A coherent structure approach for parameter estimation in Data Assimilation

John Maclean[1], Naratip Santitissadeekorn[2], Christopher KRT Jones[1]

[1]Department of Mathematics and RENCI, University of North Carolina at Chapel Hill
[2]Department of Mathematics, University of Surrey, Guildford

June 13, 2017

Data Assimilation setup

We consider x, θ to have come from a model

$$\dot{\theta} = 0,$$
$$\dot{x} = f(x, \theta, t)$$

with the interpretation that θ represents *parameters* and x represents *tracers in the flow*. Our goal is **to estimate the parameters θ affecting a flow, given observations x^o of passive tracers in the flow.**

Data Assimilation setup

We consider x, θ to have come from a model

$$\dot{\theta} = 0,$$
$$\dot{x} = f(x, \theta, t)$$

with the interpretation that θ represents *parameters* and x represents *tracers in the flow*. Our goal is **to estimate the parameters θ affecting a flow, given observations x^o of passive tracers in the flow.** As usual we employ Bayes' rule to formulate the data assimilation problem:

$$p(\theta, x | x^o) \propto p(x^o | \theta, x) p(\theta, x)$$

(Standard, naive) particle filters

The particle filter sequentially approximates the distribution of θ at time n by a set $\{\theta_n^{(i)}, w_n^{(i)}\}$, $i = 1, \ldots, N$ of particles and weights.

The weight update for the i-th particle is

$$\begin{aligned} w_n^{(i)} &\propto w_{n-1}^{(i)}\, p(x_n^o | x_n^{(i)}, \theta_n^{(i)}) \\ &\approx w_{n-1}^{(i)}\, \exp\left(-\frac{1}{2}(x_n^o - x_n^{(i)})^T \mathbf{R}^{-1}(x_n^o - x_n^{(i)})\right), \end{aligned}$$

where $\mathbf{R}$ is the covariance matrix for the observations.

The key quantity by which the particle filter gains information is the innovation, $x_n^o - x_n^{(i)}$.

Toy model

We consider a kinematic traveling wave model in the co-moving frame, perturbed by an oscillatory disturbance and stochastically perturbed in the x_1-direction:

$$dx_1 = c - A\sin(Kx_1)\cos(x_2) + \varepsilon l_1 \sin(k_1(x_1 - c_1 t))\cos(l_1 x_2) + \sigma dW \quad (1)$$

$$dx_2 = AK\cos(Kx_1)\sin(x_2) + \varepsilon k_1 \cos(k_1(x_1 - c_1 t))\sin(l_1 x_2). \quad (2)$$

We will perform experiments to attempt to 'discover' the true values of ε, and/or k_1, given all the other parameters are fixed and given observations from a run with the 'true' parameters.

Toy model

We consider a kinematic traveling wave model in the co-moving frame, perturbed by an oscillatory disturbance and stochastically perturbed in the x_1-direction:

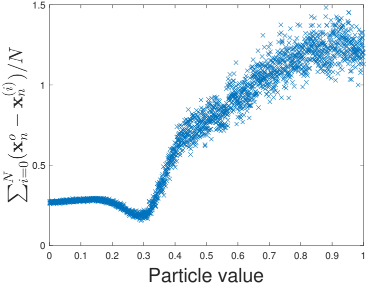
$$dx_1 = c - A\sin(Kx_1)\cos(x_2) + \varepsilon l_1 \sin(k_1(x_1 - c_1 t))\cos(l_1 x_2) + \sigma dW \qquad (1)$$

$$dx_2 = AK\cos(Kx_1)\sin(x_2) + \varepsilon k_1 \cos(k_1(x_1 - c_1 t))\sin(l_1 x_2). \qquad (2)$$

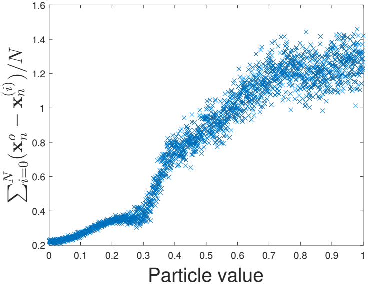
We will perform experiments to attempt to 'discover' the true values of ε, and/or k_1, given all the other parameters are fixed and given observations from a run with the 'true' parameters.

This flow contains two gyres. For the value $\varepsilon = 0.3$ that we choose to be the truth, tracer trajectories inside the gyres are dominated by chaotic advection on long time scales.

What will the particle filter do if we initialise all tracers in the boundaries of the gyres? Let us look at the innovations...



(a) Observation taken at $t = 20$.



(b) Observation taken at $t = 30$.

The figures show experiments in which we uniformly spaced 2000 guesses for ε in $[0, 1]$, numerically integrated 50 tracers using each value of ε, and calculated the innovation.

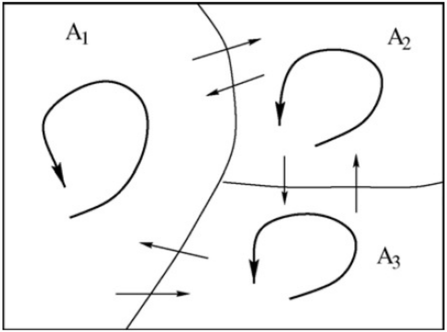
Coherent patterns

We would like to perform Data Assimilation by assimilating the pattern of the observed tracers. In so doing, we hope to exploit the robustness of coherent patterns to peturbations, while preserving whatever information the tracers carry on the model parameters.

Coherent patterns

Almost-invariant sets

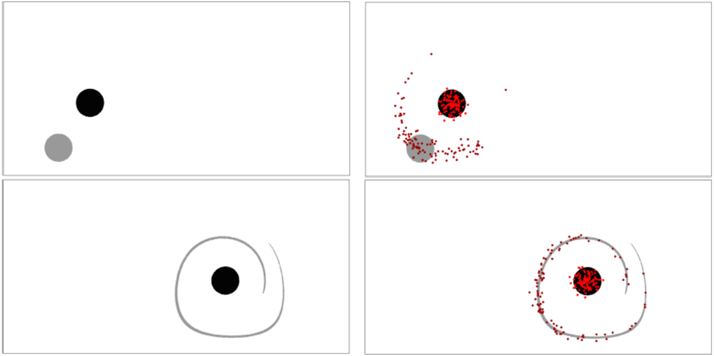
Trajectories stay in the almost-invariant sets for a comparatively long time before escaping to another region [Dellnitz and Jung 99].



Coherent patterns

Coherent sets

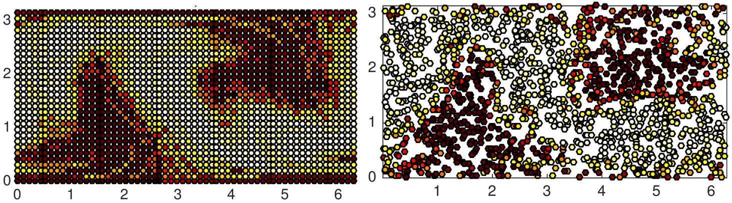
Coherent sets are regions in state space, for example a coherent vortex or nonlinear jet, that move along with the flow without dispersing. Coherent sets must also be robust under small diffusive peturbations.



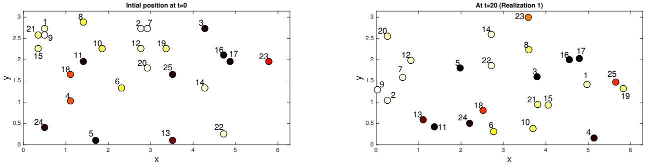
Coherent patterns

Usage in Data Assimilation

We rely on numerical methods that extract the coherent pattern from data, so that we can produce both a ‘simulated’ pattern (extracted from simulated tracer positions) and an ‘observed’ pattern (extracted from data).
We use PCA to find the almost-invariant sets, shown here for many tracers:



and for few tracers:



Assimilating patterns, p1

Suppose the random variable y^o represents the observed coherent pattern. The updated statement of Bayes' rule is that

$$p(\theta|y^o) \propto p(y^o|\theta)p(\theta),$$

where

$$p(y^o|\theta) = \int p(y^o|\theta, x_{0:n})p(x_{0:n}|\theta)dx_{0:n}$$

...is hard to evaluate.

Assimilating patterns, p1

Suppose the random variable y^o represents the observed coherent pattern. The updated statement of Bayes' rule is that

$$p(\theta|y^o) \propto p(y^o|\theta)p(\theta),$$

where

$$p(y^o|\theta) = \int p(y^o|\theta, x_{0:n})p(x_{0:n}|\theta)dx_{0:n}$$

...is hard to evaluate. We proceed by replacing $x_{0:n}$ in the likelihood function in the integral with $\hat{x}_{0:n}(\theta)$, a realisation of $x_{0:n}$ given θ, to obtain

$$p(y^o|\theta) \approx p(y^o|\theta, \hat{x}_{0:n}(\theta)).$$

Assimilating patterns, p2

Unfortunately we still cannot calculate $p(y|\theta, \hat{x}_{0:n}(\theta))$. We turn instead to an Approximate Bayesian Computation (Rubin, 1984; Sisson, Fan, Tanaka, 2007), which uses a distance function ρ to substitute for the likelihood function.

A basic algorithm description is

Algorithm 0

Step 1. Sample $\theta \sim p(\theta)$

Step 2. Sample $y \sim p(y|\theta)$

Step 3. Accept θ if $\rho(y, y^o) \leq \varepsilon$.

SMC-ABC

Algorithm 0

Step 1. Sample $\theta \sim p(\theta)$

Step 2. Sample $y \sim p(y|\theta)$

Step 3. Accept θ if $\rho(y, y^o) \leq \varepsilon$.

In this way all ABC algorithms sample from

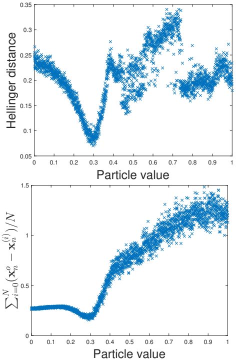
$$p_\varepsilon(\theta, y|y^o) \propto p(y|\theta)p(\theta)I_\varepsilon(y),$$

where

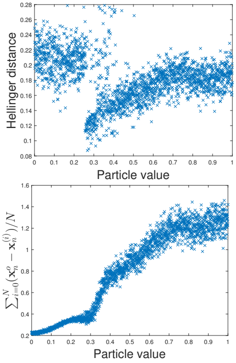
$$I_\varepsilon(y) = \begin{cases} 1 & \text{for } \rho(y, y^o) < \varepsilon, \\ 0 & \text{otherwise.} \end{cases}$$

We choose to use the Hellinger distance for ρ, so the fundamental source of information in our ABC scheme is the Hellinger distance between an observed pattern and a simulated pattern.

Building blocks for DA schemes - innovation vs Hellinger distance



(a) t=20



(b) t=30

We repeat the prior experiment, now including results for the Hellinger distance between the observed pattern and the simulated pattern at each value of ε.

Final Algorithms

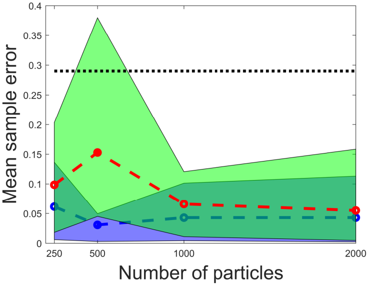
We used the merging Particle Filter from (Nakano, Ueno, Higuchi, 2007). This scheme provides a way to gain sample diversity in the resampling step, and the weights are designed to preserve the sample and variance of the original sample.

We used a Sequential Monte Carlo implementation of ABC from (Del Moral, Doucet, Jasra, 2011). This scheme employs an adaptive sequence of tolerance levels ε to control the rate of sample collapse towards the posterior. A Metropolis-Hastings algorithm is used in the SMC-ABC to search parameter space in lieu of a resampling algorithm.

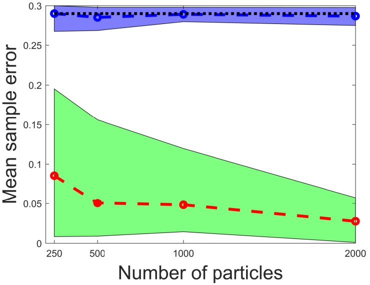
We control the computational cost of each method to be similar by limiting the number of particles in SMC-ABC.

Numerical Results

Initial tracer locations



(a) Uniform initial tracer deployment.

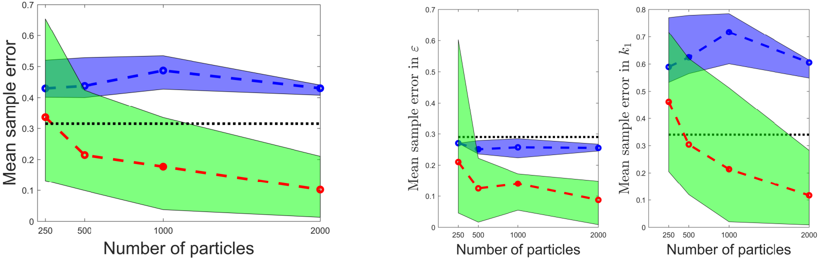


(b) Tracers deployed within gyre boundaries.

This experiment shows error results from 20 runs of the **Particle Filter (blue)** and **SMC-ABC (green)**, assimilating a single observation at $t = 30$. Dashed lines indicate mean error; patches give the 10% and 90% quartiles.

Numerical Results

Observation time step; multiple parameters



Left: Mean error in estimating parameters in the **Particle Filter (blue)** and **SMC-ABC (green)**. Dashed lines show the mean error of 20 repetitions from each numerical method, while the coloured patches show the 10% and 90% percentiles. *Left:* Mean error in estimating both ε and k_1. *Right*: Error in estimating the individual parameters.

Summary

-To our knowledge, this is the first attempt to use coherent patterns in (Lagrangian) data assimilation.

-In contrast to previous work, the tracer trajectories are not assimilated directly but instead a coherent structure or pattern is assimilated.

-our numerical results demonstrate that this new approach is remarkably superior to the trajectory-based Lagrangian DA (employing the standard/naive particle filter) in the situation where the number of tracers is large and the drifter trajectories are dominated by chaotic advection.