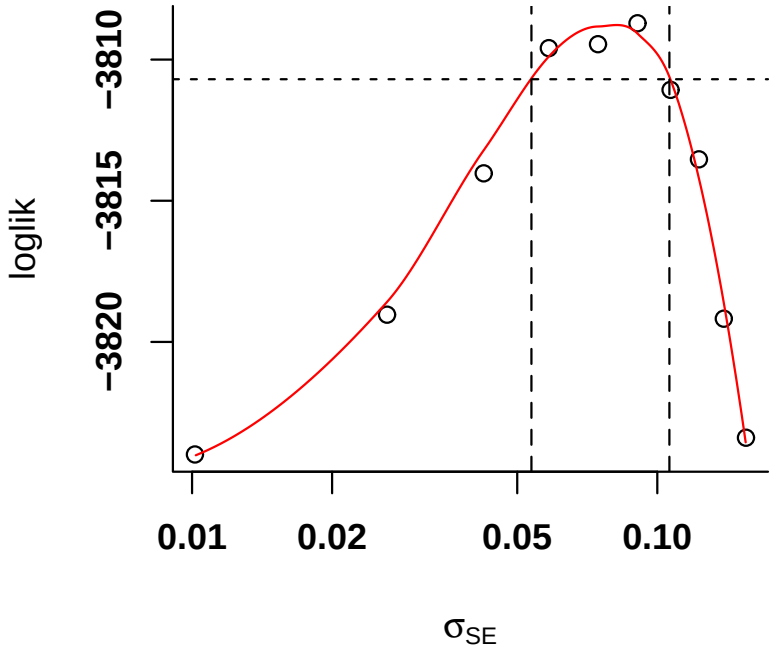
- Comprehensive doctor reports in western Europe and America before vaccination ($\approx 1968$) are textbook data.

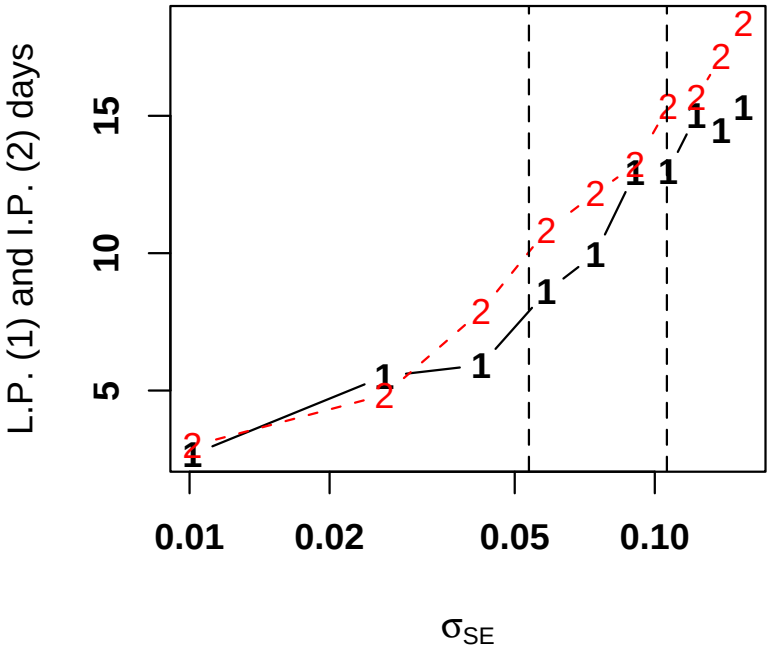



- Measles cases in London 1944–1965 (circles and red lines) and a deterministic SEIR fit (blue line) (from Grenfell *et al*, 2002).
- A deterministic fit, specified by the initial values in January 1944, captures remarkably many features.

## Is demographic stochasticity ($\sigma_{SE} = 0$) plausible?

- Profile likelihood for $\sigma_{SE}$ and effect on estimated latent period (L.P.) and infectious period (I.P.) for London, 1950–1964.
- Variability of $\approx 5\%$ per year on the force of infection substantially improves the fit, and has consequences for parameter estimates.



## Interpretation of over-dispersion

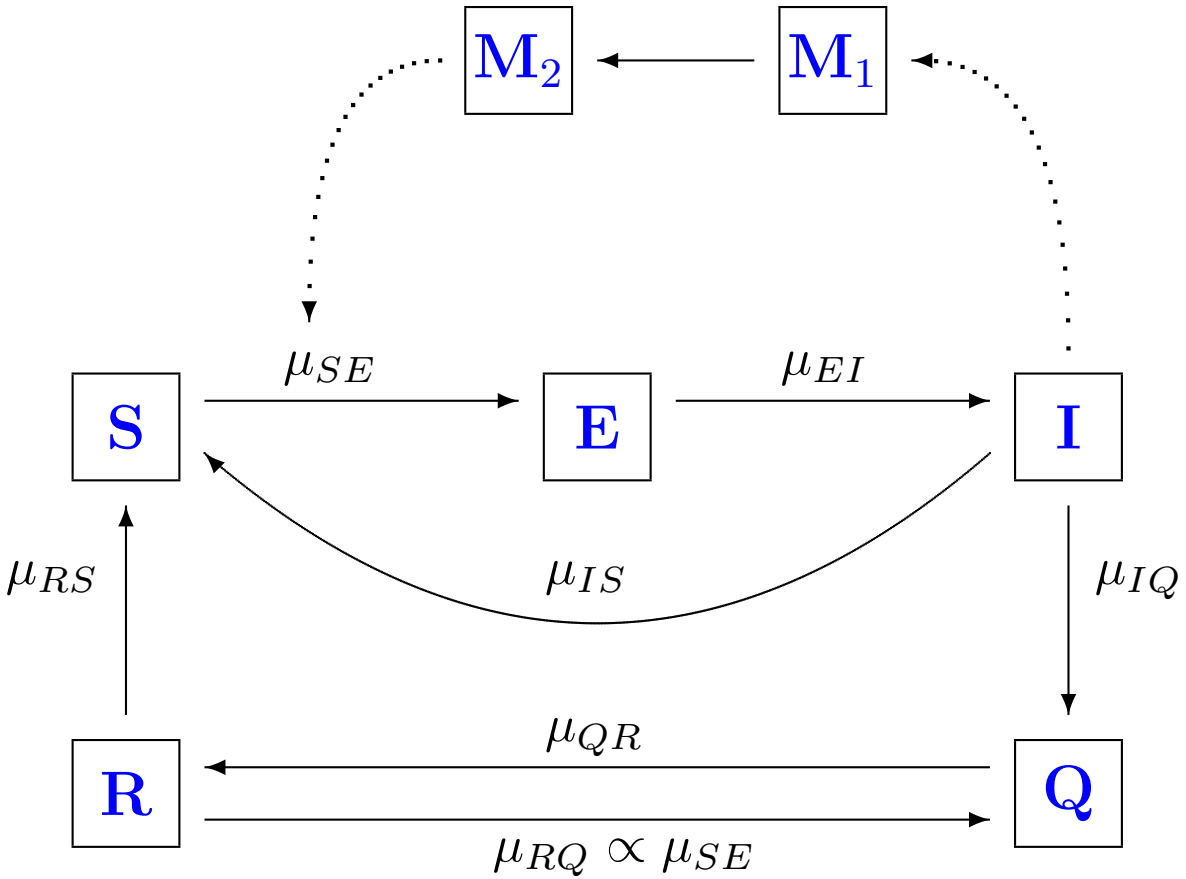
- Social and environmental events (e.g., football matches, weather) lead to stochastic variation in rates: **environmental stochasticity**.
- A catch-all for other model mis-specification? It is common practice in linear regression to bear in mind that the "error" terms contain un-modeled processes as well as truly stochastic effects. This reasoning can be applied to dynamic models as well.

## Malaria: a major global challenge

- Bill Gates would like to eradicate it, but others have tried before...
- There has been extensive debate on whether/how global climate change will affect malaria burden—a model validated by data is required.

## From the perspective of statistical methodology

- Despite the huge literature, no dynamic model of malaria transmission has previously been fitted directly to time series data.
- Incomplete and complex immunity, dynamics in both mosquito and human stages & diagnostic difficulties (the immediate symptom is non-specific fever) have put malaria beyond the scope of previous methodology developed for measles.

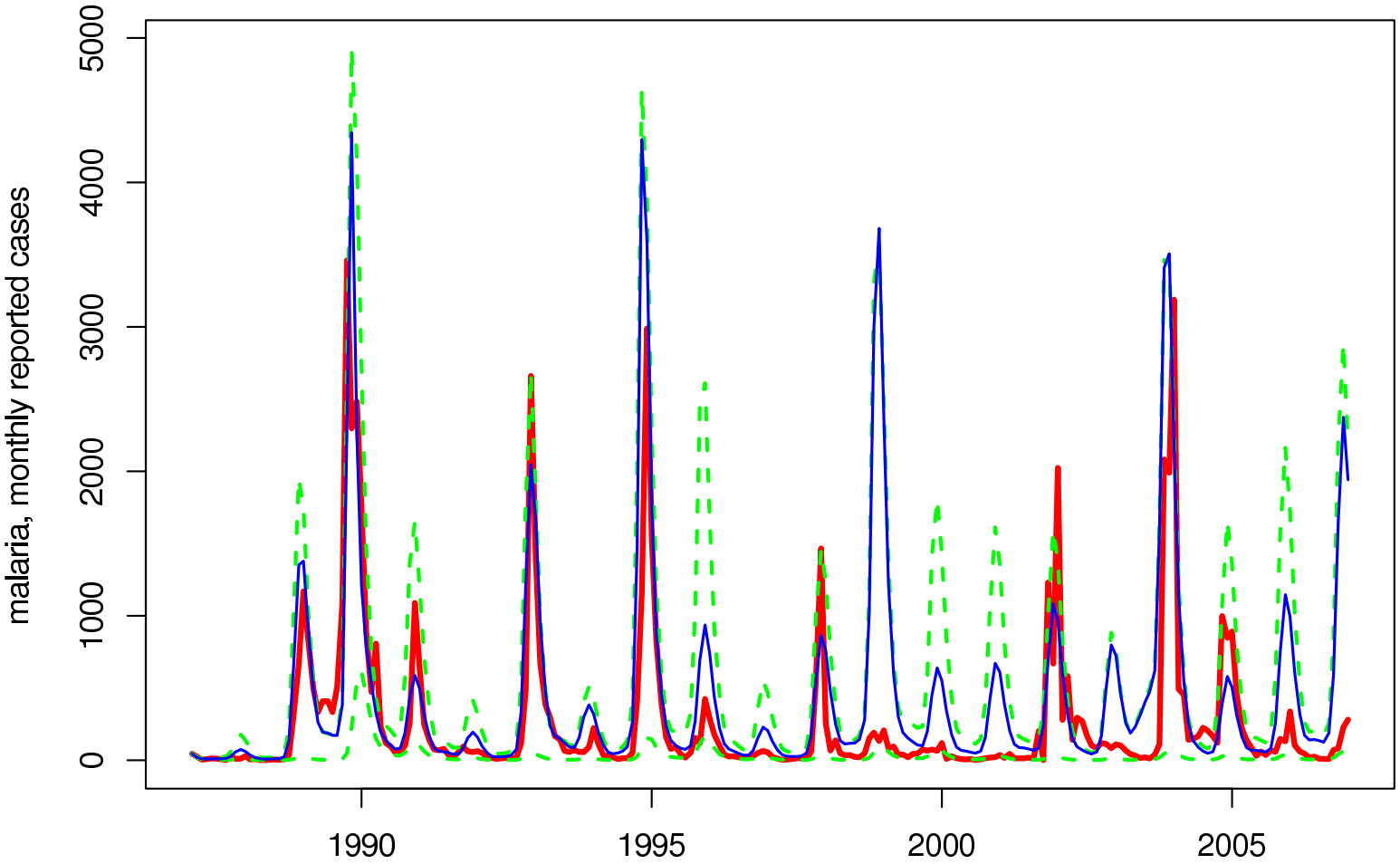



Infected mosquitoes

Disease status of human population

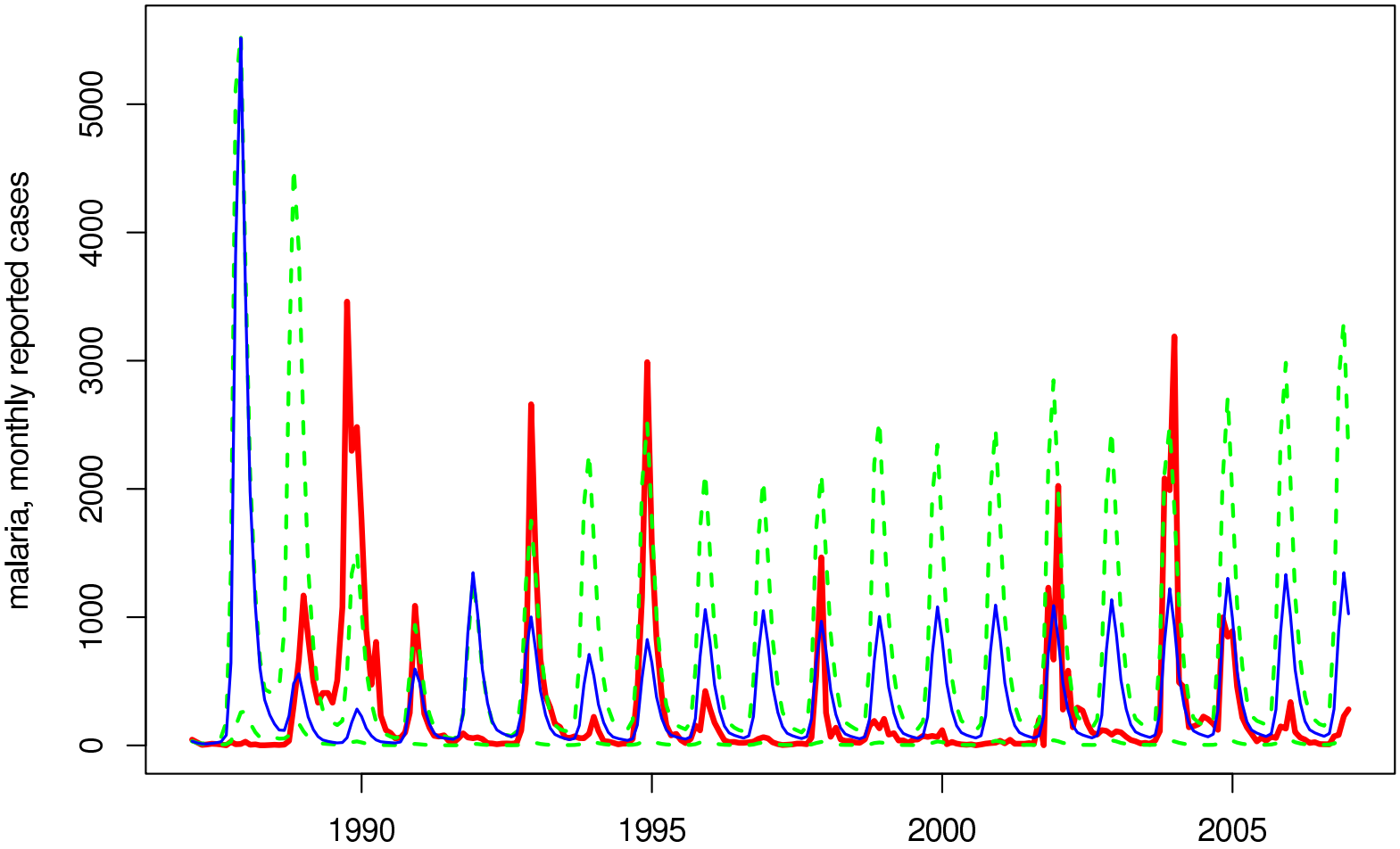
- $Q$ counts quiescent infections: no clinical symptoms and low infectivity.

# Rainfall and malaria in Kutch (NW India)




- **Cases**, **deterministic skeleton (ODE) from fit with rainfall as a covariate**, **10th and 90th percentiles of SDE model**.

# The model without rainfall doesn't capture dynamics




- **Cases**, **deterministic skeleton (ODE) from fitted model without rainfall**, **10th and 90th percentiles of SDE model**.

## Influenza

- **The threat of bird flu (H5N1 influenza) crossing to humans spurred research into the transmission and evolution of the influenza virus.**
- Descriptive statistical methods show relationships between influenza transmission, climate and dominant strains (Greene, Ionides & Wilson, *American Journal of Epidemiology*, 2006).
- Recent work has used influenza genetic sequence data to study global transmission and evolution (Rambaut et al., *Nature*, 2008; Russel et al., *Science*, 2008).
- **Building a dynamic model describing global transmission and evolution which is in statistical agreement with genetic and case report data is an open problem.**

## Influenza: individual-based models

- Influenza transmission has inspired state-of-the-art statistical methodology for individual-based population models:
  **Cauchemez et al (*Nature*, 2008) inferred transmission parameters for a partially observed Markov process with a state vector of size $\approx 10^5$, modeling every individual in a small town.**

- **In climate modeling, it is not unknown to carry out inference with a state vector of size $10^7$** ( Anderson and Collins, *Journal of Atmospheric and Oceanic Technology*, 2007). **This requires careful use of appropriate approximations.**

## Conclusions

- Plug-and-play statistical methodology permits likelihood-based analysis of flexible classes of stochastic dynamic models.
- **It is increasingly possible to carry out data analysis via nonlinear mechanistic stochastic dynamic models**. This can build links between the mathematical modeling community (within which models are typically conceptual and qualitative) and quantitative applications (testing hypotheses about mechanisms, forecasting, evaluating the consequences of interventions).
- General-purpose statistical software for partially observed Markov processes is available in the **pomp** package for R (on CRAN).

# Thank you!

These slides (including references for the citations) are available at

`www.stat.lsa.umich.edu/~ionides/pubs`

# References

Anderson, J. L. and Collins, N. (2007). Scalable implementations of ensemble filter algorithms for data assimilation. *Journal of Atmospheric and Oceanic Technology*, 24:1452–1463.

Bretó, C., He, D., Ionides, E. L., and King, A. A. (2009). Time series analysis via mechanistic models. *Annals of Applied Statistics*, 3:319–348.

Cauchemez, S., Valleron, A., Boëlle, P., Flahault, A., and Ferguson, N. M. (2008). Estimating the impact of school closure on influenza transmission from sentinel data. *Nature*, 452:750–754.

Chen, Y. and Blaser, M. J. (2008). Helicobacter pylori colonization is inversely associated with childhood asthma. *The Journal of Infectious Diseases*, 198(4):553–560.

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*, 163:316–326.

Grenfell, B. T., Bjornstad, O. N., and Finkenstädt, B. F. (2002). Dynamics of measles epidemics: Scaling noise, determinism, and predictability with the TSIR model. *Ecological Monographs*, 72(2):185–202.

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

Kendall, B. E., Briggs, C. J., Murdoch, W. W., Turchin, P., Ellner, S. P., McCauley, E., Nisbet, R. M., and Wood, S. N. (1999). Why do populations cycle? A synthesis of statistical and mechanistic modeling approaches. *Ecology*, 80:1789–1805.

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

Kevrekidis, I. G., Gear, C. W., Hyman, J. M., Kevrekidis, P. G., Runborg, O., and Theodoropoulos, C. (2003). Equation-free coarse-grained multiscale computation: Enabling microscopic simulators to perform system-level analysis. *Communications in the Mathematical Sciences*, 1:715–762.

King, A. A., Ionides, E. L., Pascual, M., and Bouma, M. J. (2008). Inapparent infections and cholera dynamics. *Nature*, 454:877–880.

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.

Rambaut, A., Pybus, O. G., Nelson, M. I., Viboud, C., Taubenberger, J. K., and Holmes, E. C. (2008). The genomic and epidemiological dynamics of human influenza A virus. *Nature*, 453:615–619.

Russell, C. A., Jones, T. C., Barr, I. G., Cox, N. J., Garten, R. J., Gregory, V., Gust, I. D., Hampson, A. W., Hay, A. J., Hurt, A. C., de Jong, J. C., Kelso, A., Klimov, A. I., Kageyama, T., Komadina, N., Lapedes, A. S., Lin, Y. P., Mosterin, A., Obuchi, M., Odagiri, T., Osterhaus, A. D. M. E., Rimmelzwaan, G. F., Shaw, M. W., Skepner, E., Stohr, K., Tashiro, M., Fouchier, R. A. M., and Smith, D. J. (2008). The global circulation of seasonal influenza A (H3N2) viruses. *Science*, 320(5874):340–346.

Sellke, T. (1983). On the asymptotic distribution of the size of a stochastic epidemic. *Journal of Applied Probability*, 20:390–394.

Sisson, S. A., Fan, Y., and Tanaka, M. M. (2007). Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the USA*, 104(6):1760–1765.

## An iterated filtering algorithm

Select $\hat{\theta}^{(1)}$, $\sigma_1$, $c$, $\alpha$ 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) For $t = 1, \ldots, T$, evaluate **filtering means** $\hat{\theta}_t^{(n)} = E[\theta_t^{(n)} | y_{1:t}]$ and **prediction variances** $V_{t,n} = \mathrm{Var}(\theta_t^{(n)} | y_{1:t-1})$.

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

- **Each iteration updates $\hat{\theta}_n$ using an average of the filtering means with weights determined by the prediction variances.**
- **$\hat{\theta}^{(N+1)}$ converges to a (local) maximum of the likelihood as $N \to \infty$, under regularity conditions.**

## Iterated filtering via sequential 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 $w_j = f(y_{t+1}|x_{t+1}{=}X^P_{t+1,j})$. 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.

## Two analogies

- Like the **EM algorithm**, iterated filtering is an optimization trick that takes advantage of a special model structure (partially observed Markov processes).
- Like **simulated annealing**, iterated filtering introduces stochasticity, resulting in “thermal fluctuations” which “cool” toward a “freezing point” at a likelihood maximum.