

# Inference for metapopulation dynamics

Edward Ionides

University of Michigan, Department of Statistics

Multiscale Microbial Communities

Dynamical Models, Ecology, and One Health

February 22, 2022

Slides are at <https://ionides.github.io/talks/imsi22.pdf>

Joint work with

Kidus Asfaw, Ning Ning, Joonha Park and Aaron King

# Inference challenges in population dynamics

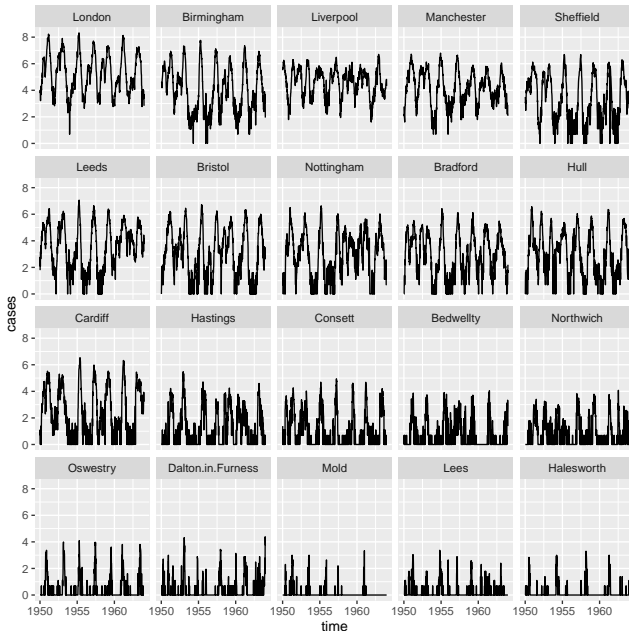
- 1 Combining measurement noise and process noise.
- 2 Including covariates in mechanistically plausible ways.
- 3 Continuous time models.
- 4 Modeling and estimating interactions in coupled systems.
- 5 Dealing with unobserved variables.
- 6 **Modeling spatial-temporal dynamics.**
- 7 **Studying population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

1–6 are from Bjornstad & Grenfell (*Science*, 2001).

7 is from Grenfell et al (*Science*, 2004).

1–5 are largely solved, from a methodological perspective.

# Example: Pre-vaccination measles in England & Wales



# Time series data, panel data & spatiotemporal data

- Looking at one unit (town) is **time series analysis**.
- Joint modeling of a few units (say, 2 or 3) is **multivariate time series analysis**.
- Analysis of many time series, without consideration of dynamic interactions, is **panel data analysis**.
- Allowing for coupling between units, we get **spatiotemporal analysis**, which in our context is **metapopulation analysis**.

Question: When should we avoid inference for spatiotemporal models?  
When do we need to consider coupling? How?

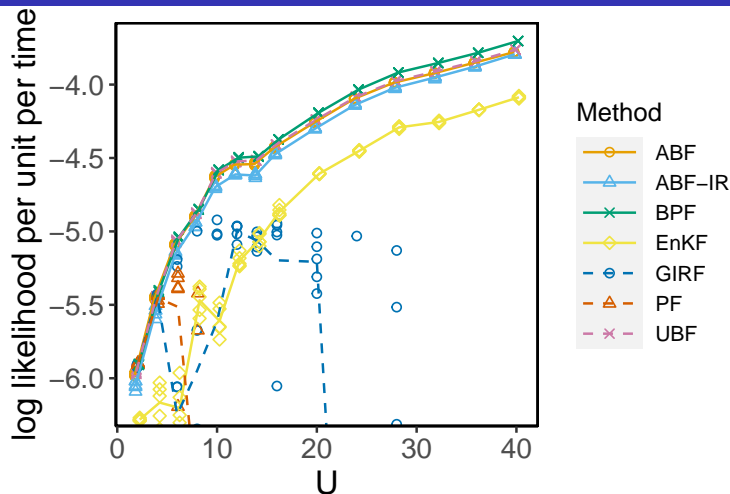
- We want to be able to fit arbitrary dynamic models. The limitations should be our scientific creativity and the information in the data.
- In practice, that means using **plug-and-play** methods which need a simulator from the model but not nice closed-form expressions for densities.
- We want statistically efficient inference, to extract all the information in the data.
- In practice, that means using likelihood-based methods.
- In the time series case, iterated particle filtering (IF2) implemented in the R package `pomp` enables Masters-level statisticians to do this (<https://ionides.github.io/531w22/>). The science may be hard, but the statistics is becoming routine.

- To investigate epidemiological dynamics in multiple cities, one can consider each city independently, perhaps modeling a background immigration rate of infections for each city.
- **Decoupling** leads to panel data analysis, by assumption. Iterated filtering methods extend to panel data (Breto et al, *Journal of the American Statistical Association*, 2019).
- We must decide which parameters should be modeled as **shared** vs **unit-specific**.
- The consequences of decoupling are becoming easier to study with the development of statistical inference methods for coupled systems, i.e., metapopulation dynamics.

# The curse of dimensionality

- Particle filter (PF) methods are effective for inference on low-dimensional nonlinear partially observed stochastic dynamic systems. They scale exponentially badly.
- Extending the successes of particle filter methods from time series data to metapopulation data is becoming possible.
- Algorithms under consideration:
  - 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.
- Each filter has an iterated version which estimates parameters by repeated filtering using stochastic parameter perturbations.
- These algorithms are all implemented in an R package, spatPomp.

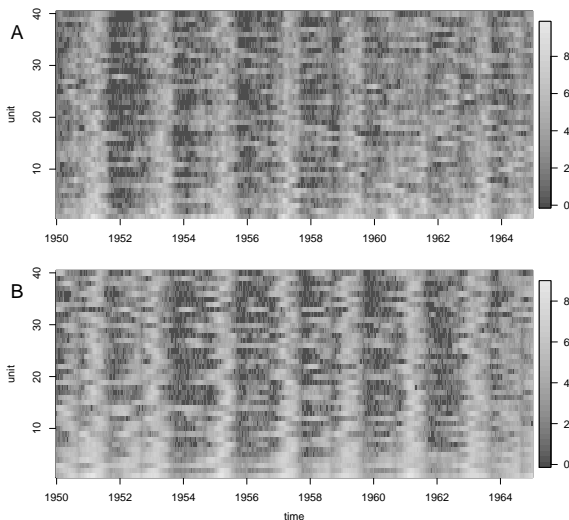
# Filtering $U$ units of a coupled measles SEIR model



Simulated data using a gravity model with geography, demography and transmission parameters corresponding to UK pre-vaccination measles (Ionides et al, JASA, 2021).



# $U = 40$ units for a coupled measles SEIR model



**A.** Simulated Susceptible-Exposed-Infected-Recovered dynamics coupled with a gravity model (log of biweekly reported cases).

**B.** Measles UK pre-vaccination case reports for the 40 largest cities.

# Parameters for the measles model

- Seasonal transmission: mean and amplitude, using school term for contact rate.
- Durations of latency and infectious period.
- Initial values: fraction susceptible, latent and infectious.
- Cohort effect: all births in an age cohort start school in September.
- Inhomogenous mixing coefficient.
- Measurement fraction.
- Transport model gravity constant.
- Dynamic noise (process overdispersion).
- Measurement overdispersion.

## More on the block particle filter

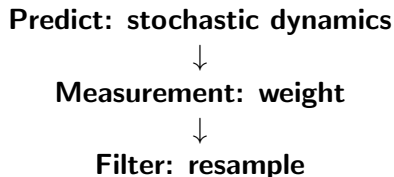
- BPF worked quickly, easily and reliably on our measles model filtering experiments.
- This motivated us to develop an IBPF for parameter estimation.
- BPF has theoretical support in some situations (Rebeschini & Van Handel, *Annals of Applied Probability*, 2015).
- BPF was independently proposed as the “factored particle filter” by Ng et al (2002, *Proc. 18th Conference on Uncertainty and Artificial Intelligence*) but not widely popularized.

# Particle filter (PF)

## Evolutionary analogy



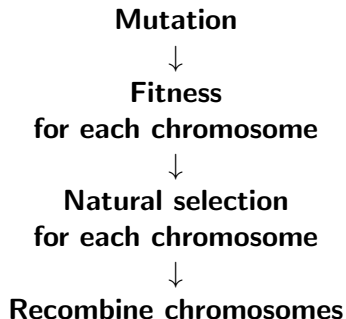
## Particle filter algorithm



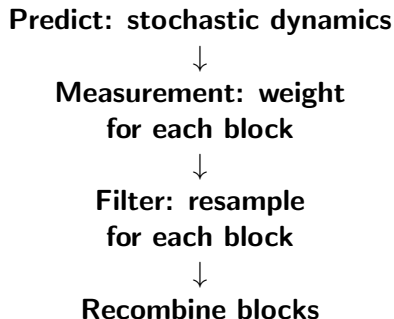
- PF is an evolutionary algorithm with good mathematical properties: an unbiased likelihood estimate and consistent latent state distribution.

# Block particle filter (BPF)

## Evolutionary analogy

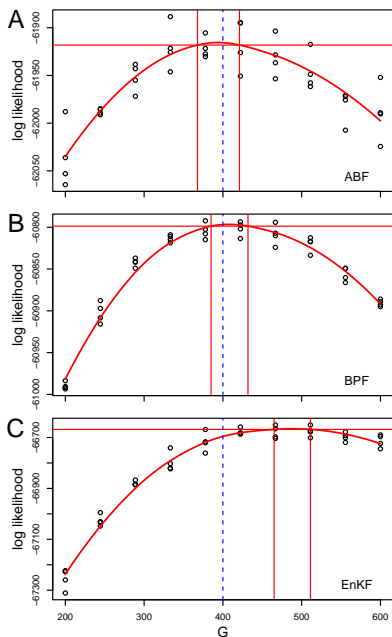


## Block particle filter



- Blocks in BPF allow recombination (reassortment of chromosomes in sexual reproduction) in the evolutionary analogy.
- Blocks are a partition of the metapopulation units. Our experiments suggest treating each sub-population (i.e., city) as a block.

# Measles likelihood slices for coupling parameter, $G$



Simulating 15 year of data from  $U = 40$  cities for the measles model. Slice likelihood, varying  $G$  with other parameters fixed at the truth.

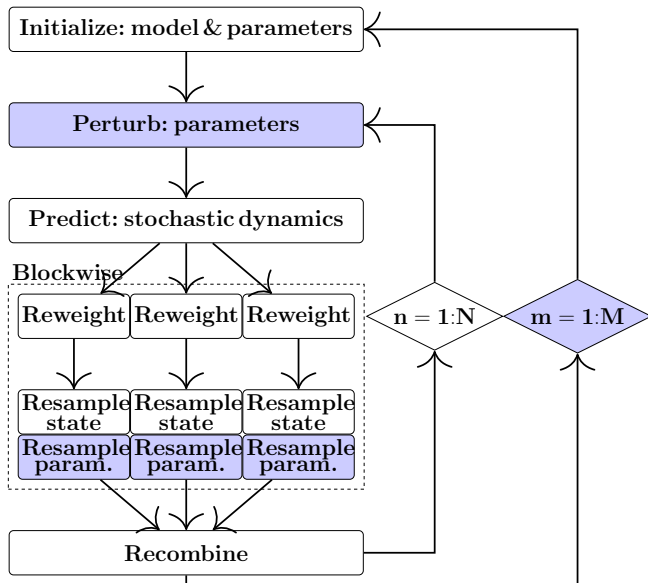
**A.** Evaluation using adapted bagged filter (ABF).

**B.** Evaluation using block particle filter (BPF).

**C.** Evaluation using EnKF.

Results from Ionides et al (2021, *JASA*). We computed a slice due to lack of good optimization algorithms to compute a profile.

# An iterated block particle filter for parameter estimation



# Scalability needed for practical inference

## Large numbers of parameters

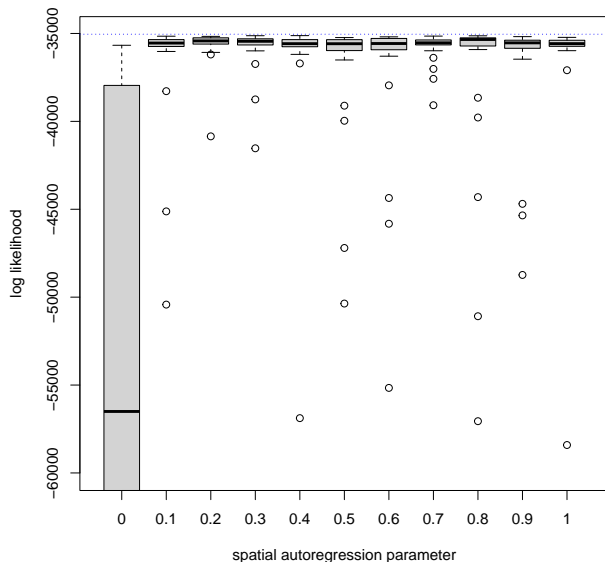
- Initial conditions will typically have to be estimated for each unit.
- Various dynamic parameters and measurement parameters (e.g., reporting rate) may also need to be unit-specific to obtain a statistical fit to the data.
- Working with hundreds of estimated parameters raises additional challenges on top of the high-dimensional coupled dynamics.

## A moderate numbers of spatial units is enough to open new possibilities.

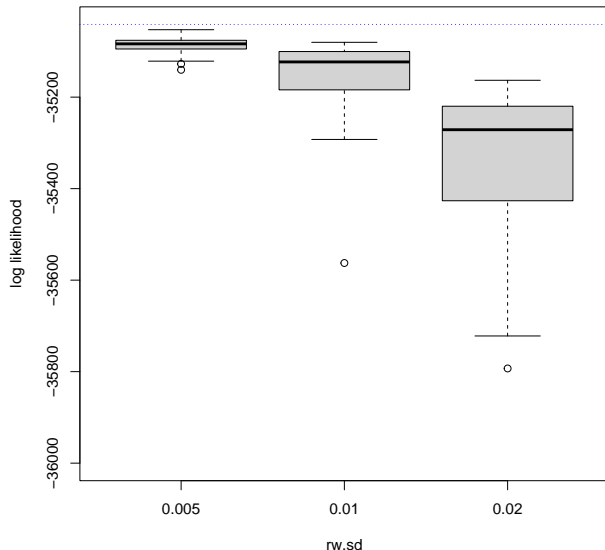
- As soon as dimension exceeds capabilities of a particle filter (say,  $U = 5$ ) we are in situations where likelihood-based inference for general models has been inaccessible.
- 10-100 coupled units is our target problem size.
- Larger problems will need numerical approximations (e.g., EnKF). Exact Monte Carlo methods help study the effect of these approximations.



# Auto-regression of spatial perturbations for shared parameters



Random perturbations must be smaller to match larger number ( $20 \times 13$ ) of parameters



## Likelihood on benchmark problems with 20 towns

	Simulation	UK measles	$p$
Benchmark	-35041	-40345	
4/13 parameters unit-specific	-35052	-43069	$4 \times 20 + 9$
12 parameters unit-specific	-35115	-40641	$12 \times 20 + 1$

- Simulated data: benchmark is likelihood at truth. Optimization used 10hr on one node.
- Actual data: benchmark is likelihood from uncoupled model with all parameters unit-specific, and a parameter for immigration rate of new cases. Optimization used  $2 \times 10$ hr on one node.

# Future work

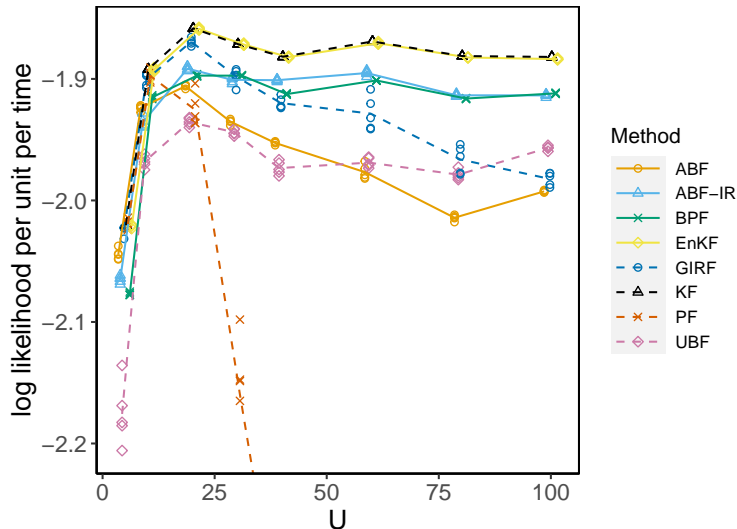
- We are getting close to the point where we can carry out likelihood-based inference for a flexible class of metapopulation models for measles. Flexibility supports generation and testing of scientific hypotheses.
- Measles was previously a motivating model system for POMP methods for single populations.
- Many systems in ecology, epidemiology and elsewhere could be studied in a SpatPOMP framework.

# References I

- Bjørnstad, O. N. and Grenfell, B. T. (2001). Noisy clockwork: Time series analysis of population fluctuations in animals. *Science*, 293:638–643.
- Bretó, C., Ionides, E. L., and King, A. A. (2019). Panel data analysis via mechanistic models. *Journal of the American Statistical Association*, 115:1178–1188.
- 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. (2021). Bagged filters for partially observed interacting systems. *Journal of the American Statistical Association*, 0(ja):1–33.
- 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.

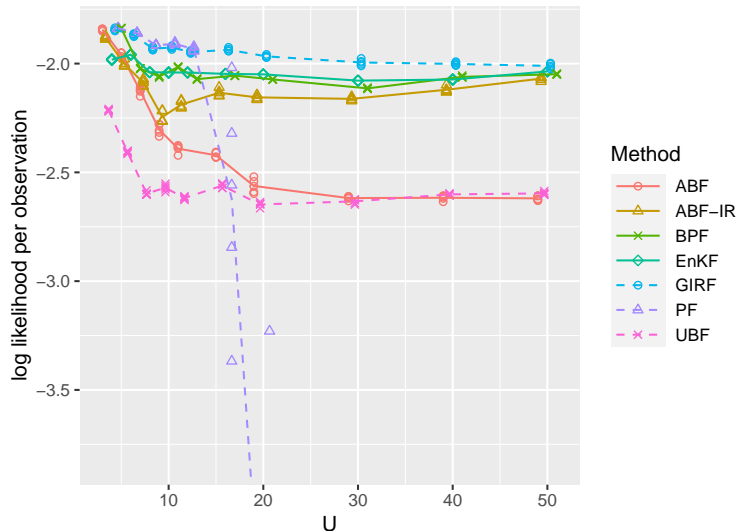
Rebeschini, P. and van Handel, R. (2015). Can local particle filters beat the curse of dimensionality? *The Annals of Applied Probability*, 25:2809–2866.

# Filtering $U$ -dimensional correlated Brownian motion



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

# Filtering $U$ units of Lorenz 96 toy atmospheric model



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