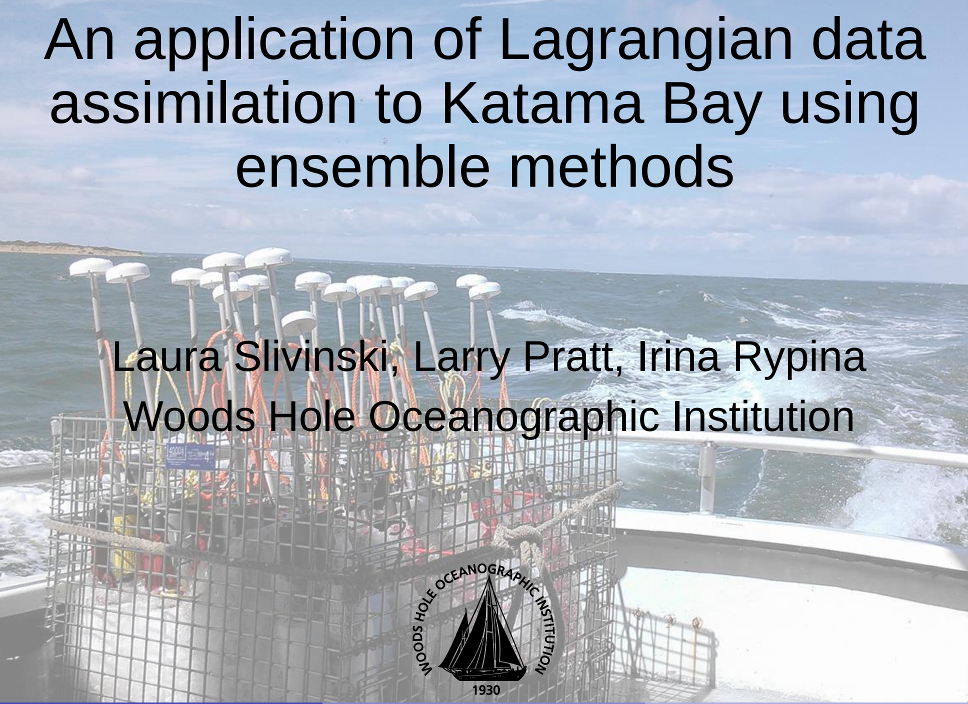


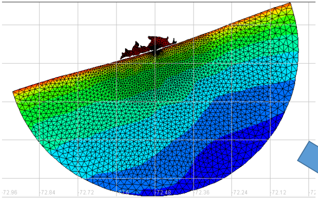
An application of Lagrangian data assimilation to Katama Bay using ensemble methods

Laura Slivinski, Larry Pratt, Irina Rypina
Woods Hole Oceanographic Institution

Outline

- Introduction to data assimilation (DA)
    - Variational vs sequential
    - Bayes' rule
- Lagrangian data assimilation (LaDA)
    - General methods
    - Complications
- Katama Bay
    - Model
    - Data

General picture



Uncertain model estimates

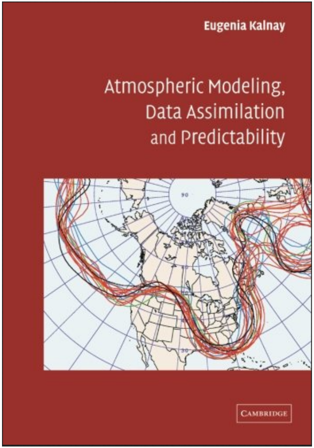
Data Assimilation

Uncertain observations

Better estimate of true state

(and uncertainty surrounding it)

Data assimilation



Setup:

- Given dynamical system $\dot{x} = f(x)$ (deterministic or stochastic)
- Uncertainty in initial conditions $x_0 = x(t_0)$ *(prior)*
- Observations at discrete times $k = 1, 2, \ldots$: $y_k = h(x_k)+$noise *(likelihood)*
- Would like to estimate state at specific time, given the knowledge of observations: $x_k = x(t_k)$ *(posterior)*

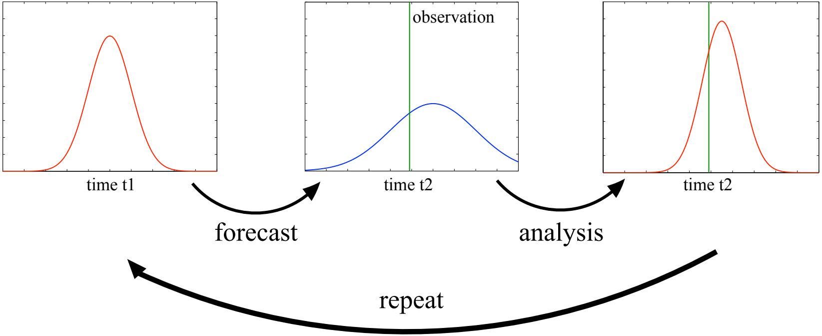
Variational vs Sequential

- **Variational methods** (3DVar, 4DVar) seek to minimize a cost function to find a best estimate of the initial condition or time series
  - $\mathcal{J}(\mathbf{x}) = \frac{1}{2}\left(\mathbf{x}-\mathbf{x}^b\right)^{\mathsf{T}}\mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^b\right) + \frac{1}{2}\left(\mathbf{y}-\mathbf{y}^o\right)^{\mathsf{T}}\mathbf{R}^{-1}\left(\mathbf{y}-\mathbf{y}^o\right)$
  - Uncertainty in estimate not calculated
  - Calculation of adjoint can be costly
- **Sequential methods** assimilate observations as they become available
  - Only use information about observations up to (and including) the current time
  - Kalman filter, extended Kalman filter
  - **Ensemble methods** (EnKF, particle filter)
    - Spread of ensemble quantitatively describes uncertainty in (posterior) estimate
    - Requires model evolution for many ensemble members

Sequential DA algorithms

Two steps:

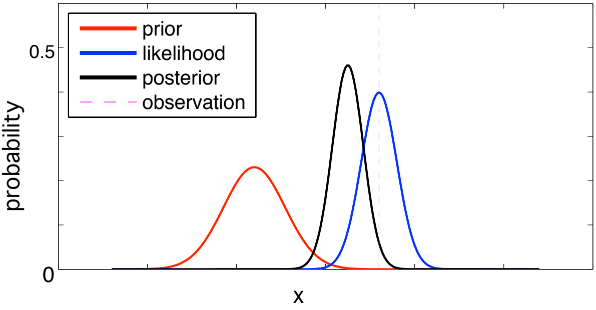
- Forecast (evolve previous estimate forward under dynamical system)
- Analysis (update current estimate with observation)



Data assimilation

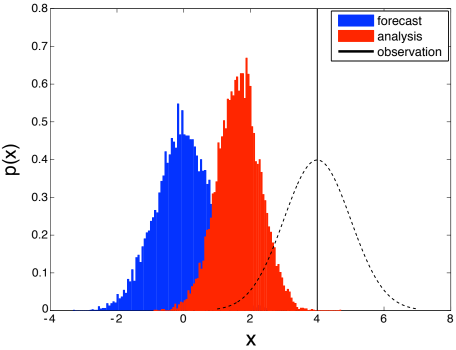
Application of Bayes' rule: $p(x|y) \propto p(x)p(y|x)$

$$p(x) : \text{prior}$$
$$p(y|x) : \text{observation likelihood}$$
$$p(x|y) : \text{posterior}$$



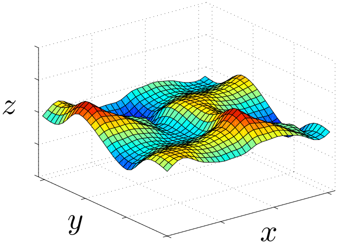
Ensemble Kalman filter

- Represent probability distribution with ensemble of state vectors $\{\mathbf{x}_i\}_{i=1,\ldots,N_e}$
- Forecast: evolve each ensemble member forward under model until observation is available.
- Analysis: update each ensemble member using observation.
  - $\mathbf{x}_i^{analysis} = \mathbf{x}_i^{forecast} + \mathbf{K}\left(\mathbf{H}\mathbf{x}_i^{forecast} - \mathbf{y}\right)$

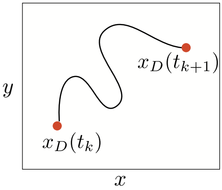


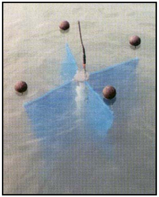
Lagrangian data assimilation

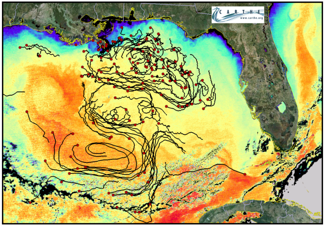
Suppose we want to estimate the (possibly high-dimensional) flow field $\mathbf{x}^F$, but the observations are of Lagrangian positions of passive drifters $\mathbf{x}^D$.



$$\dot{\mathbf{x}}^D = \mathbf{x}^F(\mathbf{x}^D, t)$$







Three approaches

- Eulerian (assimilate data from current meters, satellite obs, ...)
  - Directly measure velocities: no complicated observation operator, no strong nonlinearity in drifter trajectory, high accuracy in measurements
  - Difficult to obtain these observations: data is sparse
  - Fixed moorings cannot follow interesting structures in flow
- Pseudo-Lagrangian (approximate velocities as finite differences of subsequent drifter locations)
  - Valid for short times between observations
  - Doesn't use info about full drifter trajectory
- Fully Lagrangian (directly use drifter position data)
  - Forecasted vs observed Lagrangian velocities
  - Augmented vector approach - include drifters as variable in state vector

Fully Lagrangian approach - Molcard et al (2003)

Innovation (difference between obs and forecast): $\mathbf{v}^o(t_k) - \mathbf{v}^f(t_k)$, where

$$\mathbf{v}^o(t_k) = \frac{\mathbf{y}(t_k) - \mathbf{y}(t_{k-1})}{t_k - t_{k-1}},$$
$$\mathbf{v}^f(t_k) = \frac{\mathbf{x}_D^f(t_k) - \mathbf{x}_D^f(t_{k-1})}{t_k - t_{k-1}},$$

$\mathbf{y}$ is obs drifter position; $\mathbf{x}_D^f$ is forecasted drifter position, evolving previous observation forward using velocity estimate.

- Unlike pseudo-LaDA, compares Lagrangian velocities (instead of comparing Eulerian velocity to approximation from Lagrangian obs)
- This method does not use information about the entire particle trajectory, only the positions at two successive time steps.

Fully Lagrangian approach - Molcard et al (2003)

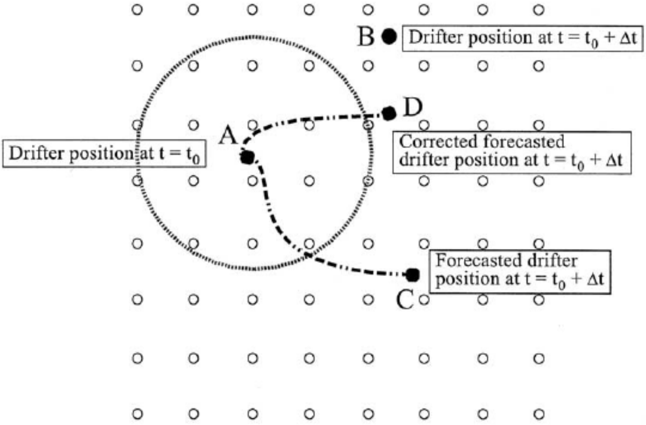


Figure 2. Schematic illustration of the assimilation algorithm. Given the drifter positions at $t = t_0$ (A) and $t = t_0 + \Delta t$ (B), the model forecast at $t = t_0 + \Delta t$ is improved from (C) to (D) by modifying the model Eulerian circulation field at $t = t_0$ within a circle of influence (model grid layout is shown in the background) using algorithm (5), which acts to minimize the distance between positions (B) and (C). The drifter position data are given at discrete time interval Δt, whereas the model simulated drifters can follow paths as shown between forecasted and corrected positions, (AC) and (AD), respectively.

Augmented vector approach

- Append drifter position $\mathbf{x}^D$ to flow state vector $\mathbf{x}^F$: $\mathbf{x} = \begin{pmatrix} \mathbf{x}^F \\ \mathbf{x}^D \end{pmatrix}$
  - $\mathbf{x}^F = (\mathbf{u}, \mathbf{v}, \mathbf{h}, [T, S, \dots])^\mathsf{T}$
  - $\mathbf{x}^D = (x_1, y_1, x_2, y_2, \dots, x_{ND}, y_{ND})^\mathsf{T}$
- Observation operator has simple, linear form: $\mathbf{H} = [\mathbf{0} \quad \mathbf{I}]$
- Flow and drifters are *both* forecasted and updated via DA scheme

The system evolution is

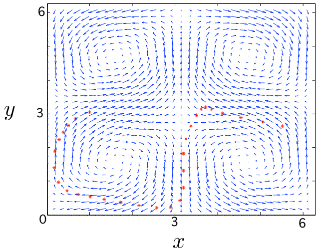
$$\begin{aligned} \dot{\mathbf{x}}^F &= f(\mathbf{x}^F) \\ \dot{\mathbf{x}}^D &= g(\mathbf{x}^F, \mathbf{x}^D) \end{aligned}$$

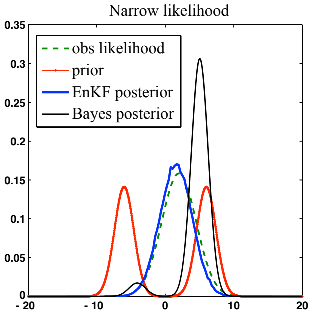
and the observations are

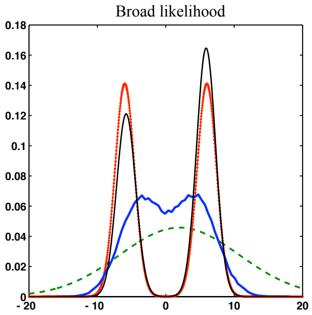
$$\begin{aligned} &\mathbf{y} = \mathbf{H}\mathbf{x} + \epsilon = \mathbf{x}^D + \epsilon, \\ &\epsilon \sim \mathcal{N}(0, \mathbf{R}) \end{aligned}$$

Nonlinearity & Non-Gaussianity

- Lagrangian data assimilation leads to non-Gaussian priors
- Flow may solve linear system, but drifters solve nonlinear system







Katama Bay

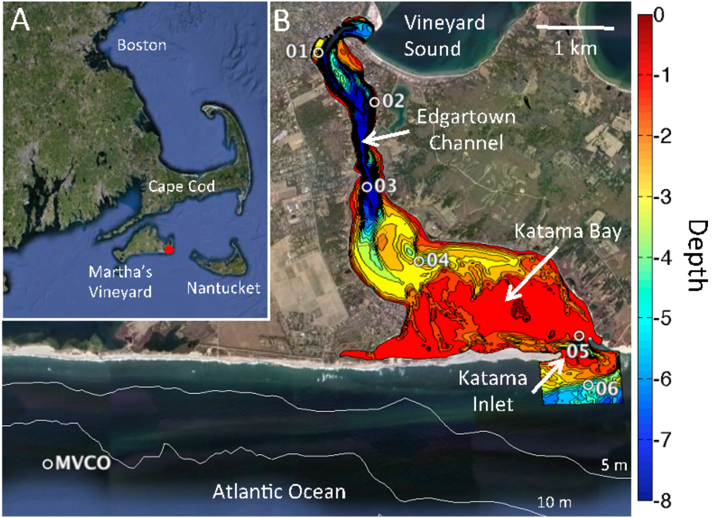


Figure: Setup of Katama Bay experiment in 2012. (Orescanin et al, 2014)

Katama Bay

Advanced Circulation Model (ADCIRC)

- Two-dimensional, hydrostatic, finite-element
- Based on shallow water dynamics: Eulerian state vector includes 2D velocity and fluid depth at each grid point
- Irregular grid developed by Mara Orescanin (MIT/WHOI)

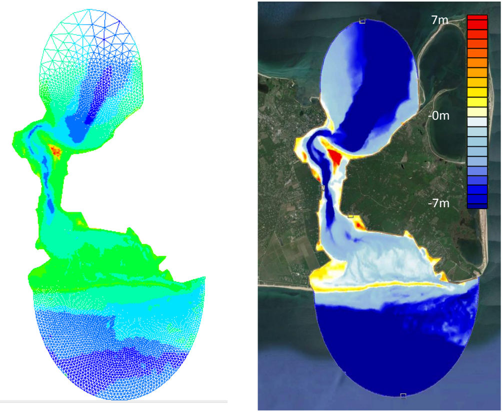


Figure: Left: bathymetry for ADCIRC, visualized with Blue Kenue. Right: domain overlaid on Google Earth (courtesy M. Orescanin)

Katama Bay

- Assimilate drifter data gathered in August 2013
- 12 drifters released for $\sim$ 2 hours, data available nearly continuously (more than 1x/second)

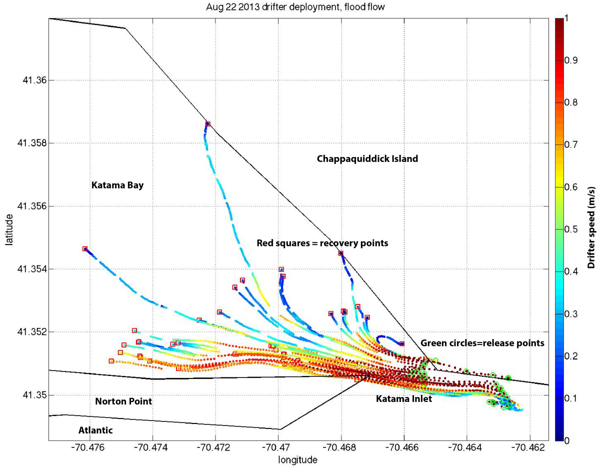


Figure: Courtesy B. Raubenheimer (WHOI), S. Elgar (WHOI), and J. MacMahan (NPS)

Summary

- Generally: LaDA involves assimilating drifter trajectories in order to estimate the velocity field
- Benefits over assimilating Eulerian (velocity) obs: drifters can track interesting structures, cover larger spatial domain; often cheaper and less sparse
- EnKF with augmented vector approach has not been applied with real Lagrangian data; will apply to small enclosed basin
- Future work: apply hybrid filter [Elaine Spiller's talk], compare results to EnKF; more drifter deployments with larger number of drifters, target specific structures; estimate bathymetry in attempt to improve model itself

References

A. Molcard et. al. (2003)
Assimilation of drifter observations for the reconstruction of the Eulerian circulation field
J. Geophys. Res. 108(C3), doi:10.1029/2001JC001240

K. Ide, L. Kuznetsov, and C.K.R.T. Jones (2002)
Lagrangian data assimilation for point vortex systems
J. Turbul.

M. Orescanin, B. Raubenheimer, and S. Elgar (2014)
Observations of wave effects on inlet circulation
Cont. Shelf Res. 82, 37–42.