Simulation-based maximum likelihood inference for partially observed Markov process models

Anindya Bhadra

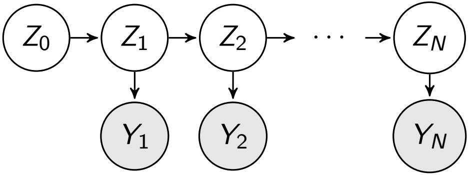
Texas A&M University

February 21, 2012

Overview

- Simulation-based inference in partially observed Markov process (POMP) models.
- An inference tool using sequential Monte Carlo (SMC): Iterated filtering and its properties.
- An application: Modeling malaria transmission in India.
- A recurring problem with SMC and a proposed improvement.

Partially Observed Markov Processes (POMP)



- Parameter vector θ.
- Markovian unobserved state process Z_n:
  $f_{Z_n|Z_{1:n-1}}(z_n|z_{1:n-1};\theta) = f_{Z_n|Z_{n-1}}(z_n|z_{n-1};\theta).$
- Conditionally independent observations $Y_n|Z_n = z_n$:
  $f_{Y_n|Z_{1:n},Y_{1:n-1}}(y_n|z_{1:n},y_{1:n-1};\theta) = f_{Y_n|Z_n}(y_n|z_n;\theta).$

Simulation-based inference

- Statistical methods for POMPs are simulation-based if they require draws from the state transition model but not its explicit evaluation.
- We just need to be able to simulate from $f_{Z_n|Z_{n-1}}(z_n|z_{n-1};\theta)$. Not being able to evaluate it is OK.
- We still need to evaluate $f_{Y_n|Z_n}(y_n|z_n;\theta)$.

Benefits: Problems with the state transition density

- May be high-dimensional.
- May not be available in closed form for non-linear continuous time models.
- If an underlying continuous time process is observed at discrete points, algorithms depending on the evaluation of the transition density - e.g. EM or MCMC can fail, depending on how much information is lost through discretization. (Roberts & Stramer, Biometrika, 2001)

Examples of simulation-based techniques

- Bayesian examples:
  1. Artificial parameter evolution (Liu and West, 2001).
  2. Approximate Bayesian computation (Marjoram et al, *PNAS*, 2003).
  3. Particle MCMC (Andrieu et al., JRSSB, 2010).

- Non-Bayesian examples:
  4. Spectral analysis via moment matching (Reumann et al, *PNAS*, 2006).
  5. Maximum likelihood via iterated filtering (Ionides, Bhadra et al, *Ann. Stat.*, 2011).

Simulation-based techniques are VERY USEFUL for investigating scientific models.

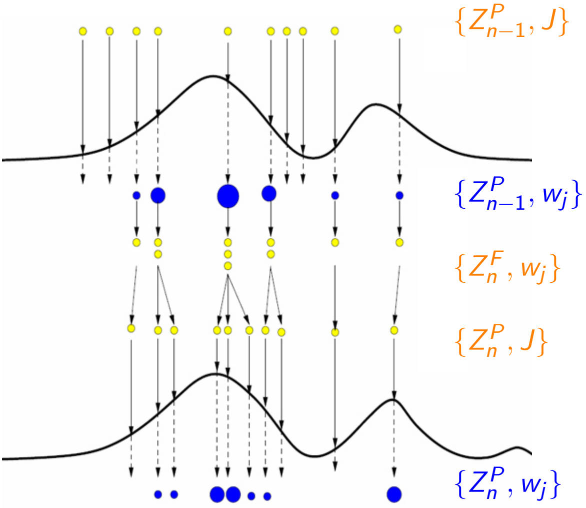
Iterated filtering

- Filtering: the conditional distribution of the unobserved state vector z_n given the observations up to that time, $y_1, y_2, \ldots, y_n$.

- Iterated filtering: an algorithm which uses a sequence of solutions to the filtering problem to maximize the likelihood function over unknown model parameters.

- If the filter is simulation-based (e.g. using standard SMC methods) this property is inherited by iterated filtering.

A pictorial representation of the SMC filter



Methematical formulation of a POMP

- Assume all joint and conditional densities of $Z_{0:N}$ and $Y_{1:N}$ exist, denoted by appropriate subscripts to the letter f.

- Unknown parameter θ takes value in $\mathbb{R}^p$.

- Introduce time varying process $\{\theta_n, 0 \leq n \leq N\}$. Then g defines a model for Markov process $\{(Z_n, \theta_n), 0 \leq n \leq N\}$ and observations $Y_{1:N}$

Key ideas of iterated filtering

- Bayesian inference for time-varying parameters becomes a solveable filtering problem. Set θ_n to be a random walk with

$$\begin{aligned} \mathrm{E}_g[\Theta_n|\Theta_{n-1}] &= \Theta_{n-1}, & \mathrm{Var}_g(\Theta_n|\Theta_{n-1}) &= \sigma^2\Sigma, \\ \mathrm{E}_g[\Theta_0] &= \theta, & \mathrm{Var}_g(\Theta_0) &= \tau^2\Sigma. \end{aligned}$$

- κ is a pdf on $\mathbb{R}^p$ that defines this random walk.

- The limit $\sigma \to 0$ and $\tau \to 0$ can be used to maximize the likelihood for fixed parameters (θ).

Iterated filtering with a perfect filter

Theorem 1. (Ionides, Bhadra, Atchadé & King (Ann. Stat., 2011))

- Suppose θ_0 and $y_{1:N}$ are fixed and define

$$\mu_n^F = \mu_n^F(\theta,\sigma,\tau) = E_g[\Theta_n \,|\, Y_{1:n} = y_{1:n}],$$
$$V_n^P = V_n^P(\theta,\sigma,\tau) = \mathrm{Var}_g(\Theta_n \,|\, Y_{1:n-1} = y_{1:n-1}),$$

- Let σ be a function of τ with $\sigma\tau^{-1} \to 0$ as $\tau \to 0$.

- Then

$$\lim_{\tau\to 0} \sum_{n=1}^{N} \left(V_n^P\right)^{-1}\left(\mu_n^F - \mu_{n-1}^F\right) = \nabla\ell(\theta).$$

Iterated filtering via an SMC filter

Theorem 2. (Ionides, Bhadra, Atchadé & King (Ann. Stat., 2011))

- Let $\tilde{\mu}^F_{n,m}$ and $\tilde{V}^P_{n,m}$ be the SMC estimates of μ^F_n and V^P_n respectively with number of particles J_m.
- Let $\tau_m \to 0$, $\sigma_m {\tau_m}^{-1} \to 0$ and $\tau_m J_m \to \infty$.
- Then

$$\lim_{m\to\infty} \widetilde{E}\Big[\sum_{n=1}^{N} \big(\widetilde{V}^P_{n,m}\big)^{-1}\big(\widetilde{\mu}^F_{n,m} - \widetilde{\mu}^F_{n-1,m}\big)\Big] \;=\; \nabla\ell(\theta),$$

$$\lim_{m\to\infty} \tau_m^2 J_m \widetilde{\mathrm{Var}}\Big(\sum_{n=1}^{N} \big(\widetilde{V}^P_{n,m}\big)^{-1}\big(\widetilde{\mu}^F_{n,m} - \widetilde{\mu}^F_{n-1,m}\big)\Big) \;<\; \infty.$$

with convergence being uniform for θ in compact sets.

Iterated filtering as a stochastic approximation

Theorem 3. (Ionides, Bhadra, Atchadé & King (Ann. Stat., 2011))

- Let $\sum_m a_m = \infty$, $a_m \to 0$ and $\sum_m a_m^2 J_m^{-1} \tau_m^2 < \infty$.

- Define a recursion by

$$\hat{\theta}_{m+1} = \hat{\theta}_m + a_m \sum_{n=1}^{N} (\tilde{V}_{n,m}^P)^{-1} (\tilde{\mu}_{n,m}^F - \tilde{\mu}_{n-1,m}^F).$$

- Then under suitable regularity assumptions for stochastic approximation (Kushner and Clark, 1978), $\lim_{m\to\infty} \hat{\theta}_m = \hat{\theta}$, the MLE, with probability 1.

Putting the pieces together: The iterated filtering algorithm

1. Select initial value $\hat{\theta}_1$ and algorithmic parameters σ_1, c, α and M

2. For m in 1:M
   (i) Let $\sigma_m = \sigma_1 \alpha^{m-1}$; initialize $E[\theta_{0,m}] = \hat{\theta}_m$, $Var[\theta_{0,m}] = c\sigma_m^2$.
   (ii) Evaluate using particle filter $\tilde{V}^P_{n,m}$, $\tilde{\mu}^F_{n,m}$.
   (iii) Update the estimate of θ by

$$\hat{\theta}_{m+1} = \hat{\theta}_m + a_m \underbrace{\sum_{n=1}^{N} (\tilde{V}^P_{n,m})^{-1} (\tilde{\mu}^F_{n,m} - \tilde{\mu}^F_{n-1,m})}_{\approx \widetilde{\nabla \ell(\theta_m)}}.$$

3. Output $\hat{\theta}_{M+1}$ as the MLE.

Why is iterated filtering simulation-based?

- A property of the SIRS sequential Monte Carlo filters (aka particle filters) in use.

- Recall from Theorem 2

$$\lim_{m\to\infty} \widetilde{E}\left[\sum_{n=1}^{N}(\tilde{V}^{P}_{n,m})^{-1}(\tilde{\mu}^{F}_{n,m} - \tilde{\mu}^{F}_{n-1,m})\right] \;=\; \nabla\ell(\theta).$$

- $\tilde{V}^{P}_{n,m}$, $\tilde{\mu}^{F}_{n,m}$ and $\tilde{\mu}^{F}_{n-1,m}$ are SMC estimates of the corresponding true quantities.

An application: Modeling malaria dynamics

- Malaria: 300 - 500 million cases each year, resulting in nearly 1 million deaths.
- Common in Sub-Saharan Africa, parts of Asia, central and south America.
- Caused by the *Plasmodium* parasite, carried by female *Anopheles* mosquitoes.
- 2008 WHO estimate of economic impact: $12 billion/year.

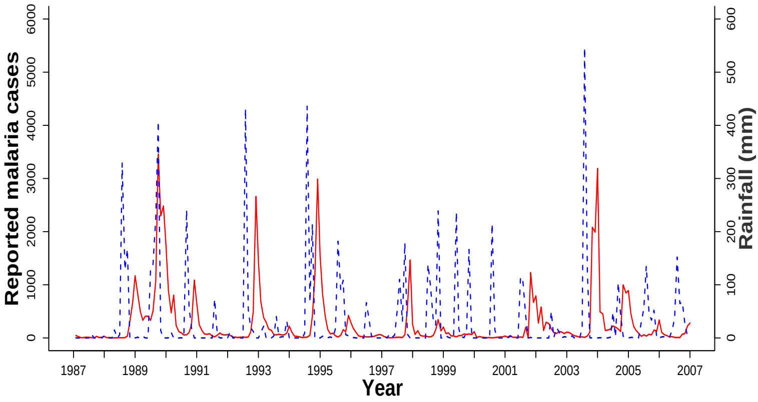
Challenges with malaria modeling

- A complex disease with incomplete immunity, mosquito-human interaction and an initial symptom of non-specific fever.
- Mechanistic inclusion of climate covariates (e.g. rainfall) have not been successful. Lags? Integrals? Threshold effects?
- State transition densities are unavailable, so MCMC and Stochastic EM are not applicable.

Previous work

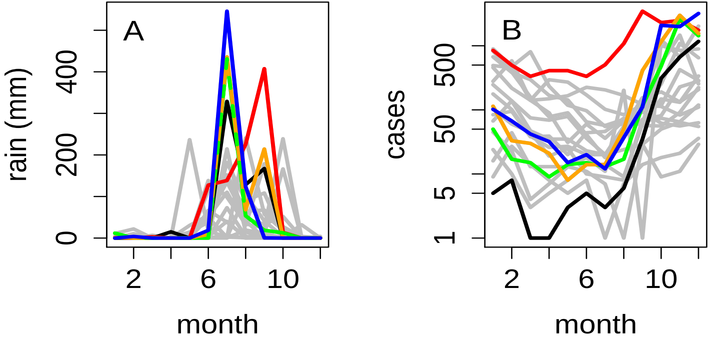
- First mathematical models by Ross (1911) and Macdonald (1957).
- Various proposed extensions since then
  - strain diversity (Gupta et al. (1994), Science).
  - drug resistance (Koella and Antia (2003), Malaria Journal).
  - partial immunity (Filipe et al. (2007), PLoS Comp. Biol.).
- But few papers on trying to fit a mechanistic model on a “real” time series data.
  - Bayesian analysis of malaria in Senegal (Cancre et al. (2000), Amer. J. Epid.).

The Data



- Monthly clinical case data and rainfall from Kutch, a district in the state of Gujarat, western India, 1987-2006.

The Data: Another look

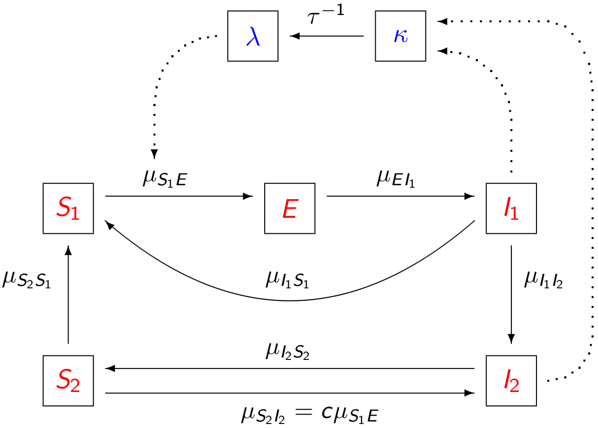


- Rain peaks around July while the epidemic peaks around December.

Debate over the role of climate factors

- Hay et al. (Nature, 2002) propose interannual variability depends on intrinsic dynamics (in E. African highlands).
- Zhou et al. (PNAS, 2004); Pascual et al. (Proc. Roy. Soc. B, 2007) think external drivers such as temperature and rainfall play a role.
- We formally test the effect of intrinsic vs. extrinsic factors on interannual variability.
- In E. Africa temperature may be the driving factor (highlands) whereas the driving factor is likely to be rainfall in NW India (desert).

Disease Transmission Model (Bhadra et al. (2011), JASA)



Force of infection taking mosquito dynamics into account

Disease status of human population

- Call this the $\mathrm{VS^2EI^2}$ model.
- If $\mu_{I_2S_2} = \infty$ and $\mu_{S_2I_2} = \mu_{I_1S_1} = 0$, call the reduced model the VSEIR model.

Model equations

- State Model

$$
\begin{aligned}
dS_1/dt &= \mu_{BS_1}P - \mu_{S_1E}S_1 + \mu_{I_1S_1}I_1 + \mu_{S_2S_1}S_2 - \mu_{S_1D}S_1 \\
dS_2/dt &= \mu_{I_2S_2}I_2 - \mu_{S_2S_1}S_2 - \mu_{S_2I_2}S_2 - \mu_{S_2D}S_2 \\
dE/dt &= \mu_{S_1E}S_1 - \mu_{EI_1}E - \mu_{ED}E \\
dI_1/dt &= \mu_{EI_1}E - \mu_{I_1S_1}I_1 - \mu_{I_1I_2}I_1 - \mu_{I_1D}I_1 \\
dI_2/dt &= \mu_{I_1I_2}I_1 + \mu_{S_2I_2}S_2 - \mu_{I_2S_2}I_2 - \mu_{I_2D}I_2 \\
d\kappa/dt &= d\lambda_0/dt \; = \; (f(t) - \kappa)\, n_\lambda\, \tau^{-1} \\
d\lambda_i/dt &= (\lambda_{i-1} - \lambda_i)\, n_\lambda\, \tau^{-1} \qquad \text{for } i = 1, \ldots, n_\lambda - 1 \\
d\lambda/dt &= d\lambda_{n_\lambda}/dt \; = \; (\lambda_{[n_\lambda - 1]} - \lambda)\, n_\lambda\, \tau^{-1}
\end{aligned}
$$

- Observation Model

$$
\begin{aligned}
M_n &= \rho \int_{t_{n-1}}^{t_n} \mu_{EI_1} E(s) ds \\
Y_n | M_n &\sim Negbin(mean = M_n, var = M_n + \sigma_{obs}^2 M_n^2)
\end{aligned}
$$

Modeling the force of infection $\mu_{S_1 E}(t)$ with rain covariate

$$\mu_{S_1 E}(t) = \int_0^\infty f(s)p(t-s)\,ds.$$

- $f(t)$ is the effective human-human transmission rate
- $p(t)$ is a delay distribution describing the vector survival

$$f(t) = \frac{I_1(t) + qI_2(t)}{N(t)}\bar{\beta}\exp\left\{\sum_{i=1}^{n_s}\beta_i s_i(t) + \beta_c C(t)\right\}\frac{d\Gamma}{dt}.$$

- For us, $C(t) = max(R(t) - 200, 0)$, where $R(t)$ is the accumulated rainfall at time t over past 5 months.
- This introduces a threshold effect of rainfall, i.e. rainfall over a certain threshold is conducive to malaria transmission.

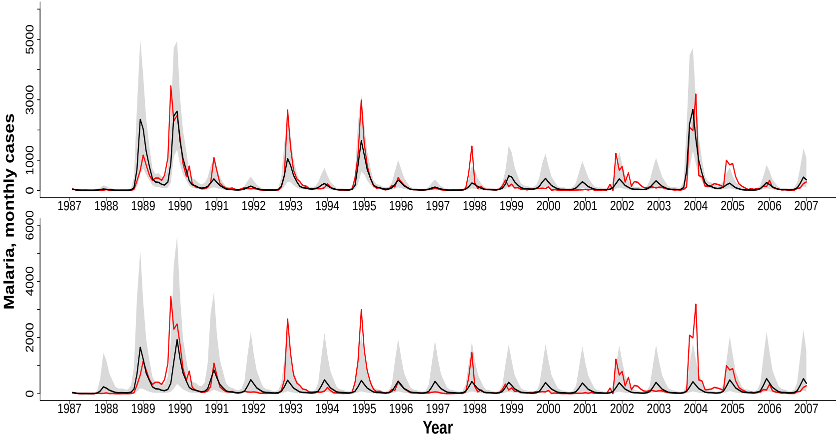
Application of iterated filtering to the malaria model : Model comparison (Bhadra et al. (2011), JASA)

model	log-likelihood	p	AIC
VSEIR without rainfall	-1275.0	19	2588.0
VSEIR with rainfall	-1265.0	20	2570.2
$\mathrm{VS^2EI^2}$ without rainfall	-1261.1	24	2570.2
$\mathrm{VS^2EI^2}$ with rainfall	-1251.0	25	2552.0
Log-SARIMA $(1,0,1) \times (1,0,1)_{12}$	-1329.0	6	2670.0
Log-SARIMA $(1,0,1) \times (1,0,1)_{12}$ with rainfall	-1322.6	7	2659.2

Table: Table of log-likelihoods and AIC of the fitted models

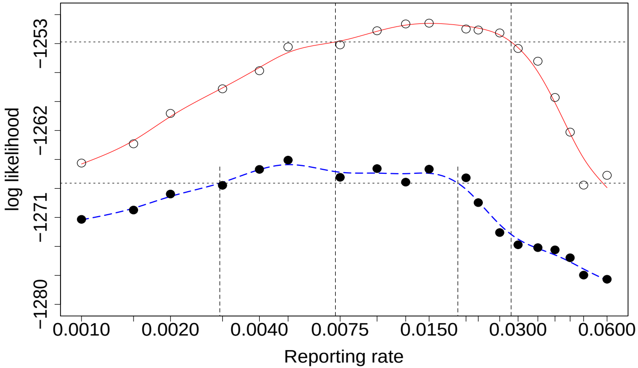
- Including rainfall properly in the model improves the likelihood.
- All the mechanistic models are significantly better than the benchmark SARIMA models.

Forward simulation using estimated parameters (Laneri, Bhadra et al. (2010), PLoS Comp. Biol.)



- Data (red) and median (solid black) of 10,000 forward simulations using estimated parameter values as initial values.
- $\mathrm{VS^2EI^2}$ model (top) and VSEIR model (bottom) with rainfall.

Reporting rate



- Profile confidence intervals for reporting rate for the $\mathrm{VS^2EI^2}$ model with rainfall and the VSEIRS model with rainfall.
- The profile plot suggests a reporting rate less than 3%.
- Indian public health officials claim this rate to be as high as 10% in Ahmedabad (Yadav et al., 2003).

Important findings for malaria in NW India

- Strong predictive relationship between rainfall and malaria incidence - this will help in developing early epidemic warnings and contribute to a scientific debate.
- Malaria shows a threshold effect to accumulated rainfall.
- Effect of rainfall robust to uncertainties in parameter estimates.
- Reporting rate may be (a) lower than what is claimed, (b) imply differences between malaria control measures in rural and urban areas.

Improvement strategy for SMC filters

- SMC filters suffer in presence of irregular data points (e.g., seen in the malaria data).
- Gives rise to more variance at the corresponding time points.
- The aim is to find an adaptive particle allocation scheme for SMC filters in an off-line, iterative setting.
- Question: Given a total number of particles (equivalently, some given computational resource), how to allocate the particles across time points in order to minimize the overall variance of the SMC likelihood estimate?

Some facts on the variance of SMC filters (Doucet and Johansen (2011))

- The asymptotic variance has an upper bound that is exponential in the number of time points.
- Under suitable mixing conditions this bound is linear in time points but it is hard to check these conditions in practice.
- The asymptotic variance has an upper bound that is exponential in the dimension of the hidden state vector.

A useful result

- Consider positive variables $M_1, \ldots, M_N$ so that $M_1 + \ldots + M_N = M$. Let $\phi_1, \ldots, \phi_N$ be known constants.

- Then the values of M_i that minimize $\sum_{i=1}^{N} \phi_i / M_i$ are given by

$$M_i^{\min} = \frac{M\sqrt{\phi_i}}{\sum_{i=1}^{N} \sqrt{\phi_i}}, \qquad \mathrm{i} = 1, \ldots, \mathrm{N}.$$

Application to SMC likelihood estimation

- The real challenge is to model SMC log likelihood.
- Actual expression of variance complicated (section 3.6, Doucet and Johansen, 2011).
- Idea: Model SMC conditional log-likelihood as an AR process and solve particle allocation for the simple model.
- AR model is motivated by geometric mixing of the SMC likelihood.

Particle allocation for AR(1) model

- Consider the AR(1) model is given by

$$x_n = \mu_n + q(x_{n-1} - \mu_{n-1}) + \epsilon_n \qquad \text{for } n = 1, \ldots, N,$$

- where $\mu_n = \mathrm{E}[x_n]$, $\epsilon_n \sim N(0, \phi_n / M_n)$.
- Minimization of $\mathrm{Var}(\sum_{n=1}^{N} x_n) = \sum_{n=1}^{N} L_n / M_n$ is achieved by

$$M_n^{\min} = \frac{M\sqrt{L_n}}{\sum_{i=1}^{N} \sqrt{L_i}},$$

where

$$\begin{aligned}
L_n &:= C_1 A_n + C_2 B_n, \\
C_1 &= (1 - q^2)^{-1}, \\
C_2 &= \frac{2q}{1-q}\left(\frac{1 - q^{2(N-1)}}{1 - q^2} - q^N \cdot \frac{1 - q^{(N-1)}}{1 - q}\right), \\
A_n &= \phi_n(1 - q^{2(N-n+1)}), B_n = \phi_n q^{-2n}.
\end{aligned}$$

Fitting an AR model to SMC conditional log-likelihood

- Define the conditional log likelihood g_n and its estimate $\hat{g}_n$ by

$$\begin{aligned} g_n &= \log f_{Y_n|Y_{1:n-1}}(y_n|y_{1:n-1},\theta), \\ \hat{g}_n &= \log \hat{f}_{Y_n|Y_{1:n-1}}(y_n|y_{1:n-1},\theta). \end{aligned}$$

- Use $\hat{g}_n$ to fit the AR model $x_n = \mu_n + q(x_{n-1} - \mu_{n-1}) + \epsilon_n$.
- Find and estimate of $\mathrm{Var}(x_n - qx_{n-1})$ as

$$\widetilde{\mathrm{Var}}(x_n - qx_{n-1}) = \frac{1}{P-1}\sum_{p=1}^{P}\{(\hat{g}_{n,p} - q\hat{g}_{n-1,p}) - \overline{(\hat{g}_n - q\hat{g}_{n-1})}\}^2.$$

- Use the relation $\mathrm{Var}(x_n - qx_{n-1}) = \phi_n/M_n$ to estimate $\tilde{\phi}_n$ and $\tilde{q}$ jointly (q is constrained in (-1,1) to make the AR model stationary).

Simulation Study

Consider the two dimensional AR(1) model

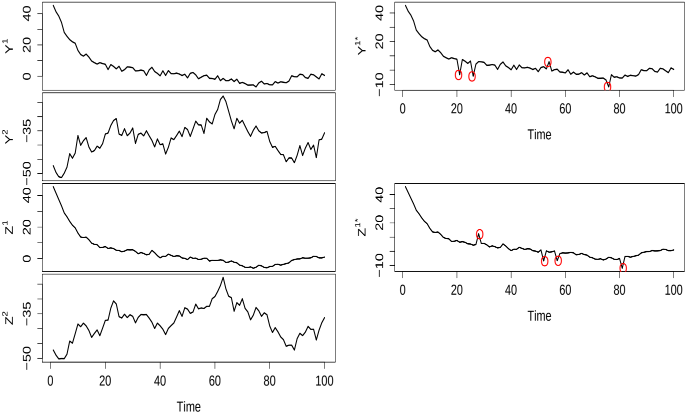
$$\begin{aligned} Z_n &= \alpha Z_{n-1} + \sigma \xi_n, && \text{(state equation)} \\ Y_n &= \beta Z_n + \tau \epsilon_n. && \text{(observation equation)} \end{aligned}$$

- $Z_n = (Z_n^1, Z_n^2)^T$ and $Y_n = (Y_n^1, Y_n^2)^T$ are in $\mathbb{R}^2$ for all $n = 1, \ldots, N$.

$$\alpha = \begin{pmatrix} 0.9 & 0.0 \\ 0.0 & 0.99 \end{pmatrix}, \qquad \sigma = \begin{pmatrix} 1.00 & 0 \\ 0 & 2.00 \end{pmatrix},$$

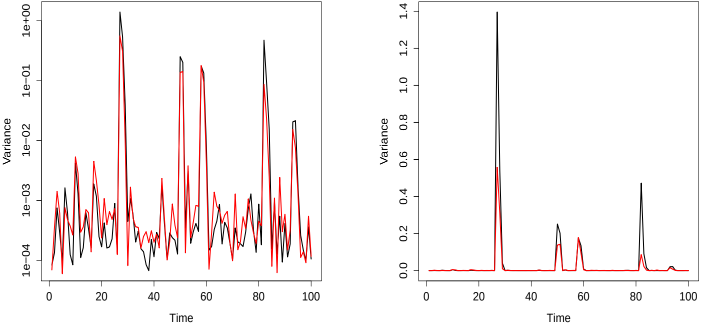
$$\beta = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \tau = \begin{pmatrix} 1 & 1 \end{pmatrix}.$$

Simulation Study



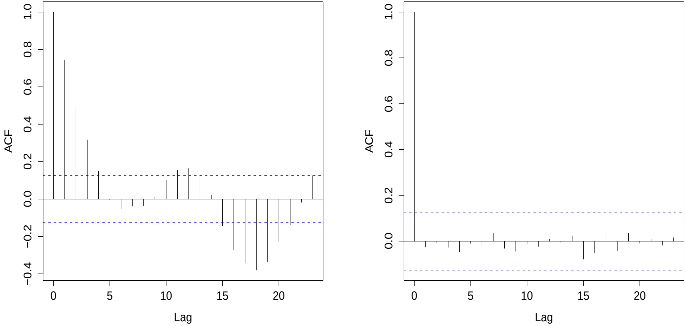
- Plot of the two dimensional AR(1) process.
- Outliers are introduced only in Y^1 at time points 21, 25, 56 and 78 and X^1 at time points 26, 53, 58 and 82.

Simulation Study



- Variance of the conditional log likelihood estimate for the data, in log scale (left) and natural scale (right).
- Black: ordinary particle filter, Red: adaptive particle filter.
- Adaptive particle filter results in noticeably lower variances in time points 26, 53 and 82.

Is AR(1) a good model for SMC conditional log-likelihood?



- Autocorrelation function of the SMC error in evaluating the conditional log likelihood via adaptive filtering (left) and the residuals from the fitted model (right).
- An autoregressive pattern is clearly visible on the left.

Reduction in variance with (almost) equal computing time

Table: Comparison of adaptive and non-adaptive particle filters. There is a total of 100 time points. The reduction in variance is 54.827%.

	Adaptive filter (a)	Ordinary filter (o)
Number of filtering (F)	10	10
Number of total particles (M)	15000*100	15000*100
Time taken per filtering (sec) (T)	23.58	22.87
Average Monte Carlo variance ($\hat{V}$)	0.766	1.692

Conclusions

- Simulation-based statistical methodology permits analysis of a flexible class of nonlinear, non-Gaussian dynamic models.
- This technique allows us to investigate the role of climate covariates on malaria while taking into account intrinsic disease dynamics.
- For simulation-based inference in complex models, efficient particle allocation strategies are crucial for particle filters, keeping in mind the constraints on computing power.

References

1. Ionides, E. L., **Bhadra A.**, Atchadé Y., and King, A. A. (2011). Iterated Filtering. *Ann. Statist., 39(3): 1776 - 1802.*
2. **Bhadra, A.**, Ionides, E., L., Laneri, K., Bouma, M., Dhiman, R., and Pascual, M. (2011). Malaria in Northwest India: Data analysis via partially observed stochastic differential equation models driven by Lévy noise. *JASA, 106(494): 440 - 451.*
3. Laneri, K., **Bhadra, A.**, Ionides, E., L., Bouma, M., Dhiman, R., Yadav, R., and Pascual, M. (2010). Forcing versus feedback: Epidemic malaria and monsoon rains in NW India. *PLoS Comput. Biol., 6(9) e1000898*
4. **Bhadra, A.**, Ionides, E., L. (2011+). An adaptive particle allocation scheme for off-line, iterated SMC schemes. (*in preparation*)