Simulation-based inference for partially observed spatiotemporal systems

Bayes4Health/CoSinES workshop
September 19, 2019

Edward Ionides
University of Michigan, Department of Statistics

Joint work with Joonha Park

Slides online:
`http://dept.stat.lsa.umich.edu/~ionides/talks/uk19.pdf`

Arxiv preprint: `https://arxiv.org/abs/1708.08543`

We seek a “mechanistic” approach to modeling and inference for dynamic systems. What does this mean?

- Write down equations, based on scientific understanding of a dynamic system, which describe how it evolves with time.
- Further equations describe the relationship of the state of the system to available observations on it.
- Data analysis via mechanistic models concerns drawing inferences from the available data about the hypothesized equations.
- Questions of general interest: Are the data consistent with a particular model? If so, for what range of values of model parameters? Does one mechanistic model describe the data better than another?
- A defining principle: the model structure should be chosen based on scientific considerations, rather than statistical convenience.

Statistical challenges for nonlinear mechanistic modeling in ecology and epidemiology

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. **Spatiotemporal data and models.**
7. Inference on population dynamics from genetic sequence data.

(1–6) were enumerated by Bjornstad and Grenfell (*Science*, 2001).

(1–5) are now routinely solved using modern methods for nonlinear partially observed Markov process (POMP) models (Ionides et al., 2015; King et al., 2016).

(7) was described by Grenfell et al (Science, 2004) and a general POMP solution was shown by Smith et al (*Molecular Biology & Evolution*, 2017).

Overview of sequential Monte Carlo

- **Sequential Monte Carlo (SMC)**, a.k.a. **the particle filter**, is a standard tool for fitting mechanistic dynamic models to nonlinear non-Gaussian time series.
- SMC allows full-information statistical inference. Standard MCMC methods struggle with POMP models; many other methods involve information loss or approximations.
- SMC struggles with a **curse of dimensionality** preventing the use of the basic algorithm when the dimension of the dynamic system gets large (in practice, say, more than 5–10 dimensions).
- Theoretical results suggest that, in some situations, this curse can avoided (Rebeschini and van Handel, 2015).
- We have a method that partially avoids the curse and is practical on some problems with 160 latent dynamic dimensions and 40 measured dimensions: Susceptible-Exposed-Infected-Recovered (SEIR) dynamics for measles in 40 connected cities.

A modified SMC algorithm for spatiotemporal data

- A **Guided intermediate resampling filter (GIRF)** breaks up the information in the data into small pieces that are used incrementally to inform the particles and "guide" them toward the next observation.
- Observations $y^*_{1:N} = (y^*_1, \ldots, y^*_N)$ are collected at times $t_{1:N}$.
- We require a latent Markov process $\{X(t), t_0 \leq t \leq t_N\}$ to be defined in continuous time. We assess particles at S **intermediate times**

$$t_{n,s} = \left(1 - \frac{s}{S}\right) t_n + \frac{s}{S} t_{n+1}$$

  using a **guide function**

$$u_{n,s}(x).$$

- Our algorithm works asymptotically for general $u_{n,s}(x)$, but gains numerical efficiency if this guide function approximates the forecast likelihood of subsequence measurements.

Construction of the guide function

- An ideal guide function for GIRF is $u_{n,s}(x) = f_{Y_{n+1:N}|X_{n,s}}(y_{n+1:N}|x)$.
- In practice, we use a simulation-based approximation to

$$u^L_{n,s}(x) \approx f_{Y_{n+1:\min(n+L,N)}|X_{n,s}}(y_{n+1:\min(n+L,N)}|x),$$

where L **is the lookahead**.

Innovations of GIRF methodology

- GIRF combines two existing approaches:
  lookahead SMC (Lin et al., 2013; Guarniero et al., 2017)
  intermediate resampling (Del Moral and Murray, 2015).
- Theoretically, we show GIRF has favorable scaling properties not possessed by either previous approach alone.
- Extendable to parameter inference using an iterated perturbed Bayes map.
- Provides full-information plug-and-play inference on spatiotemporal models of scientific interest.

Special cases of GIRF

Lookhead observations, L	Intermediate steps, S	guide function, u	algorithm
1	1	$f_{Y_{n+1}\|X_{n+1}}$	vanilla particle filter
2	1	simulated†	auxiliary particle filter
1	>1	simulated†	bridge particle filter
>1	1	simulated†	iterated auxiliary filter
>1	>1	simulated†	our implementation of GIRF

† some estimate of $f_{Y_{n+1:\min(n+L,N)}|X_{n,s}}(y_{n+1:\min(n+L,N)}|x)$ based on simulations and/or other approximations.

A guided intermediate resampling filter (GIRF)

input:

Simulator for latent process initial density, $f_{X_0}(x_0\,;\theta)$

Simulator for transition density, $f_{X(t)|X(s)}(\cdot\,|\,\cdot\,;\theta)$, $t_0 \le s < t \le t_N$

Evaluator for measurement density, $f_{Y_n|X_n}(\cdot\,|\,\cdot\,;\theta)$, $n \in 1\!:\!N$

Data, $y^*_{1:N}$. Parameter vector, θ. Number of particles, J.

Number of intermediate reweighting steps, S.

Simulation-based evaluator for the guide function, $u_t(x_t)$.

Number of lookahead observations, L, **for the guide function**.

output:

Filtered particles, $\{X_N^{F,j}, j \in 1\!:\!J\}$.

Log likelihood estimate, $\hat{\ell} \approx \log f_{Y_{1:N}}(y^*_{1:N}\,;\theta)$.

Algorithms based on a simulator of the dynamic model are **plug-and-play**. This property ensures broad applicability.

Strong and weak plug-and-play for algorithms acting on POMP models

- **Strong plug-and-play**: The only algorithmic input dependent on the latent dynamic process is a simulator.
- **Weak plug-and-play**: The transition density of the latent dynamic process is not required as an algorithmic input. Inputs may include model-dependent quantities such as local mean and variance approximations for transitions.
- The vanilla particle filter is strong plug-and-play.
- Our GIRF implementation is weak plug-and-play since a local mean approximation is used, together with simulated sample paths, to build the a guide function.
- Approximations used for a guide function do not affect the consistency of GIRF.
- Weak plug-and-play is enough for many applications.

The guided intermediate resampling filter (GIRF)

Initialize: $\hat{\ell} = 0$, $\quad X_{0,0}^{F,j} \sim f_{X_0}(x_0\,;\theta)$, $\quad u_j = 1$ for j in $1\!:\!J$.

For n in $0\!:\!N-1$

 if $n \geq 1$ then $u^j = u^j \big/ f_{Y_n|X_n}(y_n^* \,|\, X_{n,0}^{F,j}\,;\theta)$

 For s in $1\!:\!S$

  $X_{n,s,j}^{P} \sim f_{X_{n,s}|X_{n,s-1}}(x_{n,s} \,|\, X_{n,s-1,j}^{F};\theta)$ for j in $1\!:\!J$

  Construct guide weights: $v_j = u_{n,s}(X_{n,s,j}^{P})$

  $w_j = v_j \big/ u_j$

  $\hat{\ell} = \hat{\ell} + \log\left\{\frac{1}{J}\sum_{j=1}^{J} w_j\right\}$

  Draw $a_{1:J}$ with $\mathbb{P}(a_j = i) = w_{n,i}^{m} \Big/ \sum_{i'=1}^{J} w_{n,i'}^{m}$

  $X_{n,s,j}^{F} = X_{n,s,a_j}^{P}$ and $u_j = v_{a_j}$ for j in $1\!:\!J$

 End For

 $X_{n+1,0,j}^{F} = X_{n,S,j}^{F}$

End For

Software for GIRF

- An implementation by Joonha Park is at `github.com/joonhap/GIRF.git`.
- The `spatPomp` R package is in development. It adds spatiotemporal structure to the `pomp` R package.
- `spatPomp` provides tools for development of models and methods.
- `spatPomp` focuses on plug-and-play methods since this enables implementations applicable to a general class of useful models.
- `girf()` is currently the best algorithm in `spatPomp` for highly nonlinear systems with ≈ 50 coupled spatial units.

An implementation of GIRF in the R package spatPomp

`girf(P, Np =` J, `Ninter =` S, `Nguide =` K, `Lookahead =` L, `h =` h_u `,`
`theta_to_V =` $\overrightarrow{\mathrm{v}}$, `V_to_theta =` $\overleftarrow{\mathrm{v}}$`)` , where `P` is a class '`spatPomp`' object.

This implementation has access to the following:

1. simulation of $f_{\boldsymbol{X}_n|\boldsymbol{X}_{n-1}}(\boldsymbol{x}_n \,|\, \boldsymbol{x}_{n-1}\,;\theta)$
2. evaluation of $f_{Y_{u,n}|X_{u,n}}(y_{u,n} \,|\, x_{u,n}\,;\theta)$
3. simulation of $f_{\boldsymbol{X}_0}(\boldsymbol{x}_0\,;\theta)$
4. skeleton numerical integrator, $\boldsymbol{\mu}(\boldsymbol{x}, s, t\,;\theta)$
5. parameter, θ, and data, $\boldsymbol{y}^*_{1:N}$
6. number of particles, J
7. number of guide simulations, K, and number of lookahead lags, L
8. number of intermediate timesteps, S
9. measurement mean, $h_u(x,\theta)$, and variance, $\overrightarrow{\mathrm{v}}_u(x,\theta)$
10. measurement parameters from moments, $\overleftarrow{\mathrm{v}}_u(V,x,\theta)$

Initialize: simulate $\boldsymbol{X}^F_{0,0,j} \sim f_{\boldsymbol{X}_0}(\,\cdot\,;\theta)$ and set $g^F_{0,0,j}=1$

For n in $0\!:\!N-1$

Guide simulations: $\boldsymbol{X}^G_{n+1:n+L,j,k} \sim f_{\boldsymbol{X}_{n+1:n+L}|\boldsymbol{X}_n}\big(\,\cdot\,|\boldsymbol{X}^F_{n,0,j};\theta\big)$

Guide sample variance: $V^G_{u,n,0,\ell,j} = \mathrm{Var}\big\{h_u\big(X^G_{u,\ell,j,k}\big), k\in 1\!:\!K\big\}$

For s in $1\!:\!S$

Prediction simulations: $\boldsymbol{X}^P_{n,s,j} \sim f_{\boldsymbol{X}_{n,s}|\boldsymbol{X}_{n,s-1}}\big(\,\cdot\,|\boldsymbol{X}^F_{n,s-1,j};\theta\big)$

Skeleton: $\boldsymbol{\mu}^P_{n,s,\ell,j} = \boldsymbol{\mu}\big(\boldsymbol{X}^P_{n,s,j}, t_{n,s}, t_\ell\,;\theta\big)$

Measurement variance at skeleton: $V^M_{u,n,s,\ell,j} = \overrightarrow{\mathrm{v}}_u(\theta, \mu^P_{u,n,s,\ell,j})$

Forecast variance: $V^P_{u,n,s,\ell,j} = V^G_{u,n,s-1,\ell,j}\,\big(t_\ell - t_{n,s}\big)\Big/\big(t_\ell - t_{n,0}\big)$

Moment match: $\theta_{u,n,s,\ell,j} = \overleftarrow{\mathrm{v}}_u\big(V^M_{u,n,s,\ell,j} + V^P_{u,n,s,\ell,j},\, \mu^P_{u,n,s,\ell,j}\big)$

Guide: $g^P_{n,s,j} = \displaystyle\prod_{\ell=n+1}^{\min(n+L,N)} \prod_{u=1}^{U} f_{Y_{u,\ell}|X_{u,\ell}}\big(y^*_{u,\ell}\,|\,\mu^P_{u,n,s,\ell,j}\,;\theta_{u,n,s,\ell,j}\big)$

$$w(n,s,j) = \begin{cases} f_{\boldsymbol{Y}_n|\boldsymbol{X}_n}\big(\boldsymbol{y}_n|\boldsymbol{X}^F_{n,s-1,j}\,;\theta\big)\ \frac{g^P_{n,s,j}}{g^F_{n,s-1,j}} & \text{if } s=1,\ n\neq 0 \\ \frac{g^P_{n,s,j}}{g^F_{n,s-1,j}} & \text{else} \end{cases}$$

log likelihood component: $c_{n,s} = \log\Big(J^{-1}\,\sum_{i=1}^J w(n,s,i)\Big)$

Normalized weights: $\tilde{w}(n,s,j) = w(n,s,j)\Big/\sum_{i=1}^J w(n,s,i)$

Systematic resampling indices: $r_{1:J}$ with $\mathbb{P}\left[r_j = i\right] = \tilde{w}(n,s,i)$
Set $\boldsymbol{X}^F_{n,s,j} = \boldsymbol{X}^P_{n,s,r_j}$, $g^F_{n,s,j} = g^P_{n,s,r_j}$, $V^G_{u,n,s,\ell,j} = V^G_{u,n,s-1,\ell,r_j}$
end For
Set $\boldsymbol{X}^F_{n+1,0,j} = \boldsymbol{X}^F_{n,S,j}$ and $g^F_{n+1,0,j} = g^F_{n,S,j}$
end For

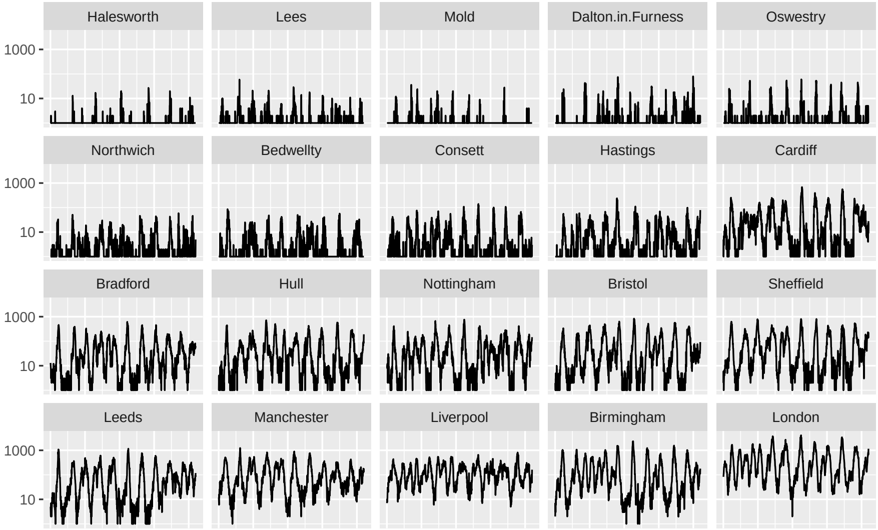
Weak coupling and its role in this GIRF implementation

- We used a simulation-based product approximation for the guide function.
- Intuitively, this is appropriate when the model has weak coupling: most of the interactions and stochasticity are localized in space over short time intervals.

From filtering to inference

- Filtering estimates latent states given data and a model. It doesn't estimate parameters.
- Iterated filtering algorithms (Ionides et al., 2015) filter repeatedly with perturbed parameters, approximating an iterated Bayes may that approaches the maximum of the likelihood.
- Iterated filtering has proved effective on challenging POMP models for which alternative methods (e.g., Particle Markov chain Monte Carlo, Monte Carlo Expectation-Maximization) have proved numerically intractable.
- To implement an iterated filtering algorithm, we iterate GIRF on an extended POMP model where parameters make a random walk.
- This is called **iGIRF**, implemented by `igirf()` in `spatPomp`.

Measles in 20 UK cities, 1944–1965



Motivation for studying measles

- Standard models have arisen from the extensive study of pre-vaccination measles. To demonstrate the new methodology, we chose a situation where basic models for transmission within a city are well established.
- Coupling between cities is less well understood. We analyze cases in the forty largest UK cities.
- Coupling between neighborhoods of a single city, or aggretated at county or state level gives rise to similar numbers of spatial units.
- Measles remains a major cause of morbidity and mortality globally. It may be an upcoming target for global eradication.

A spatiotemporal model for measles

- We start with the Susceptible-Exposed-Infected-Recovered (SEIR) measles model of He et al. (2010) and add spatial interaction.
- For each city k, the population dynamics satisfy a set of equations,

$$\begin{aligned} \frac{dS_k}{dt} &= -\frac{dN_{SE,k}(t)}{dt} - \mu S_k(t) + r_k(t) \\ \frac{dE_k}{dt} &= \frac{dN_{SE,k}(t)}{dt} - \frac{dN_{EI,k}(t)}{dt} - \mu E_k(t) \qquad\qquad k = 1, \cdots, d, \\ \frac{dI_k}{dt} &= \frac{dN_{EI,k}(t)}{dt} - \frac{dN_{IR,k}(t)}{dt} - \mu I_k(t), \end{aligned}$$

where, $N_{SE,k}(t), N_{EI,k}(t), N_{IR,k}(t)$ denote the cumulative number of transitions between the compartments up to time t in city k, μ is the per-capita mortality rate, and r_k the susceptible recruitment rate.

The cumulative transitions were modelled as negative binomial processes, following the construction of Bretó et al. (2009). Specifically,

$$
\begin{aligned}
&\mathbb{E}\left[N_{SE,k}(t+dt) - N_{SE,k}(t)\right] \\
&\quad = \beta(t) \cdot S_k(t) \cdot \left[\left(\frac{I_k}{P_k}\right)^{\alpha} + \sum_m \frac{v_{km}}{P_k} \left(\left(\frac{I_m}{P_m}\right)^{\alpha} - \left(\frac{I_k}{P_k}\right)^{\alpha} \right) \right] dt + o(dt),
\end{aligned}
$$

where $\beta(t)$ is transmission coefficient with time dependence due to seasonality, α is a mixing coefficient, P_k is the population at city k, and v_{kl} the number of travelers from city k to l.

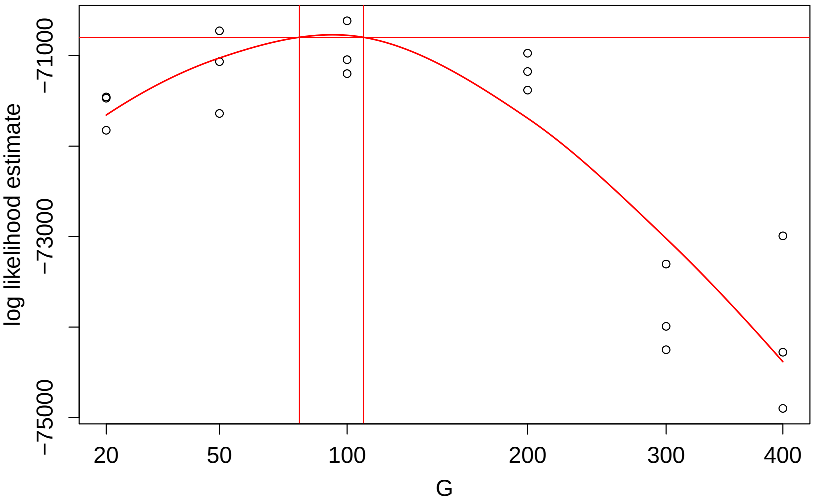
The gravity model of Xia et al. (2004) describes the number of travelers:

$$
v_{kl} = G \cdot \frac{\bar{d}}{\bar{P}^2} \cdot \frac{P_k \cdot P_l}{d_{kl}}.
$$

Here, the gravitation constant G was rescaled using the average population of all 20 cities, $\bar{P}$, and the average distance of all pairs of cities, $\bar{d}$.

Data are weekly cumulative reported cases, modeled using an overdispersed binomial distribution.

Profiling the coupling constant for measles in 40 cities



A Monte Carlo adjusted profile (MCAP) confidence interval (Ionides et al., 2017) uses a cutoff of 35.1, rather than the usual 1.92.

Benchmarking on a linear Gaussian model

d	APF $S{=}1, L{=}2$	2-lookahead $S{=}1, L{=}3$	GIRF $S{=}d, L{=}2$	GIRF $S{=}d, L{=}3$	Kalman filter $\log \ell$
5	-0.001 (0.53)	-0.07 (0.46)	-0.32 (0.49)	-0.06 (0.62)	-485.6
20	-37.3 (9.1)	-24.8 (8.6)	-1.1 (1.1)	+0.26 (0.86)	-1904.0
50	-1366 (144)	-1146 (119)	-5.6 (5.4)	-0.6 (1.8)	-4790.2
100	-7096 (424)	-6717 (366)	-73 (10)	-7.7 (3.4)	-9499.1
200	-30688 (1323)	-29544 (1333)	-277 (27)	-23 (7.2)	-18909
500	– –	– –	-1282 **(56)**	-162 **(16)**	-47415

- $\log \hat{\ell} - \log \ell$ (with Monte Carlo s.d.) for four special cases of GIRF.

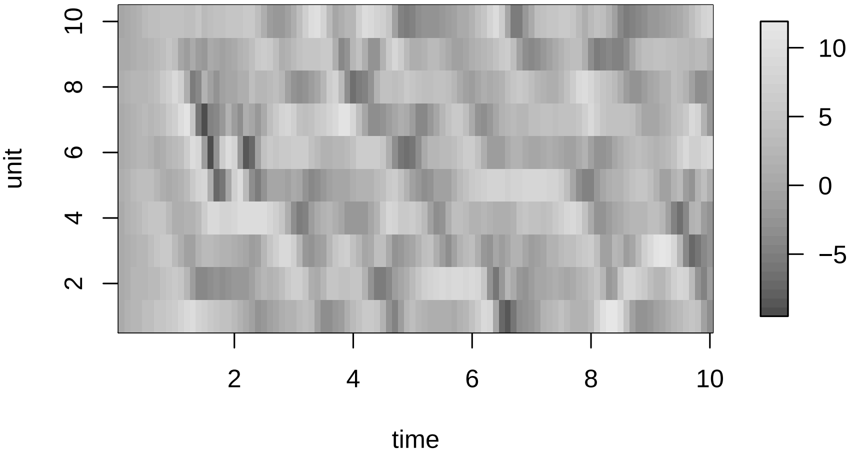
Benchmarking on the Lorenz 96 model

- The Lorenz 96 model is a nonlinear chaotic system providing a toy representation of global atmospheric circulation (Lorenz, 1996).
- Stochastic versions of this model have been used to support the increased use of non-deterministic models for atmospheric science.
- We tested a stochastic Lorenz 96 model with Gaussian process noise:

$$dX_t^{[i]} = \Big\{ X_t^{[i-1]} \Big(X_t^{[i+1]} - X_t^{[i-2]} \Big) - X_t^{[i]} + F \Big\} dt + \sigma_p dB_t^{[i]}, \quad i \in 1{:}d.$$

- A standard data assimilation tool is the ensemble Kalman filter (EnKF) which simulated to propagates particles but adjusts using a Gaussian approximation.
- When measurements are infrequent, or have high error, EnKF can fail when GIRF succeeds.

A Lorenz 96 simulation with $d = 10$, $F = 8$



A theoretical result (Park and Ionides, 2019)

- Assumption 1. The predictive likelihood can be closely approximated.
- Assumption 2. The length of subinterval is sufficiently small.
- Assumption 3. The POMP possesses conditional mixing property given data.

Sketch of Theorem: Under assumptions 1, 2, and 3, for any h with $\|h\|_\infty \le 1$,

$$\left| \frac{1}{J} \sum_{j=1}^{J} h\big(X_{t_N}^{F,j}\big) - \mathbb{E}\Big[h\big(X_N\big) \Big| Y_{1:N} = y_{1:N}^*\Big] \right| \le a_1 + \frac{a_2(d)}{\sqrt{J}}$$

with high probability. The constant $a_2(d)$ is dependent on the accuracy of $u_{n,s}$ as an approximation to the predictive likelihood. It can increase slowly with d.

Conclusions

- Guided intermediate resampling filter (GIRF) methodology can permit statistical inference for coupled nonlinear partially observed stochastic dynamic systems of moderate dimension.
- GIRF enables likelihood-based inference for a spatiotemporal SEIR model with 40 coupled cities.
- GIRF is weakly plug-and-play, therefore widely applicable.
- Techniques assisting the use of a Monte Carlo filter for parameter estimation and hypothesis testing include:

  (i) Iterated filtering methodology to adapt a successful filter for maximum likelihood estimation.

  (ii) Monte Carlo adjusted profile (MCAP) methodology to enable proper inference despite non-negligible Monte Carlo error.

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., He, D., Ionides, E. L., and King, A. A. (2009). Time series analysis via mechanistic models. *Ann. Appl. Stat.*, 3:319–348.

Del Moral, P. and Murray, L. M. (2015). Sequential Monte Carlo with highly informative observations. *J. Uncertainty Quant.*, 3:969–997.

Guarniero, P., Johansen, A. M., and Lee, A. (2017). The iterated auxiliary particle filter. *Journal of the American Statistical Association*, 112(520):1636–1647.

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. *J. R. Soc. Interface*, 7:271–283.

Ionides, E. L., Breto, C., Park, J., Smith, R. A., and King, A. A. (2017). Monte Carlo profile confidence intervals for dynamic systems. *J. R. Soc. Interface*, 14:20170126.

References II

Ionides, E. L., Nguyen, D., Atchadé, Y., Stoev, S., and King, A. A. (2015). Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. *Proc. Natl. Acad. Sci. USA*.

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.

Lin, M., Chen, R., and Liu, J. S. (2013). Lookahead strategies for sequential Monte Carlo. *Statistical Science*, 28:69–94.

Lorenz, E. N. (1996). Predictability: A problem partly solved. *Proceedings of the Seminar on Predictability*, 1:1–18.

Park, J. and Ionides, E. L. (2019). A guided intermediate resampling particle filter for inference on high dimensional systems. *Arxiv:1708.08543v2*.

Rebeschini, P. and van Handel, R. (2015). Can local particle filters beat the curse of dimensionality? *Ann. Appl. Prob.*, 25:2809–2866.

References III

Smith, R. A., Ionides, E. L., and King, A. A. (2017). Infectious disease dynamics inferred from genetic data via sequential Monte Carlo. *Molecular Biology and Evolution*, 34:2065–2084.

Xia, Y., Bjørnstad, O. N., and Grenfell, B. T. (2004). Measles metapopulation dynamics: A gravity model for epidemiological coupling and dynamics. *Am. Nat.*, 164:267–281.