Advances in simulation-based inference for stochastic dynamic systems

Midwest Statistics Research Colloquium

March 16, 2013

Edward Ionides

The University of Michigan, Department of Statistics

Outline

1. Historical overview: a decade of progress on time series analysis via mechanistic models.
2. Some practical considerations: relationship between statistical methodology and software.
3. The *plug-and-play* property: iterated filtering and other plug-and-play approaches.
4. Case studies: malaria, measles and HIV.
5. Outstanding challenges.

Six problems of Bjørnstad and Grenfell (Science, 2001)

Obstacles for *ecological* 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.

Partially observed Markov process (POMP) models have been repeatedly proposed as an approach to combining modeling and inference for biological systems

- The Markov property—all information about future dynamics of the system is in the current state—is natural for mechanistic modeling. If some variable is relevant to the dynamics, add it to the state!
- General-purpose software has been a challenge for stastistcal inference using non-linear non-Gaussian POMPs:
  - ◇ WinBUGS performs poorly on these models.
  - ◇ **pomp**, an R package for POMPs, is recently available.

Partially Observed Markov Process (POMP) notation

The unobserved Markov dynamic process is denoted $X(t)$. For observation times $t_1, \ldots, t_N$ we write $X_n = X(t_n)$. The observable variables $Y_1, \ldots, Y_N$ are conditionally independent given $X_1, \ldots, X_N$. The model depends on an unknown parameter vector θ.

- To think algorithmically, we define some function calls:

  `rprocess( )`: a draw from $f_{X_n|X_{n-1}}(x_n \,|\, x_{n-1}\,;\theta)$

  `dprocess( )`: evaluation of $f_{X_n|X_{n-1}}(x_n \,|\, x_{n-1}\,;\theta)$

  `rmeasure( )`: a draw from $f_{Y_n|X_n}(y_n \,|\, x_n\,;\theta)$

  `dmeasure( )`: evaluation of $f_{Y_n|X_n}(y_n \,|\, x_n\,;\theta)$

Plug-and-play inference for POMP models

- An algorithm operating on a POMP is **plug-and-play** if it involves calls to **`rprocess`** but not to **`dprocess`**, and so code simulating sample paths is 'plugged' into the inference software.
- Bayesian plug-and-play:
  1. Particle MCMC (Andrieu et al, *J. Roy. Statist. Soc. B*, 2010)
  2. ABC (Approximate Bayesian Computation; Toni et al, *Interface*, 2009)
  3. Artificial parameter evolution (Liu and West, 2001)
- Non-Bayesian plug-and-play:
  4. Simulation-based prediction rules (Kendall et al, *Ecology*, 1999)
  5. Simulated likelihood of summary statistics (Wood, *Nature*, 2010)
  6. Iterated filtering (Ionides et al, *PNAS*, 2006)

Plug-and-play is a VERY USEFUL PROPERTY for scientific work.

Classification of methodologies by required operations

	`rprocess`	`dprocess`	`rmeasure`	`dmeasure`
EM via SMC	✔	✔	✗	✔
MCMC	✗	✔	✗	✔
Iterated filtering	✔	✗	✗	✔
Particle MCMC	✔	✗	✗	✔
Liu-West SMC	✔	✗	✗	✔
Nonlinear forecasting	✔	✗	✔	✗
ABC	✔	✗	✔	✗
Probe matching	✔	✗	✔	✗

- Textbook EM and MCMC methods are not plug-and-play.
- Nonlinear forecasting and probe matching are simulation-based techniques developed by scientists, likely due to the inapplicability of standard EM and MCMC techniques.

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 numerical simulations (e.g., molecular models for phase transitions) that only employ the underlying code as a "black box" to generate simulations are called **equation-free** (Kevrekidis et al., 2003, 2004).

- **ABC and MCMC**. Plug-and-play methods have recently been called **likelihood-free**. In this terminology, iterated filtering does likelihood-free likelihood-based inference.

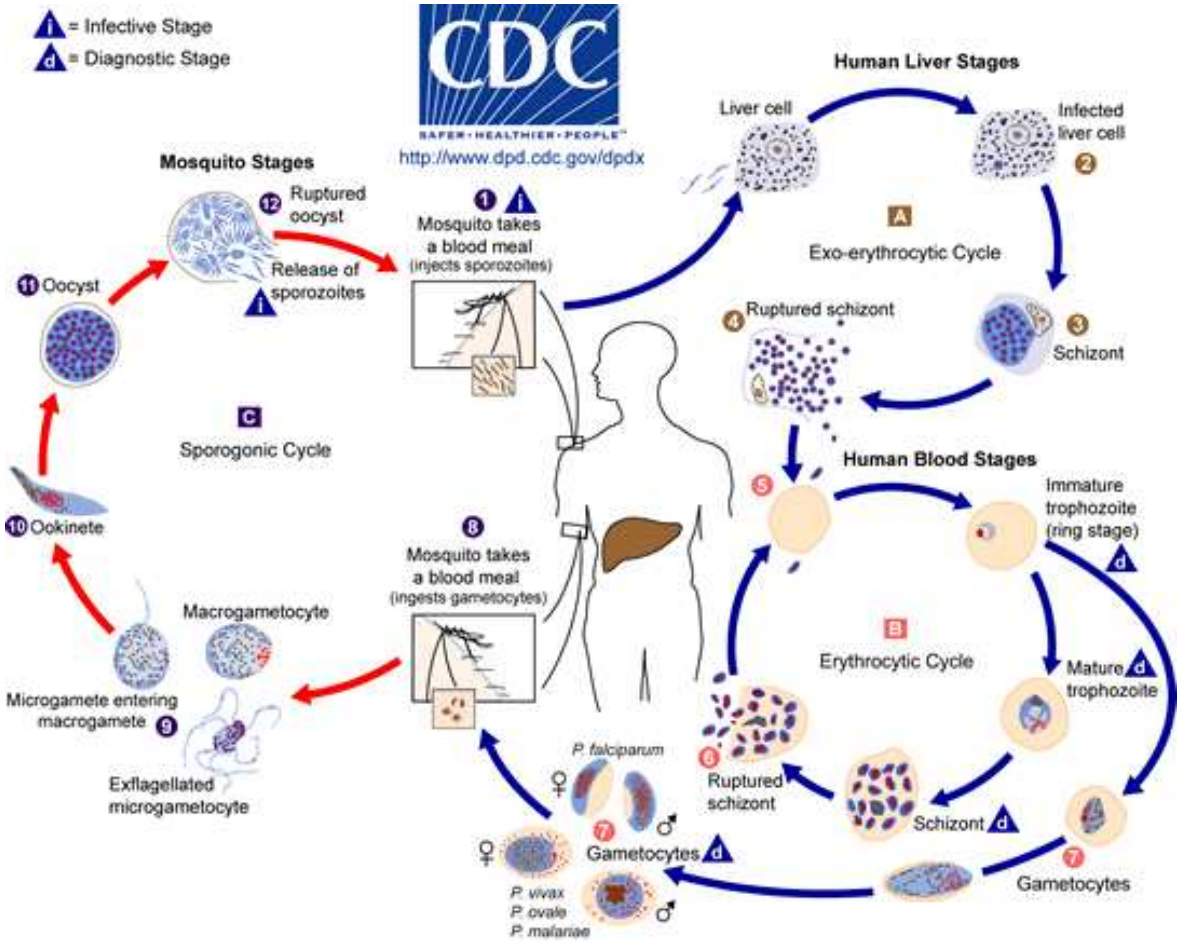
The cost of plug-and-play

- 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** algorithms use a sequence of solutions to the filtering problem to maximize the likelihood function over unknown model parameters (proposed by Ionides, Bretó & King; *PNAS*, 2006).
- **Sequential Monte Carlo (SMC)** provides a plug-and-play filter. The plug-and-play property property is inherited by SMC-based implementations of iterated filtering and PMCMC.

Example: malaria (mosquito-transmitted *Plasmodium* infection)



Despite extensive study of the disease system (mosquito, *Plasmodium* & human immunology) malaria epidemiology remains hotly debated.

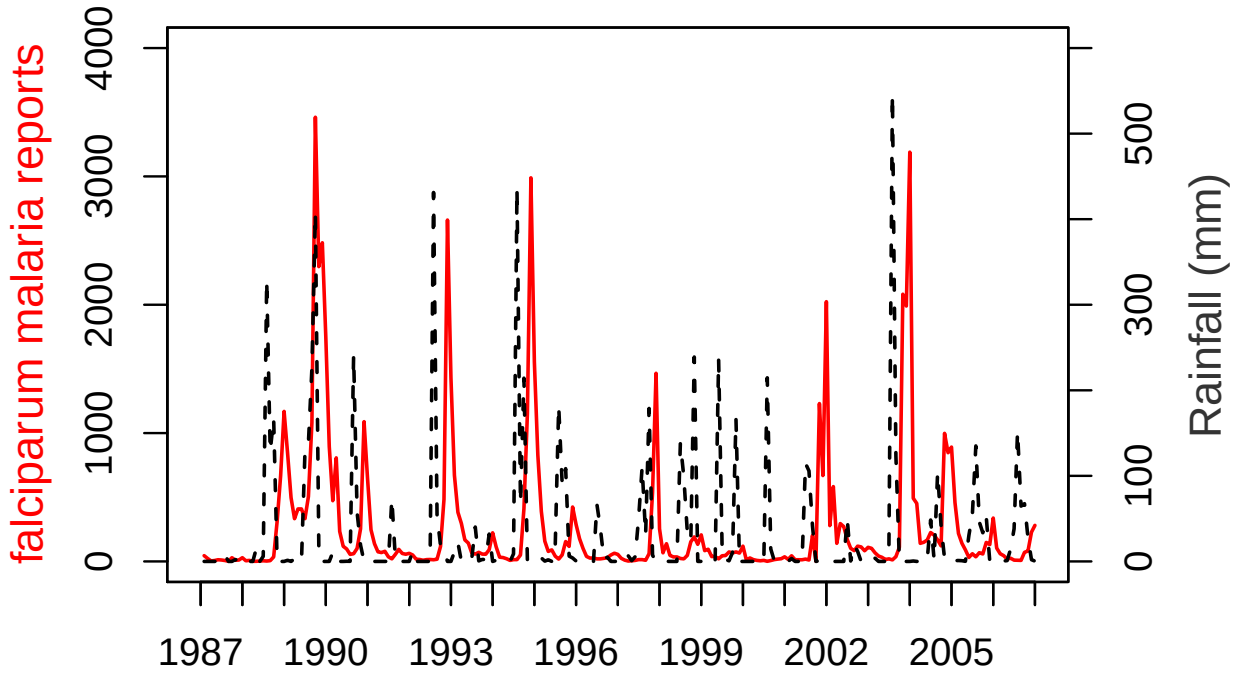
Malaria: A global challenge

- The Gates Foundation targets eradication. The previous Global Malaria Eradication Program (1955-1969) ultimately failed, though with some lasting local successes.
- Malaria transmission dynamics have much local variation (vectors and their ecology; human behaviors).

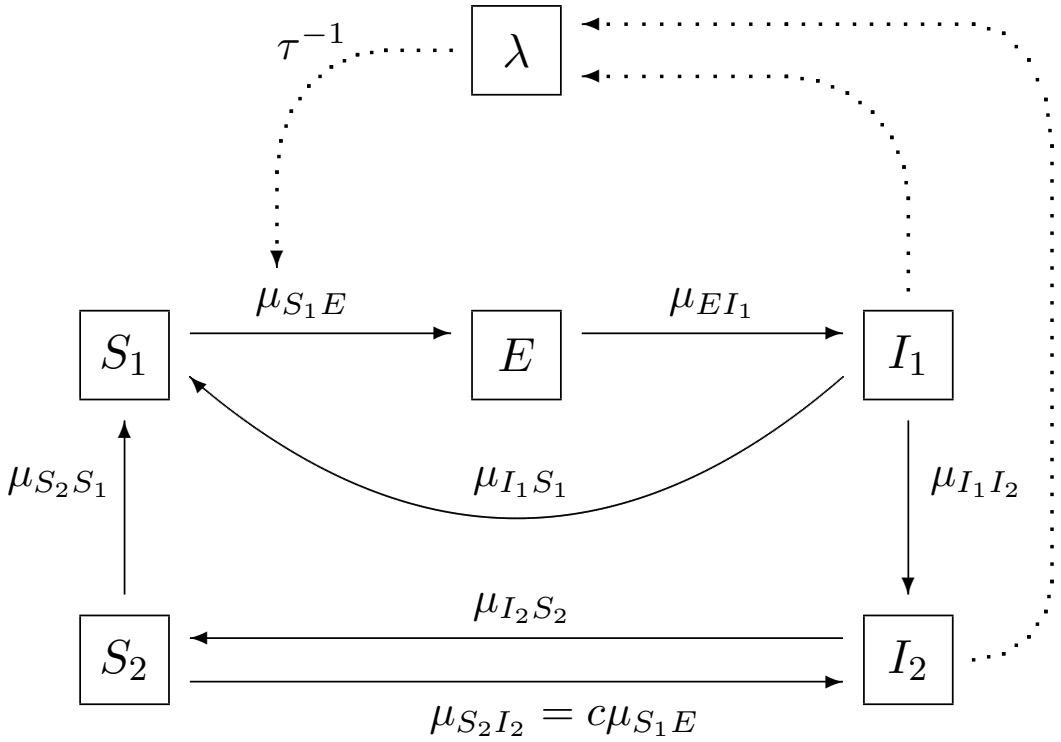
From the perspective of statistical methodology

- Despite the huge literature, no dynamic model of malaria transmission has previously been fitted directly to time series data.
- Difficulties include: Incomplete and complex immunity; dynamics in both mosquito and human stages; non-specific diagnosis via fever.
- Malaria is beyond the scope of methods developed for simpler diseases.

Malaria and rainfall in Kutch (an arid region of NW India)



- To what extent are cycles driven by immunity rising and falling? To what extent are they driven by rainfall?



Latent force of infection

Disease status of the human population

(Laneri et al, *PLoS Comp. Biol.*, 2010; Bhadra et al, *JASA*, 2011)

μ_{S_1E}, force of infection; λ, latent force of infection; S_1, fully susceptible humans; S_2 clinically protected (partially immune); I_1, clinically infected; I_2, asymptomatically infected.

Minimal complexity acceptable to scientists

$\approx$ **Maximal complexity acceptable to available data**

Model representation: coupled SDEs driven by Lévy noise

$$
\begin{aligned}
dS_1/dt &= \mu_{BS_1} P - \mu_{S_1E} S_1 + \mu_{I_1S_1} I_1 + \mu_{S_2S_1} S_2 - \mu_{S_1D} S_1 \\
dS_2/dt &= \mu_{I_2S_2} I_2 - \mu_{S_2S_1} S_2 - \mu_{S_2I_2} S_2 - \mu_{S_2D} S_2 \\
dE/dt &= \mu_{S_1E} S_1 - \mu_{EI_1} E - \mu_{ED} E \\
dI_1/dt &= \mu_{EI_1} E - \mu_{I_1S_1} I_1 - \mu_{I_1I_2} I_1 - \mu_{I_1D} I_1 \\
dI_2/dt &= \mu_{I_1I_2} I_1 + \mu_{S_2I_2} S_2 - \mu_{I_2S_2} I_2 - \mu_{I_2D} I_2 \\
d\lambda_i/dt &= (\lambda_{i-1} - \lambda_i)\, k\, \tau^{-1} \qquad \text{for } i = 1, \ldots, k \\
\mu_{S_1E}(t) &= \lambda_k(t) \\
\lambda(t) &= \lambda_0(t) = \frac{I_1(t) + qI_2(t)}{N(t)} \exp\left\{ \sum_{i=1}^{n_s} \beta_i s_i(t) + Z_t\beta \right\} \frac{d\Gamma}{dt}.
\end{aligned}
$$

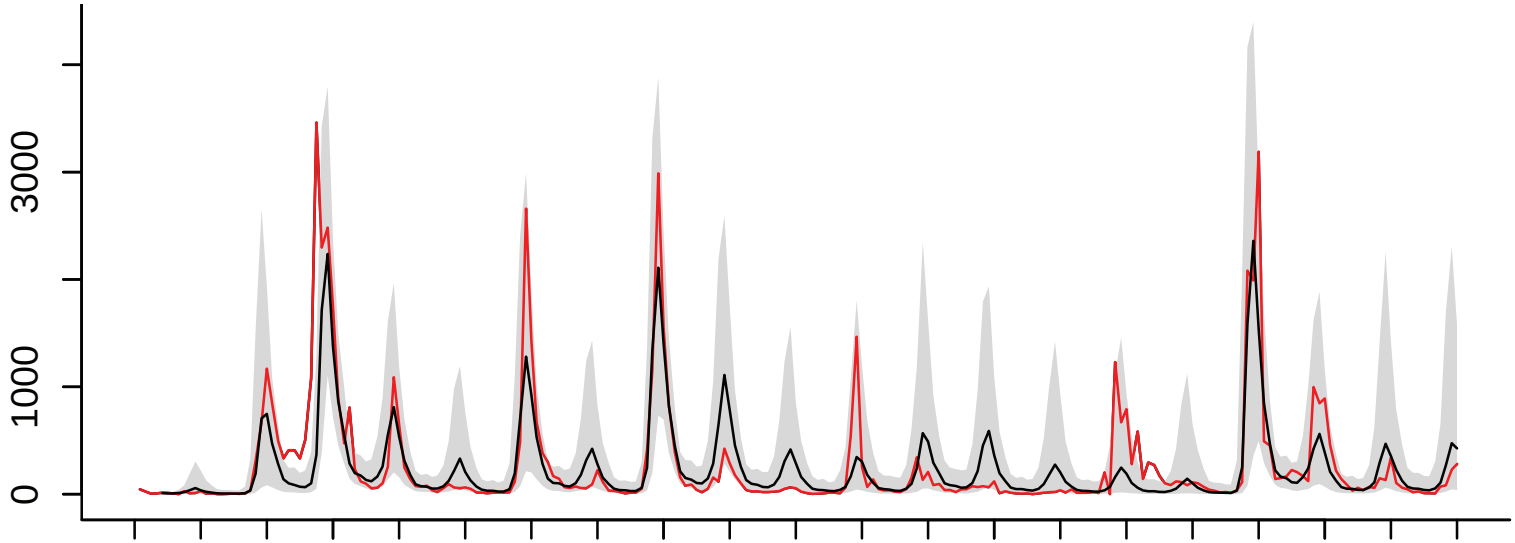
Z_t is a vector of climate covariates (here, rainfall).

$\sum_{i=1}^{n_s} \beta_i s_i(t)$ is a spline representation of seasonality.

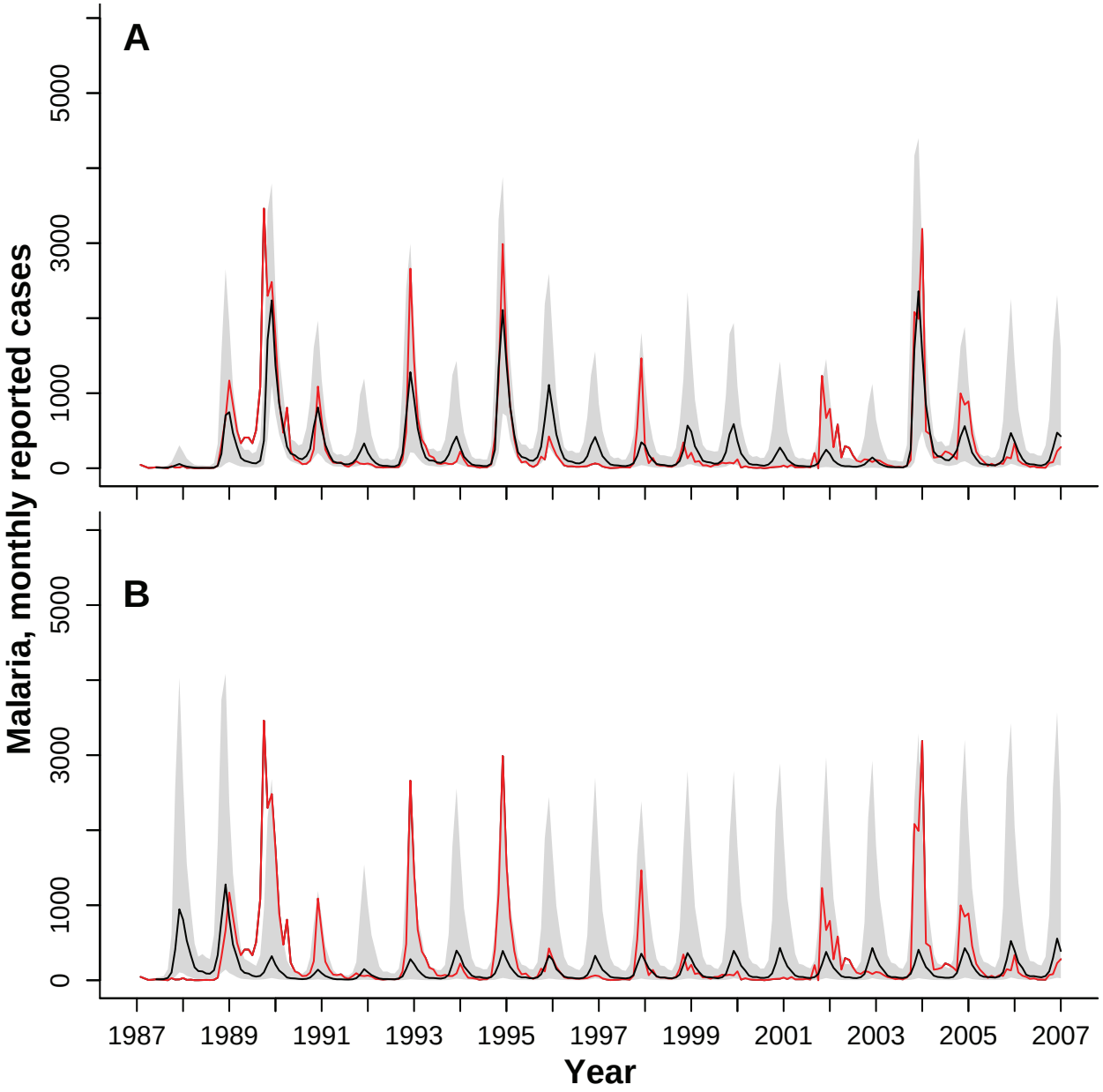
Parasite latency within the vector has mean τ and shape parameter k.

Conclusions from malaria data analysis

- Rainfall (with an appropriate delay and threshold) a critical role in determining interannual cycles.
- Immunity has a minor role, at a fast timescale (limiting annual peaks)



Simulations forward from 1987 to 2007, from the MLE, with prescribed rainfall. Showing monthly case reports (red), simulation median (black) and 10th to 90th percentiles (grey). Without rainfall, the model cannot come close to this.



Simulations forward from 1987 to 2007 from fitted models (A) with rainfall; (B) without rainfall. Showing monthly case reports (red), simulation median (black) and 10th to 90th percentiles (grey).

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.
- This motivates the consideration of discrete population, continuous time POMP models (Markov chains).

Over-dispersion in Markov chain models of populations

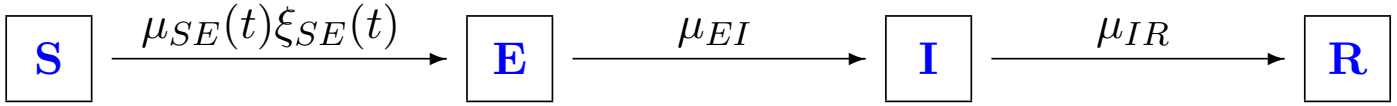
- **Remarkably, in the vast literatures on continuous-time individual-based Markov chains for population dynamics (e.g. applied to ecology and chemical reactions) no-one has previously proposed models capable of over-dispersion.**
- It turns out that the usual assumption that no events occur simultaneously 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 call such models *implicit*.** 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 in Bretó & Ionides (2011, *Stoc. Proc. Appl.*).

Measles: an exhaustively studied system

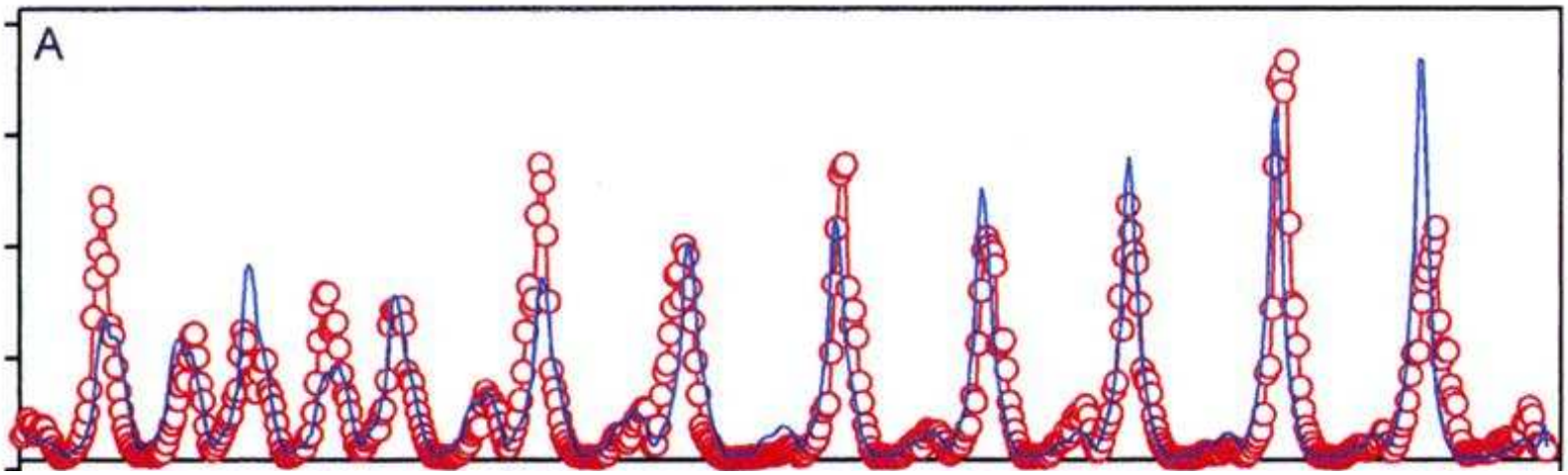
- Measles is simple: direct infection of susceptibles by infecteds; characteristic symptoms leading to accurate clinical diagnosis; life-long immunity following infection.



Susceptible $\longrightarrow$ Exposed (latent) $\longrightarrow$ Infected $\longrightarrow$ Recovered,

with noise intensity σ_{SE} on the force of infection.

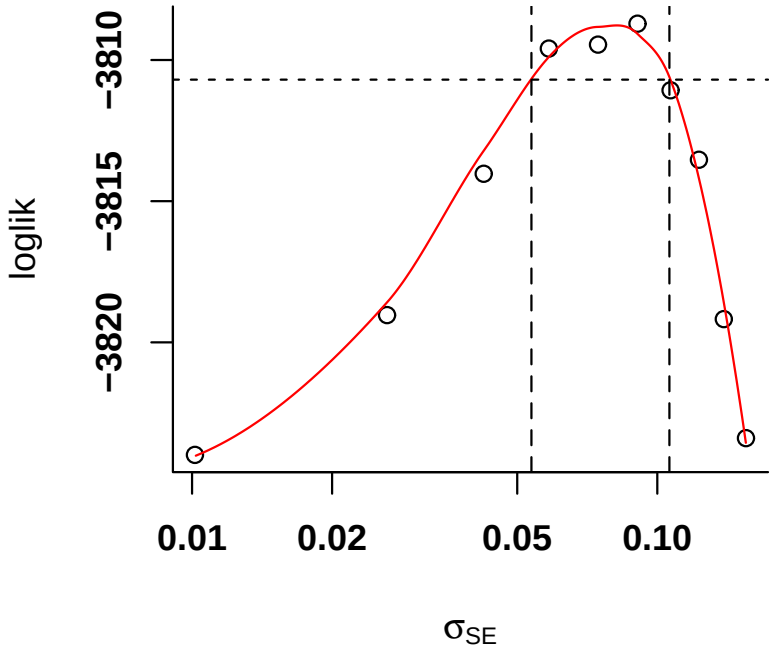
- Measles is still a substantial health issue in sub-Saharan Africa.
- A global eradication program is under debate.
- Comprehensive doctor reports in western Europe and America before vaccination (≈ 1968) are textbook data.

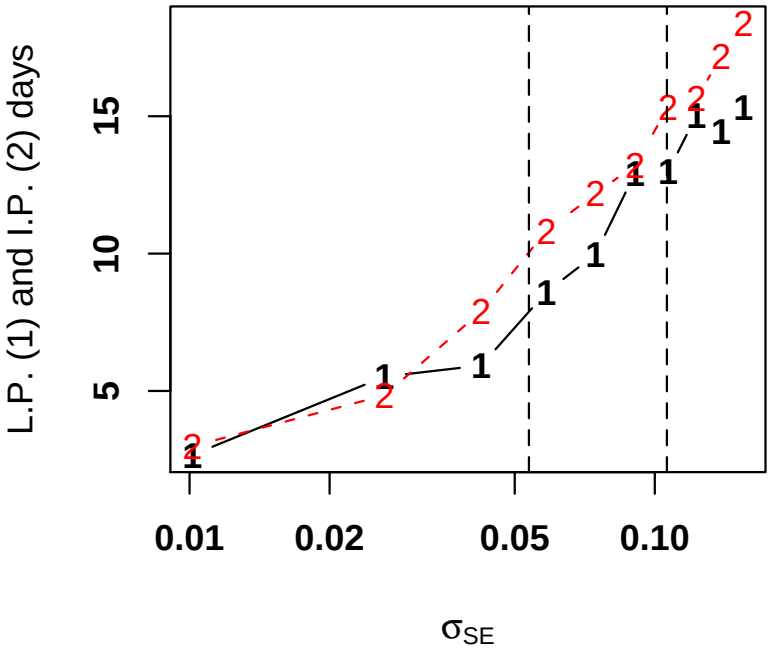


- Measles cases in London 1944–1965 (circles and red lines) and a deterministic SEIR fit (blue line) (from Grenfell *et al*, 2002).
- A deterministic fit, specified by the initial values in January 1944, captures remarkably many features.

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

- Profile likelihood for σ_{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 infection rate substantially improves the fit, and affects scientific conclusions (He et al, *JRSI*, 2010).





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 **model misspecification**? 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.

Longitudinal analysis via mechanistic models

- Multiple short time series can also be related to mechanistic hypotheses via POMP models, e.g.
  - ◇ within-host pathogen dynamics.
  - ◇ progression of chronic diseases.
  - ◇ behavorial studies.
- Plug-and-play methods facilitate the develpment and analysis of novel models, though alternatives such as MCMC and EM may also be viable for short time series.

Dynamics of sexual behaviors related to HIV risk

- Models with dynamic variation in sexual behaviors, with individuals having high and low risk episodes, can lead to much higher transmission than static heterogeneity models (Alam et al. 2010, *Epidemiology*)
- We used a POMP framework to re-analyze data from a cohort study of men who have sex with men (MSM), looking to quantify empirical evidence for dynamic variation.
- Behavioral data count reported contacts during 4 intervals of 6 months each. We studied either total contacts or disaggregation by contact type.

Data. Reported MSM contacts y_{ij} for individual i from time t_{j-1} to t_j.

Individual-level random effects:

$R_i \sim \Gamma(\mu_R, \sigma_R)$, rate of changing behavior.

$D_i \sim \Gamma(\mu_D, \sigma_D)$, over-dispersion of behavior.

Latent dynamic behavior process:

$X_i(t)$ is piecewise constant. After a $\text{Exponential}(R_i)$ time interval, $X_i(t)$ jumps to an independent $\Gamma(\mu_X, \sigma_X)$ value.

Measurement model:

$\mu_{ij} = \alpha^{j-1} \int_{t_{j-1}}^{t_j} X_i(t)\, dt$, where α models the decreasing contact rate.

$y_{ij} \sim \text{NegBin}(\mu_{ij}, D_i)$.

$\Gamma(\mu, \sigma)$ is the Gamma distribution with mean μ and variance σ^2.

$\text{NegBin}(\mu, D)$ is negative binomial with mean μ and variance $\mu + \mu^2 D^2$.

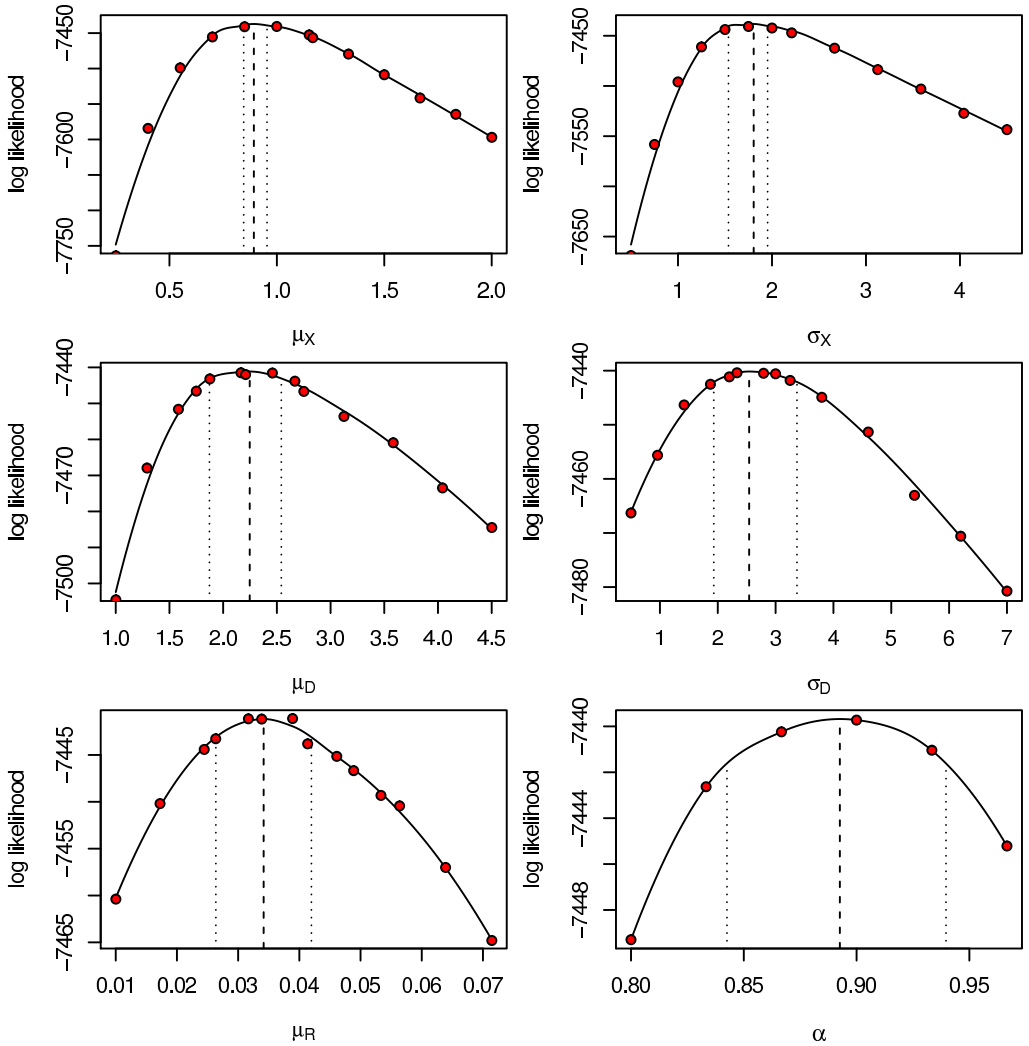
Maximum likelihood estimates for six models of total contacts

Model	μ_X	σ_X	μ_D	σ_D	μ_R	σ_R	α	Log Lik.
1	1.61	–	0.31	–	–	–	1	-10288.3
2	1.73	–	0.6	0.67	–	–	0.99	-9935.5
3	1.62	2.11	0.76	–	–	–	0.99	-9772.9
4	1.73	1.89	1.53	1.82	–	–	0.99	-9605.6
5	1.82	2.66	3.63	4.32	0.04	–	0.94	-9555.8
6	1.73	2.6	3.34	3.68	0.04	0.01	0.96	-9557.4

- $\mu_R \approx 0.04$ gives a mean episode duration of 25 months, implying ≈ 1 transition per individual over the study period.
- The likelihoods imply that the only unnecessary parameter is σ_R. The small inconsistency of -9557.4 with the nesting of model 5 within 6 is Monte Carlo error in likelihood optimization and/or evaluation.

Profile likelihoods show consistent results across contact types

- e.g., profiles for parameters of model 5, fitted to insertive contacts.



- As expected, data are highly informative about μ_X and σ_X. Other parameters are identified with adequate precision.

Conclusions and outstanding challenges

- Plug-and-play statistical methodology permits likelihood-based analysis of flexible classes of stochastic dynamic models.
- Many models of interest are beyond current algorithms & computational resources. Much work remains to be done!
- New data types (e.g., genetic sequence data on pathogens for some or all infected hosts) both enable and require the fitting of more complex models.
- Spatio-temporal models and individual-level models in large populations are typically beyond the scope of current plug-and-play methods, unless some special model structure can be exploited.

Collaborators: Yves Atchadé, Anindya Bhadra, Carles Bretó, Daihai He, Aaron King, Jim Koopman, Karina Laneri, Dao Nguyen, Mercedes Pascual, Ethan Romero-Severson, Manojit Roy.

Thank you!

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

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

References

Alam, S. J., Romero-Severson, E., Kim, J.-H., Emond, G., and Koopman, J. S. (2010). Dynamic sex roles among men who have sex with men and transmissions from primary HIV infection. *Epidemiology*, 21(5):669–675.

Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 72:269–342.

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.

Bretó, C. and Ionides, E. L. (2011). Compound markov counting processes and their applications to modeling infinitesimally over-dispersed systems. *Stochastic Processes and their Applications*, 121:2571–2591.

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.

He, D., Ionides, E. L., and King, A. A. (2010). Plug-and-play inference for disease dynamics: Measles in large and small towns as a case study. *Journal of the Royal Society Interface*, 7:271–283.

Ionides, E. L., Bhadra, A., Atchadé, Y., and King, A. A. (2011). Iterated filtering. *Annals of Statistics*, 39:1776–1802.

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.

Laneri, K., Bhadra, A., Ionides, E. L., Bouma, M., Yadav, R., Dhiman, R., and Pascual, M. (2010). Forcing versus feedback: Epidemic malaria and monsoon rains in NW India. *PLoS Computational Biology*, 6:e1000898.

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.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., and Stumpf, M. P. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6:187–202.

Wood, S. N. (2010). Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466:1102–1104.