Applied Mathmetics Reading Seminar

Han Yong Wunrow

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

October 30, 2023

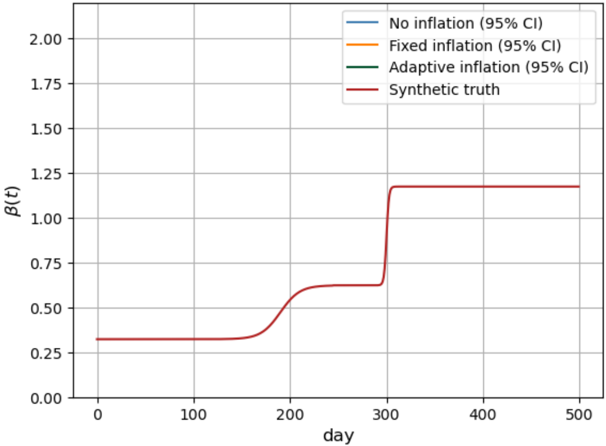
Outline

- Introduction
- Ensemble Adjustment Kalman Filter
- Adaptive Inflation
- Numerical Results

Motivation

- Time-varying reproduction number R_t is the average number of secondary cases caused by an infectious individual in a well-mixed population of susceptible individuals at time t
- Online estimation of time-varying parameters (e.g., R_t) can identify periods of heightened transmission and inform control strategies
- Current data assimilation methods often face challenges with **underestimating uncertainty** and capturing temporal trends when there are abrupt changes in parameters

Covariance Inflation



Data Assimilation

State Space Model

$$\begin{aligned}
\text{Dynamics models:} \quad & v_{n+1} = \Psi\left(v_n\right) + \xi_n,\ n \in \{1, \ldots, N\} \\
\text{Data Model:} \quad & y_{n+1} = h(v_{n+1}) + \eta_{n+1},\ n \in \{1, \ldots, N\} \\
\text{Probabilistic Structure:} \quad & v_0 \sim \mathcal{N}\left(m_0, C_0\right),\ \xi_n \sim \mathcal{N}(0, \Sigma),\ \eta_n \sim \mathcal{N}(0, \Gamma) \text{ i.i.d.} \\
\text{Probabilistic Structure:} \quad & v_0 \perp \{\xi_n\} \perp \{\eta_n\}
\end{aligned}$$

States (and parameters): $v_n \in \mathbb{R}^d$. *Observations*: $y_n \in \mathbb{R}^k$.
Assume $C_0, \Sigma, \Gamma > 0$, $\Psi \in C(\mathbb{R}^d, \mathbb{R}^d)$,and $h \in C(\mathbb{R}^d, \mathbb{R}^k)$ are all known.
Observed data:

$$Y_n^\dagger := \left\{y_1^\dagger, \ldots, y_n^\dagger\right\}$$

Goal: iteratively compute posterior distribution

$$\mathbb{P}(v_n \mid Y_n^\dagger),\ n = 1, 2, \ldots, N$$

Ensemble Kalman Methods

Represent distribution with an ensemble of particles $\{v_n^{(j)}\}_{j=1}^J$ so

$$\mathbb{P}(v_n \mid Y_n^\dagger) \approx \frac{1}{J} \sum_{j=1}^{J} \delta(v_n - v_n^{(j)}).$$

Bayes Rule

$$\mathbb{P}(v_n \mid Y_n^\dagger) \propto \underbrace{\mathbb{P}(v_n \mid Y_{n-1}^\dagger)}_{\text{prior}} \underbrace{\mathbb{P}(y_n \mid v_n)}_{\textit{likelihood}}$$

EAKF

EAKF

(1)

$$
\text{Forecast}\begin{cases}
\widehat{v}_{n+1}^{(j)} = \Psi\left(v_n^{(j)}\right) + \xi_n^{(j)}, j = 1, \dots, J, \\
\widehat{m}_{n+1} = \frac{1}{J}\sum_{j=1}^{J} \widehat{v}_{n+1}^{(j)}, \quad \widehat{C}_{n+1} = \frac{1}{J-1}\sum_{j=1}^{J}\left(\widehat{v}_{n+1}^{(j)} - \widehat{m}_{n+1}\right)\left(\widehat{v}_{n+1}^{(j)} - \widehat{m}_{n+1}\right)^{\top}, \\
\widehat{H}_{n+1} = \frac{1}{J-1}\sum_{j=1}^{J}\left(h(\widehat{v}_{n+1}^{(j)}) - \sum_{j=1}^{J} h(\widehat{v}_{n+1}^{(j)})\right)\left(h(\widehat{v}_{n+1}^{(j)}) - \sum_{j=1}^{J} h(\widehat{v}_{n+1}^{(j)})\right)^{\top}
\end{cases}
$$

(2)

$$
\text{Analysis}\begin{cases}
K_{n+1} = \widehat{C}_{n+1}^{\frac{1}{2}} \widehat{H}_{n+1}^{\frac{1}{2}\,\top}\left(\widehat{H}_{n+1} + \Gamma\right)^{-1}, \\
m_{n+1} = \widehat{m}_{n+1} + K_{n+1}\left(y_{n+1}^{\dagger} - \frac{1}{J}\sum_{j=1}^{J} h(\widehat{v}_{n+1}^{(j)})\right), \\
C_{n+1} = \widehat{C}_{n+1}^{\frac{1}{2}}\left(I + \widehat{H}_{n+1}^{\top}\Gamma^{-1}\widehat{H}_{n+1}\right)^{-1}\widehat{C}_{n+1}^{\frac{1}{2}\,\top} \\
v_{n+1}^{(j)} = m_{n+1} + C_{n+1}^{\frac{1}{2}}\widehat{C}_{n+1}^{-\frac{1}{2}}\left(\widehat{v}_{n+1}^{(j)} - \widehat{m}_{n+1}\right)
\end{cases}
$$

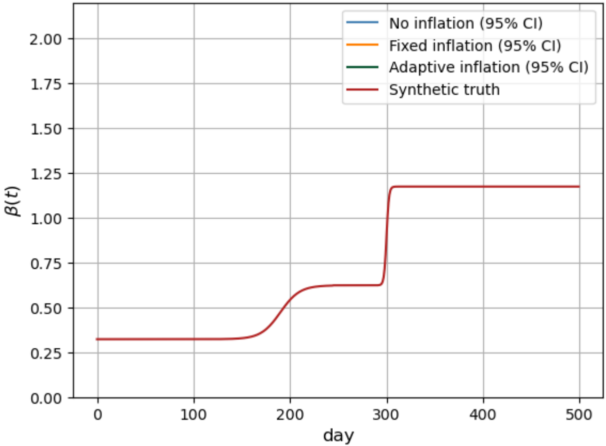
Covariance Inflation

To correct the underestimated the uncertainty of ensemble $\{v_n^{(j)}\}_{j=1}^J$ we can inflate the variance by a factor λ_n

$$v_n^{(j)} \leftarrow \sqrt{\lambda_n}(v_n^{(j)} - m_n) + m_n$$

- No inflation $\lambda_n = 1$
- Fixed inflation $\lambda_n = 1.02$
- Adaptive inflation $\lambda_n = \underset{\lambda}{\operatorname{argmax}}\ p(\lambda \mid Y_n^\dagger)$

Covariance Inflation



Adaptive Inflation

Laplace Approximation

$$p(\lambda \mid y_n) \sim N(\mu, \Sigma)$$

$$p(\lambda \mid y_n) = \frac{p(y_n, \lambda)}{\int p(y_n, \lambda) d\lambda}$$

$$= \frac{e^{\ln p(y_n, \lambda)}}{\int e^{\ln p(y_n, \lambda)} d\lambda}$$

Choose $\lambda_0 = \lambda_{MAP}$.

$$p(y_n, \lambda) \approx p(y_n, \lambda_{MAP}) + (\lambda - \lambda_0)^\top \nabla p(y_n, \lambda_{MAP}) + \frac{1}{2}(\lambda - \lambda_0)^\top \nabla^2 p(y_n, \lambda_{MAP})(\lambda - \lambda_0)$$

Adaptive Inflation

See board...

SIR Model

S(Susceptible) I(Infected) R(Recovered) N (fixed population)

$$
\begin{aligned}
\frac{dS}{dt} &= -\frac{\beta(t)SI}{N} \\
\frac{dI}{dt} &= \frac{\beta(t)SI}{N} - \gamma I \\
\frac{dR}{dt} &= \gamma I
\end{aligned} \tag{3}
$$

$$\beta(t) = \beta_0 + \frac{\beta_1 - \beta_0}{1 + e^{-k(t-m)}}$$

Stochastic SIR Model

S(Susceptible) I(Infected) R(Recovered)

$$dSI \sim \text{Pois}\left(\frac{\beta_t S_t I_t}{N}\right)$$
$$dIR \sim \text{Pois}\left(\gamma I_t\right)$$

$$(4) \quad \left.\begin{aligned} S_{t+1} &= S_t - dSI \\ I_{t+1} &= I_t + dSI - dIR \\ R_{t+1} &= R_t + dIR \\ i_{t+1} &= dSI \end{aligned}\right\} \psi$$

$v_t = \{S_t, I_t, \gamma, \beta_t, i_t\}$. $h = H = [0, 0, 0, 0, 1]^\top$.

Numerical Results

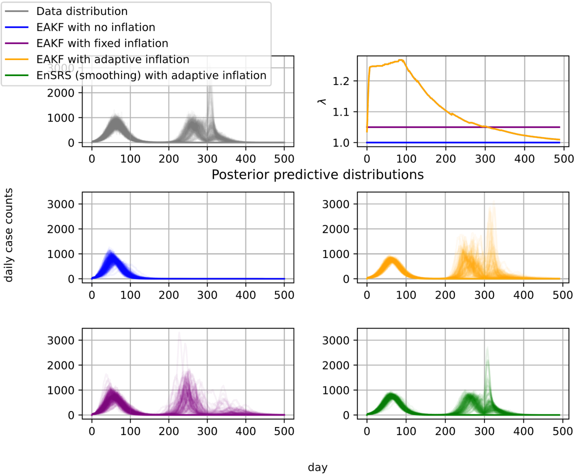


Figure 1: Underlying data distribution, inflation λ, and posterior predictive distributions for each of the 4 methods. Each line in the posterior predictive distribution represents one realization.

Numerical Results

Ran each method for 100,000 parameter scenarios with 100 realizations for each method.

Method	$D_{KL}(p\|\|q)$	$W_2(p, q)$	Prop of true R_t within 95% CI
EAKF no inflation	7.89	1.08	0.00
EAKF fixed inflation	1.75	0.72	0.15
EAKF adaptive inflation	0.63	0.52	0.97
EnSRS adaptive inflation	0.50	0.39	0.97

Table 1: Performance metrics for posterior predictive distributions p and data distribution q. Proportion of true $\mathrm{R_t}$ within 95% CI was computed on the day the sigmoid curve had reached 99% of its final value after the last change point.

Future Directions

- PAPER
- Investigate other data assimilation methods
- Adaptive assimilation window