Inference on spatiotemporal infectious disease dynamics

Edward Ionides
University of Michigan, Department of Statistics

Workshop on *Mathematical and Statistical Challenges in Post-Pandemic Epidemiology and Public Health*
Casa Matemática Oaxaca
June 16, 2025

Slides are at `https://ionides.github.io/talks/cmo25.pdf`

Joint work with Ning (Patricia) Ning, Jesse Wheeler, Kidus Asfaw, Jifan Li, Joonha Park, Aaron King, Mercedes Pascual

Question to be addressed

1. When can we carry out full-information likelihood-based inference on nonlinear non-Gaussian spatiotemporal partially observed Markov process (SpatPOMP) models? In particular, models for networks of interacting biological population dynamics.
2. We introduce the **iterated block particle filter**, currently the most effective algorithm in the `spatPomp` R package.
3. Bonus question: How do we know if our model is statistically adequate, or needs more work?
   - Hint: we need **performance benchmarks**, i.e., comparison with non-mechanistic machine learning methods. If there is performance shortfall, we need diagnostics to figure out where and how.

Inference challenges in population dynamics: A brief history

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 spatiotemporal dynamics**.
7. Studying population dynamics via genetic sequence data.

1–5 are largely solved, from a methodological perspective.
6 is our immediate topic.
7 is exciting but not the focus of this talk.

Reviews: Bjornstad & Grenfell (*Science*, 2001); Grenfell et al (*Science*, 2004)

Desiderata

- Consideration of arbitrary dynamic models. The limitations should be our scientific creativity and the information in the data.

  Hence, **plug-and-play** methods which need a simulator from the model but not nice closed-form expressions for densities.

- Statistically efficient inference, to extract all the information in the data.

  Hence, **likelihood-based** methods.

Fitting mechanistic models to time series

- **Iterated particle filtering** via `mif2` in the R package `pomp` (King et al., 2016) enables plug-and-play likelihood-based inference for nonlinear, non-Gaussian, partially observed Markov process (POMP) models.
- Allows consideration of arbitrary POMP models for low-dimensional time series data.
- Accessible for Masters-level statisticians:

  `https://ionides.github.io/531w24/`

- The science may be hard, but the statistics is becoming routine.

The curse of dimensionality

- Particle filter (PF) methods fail for high-dimensional systems. They scale exponentially badly.
- Algorithms with improved scalability include:
  Bagged filters (BF, IBF)
  Ensemble Kalman filter (EnKF, IEnKF)
  Guided intermediate resampling filter (GIRF, IGIRF)
  Block particle filter (BPF, IBPF)
- Filters estimate latent states and evaluate the likelihood.
- Iterated filters estimate parameters using stochastic parameter perturbations.
- These algorithms are all implemented in the `spatPomp` R package.

spatPomp: an R package for SpatPOMP models

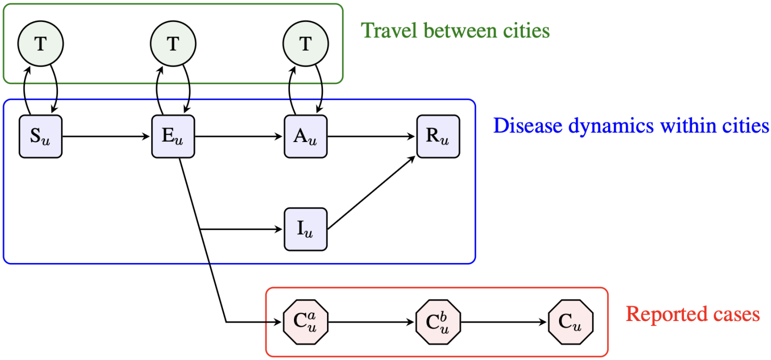
Asfaw et al (2024, *J. Open Source Software*)
https://github.com/spatPomp-org

- Permits specification of arbitrary non-linear non-Gaussian spatiotemporal POMP (SpatPOMP) models.
- spatPomp spatial structure to pomp.
- Includes Monte Carlo algorithms designed to address the **curse of dimensionality**.
- Algorithms compile via C to assist performance.
- Statistically efficient, likelihood-based inference despite methods being simulation-based.

Future goals: add GPU and autodiff support via pypomp,
https://github.com/pypomp

COVID in 373 cities in China, Jan 10 to Feb 8, 2020

- Metapopulation data were used to infer the fraction of asymptomatic cases and their contagiousness (Li et al, *Science*, May 2020).
- SEIR (susceptible-exposed-infected-removed) model with asymptomatics, reporting delay, and coupling based on cell phone data.
- Li et al (2020) used iterated EnKF for inference.
- The time interval covers the initial China lockdown.
- We present re-analysis of this model and data (Li et al, J. Roy. Soc. Interface, 2024).



- $U = 373$ cities (spatial units) labeled $u = 1, \ldots, U$.
- Reportably infectious individuals, I_u for city u, are included in the delayed reporting compartment, $\mathrm{C}^{\mathrm{a}}_{\mathrm{u}}$.
- An individual arriving at C_u is a case report for city u.
- Individuals in A_u are not reportable and transmit at a reduced rate.
- Travel occurs to and from T, based on 2018 data from Tencent.

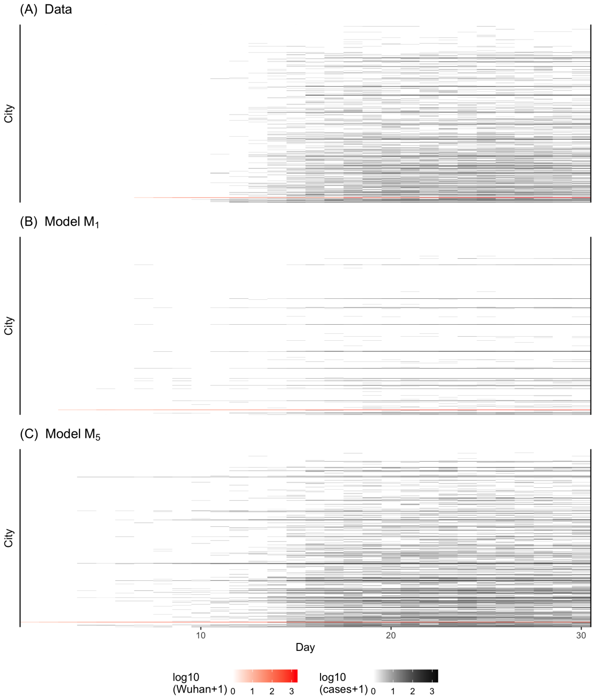
Log-likelihood values for six models

	loglik	df	description
M_1	$-\infty$	10	SEAIR model, parameter values and mobility of Li et al (2020)
M_2	-14240.5	10	Adjusted mobility and measurement in M_1
M_3	-11257.9	374	Independent identically distributed negative binomial
M_4	-10825.3	375	Autoregressive negative binomial
M_5	-9088.2	12	Adding overdispersed dynamics to M_2
M_6	-9116.5	10	Latent and infectious periods unchanged by lockdown in M_5

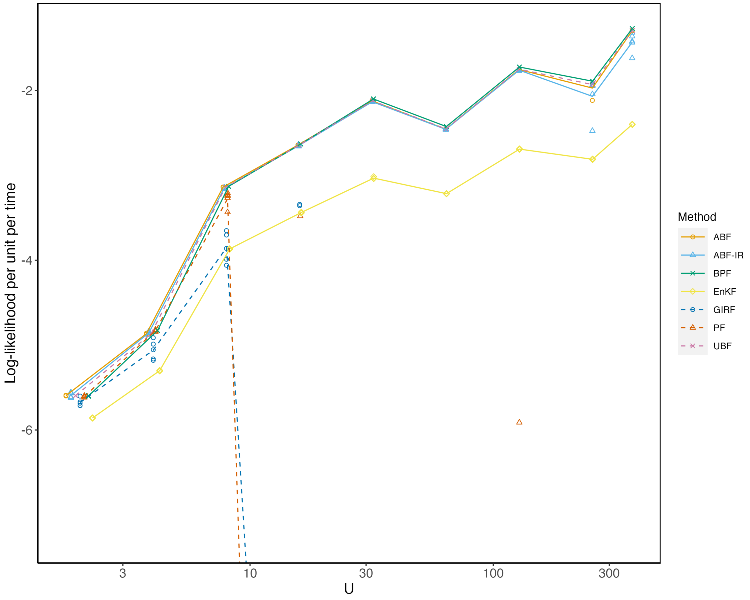
Model comparisons by log-likelihood. The degrees of freedom (df) is the number of estimated parameters.

Data analysis progression

- Some data errors meant that the original data was incompatible with the model.
- Tracked down by assessing conditional log-likelihood for each unit and each time.
- Even then, the likelihood was poor compared to non-mechanistic benchmarks. Including over-dispersion fixed that.
- Some weak identifiability remained. extra assumptions can lead to plausible parameter estimates without much loss of model fit.
- Less severe outliers remain, e.g., Wuhan on day 18.



(A) Data
City
(B) Model M_1
City
(C) Model M_5
City
10
20
30
Day
log10
(Wuhan+1)
0 1 2 3
log10
(cases+1)
0 1 2 3



- Block particle filter (BPF) beats alternatives for this COVID model.
- Similar results hold for other models (Ionides et al, *JASA*, 2023).

More on the block particle filter (BPF)

- BPF also worked quickly, easily and reliably on a measles metapopulation, as well as various toy benchmark problems (Ionides et al, *JASA*, 2023).
- BPF has theoretical support in some situations (Rebeschini & Van Handel, *Annals of Applied Probability*, 2015).
- This motivated us to develop an iterated BPF (IBPF) for parameter estimation.
- IBPF has theoretical guarantees similar to BPF (Ning & Ionides, *JMLR*, 2023).
- BPF was independently proposed as the "factored particle filter" by Ng et al (2002).

Particle filter (PF)

Evolutionary analogy

Mutation

$\downarrow$

Fitness

$\downarrow$

Natural selection

Particle filter algorithm

Predict: stochastic dynamics

$\downarrow$

Measurement: weight

$\downarrow$

Filter: resample

- PF is an evolutionary algorithm with good mathematical properties: an unbiased likelihood estimate and consistent latent state distribution.

Block particle filter (BPF)

- Blocks are a partition of the metapopulation units.
- For measles, we use each city as a block.

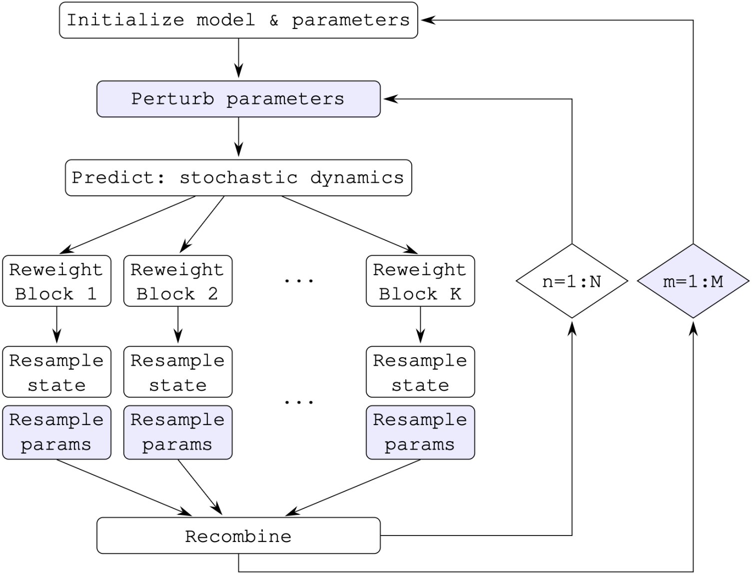
Evolutionary analogy	**Block particle filter**
Mutation	**Predict: stochastic dynamics**
$\downarrow$	$\downarrow$
Fitness for each chromosome	**Measurement: weight for each block**
$\downarrow$	$\downarrow$
Natural selection for each chromosome	**Filter: resample for each block**
$\downarrow$	$\downarrow$
Recombine chromosomes	**Recombine blocks**

- Blocks are segments of the full state which can be reassorted between particles at the resampling step.

Comments on the Ensemble Kalman Filter (EnKF)

- EnKF is more dependent on approximate Gaussianity than is sometimes supposed.
- The Gaussian-inspired update rule is similar to the extended Kalman filter (EKF), which has largely been superseded by particle filter methods for low-dimensional nonlinear biological dynamics.
- Simple systems can defeat EnKF: the linear Gaussian update is helpless when data inform the conditional variance rather than the conditional mean.
- Big systems need computationally tractable analysis. EnKF may sometimes be the best solution available, but be aware of its limitations.

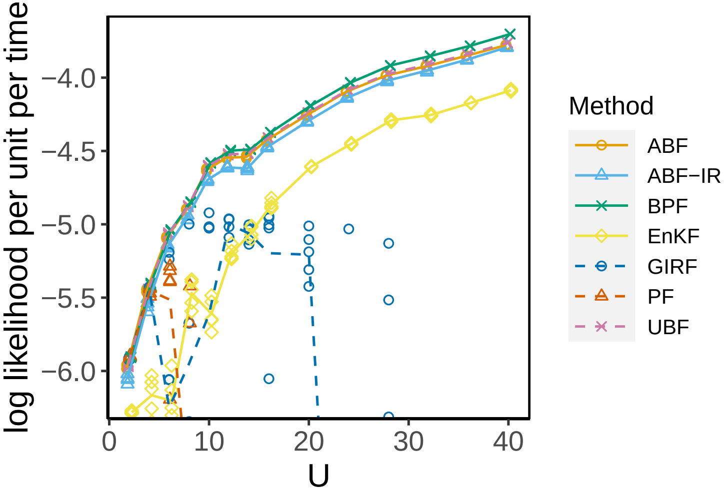
An iterated block particle filter for parameter estimation



Practical inference using IBPF

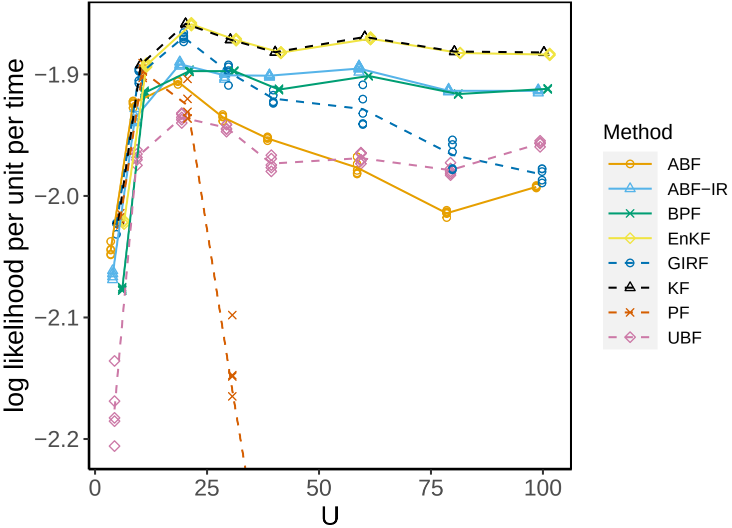
1. Monte Carlo adjusted profile likelihood (Ionides et al., 2017) obtains confidence intervals that accommodate Monte Carlo error.
2. Comparing the log-likelihood with an autoregressive model (or other simple statistical model) provides a check of model fit.
3. Comparing the block log-likelihood against the benchmark provides insight into problematic units.
4. Comparing the conditional log-likelihood for each observation against the benchmark helps to identify outliers.
5. Two recent case studies (Wheeler et al, 2024; Li et al, 2024) demonstrate data analysis using IBPF. Code and data are provided via R packages extending `spatPomp`.

Filtering U units of a coupled measles SEIR model



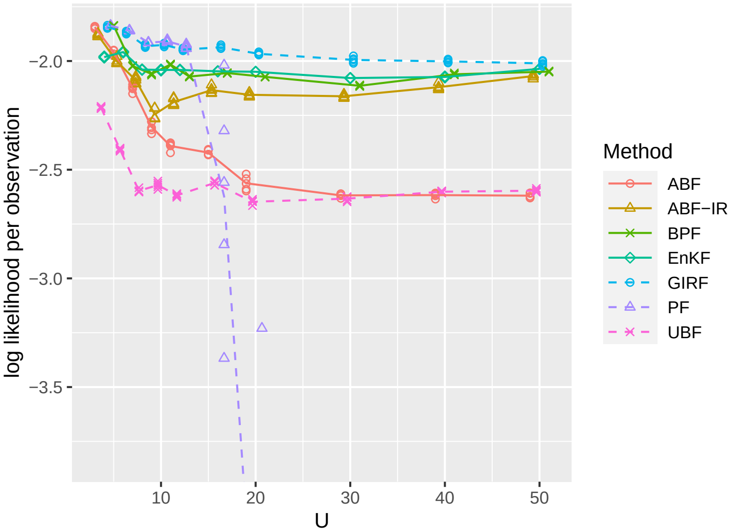
Simulated data using a gravity model with geography, demography and transmssion parameters corresponding to UK pre-vaccination measles (Ionides et al, JASA, 2023).

Filtering U-dimensional correlated Brownian motion



$$\mathrm{Cov}\big(X_{u,n} - X_{u,n-1}, X_{\tilde{u},n} - X_{\tilde{u},n-1}\big) \sim 0.4^{|u-\tilde{u}|}$$

Filtering U units of Lorenz 96 toy atmospheric model



$$dX_u(t) = \Big\{X_{u-1}(t)\big(X_{u+1}(t) - X_{u-2}(t)\big) - X_u(t) + F\Big\}dt + \sigma\, dB_u(t)$$

References I

Asfaw, K., Park, J., King, A. A., and Ionides, E. L. (2024a). spatPomp: An R package for spatiotemporal partially observed Markov process models. *Journal of Open Source Software*, 9:7008.

Asfaw, K., Park, J., King, A. A., and Ionides, E. L. (2024b). A tutorial on spatiotemporal partially observed Markov process models via the R package spatPomp. *arXiv:2101.01157v4*.

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

Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L. N., Daly, J. M., Mumford, J. A., and Holmes, E. C. (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*, 303:327–332.

Ionides, E. L., Asfaw, K., Park, J., and King, A. A. (2023). Bagged filters for partially observed interacting systems. *Journal of the American Statistical Association*, 118(542):1078–1089.

References II

Ionides, E. L., Breto, C., Park, J., Smith, R. A., and King, A. A. (2017). Monte Carlo profile confidence intervals for dynamic systems. *Journal of the Royal Society Interface*.

Ionides, E. L., Ning, N., and Wheeler, J. (2024). An iterated block particle filter for inference on coupled dynamic systems with shared and unit-specific parameters. *Statistica Sinica*, 34:1145–1166.

King, A. A., Nguyen, D., and Ionides, E. L. (2016). Statistical inference for partially observed Markov processes via the R package pomp. *Journal of Statistical Software*, 69:1–43.

Li, J., Ionides, E. L., King, A. A., Pascual, M., and Ning, N. (2024). Inference on spatiotemporal dynamics for coupled biological populations. *Journal of the Royal Society Interface*, 21(216):20240217.

References III

Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., and Shaman, J. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*, 368(6490):489–493.

Ng, B., Peshkin, L., and Pfeffer, A. (2002). Factored particles for scalable monitoring. *Proceedings of the 18th Conference on Uncertainty and Artificial Intelligence*, pages 370–377.

Ning, N. and Ionides, E. L. (2023). Iterated block particle filter for high-dimensional parameter learning: Beating the curse of dimensionality. *Journal of Machine Learning Research*, 24:1–76.

Park, J. and Ionides, E. L. (2020). Inference on high-dimensional implicit dynamic models using a guided intermediate resampling filter. *Statistics and Computing*, 30(5):1497–1522.

References IV

Rebeschini, P. and van Handel, R. (2015). Can local particle filters beat the curse of dimensionality? *The Annals of Applied Probability*, 25:2809–2866.

Wheeler, J., Rosengart, A., Zhuxun, J., Tan, K. H. E., Treutle, N., and Ionides, E. L. (2024). Informing policy via dynamic models: Cholera in Haiti. *PLoS Computational Biology*, 20:e1012032.