Time series analysis of infectious disease dynamics:
State of the art and future challenges

Epidemic model hierarchies and model validation workshop

RAPIDD: Research and Policy in Infectious Disease Dynamics

Center for Infectious Disease Dynamics

Penn State University

9-11 April, 2008

Ed Ionides

The University of Michigan, Department of Statistics

Collaborators

Carles Breto (Universidad Carlos III de Madrid)

Daihai He (UM Ecology & Evolutionary Biology)

Aaron King (UM Ecology & Evolutionary Biology).

> ## Methodological challenge:
>
> A general approach to answering (and therefore asking) questions formulated as hypotheses about **partially observed nonlinear stochastic dynamic systems**.

- Partially observed: some variables unmeasured, others measured with error.

- Stochastic dynamic systems: random processes in discrete or continuous time, with discrete or continuous populations.

- Such situations arise throughout ecology and epidemiology. Maybe a characteristic feature of disease dynamics is that continuous time models are often appropriate.

> **Practical issue:**
>
> To develop approaches which are accessible to a wider community. Formal statistical inference via mechanistically-motivated models should aspire to be accessible to all those with a quantitative interest in disease dynamics.

- **Plug and play** inference methods are those which require only code to simulate from the dynamic system—and specifically do not require analytic expressions for transition probabilities (Bretó et al., 2008; Ionides et al., 2006).

- `pomp`: an R package for statistical analysis via partially observed Markov processes (King et al., 2007).

An issue of model complexity: Over-dispersion

- Adding arbitrary model complexity to an individual-based model (with only demographic stochasticity) ALWAYS gives a model for which the variance of the process increments equals the mean, over short time intervals (Bretó et al., 2008). We call this **equidispersion**.

- **over-dispersion** can be specified by one parameter: the intensity of stochastic variation in the force of infection (Bretó et al., 2008).

- Is it important to include over-dispersion in a dynamic model, to the extent that it is present in data? In generalized linear models (Poisson, binomial and negative binomial regression) it is widely recognized that (a) over-dispersion is common; (b) failure to allow for it can lead to highly misleading conclusions (McCullagh and Nelder, 1989).

Over-dispersion in disease dynamics

- Goodness of fit:

  - A model can capture the dynamics of a disease system well, and yet drastically fail to fit the data in a statistical sense.

  - Models which fit are necessary to make statistically correct forecasts, and help when making model-based inferences.

- Example: Susceptible–Exposed–Infected–Recovered (SEIR) models for measles.
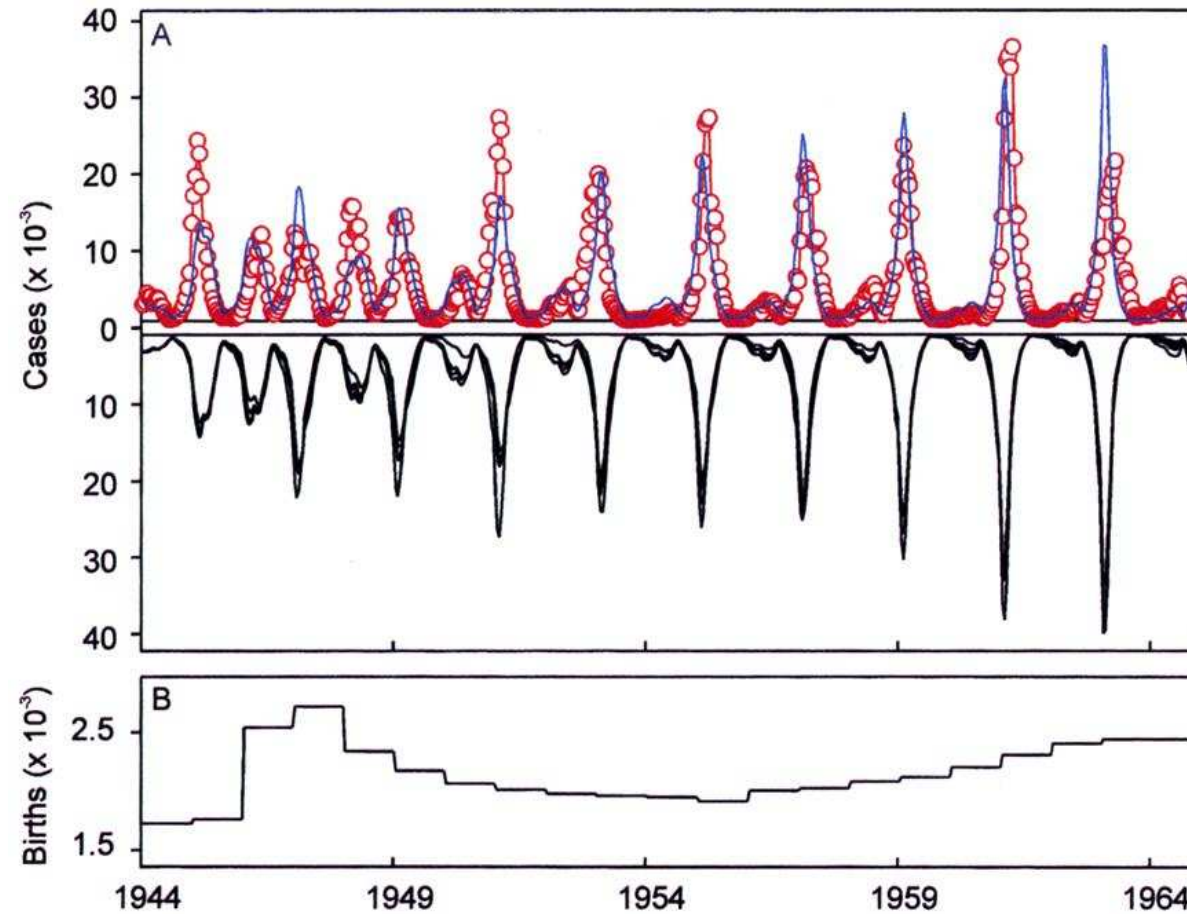
FIG. 1. (A) Measles incidence in 2-wk periods (in hundreds) in London from 1944 to 1965. The circles and the red line represent observed incidence (corrected for underreporting). The blue line represents the deterministic prediction from the TSIR model (using the susceptible and infected density in the first 2-wk period of 1944 as initial conditions). The black lines (and inverted scale) represent five stochastic realizations of the TSIR model. (B) The biweekly number of births (in hundreds) in London. The numbers are averaged within each year. The post-World War II baby boom in the late 1940s is associated with a period of annual cycles in measles incidence.

(Reproduced from Grenfell et al., 2002)

Three SEIR data analyses for measles

- **Bjørnstad et al. (2002)** studied a discrete-time SIR model via susceptible reconstruction and log-linear regression.

- **Cauchemez and Ferguson (2008)** studied a continuous-time SIR model, based on demographic stochasticity, reconstructing the unobserved variables via Monte Carlo Markov chain methods.

- **Bretó et al. (2008)** and **He et al. (2008, in preparation)** studied a continuous time SEIR model with demographic and environmental stochasticity, reconstructing unobserved variables via Sequential Monte Carlo (leading to a plug-and-play inference procedure).

## Over-dispersion in Bjørnstad et al. (2002)

- The model has negative binomial transition probabilities, corresponding to demographic stochasticity in continuous time.

- Parameters are estimated via log-linear least squares.

- Exactly the same log-linear least squares method could be used to estimate parameters for an over-dispersed negative binomial model. So the model fitting approach itself does not depend on the restriction to demographic stochasticity.

- Simulations from the fitted model (with demographic stochasticity alone) have some impressive similarities with data.

Over-dispersion in Cauchemez and Ferguson (2008)

- This Monte Carlo Markov chain approach, unlike Bjørnstad et al. (2002), makes explicit use of demographic stochasticity in the estimation procedure.

- That may explain the very low log likelihood of -7011 obtained for biweekly cases in London 1948–1964; compare with -2505 obtained for the same data by Bretó et al. (2008).

- Nevertheless, Cauchemez and Ferguson (2008) showed impressive simulation results for their fitted model. Does it even matter that the model does not fit well (as measured by likelihood) when it clearly captures the key dynamic features?

Over-dispersion in He et al. (2008)

- Profile likelihood for environmental stochasticity σ_{SE} and effect on estimated latent and infectious periods (London, 1950–1964).

- Variability of $\approx 5\%$ per year substantially improves the fit, and has consequences for parameter estimates.
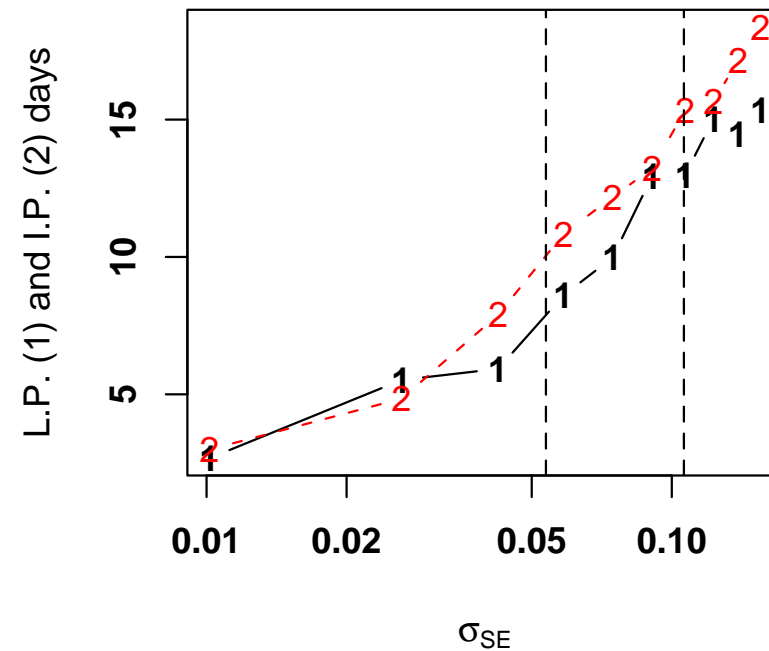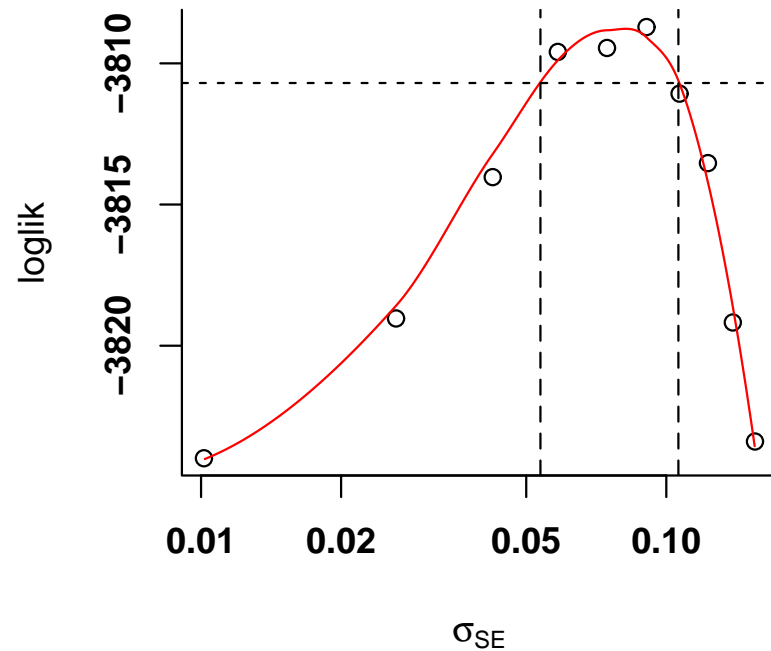
Interpretation of over-dispersion

- Social and environmental events (e.g., football matches, weather) lead to stochastic variation in rates: **environmental stochasticity**.

- Perhaps over-dispersion gives a way to measure non-homogeneous mixing.

  - Non-homogeneous mixing undoubtedly occurs.

  - It does not appear to be well-modeled by the tradition exponents (α) which have been repeatedly found to be close to one in data analyses (Bjørnstad et al., 2002; He et al., 2008).

- 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.

Beyond time series analysis

- Time series techniques are a launching pad for analysis of dynamic

  processes with more complex structures:

  - spatio-temporal.

  - dynamic trees and networks.

  - multiple related time series (i.e., longitudinal analysis with many

    observations on each subject).

Thank you!

These slides are available at

`www.stat.lsa.umich.edu/~ionides/pubs/rapidd08.pdf`

References

Bjørnstad, O. N., Finkenstädt, B. F., and Grenfell, B. T. (2002). Dynamics of measles epidemics: Estimating scaling of transmission rates using a time series SIR model. *Ecological Monographs*, 72(2):169–184.

Bretó, C., He, D., Ionides, E. L., and King, A. A. (2008). Time series analysis via mechanistic models. *In review, Annals of Statistics; pre-published at arxiv.org/abs/0802.0021*.

Cauchemez, S. and Ferguson, N. M. (2008). Likelihood-based estimation of continuous-time epidemic models from time-series data: Application to measles transmission in London. *J. R. Soc. Interface*, pre-published online, doi:10.1098/rsif.2007.1292.

Grenfell, B. T., Bjørnstad, 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. (2008). Inference on infectious disease dynamics: Measles as a case study. *In Preparation*.

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

King, A. A., Ionides, E. L., and Breto, C. (2007). `pomp`: Inference methods for partially-observed markov processes. R package, available at `http://cran.r-project.org/web/packages/pomp/index.html`.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman & Hall, London, 2nd edition.