Parameter estimation and Optimal sensor placement for Data Assimilation problems

Nikolas Kantas

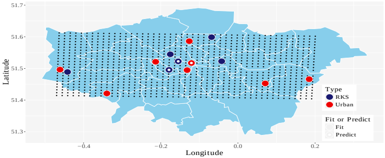
Imperial College London

joint work with Louis Sharrock

2021 EnKF workshop

Introduction - motivation

- Environmental monitoring for emissions and air quality



- Main object of interest
  - $v(x, t)$: is a <u>scalar field</u> of pollutant concentration on a <u>bounded domain</u>
- Data: measurements of dozens pollutants and weather related quantities at various times and locations
- Funding & collaborators: Alastair Forbes - **NPL**

Introduction - problem structure

- Data assimilation: Estimate $v(x, t)$ given data $\mathcal{Y}_t$ obtained at different locations
- Inference procedure
  - α) Choose model for $v(x, t)$
  - β) Fit models to data to get **model parameters** θ
  - γ) Improve our sensing capabilities,
    - move sensors to better locations, possibly on-line
- This talk:
  - α) continuous time Linear Gaussian model - Kalman filter
  - β) and γ) performed jointly using on-line gradient methods

Outline

- Model

$$dV_t = \mathcal{B} V_t dt + Q^{\frac{1}{2}} dW_t$$
$$dY_t = \mathscr{F} V_t dt + \tau dZ_t$$

  typically $V_t \in$ some Hilbert space U, is unknown and observations $Y_t \in \mathbb{R}^{d_y}$
  - example for V_t: the advection diffusion equation

- Filtering and parameter estimation
- Optimal sensor placement
- Joint parameter estimation and sensor placement
- Numerical results and discussion

Modelling for space time processes

- Various approaches for space time processes
  - Large scale regression of Gaussian Processes:
    - Banerjee, Gelfand, Finley, Sang 08, Rue, Martino, Chopin 09, Lindgren, Rue, Lindström 11
  - Linear state space models, GPs & Kalman filters
    - Wikle, Cressie 99, Sahu et. al. 05, 07,....,Duan, Gelfand, Sirmans 09, Sarkka et. al. 12, 13, ...
  - ... and many more
- Linear SPDE approach
  - [Sigrist, Künsch & Stahel, JRSSB 16]
  - tractable space time covariance properties (“non separable”)
  - efficient inference:
    - Kalman filtering, and MCMC for estimating θ

Model

- Stochastic Advection-Diffusion

$$\partial_t v + \zeta v - \nabla \cdot \Sigma \nabla v + \mu^T \nabla v = \epsilon$$

- 2D bounded domain, periodic boundaries
- ϵ noise
- Parameters: $\theta = (\zeta, \mu, \Sigma, ...)$
- Whittle 54, 63,...,Sigrist, Künsch & Stahel 16

Some particulars

- Σ is composed of a rotation & translations

$$\Sigma = \frac{1}{\rho_1^2} \begin{bmatrix} \cos\psi & \sin\psi \\ -\gamma\sin\psi & \gamma\cos\psi \end{bmatrix}^T \begin{bmatrix} \cos\psi & \sin\psi \\ -\gamma\sin\psi & \gamma\cos\psi \end{bmatrix}$$

- SPDE with "Matern" type noise for ϵ

$$d\epsilon(t) = \underbrace{\sigma\left(\triangle - \frac{1}{\rho_0^2} I\right)^{-1}}_{:=Q^{1/2}} dW(t)$$

with W being space time Brownian motion.

Basis and projections

- $V_t \in$ Hilbert space, U, with standard Fourier basis
  - zero mean functions: $\int v dx = 0$.
  - and span of $\psi_k(x) = \frac{1}{2\pi}\exp(ik \cdot x)$ where $k \in \mathbb{Z}^2 \setminus \{0\}$
- Decomposition of v:

$$v(x,t) = \sum_{k \in \mathbb{Z}^2/\{0\}} v_k(t)\psi_k(x)$$

with $v_k = \langle v, \psi_k \rangle = \int_{\mathbb{T}} u\psi_k(x)dx$.

- Noise process:

$$W_t = \sum_{k \in \mathbb{Z}^2/\{0\}} \sigma_k W_k(t)\, \psi_k(x),$$

with $W_k(t)$ are i.i.d. Brownian motions, $\sum_{k \in \mathbb{Z}^2/\{0\}} {\sigma_k}^2 < \infty$.

SDE form of dynamics

- SPDE (on U) is Ornstein Uhlenbeck (OU)

$$dV_t = \mathcal{B} V_t dt + Q^{\frac{1}{2}} dW_t$$

- Each $v_k(t)$ is a scalar OU process.

$$dv_k(t) = -b_k(\psi, \gamma, \rho_1, \mu) v_k dt + \sigma_k(\rho_0, \sigma) dW_k(t)$$

with coefficients depending on parameters

$$\begin{aligned} b_k &= \zeta + \frac{1}{\rho_1} \Sigma^{11}(\gamma, \psi) k_1^2 + \frac{1}{\rho_1} 2\Sigma^{12}(\gamma, \psi) k_1 k_2 \\ &+ \frac{1}{\rho_1} \Sigma^{22}(\gamma, \psi) k_2^2 + \mu_1 k_1 + \mu_2 k_2 \\ \sigma_k &= \frac{\sigma}{2\pi} (|k|^2 + \frac{1}{\rho})^{-1} \end{aligned}$$

Observations

- V_t is latent/unknown
- Can model observation as a linear projection $\mathscr{F} : U \to \mathbb{R}^{d_y}$.
- At a fixed location o_l:

$$\mathscr{F} V(o_l, t) = \frac{1}{|B_{o_l}(r)|} \int_{B_{o_l}(r)} V(t, x) dx$$

- Add noise either in:
  - discrete time:

$$Y_n = \mathscr{F} V_{t_n} + Z_n, \quad Z_n \sim \mathcal{N}(0, \tau^2 I),$$

or continuous time:

$$dY_s = \mathscr{F} V_s ds + \tau dZ_s$$

Filtering

- Conditional distr. or **Filter**

$$\pi_t(\cdot) = P(V_t \in \cdot \,|\, \mathcal{Y}_t, \theta, o) \quad \text{here} = \mathcal{N}(m_t, P_t)$$

where $\mathcal{Y}_t = \sigma(Y_s; s \leq t)$.

- Bayes rule (or Kallianpur-Striebel)

$$\pi_t(\varphi) = \frac{\rho_t(\varphi)}{\rho_t(1)}$$

Discr. time $\rho_n(\varphi) = E_X\left[\varphi(X_n)\exp\left(-\frac{1}{2\tau^2}\sum_{l=1}^{n}\left(Y_l - \mathscr{F}(V_{t_l})\right)^2\right)\right]$

Cont. time $\rho_t(\varphi) =$

$$E_X\left[\varphi(X_t)\exp\left(\frac{1}{\tau^2}\int_0^t \mathscr{F}(V_s)^T dY_s - \frac{1}{2\tau^2}\int_0^t |\mathscr{F}(V_s)|^2\, ds\right)\right]$$

Kalman filter

- In discrete time there are standard recursions for $m_n, P_n, \rho_n(1) = \mathcal{N}(c_n, \Upsilon_n)$

$$
\begin{aligned}
\mu_n &= A_{t_n} m_{n-1} \\
\Sigma_n &= A_{t_n} P_{n-1} A_{t_n}^* + \int_{t_{n-1}}^{t_n} A_t Q A_t^* dt \\
c_n &= \mathscr{F} \mu_n \\
\Upsilon_n &= \mathscr{F} \Sigma_n \mathscr{F}^* + \tau^2 I \\
K_n &= \Sigma_n \mathscr{F}^* \Upsilon_n^{-1} \\
m_n &= \mu_n + K_n (Y_n - c_n) \\
P_n &= (I - K_n \mathscr{F}) \Sigma_n
\end{aligned}
$$

$$
A(t) = \exp\left(\mathcal{B}(t - t_{n-1})\right), \quad t > t_n
$$

Kalman filter

- In continuous time:

$$dm_t = -\mathcal{B} m_t dt + \frac{1}{\tau^2} P_t \mathscr{F}^* \left(dY_t - \mathscr{F} m_t \right)$$

  P comes from **Riccatti** equation

$$\dot{P}_t = \mathcal{B} P_t + P_t \mathcal{B}^* + Q - \frac{1}{\tau^2} P_t \mathscr{F}^* \mathscr{F} P_t^*$$

- Marginal likelihood:

$$\rho_t(1) = \exp\left(\frac{1}{\tau^2} \int_0^t \mathscr{F}(m_s)^T dY_s - \frac{1}{2\tau^2} \int_0^t |\mathscr{F}(m_s)|^2 \, ds \right)$$

On the Riccati equation and stability

Problem well studied in inf. dim./Hilbert space setting: Athans, Falb 60-s, Curtain, Bensoussan 70-s, Khapalov 80-s, ...

- P_t is unique continuous mild solution

(Curtain 75, 78...) **IF**

- $(\mathcal{B}, Q^{1/2})$ exp. stabilisable
  - there is a K s.t. $\mathcal{B} - Q^{1/2}K$ generates stable semigroup
- $(\mathcal{B}, \mathscr{F})$ exp. detectable
  - as above for $\mathcal{B} - K\mathscr{F}$
- Then KF stable and

$$P_t \to P_\infty$$

Recursive Maximum likelihood - discrete time

- Suppose sensor positions o are fixed.
- Recall $\theta = (\zeta, \mu, \psi, \gamma, \rho_0, \rho_1, \tau, \sigma, ...)$

- Discrete time on-line gradient update

$$\theta_n = \theta_{n-1} + \gamma_n \nabla \log p_{\theta_{0:n-1}}(Y_n | Y_{1:n-1})$$

where

$$\nabla \log p_{\theta_{0:n-1}}(Y_n | Y_{1:n-1}) = -\frac{1}{2} \nabla_{\theta_{n-1}} \log \det(\Upsilon_n) - \frac{1}{2} \nabla_{\theta_{n-1}} \left((Y_n - c_n)^* \Upsilon_n^{-1} (Y_n - c_n) \right)$$

- Need the tangent filter

Recursive Maximum likelihood - discrete time

- In parallel to Kalman filter compute:

$$\begin{aligned}
\nabla_{\theta_{n-1}}\mu_n &= A_{t_n}\nabla_{\theta_{n-1}}m_{n-1} + \left(\nabla_{\theta_{n-1}}A_{t_n}\right)m_{n-1}\\
\nabla_{\theta_{n-1}}\Sigma_n &= ....\\
\nabla_{\theta_{n-1}}c_n &= \mathscr{F}\nabla_{\theta_{n-1}}\mu_n\\
\nabla_{\theta_{n-1}}\Upsilon_n &= \mathscr{F}\nabla_{\theta_{n-1}}\Sigma_n\mathscr{F}^*
\end{aligned}$$

- Tangent update

$$\begin{aligned}
\nabla_{\theta_{n-1}}K_n &= \nabla_{\theta_{n-1}}\left(\Sigma_n\mathscr{F}^*\Upsilon_n^{-1}\right)\\
\nabla_{\theta_{n-1}}m_n &= \nabla_{\theta_{n-1}}\mu_n + \nabla_{\theta_{n-1}}K_n(Y_n - c_n) + K_n(Y_n - \nabla_{\theta_{n-1}}c_n)\\
\nabla_{\theta_{n-1}}P_n &= (I - \nabla_{\theta_{n-1}}K_n\mathscr{F})\Sigma_n + (I - K_n\mathscr{F})\nabla_{\theta_{n-1}}\Sigma_n
\end{aligned}$$

Recursive Maximum likelihood - discrete time

- Recall $\rho_n(1) = p(Y_{1:n}|\theta, o)$
- Approch based on ergodicity of $\frac{1}{n}\nabla_\theta \log \rho_n(1)$,

$$\frac{1}{n}\sum_{n\geq 1} \nabla_\theta \log p(Y_n|Y_{1:n-1}) \to \int \nabla_\theta \log p(Y_n|Y_{1:n-1})\nu_{\theta,o}(dm, dP, dY)$$

- Stochastic gradient descent
  - ..., Legland & Mevel 99, Doucet & Tadic 04, ..., 18

Recursive Maximum likelihood - cont. time

- Want to write something like:

$$"\dot{\theta}_t = \gamma(t)\nabla_{\theta_t}\left(\frac{1}{t}\log\rho_t(1)\right)"$$

  to get an explicit recursion

$$d\theta_t = \frac{\gamma(t)}{\tau^2}(\mathscr{F}\mathring{m}_t)^*\left(dY_t - \mathscr{F}(m_t)dt\right)$$

  with $\mathring{m}_t = "\nabla_{\theta_t} m_t"$ obeying a SDE derived from m_t, P_t

- some analysis:
  - recent: Surace & Pfister 18, using Sirignano & Spiliopoulos 17
  - older: Sen & Athreya 77, Ljung 78,..., Levanony, Shwartz, Zeitouini 93,...

Optimal Sensor placement

- Suppose θ is known and fixed.
- Uncertainty in π_t depends on sensor locations via $\mathscr{F}$
- Optimise locations to minimise uncertainty in P_t or P_∞?
- Many approaches:
  - Burns & Rautenberg 15, Hintermuller et. al. 17, Herzog, Riedel, Ucinski, 17, Zhuk et. al. 16 , Walter 19, Zhang & Morris 18, Demetriou et. al, 04,...
- Ideas very similar to experiment design
  - Chaloner & Verdinelli 95

Optimal Sensor placement

- Find sensor locations $o = (o_1, \ldots, o_m)$ that minimise:

$$\lim_t \frac{1}{t} \int_0^t TrJP_t \quad \text{or} \quad TrJP_\infty$$

with J is an optional operator to emphasise on particular areas

- Control or Optimisation problem of the Riccati equation
- Average cost case: Burns & Rautenberg 15,
  - fixed or moving sensors
  - Problem has a solution, P_t Frechet differentiable, Galerkin convergence
- Using P_∞: Morris 11, Zhang & Morris 18
  - higher $\frac{1}{\tau^2}\mathscr{F}^*\mathscr{F}$ means lower TrP_∞

Optimal Sensor placement

- Problem well posed for our model setting
- Can write an online gradient as an ODE

$$do_t = -\beta_t \nabla_{o_t} \left(TrJP_t \right) dt$$

- We are controlling the Riccati equation to optimise steady state
- Existence of P_∞ allows to extend arguments in RML
  - for fixed θ there is an optimal solution $o^*(\theta)$

Joint parameter estimation and optimal sensor placement

- We want to combine both gradients

$$d\theta_t = \frac{\gamma_t}{\tau^2}(\mathscr{F}\mathring{m}_t)^* \left(dY_t - \mathscr{F}^{o_t}(m_t)dt\right)$$

$$do_t = -\beta_t \nabla_{o_t}\left(TrJP_t\right)dt$$

  while propagating KF and its θ and o gradients

- Note
  - observation model non-homogeneous
  - γ_t and β_t need to have different time scales

- Works very well in practice!

Convergence results

- Ultimate aim is to solve

$$\hat{\theta} = \arg\max_{\theta} \tilde{\mathcal{L}}\Big(\theta, \arg\min_{o} \tilde{\mathcal{J}}(\theta, o)\Big), \quad \hat{o} = \arg\min_{o} \tilde{\mathcal{J}}\big(\hat{\theta}, o\big).$$

- We can establish weaker

$$\lim_{t\to\infty} \nabla_\theta \tilde{\mathcal{L}}(\theta(t), o(t)) = \lim_{t\to\infty} \nabla_o \tilde{\mathcal{J}}(\theta(t), o(t)) = 0.$$

- Convergence results
  - are formulated for general state space models
  - verified also for the models described here

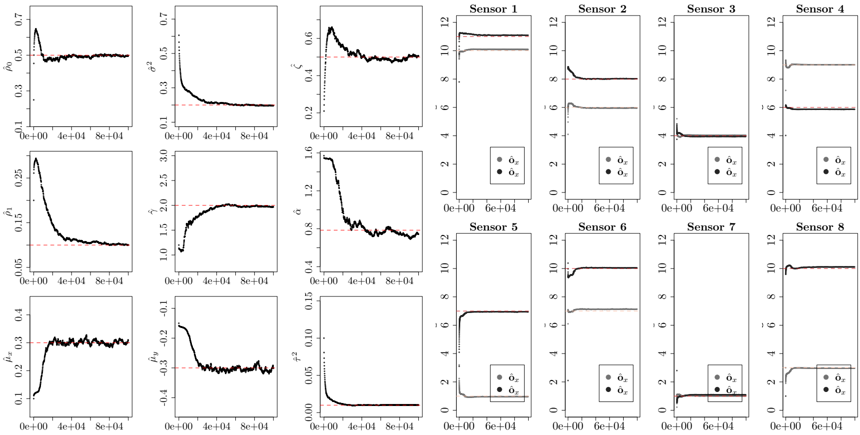
Convergence results - approach

- Extend Borkar's two time scale stochastic approximation
- ODE method and Benaim's asymptotic pseudo trajectory method

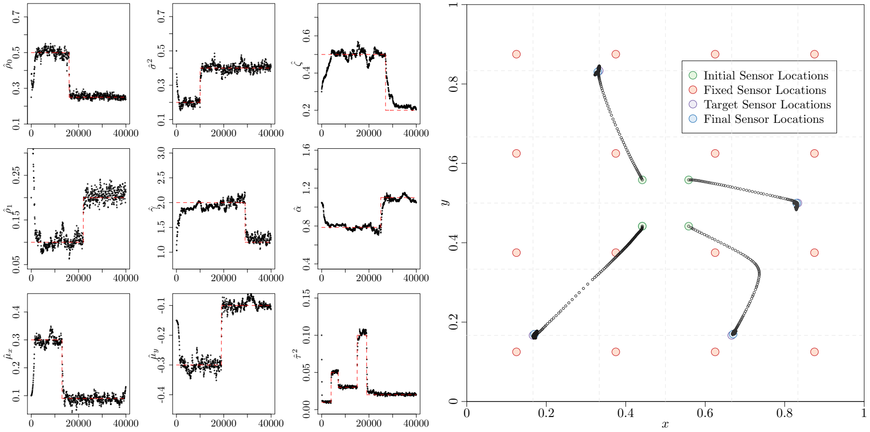
- Convergence of θ_t: $\nabla_{\theta_t} \log \tilde{\mathcal{L}}(m, P; \theta_t, o^*(\theta_t)) \to 0$
- Convergence of o_t to $o^*(\theta)$

- Specific requirements:
  - ergodicity for m_t, P_t and gradient dynamics for all θ, o,
  - moment conditions on invariant measure ν and $\nabla\nu$,
  - regularity of solutions of Poisson equation and gradients

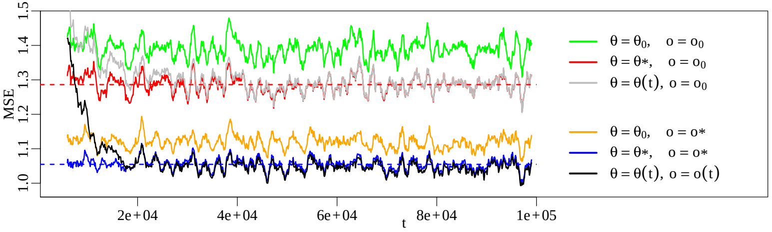
Convergence of parameters and sensor locations



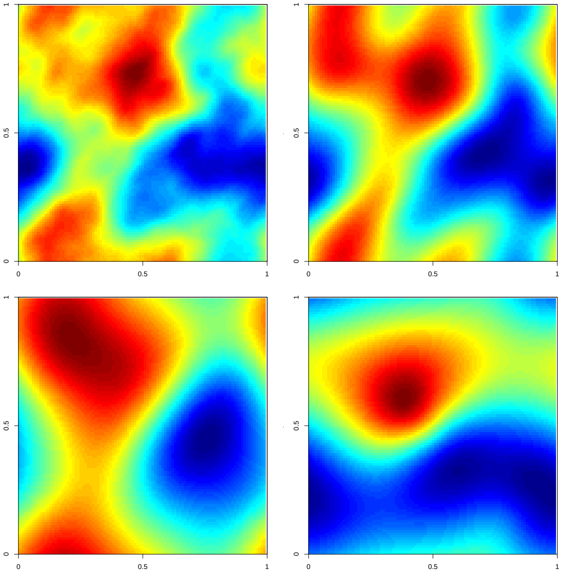
Adapting to changes of parameters (fixed step size)



Mean square error



Estimation of $v(x, t)$



Discussion

- Method is provably convergent to stationary points of a bilevel optimisation problem
- For simple linear Gaussian models allows scalable practical inference, sensor control

- Possible extensions; non-linear/non Gaussian models:
  - EnKFs:
    - D. Crisan, P. Del Moral, A. Jasra, H. Ruzayqat, Log-Normalization Constant Estimation using the Ensemble Kalman-Bucy Filter with Application to High-Dimensional Models, 2021
  - Particle Filters: low dimensional problems
    - A. Beskos, D. Crisan, A. Jasra, N. K., H. Ruzayqat, Score-Based Parameter Estimation for a Class of Continuous-Time State Space Models, SIAM SISC 2021.

Preprints

- Case studies in this talk:
  - L. Sharrock, N. K. Joint Online Parameter Estimation and Optimal Sensor Placement for the Partially Observed Stochastic Advection-Diffusion Equation, 2020.
- General theoretical results:
  - L. Sharrock, N. K. Two-Timescale Stochastic Gradient Descent in Continuous Time with Applications to Joint Online Parameter Estimation and Optimal Sensor Placement, 2020.