

Data Assimilation for Stochastic Transport Models

Dan Crisan

Imperial College London

12TH INTERNATIONAL ENKF WORKSHOP JUNE 12-14, 2017
BERGEN, NORWAY

- What is Data Assimilation ?
- What is Stochastic Filtering ?
- Data Assimilation \Leftrightarrow Stochastic Filtering: a dictionary.
- Why is the high-dimensional problem hard ?
- Stochastic transport models
- Data assimilation for STMs

What is data assimilation ?

Jana de Wiljes, Sebastian Reich, Andrew Stuart "Data Assimilation: The Mathematics of Connecting Dynamical Systems to Data, MFO"

The seamless integration of large data sets into computational models provides one of the central challenges for the mathematical sciences of the 21st century. When the computational model is based on dynamical systems and the data is time ordered, the process of combining data and models is called data assimilation.

Mark Asch, Marc Bocquet, Ma lle Nodet, "Data Assimilation: Methods, Algorithms, and Applications"

Data assimilation is an approach that combines observations and model output, with the objective of improving the latter.

Peter Jan van Leeuwen, "Nonlinear Data Assimilation for high-dimensional systems with geophysical applications.

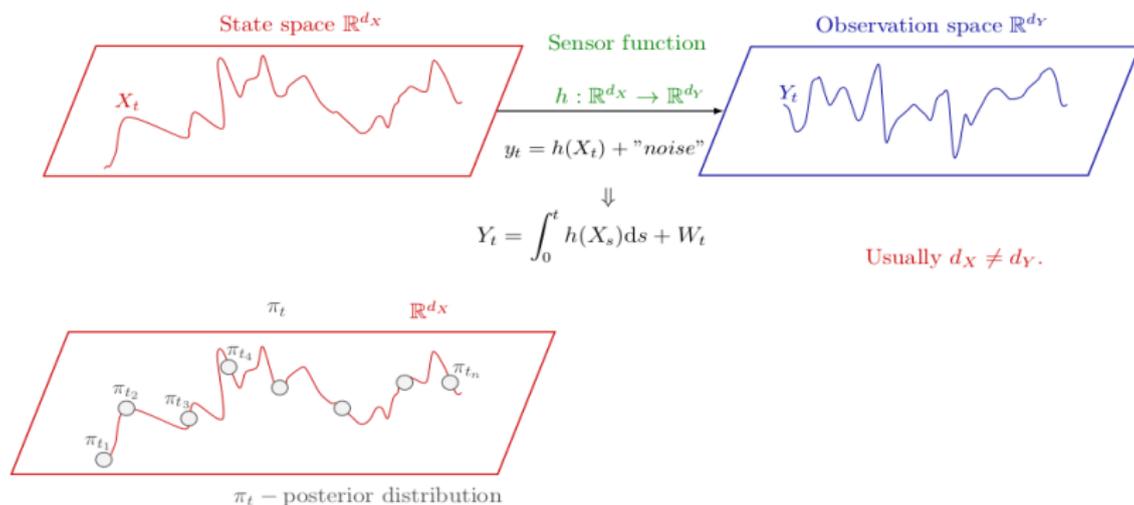
Data assimilation combines past knowledge of a system in the form of a numerical model with new information about that system in the form of observations of that system.

Sebastian Reich, Colin Cotter, "Probabilistic Forecasting and Bayesian Data Assimilation"

Data assimilation was coined in the computational geoscience community to describe methodologies for improving forecasting skill by combining measured data with computer generated forecasts. More specifically, data assimilation algorithms meld computational models with sets of observations in order to, for example, reduce uncertainties in the model forecasts or to adjust model parameters.

Stochastic filtering

The process of using partial observations and a stochastic model to make inferences about an evolving dynamical system.



Courtesy of Oana Lang (Imperial College London)

- X the signal process - “hidden component”
- Y the observation process - “the data”

The filtering problem: Find the conditional distribution of the signal X_t given $\mathcal{Y}_t = \sigma(Y_s, s \in [0, t])$, i.e.,

$$\pi_t(A) = \mathbb{P}(X_t \in A | \mathcal{Y}_t), \quad t \geq 0, \quad A \in \mathcal{B}(\mathbb{R}^d).$$

Discrete framework:

$$\begin{aligned} \{X_t\}_{t \geq 0} \quad & \text{Markov chain} \quad \mathbb{P}(X_t \in dx_t | X_{t-1} = x_{t-1}) = f_t(x_t | x_{t-1}) dt, \\ \{X_t, Y_t\}_{t \geq 0} \quad & \mathbb{P}(Y_t \in dy | X_t = x_t) = g_t(y | x_t) dy \end{aligned}$$

Continuous framework:

$$\begin{aligned} dX_t &= f(X_t)dt + \sigma(X_t)dV_t, \\ dY_t &= h(X_t)dt + dW_t. \end{aligned}$$

The description of a numerical approximation for the solution of the filtering problem should contain three parts:

particle approximations	Gaussian approximations
$(\underbrace{a_j(t)}_{\text{weight}}, \underbrace{v_j^1(t), \dots, v_j^d(t)}_{\text{position}})_{j=1}^n$	$(\underbrace{a_j(t)}_{\text{weight}}, \underbrace{v_j^1(t), \dots, v_j^d(t)}_{\text{mean}}, \underbrace{\omega_j^{11}(t), \dots, \omega_j^{dd}(t)}_{\text{covariance matrix}})_{j=1}^n$
$\pi_t \rightsquigarrow \pi_t^n = \sum_{j=1}^n a_j(t) \delta_{v_j(t)}$	$\pi_t \rightsquigarrow \pi_t^n = \sum_{j=1}^n a_j(t) N(v_j(t), \omega_j(t))$

2. The law of evolution of the approximation:

particle approximations	Gaussian approximations
$\pi_t^n \xrightarrow[\text{model}]{\text{mutation}} \bar{\pi}_{t+\delta}^n \xrightarrow[\{Y_s\}_{s \in [t, t+\delta]}]{\text{selection}} \pi_{t+\delta}^n$	$\pi_t^n \xrightarrow[\text{model}]{\text{forecast}} \bar{\pi}_{t+\delta}^n \xrightarrow[\{Y_s\}_{s \in [t, t+\delta]}]{\text{assimilation}} \pi_{t+\delta}^n$

3. The measure of the approximating error:

$$\sup_{\varphi \in C_b} \mathbb{E} [|\pi_t^n(\varphi) - \pi_t(\varphi)|], \quad \hat{\pi}_t - \hat{\pi}_t^n, \quad \|\pi_t^n - \pi_t\|_{\text{TV}}.$$

The quantized information is modelled by n stochastic processes

$$\{p_i(t), t > 0\} \quad i = 1, \dots, n, \quad p_i(t) \in \mathbb{R}^N.$$

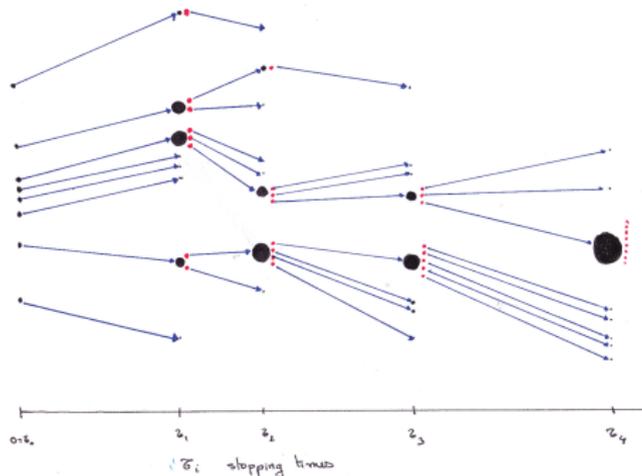
- We think of the processes p_i as the trajectories of n (generalized) particles.
- Typically $N > d$, where d is the dimension of the state space.

$$\pi_t^n = \Lambda_t^n(p_i(t), t > 0 \quad i = 1, \dots, n).$$

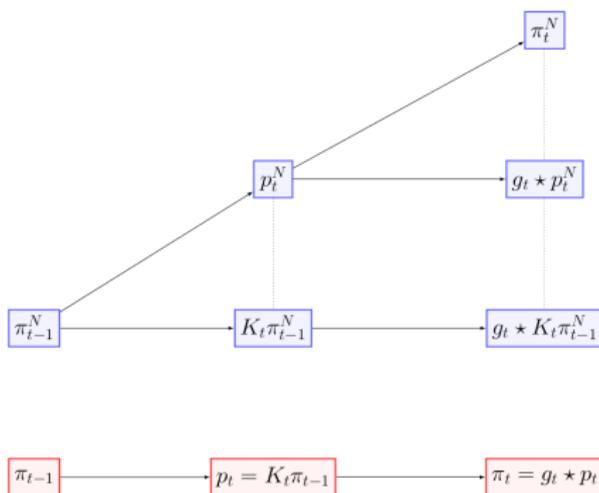
- Generalized particle filters:
 - classical particle filters
 - gaussian approximations
 - wavelets
 - grid methods

The classical/standard/bootstrap/garden-variety particle filter

$\pi_n = \{\pi_n(t), t \geq 0\}$ the occupation measure/empirical distribution of a sequence of a system of weighted particles



$$\pi_n(0) = \sum_{i=1}^n \frac{1}{n} \delta_{x_i^n} \quad \longrightarrow \quad \pi_n(t) = \sum_{i=1}^n \bar{a}_i^n(t) \delta_{V_i^n(t)}.$$



Courtesy of Oana Lang (Imperial College London)

Theorem

π_n converges to π . Moreover

$$\sup_{t \in [0, T]} \sup_{\{\varphi \in C_b^1(\mathbb{R}^d), \|\varphi\|_{1, \infty} \leq 1\}} \|\pi_t^n(\varphi) - \pi_t(\varphi)\|_P \leq \frac{\alpha(T, p)}{\sqrt{n}}.$$

Remarks:

- If d is small to moderate, then the standard particle filter can perform very well in the time parameter n .
- Under certain conditions, the Monte Carlo error of the estimate of the filter can be uniform with respect to the time parameter.
- The function $x_k \mapsto g(x_k, y_k)$ can convey a lot of information about the hidden state, especially so in high dimensions. If this is the case, using the prior transition kernel $f(x_{k-1}, x_k)$ as proposal will be ineffective.
- It is then known that the standard particle filter will typically perform poorly in this context, often requiring that $N = \mathcal{O}(\kappa^d)$.

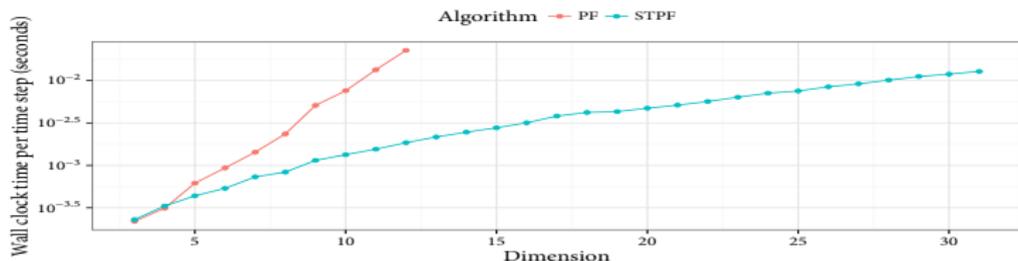
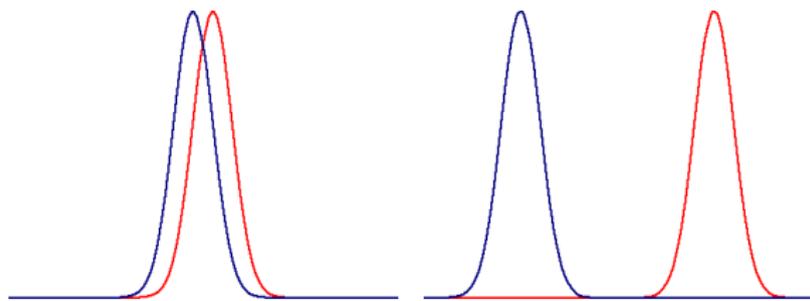


Figure: Computational cost per time step to achieve a predetermined RMSE versus model dimension, for standard particle filter (PF) and STPF.

Consider

- $\Pi_0 = \mathcal{N}(0, 1)$ (mean 0 and variance matrix 1).
- $\Pi_1 = \mathcal{N}(1, 1)$ (mean 1 and variance matrix 1).
- $\Pi_d = \mathcal{N}(d, 1)$ (mean d and variance matrix 1).
- $d(\Pi_0, \Pi_1)_{\text{TV}} = 2\mathbb{P}[|X| \leq 1/2], X \sim \mathcal{N}(0, 1)$.
- $d(\Pi_0, \Pi_d)_{\text{TV}} = 2\mathbb{P}[|X| \leq d/2], X \sim \mathcal{N}(0, 1)$.
- as d increases, the two measures get further and further apart, becoming singular w.r.t. each other.
- as d increases, it becomes increasingly harder to use standard importance sampling, to construct a sample from Π_3 by using a proposal from Π_1 , weighting it using $\frac{d\Pi_d}{d\Pi_0}$ and (possibly) resample from it.



Consider

- $\Pi_0 = \mathcal{N}((0, \dots, 0), \mathbf{I}_d)$ (mean $(0, \dots, 0)$ and covariance matrix \mathbf{I}_d).
- $\Pi_d = \mathcal{N}((1, \dots, 1), \mathbf{I}_d)$ (mean $(1, \dots, 1)$ and covariance matrix \mathbf{I}_d).
- $d(\Pi_0, \Pi_d)_{\text{TV}} = 2\mathbb{P}[|X| \leq d/2], X \sim \mathcal{N}(0, 1)$.
- as d increases, the two measures get further and further apart, becoming singular w.r.t. each other exponentially fast.
- it becomes increasingly harder to use standard importance sampling, to construct a sample from Π_d by using a proposal from Π_0 .
- ‘Moving’ from Π_0 to Π_d is equivalent to moving from a standard normal distribution $\mathcal{N}(0, 1)$ to a normal distribution $\mathcal{N}(d, 1)$ (the total variation distance between $\mathcal{N}(0, 1)$ and $\mathcal{N}(d, 1)$ is the same as that between Π_1 and Π_2).

Remedies:

- Tempering *
- Sequential data assimilation in space *
- Model Reduction (High \mapsto Low Res)*
- Using hybrid models
- Hamiltonian Monte Carlo
- Jittering
- Nudging
- Localization

Discrete framework:

$$\begin{aligned} \{X_t\}_{t \geq 0} & \text{ Markov chain } \mathbb{P}(X_t \in dx_t | X_{t-1} = x_{t-1}) = f_t(x_t | x_{t-1}) dx_t, \\ \{X_t, Y_t\}_{t \geq 0} & \mathbb{P}(Y_t \in dy_t | X_t = x_t) = g_t(y_t | x_t) dy_t \end{aligned}$$

Remedies:

- **a tempering procedure** For $i = 1$ to d
 - reweight the particle using $g_t^{\frac{1}{d}}$ and (possibly) resample from it
 - move particles using an MCMC that leaves $g_t^{\frac{k}{d}} f_t \Pi_{[0, t-1]}$ invariant

Beskos, DC, Jasra, *On the stability of SMC methods in high dimensions*, 2014.
 Kantas, Beskos, Jasra, *Sequential Monte Carlo for inverse problems*, 2014.

- **Sequential data assimilation in space**

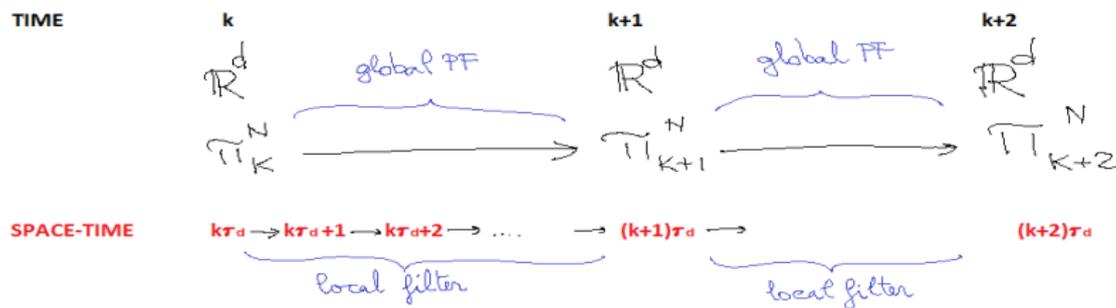
Assume that there exists an increasing sequence of sets $\{\mathcal{A}_{k,j}\}_{j=1}^{\tau_{k,d}}$, with $\mathcal{A}_{k,1} \subset \mathcal{A}_{k,2} \subset \dots \subset \mathcal{A}_{k,\tau_{k,d}} = \{1 : d\}$, for some integer $0 < \tau_{k,d} \leq d$, such that we can factorize:

$$g(x_k, y_k) f(x_{k-1}, x_k) = \prod_{j=1}^{\tau_{k,d}} \alpha_{k,j}(y_k, x_{k-1}, x_k(\mathcal{A}_{k,j})),$$

for appropriate functions $\alpha_{k,j}(\cdot)$, where $x_k(\mathcal{A}) = \{x_k(j) : j \in \mathcal{A}\} \in \mathbb{R}^{|\mathcal{A}|}$.

This holds when:

- one can obtain a factorization for the prior term $f(x_{k-1}, x_k)$ by marginalising over subsets of co-ordinates.
- the likelihood component $g(x_k, y_k)$ can be factorized when the model assumes a local dependence structure for the observations.



For $j = 1$ to $\tau_d - 1$

- Move particle according to $q_{k+1,j}(x_{k+1}(\mathcal{A}_{k+1,j}) | x_k, x_{k+1}(\mathcal{A}_{k+1,j-1}))$.
- weight the particle using $\frac{\alpha_{k+1,j}(y_{k+1}, x_k, x_{k+1}(\mathcal{A}_{k+1,j-1}))}{q_{k+1,j}(x_{k+1}(\mathcal{A}_{k+1,j}) | x_k, x_{k+1}(\mathcal{A}_{k+1,j-1}))}$ and (possibly) resample from it.

Beskos, CD, Jasra, Kamatani, Zhou, A Stable Particle Filter in High-Dimensions, 2017

- jitter the particle

- reweight the particle using $g_t^{\frac{1}{d}}$ and (possibly) resample
- move particles using a suitable chosen kernel

D. C., Joaquin Miguez, Nested particle filters for online parameter estimation in discrete-time state-space Markov models, <http://arxiv.org/abs/1308.1883>.

D. C., Joaquin Miguez, Uniform convergence over time of a nested particle filtering scheme for recursive parameter estimation in state-space Markov models, <https://arxiv.org/abs/1603.09005>.

- stay on the typical set by using Hamiltonian Monte Carlo

The stochastic transport models we study are inspired from geometric mechanics. Holm (2015) introduced a new set of stochastic PDEs modelling the motion of an either compressible, or incompressible fluid in \mathbb{R}^3 resulting from a stochastically constrained variational principle $\delta S = 0$, with action, S , given by

$$S(u, p, q) = \int \left(\ell(u, q) dt + \langle p, dq + \mathcal{L}_{dx_t} q \rangle_V \right), \quad (1)$$

- $\ell(u, q)$ unperturbed deterministic fluid Lagrangian, written as a functional of velocity vector field u and advected quantities q .
- $\langle p, q \rangle_V := \int \langle p(x), q(x, t) \rangle dx$, $q \in V$ and their dual elements $p \in V^*$.
- $\mathcal{L}_{dx_t} q$ is the Lie derivative of the advected quantity $q \in V$, along a vector field dx_t

$$dx_t(x) = u(x, t) dt - \sum_i \xi_i(x) \circ dW_i(t). \quad (2)$$

- the methodology incorporates physically meaningful stochastic perturbations in fluid dynamics equations.

DD Holm, Variational principles for stochastic fluid dynamics, Proc. R. Soc. A 471, 2015.

The SPDEs resulting from the SVP $\delta S = 0$ has the form

$$d \frac{\delta \ell}{\delta \mathbf{u}} + \mathcal{L}_{d\mathbf{x}_t} \frac{\delta \ell}{\delta \mathbf{u}} - \frac{\delta \ell}{\delta \mathbf{q}} \diamond \mathbf{q} dt = 0, \quad \text{and} \quad d\mathbf{q} + \mathcal{L}_{d\mathbf{x}_t} \mathbf{q} = 0, \quad (3)$$

The diamond operation $(\diamond) : T^*V \rightarrow \mathfrak{X}^*$ is defined for a vector space V with $(\mathbf{q}, \mathbf{p}) \in T^*V$ and vector field $\xi \in \mathfrak{X}$ is given by $\langle \mathbf{p} \diamond \mathbf{q}, \xi \rangle_{\mathfrak{X}} := \langle \mathbf{p}, -\mathcal{L}_{\xi} \mathbf{q} \rangle_V$ for the pairings $\langle \cdot, \cdot \rangle_V : T^*V \times TV \rightarrow \mathbb{R}$ and $\langle \cdot, \cdot \rangle_{\mathfrak{X}} : \mathfrak{X}^* \times \mathfrak{X} \rightarrow \mathbb{R}$ with $\mathbf{p} \diamond \mathbf{q} \in \mathfrak{X}^*$.

If we choose in equation (3) the Lagrangian

$$l(\mathbf{u}) = \frac{1}{2} \|\mathbf{u}\|_{L^2}^2 = \frac{1}{2} \int |\mathbf{u}|^2 d^3x$$

that is, the kinetic energy of the incompressible fluid, constrained to only allow divergence free velocity vector fields, independent of the advected variable \mathbf{q} , and compute the curl of right hand side we obtain the **stochastic Euler equation**.

Two dimensional incompressible fluid flow u defined on 2D-torus $\Omega = [0, L_x] \times [0, L_y]$ modelled by the two-dimensional Euler equations with forcing and dampening. Let $q = \hat{z} \times \text{curl } u$ denote the vorticity of u , where \hat{z} denotes the z -axis. For a scalar field $g : \Omega \rightarrow \mathbb{R}$, we write $\nabla^\perp g = (-\partial_y g, \partial_x g)^T$. Let $\psi : \Omega \times [0, \infty) \rightarrow \mathbb{R}$ denote the stream function.

$$\begin{aligned} \partial_t q + (u \cdot \nabla) q &= Q - r q \\ u &= \nabla^\perp \psi \\ \Delta \psi &= q. \end{aligned}$$

where

- Q is the forcing term given by $Q = 0.1 \sin(8\pi x)$
- r is a positive constant that can be interpreted as defining the large scale dissipation time scale.
- we consider slip flow boundary condition $\psi|_{\partial\Omega} = 0$.
- we use FireDrake to solve the PDE.

Observations:

- u is observed on a subgrid of the signal grid (t continuous/discrete)

$$y_i(t) = u(x_i, t) + \varepsilon \xi_i. \quad \xi_i \sim N(0, 1).$$

- ε is calibrated to the standard deviation of the true solution over a coarse grid cell.

Initialization (1000 particles):

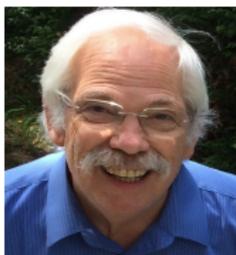
- run the SPDE on the interval $[-a, 0]$ initialized at time $-a$ with the projection of the true (deterministic) solution on the coarse grid.
- evolve the projection of the true (deterministic) solution on the coarse grid with a random velocity.

Data Assimilation Algorithm:

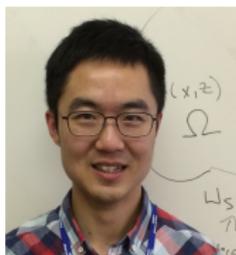
- Forecast: Particles are evolved according to the SPDE with informed importance sampling step.
- Analysis: Adapted tempering procedure.



Colin



Darryl



Wei



Igor



Oana



Peter Jan



Roland



Ajay



Nikolas



Francesc



Alex



Joaquin