Nonlinearity and stability of Lagrangian data assimilation

Amit Apte[1]

[1] Tata Institute of Fundamental Research, Centre for Applicable Mathematics (TIFR-CAM), Bangalore

Clouds, Convection, and Tropical Meteorology discussion meeting
ICTS Bangalore - 24 Jan 2013

Outline

1. Quick recap of the probabilistic formulation of data assimilation
2. What is Lagrangian data assimilation?
3. Data assimilation for skew-product systems
4. Effects of nonlinearity
5. En-route data assimilation

Outline

1. Quick recap of the probabilistic formulation of data assimilation
2. What is Lagrangian data assimilation?
3. Data assimilation for skew-product systems
4. Effects of nonlinearity
5. En-route data assimilation

Data assimilation is a state estimation problem

- Dynamical model for the state vector $x \in X$

$$\frac{dx}{dt} = f(x) \quad \text{with} \quad x(0) = x_0$$

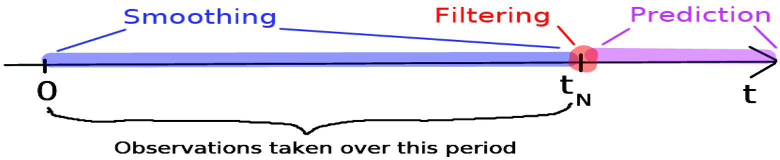
- The solution denoted by $x(t) = \Phi(x_0, t)$
- Given some noisy observations of the system at times $0 < t_1 < t_2 < \cdots < t_N$, we can consider three problems:
  - Smoothing (S_N): Obtain a state estimate $x_s(t)$ for $t < t_N$ using all the observations up to time t_N; In particular, determine $x_s(0)$
  - Filtering (F_i): Obtain a state estimate $x_f(t_i)$ using observations up to time t_i
  - Prediction (P_N): Obtain a state estimate $x_p(t)$ for $t > t_N$ (the time horizon of prediction is important).

Data assimilation is a state estimation problem

- Dynamical model for the state vector $x \in X$

$$\frac{dx}{dt} = f(x) \quad \text{with} \quad x(0) = x_0$$

- The solution denoted by $x(t) = \Phi(x_0, t)$
- Given some noisy observations of the system at times $0 < t_1 < t_2 < \cdots < t_N$, we can consider three problems:



Or data assimilation ≡ filtering ≡ determination of posterior distribution

Observations $y_i \in Y$ at time t_i depend on the state at that time.

$$y_i = H(x(t_i)) + \eta_k = H(\Phi(x_0, t_i)) + \eta_i, \quad i = 1, \dots, N$$

η_i is observational noise on Y which is usually finite dimensional.

$H : X \to Y$ is the observation operator

$H \circ \Phi_{t_i}$ is the observation operator from initial conditions x_0 to observation y_i at time t_i.

Probabilistic statement of Data assimilation problem: find the posterior distribution of the state conditioned on the observations

- $p(x(t)|y_1, y_2, \dots, y_N)$ for $t < t_N$: smoothing
- $p(x(t_N)|y_1, y_2, \dots, y_N)$: filtering
- $p(x(t)|y_1, y_2, \dots, y_N)$ for $t > t_N$: prediction

Bayes rule gives the posterior

Assume a *prior* probability distribution $\zeta(x_0)$ for x_0

- The smoothing density for the initial condition is given by Bayes theorem:

$$p(x_0|y_1, y_2, \dots, y_N) \propto \zeta(x_0) \prod_{k=1}^{K} p_\eta \left(y_k - H\left(\Phi(x_0, t_k)\right)\right)$$

- Filtering (and prediction) distribution can be obtained by pushing this forward to the final observation time t_N
- We use Markov Chain Monte Carlo (MCMC) methods to sample this density – *only applicable in small model problems - e.g. ODE models shown here, or 2D Navier-Stokes, etc.*
- Most deterministic methods can be related to finding the mode or the mean of this posterior

We will compare EnKF with the exact sampling

Recall the posterior density function

$$\pi(x_0) \propto \zeta(x_0) \prod_{k=1}^{K} p_\eta \left(y_k - h\left(\Phi(x_0, t_k)\right)\right)$$

- Metropolis-Hastings methods will be used to sample this density
- The Enesmble Kalman Filter EnKF tries to sample this distribution function (rather, its push forward) sequentially in time – exactly for linear systems, approximately for nonlinear systems

Main objective: Comparison of the "exact distribution functions" given by Metropolis-Hastings methods with the "EnKF distribution function"

- Rationale: Difference between EnKF and exact distribution is ONLY because of nonlinearities

We will also compare the exact posterior with the "truth"

- Compare the true values of the observed and unobserved variables with the means of the exact and the EnKF distribution
- Note: determining the initial conditions of long trajectories become exceedingly difficult because
  - The MCMC sampling method is a smoother, and
  - The drifter dynamics is chaotic
- A note about yesterday's discussion:
  - "Filter stability" refers to the situation when the filter estimate approaches the true state of the system (which is not accessible except through observations); Thus checking stability of filters in real life situations is possible only by comparing with observations that have not been assimilated
  - Even if "forecast errors" (or ensemble spread) decreases, the filter may not be stable (in fact, in many cases, too small an ensemble spread IS the main cause of filter instability)

Outline

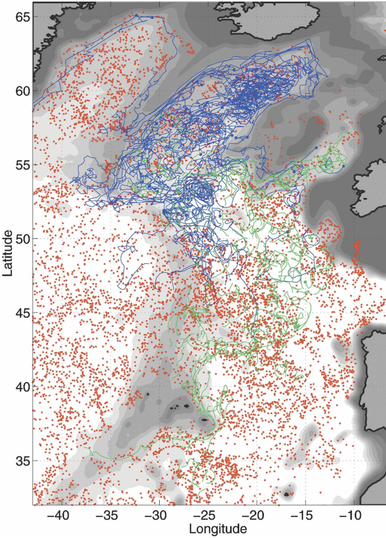
Lagrangian data assimilation (LaDA)

We have available with us the following

- Observations from floats or drifters
  - positions of the drifters
  - prognostic variables such as temperature and salinity of ocean at the location of the drifter
- A model for the velocity field and the other coupled dynamical variables (temperature, salinity, etc.)
- A model for the drifters (typically Lagrangian)

We are mainly interested in

- Estimate of the prognostic variables
- the positions of the drifters



from M. Lankhorst, W. Zenk (2006) JPO, **36**, p.43

Skew-product structure of the LaDA problem

An approach to assimilating the information about the observations of positions of drifters is to combine

- the prognostic variables (collectively denoted by x_v) and
- the positions of the drifters (denoted by x_d)

into the state vector:

$$x = (x_v, x_d)^T$$

The model has a skew-product structure, $\dot{x} = f(x) = (m_v(x), m_d(x))^T$, in the case of passive drifters (which is what we will assume):

$$\frac{dx_v}{dt} = m_v(x_v)\,, \qquad \frac{dx_d}{dt} = m_d(x_v, x_d) = V(x_d, x_v)\,,$$

where V is the velocity of the fluid flow at the point x_d.

Originated and studied extensively in the work of Ide, Jones, Kuznetsov, Salman, Spiller, ...

Observation operator is the projection on x_d variables

- *If the only observations are drifter locations*, then the observations at time t can be written as

$$y(t) = Hx(t) + \eta$$

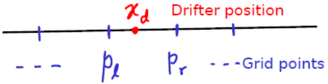
  where $x = (x_v, x_d)$, and thus $H = [0 \quad I]$ is just a projection; and η is the observational noise.
- Observation operator is linear in this case

Drifter model m_d is always nonlinear

Two main cases are the following:

- When using discretized velocity field, $x_v = (\dots, v_l, v_r, \dots)$; So velocity V at the position of drifter $x_d \in [p_l, p_r]$ is obtained by some interpolation $\implies$ at least quadratic non-linearity:

$$V(x_d, x_v) \propto (p_r - x_d)v_l + (x_d - p_l)v_r$$



- When using spectral methods, $x_v = (\dots, v_1, v_2, \dots)$ containing the Fourier modes of velocity $\implies$

$$V(x_d, x_v) \propto v_1 e^{ik_1 x_d} + v_2 e^{ik_2 x_d} + \dots$$

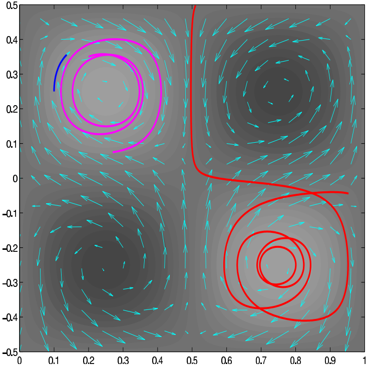
Later part of the talk will focus on the implications of:

- The skew-product structure of the model and
- the nonlinearity of the drifter model

Linear shallow water equations with Lagrangian data

For two dimensional velocity (u, v) and height h fields:

$$\frac{\partial u}{\partial t} = v - \frac{\partial h}{\partial s_1},$$
$$\frac{\partial v}{\partial t} = -u - \frac{\partial h}{\partial s_2},$$
$$\frac{\partial h}{\partial t} = -\frac{\partial u}{\partial s_1} - \frac{\partial v}{\partial s_2},$$



We seek periodic solutions on $\mathbb{R}^2$ in u, h, specifically, the following Fourier modes:

$$u(s_1, s_2, t) = -2\pi l \sin(2\pi k s_1)\cos(2\pi l s_2) u_0 + \cos(2\pi m s_2) u_1(t)$$
$$v(s_1, s_2, t) = 2\pi k \cos(2\pi k s_1)\sin(2\pi l s_2) u_0 + \cos(2\pi m s_2) v_1(t)$$
$$h(s_1, s_2, t) = \sin(2\pi k s_1)\sin(2\pi l s_2) u_0 + \sin(2\pi m s_2) h_1(t)$$

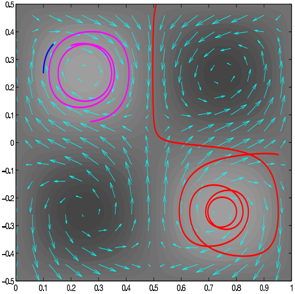
Linear shallow water equations with Lagrangian data

The amplitudes satisfy the following:

$$\dot{u}_0 = 0, \qquad \dot{u}_1 = v_1,$$
$$\dot{v}_1 = -u_1 \quad -2\pi m h_1, \quad \dot{h}_1 = 2\pi m v_1$$

The observations are the positions of the drifters that satisfy:

$$\dot{s_1}(t) = u(s_1(t), s_2(t), t), \quad \dot{s_2}(t) \quad = v(s_1(t), s_2(t), t)$$

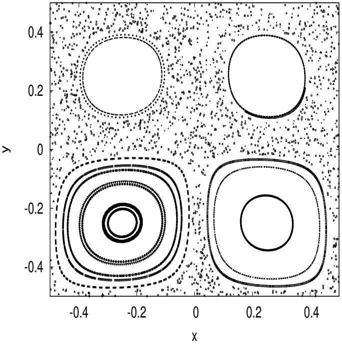


- Observations of drifter positions alone: Lagrangian data assimilation
- Observations of height field along the Lagrangian path: en-route data assimilation
- Main points of interest: this flow has
  - Nonlinear centre with shear (differential rotation) around it and
  - The unstable fixed points have chaotic regions near the separatrices
  - Velocity field is coupled to an additional variable (height)

Linear shallow water equations with Lagrangian data

A few more properties of the drifter dynamics:

- No attractor (the unperturbed flow is Hamiltonian)
- Some regions with regular trajectories (periodic / quasi-periodic)
- Some regions with chaotic trajectories



In the case of the model above, the velocity flow itself:

- has no attractor or chaotic dynamics
- is purely periodic

Thus, the nonlinearity is entirely in the drifter dynamics, which are the observed variables.

Outline

A linear(ized) discrete model for Lagrangian data assimilation

We will use the following linear model to illustrate the implications of skew-product structure for the forecast and update step of Kalman filter:

$$\begin{pmatrix} x_{v,k+1} \\ x_{d,k+1} \end{pmatrix} = \begin{pmatrix} Ax_{v,k} \\ M_v x_{d,k} + M_d x_{d,k} \end{pmatrix} = \begin{pmatrix} A & 0 \\ M_v & M_d \end{pmatrix} \begin{pmatrix} x_{v,k} \\ x_{d,k} \end{pmatrix} = M z_k$$

where $z_k = (x_{v,k}, x_{d,k})^T$.

- Observation operator is $H = [0 \quad I]^T$.
- Prior at initial time is Gaussian, uncorrelated:

$$x_{v,0}, x_{d,0} \sim N\left(\bar{z}_0, \begin{pmatrix} B_v & 0 \\ 0 & B_d \end{pmatrix}\right)$$

Forecast step leads to cross-correlations between Lagrangian and Eulerian variables

- Prior at the first time $k = 1$ is

$$P_1^f = \begin{pmatrix} AB_vA^T & AB_vM_v^T \\ M_vB_vA^T & M_VB_vM_V^T + M_dB_dM_d^T \end{pmatrix} = \begin{pmatrix} P_v^f & P_c^f \\ P_c^{fT} & P_d^f \end{pmatrix}$$

  The forecast covariance has cross-correlations which are entirely dependent on the covariance of the Eulerian variables B_v.

- The forecast is

$$z_1 = (x_{v,1}, x_{d,1}) = (Ax_{v,0}, M_vx_{v,0} + M_dx_{d,0})$$

The cross-correlations lead to update of Eulerian variables with observation of Lagrangian variables

Recall the observations are $y_k = Hz_k + \eta$ with $\eta \sim N(0, R)$ and $H = [0 \;\; I]$. The Kalman update equations are:

$$x^a_{v,1} = x^f_{v,1} + P^f_c (P^f_d + R)^{-1} (y - x^f_{d,1})$$
$$P^a_v = P^f_v - P^f_c (P^f_d + R)^{-1} P^{fT}_c$$

$$x^a_{d,1} = x^f_{d,1} + P^f_d (P^f_d + R)^{-1} (y - x^f_{d,1})$$
$$P^a_x = P^f_x - P^f_x (P^f_d + R)^{-1} P^{fT}_x$$

$$P^a_c = P^f_c - P^f_c (P^f_d + R)^{-1} P^{fT}_c$$

The skew-product leads to the following structure for the Kalman filter

- The update of the Lagrangian component is entirely independent of the Eulerian component (as if the state vector consisted of $x_{d,1}$ alone).
- The update of the Eulerian component of course involves the cross-correlations P_c^f.
- Thus,
  - the model creates the cross-correlations which
  - update the Eulerian components

Outline

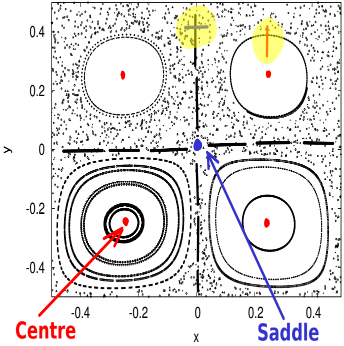
The drifter flow is highly nonlinear

Recall the velocity field [I will replace (s_1, s_2) by (x, y)]:

$$\dot{x} = u(x, y, t) = -2\pi l \sin(2\pi kx) \cos(2\pi ly) u_0 + \cos(2\pi my) u_1(t)$$
$$\dot{y} = v(x, y, t) = 2\pi k \cos(2\pi kx) \sin(2\pi ly) u_0 + \cos(2\pi my) v_1(t)$$

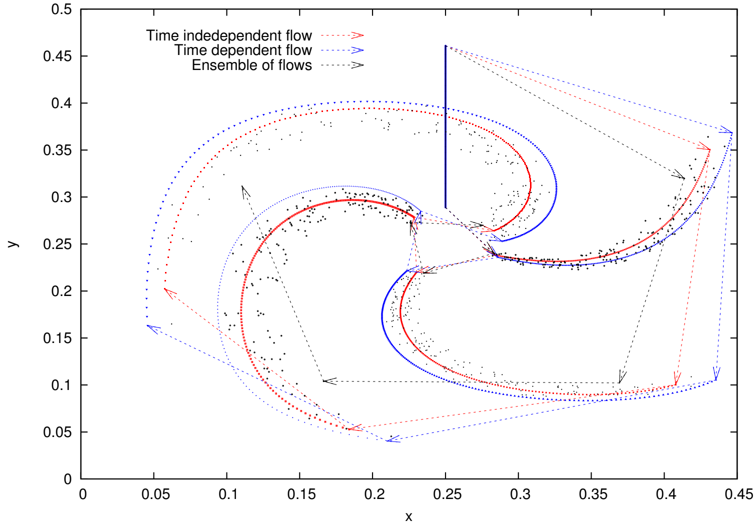
In order to illustrate the effects of this non-linearity, we will look at the time evolution of a straight line of drifters in following three cases:

- Unperturbed flow: $u_1 = v_1 = 0$
- Perturbations u_1, v_1 of $O(1)$
- An ensemble of perturbations u_1, v_1 drawn uniformly over $[0, 1]$



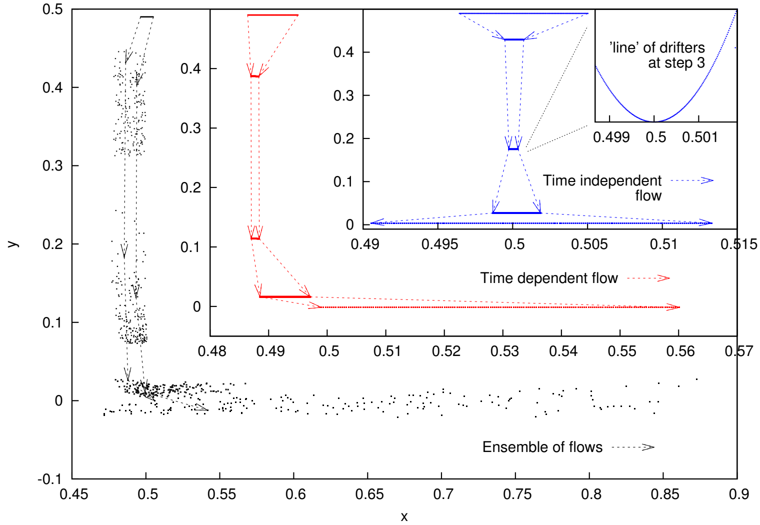
Near the centre, the effect of perturbation is not significant

Positions of drifters up to $t = 0.05$ (recall period of the flow $= 1$)



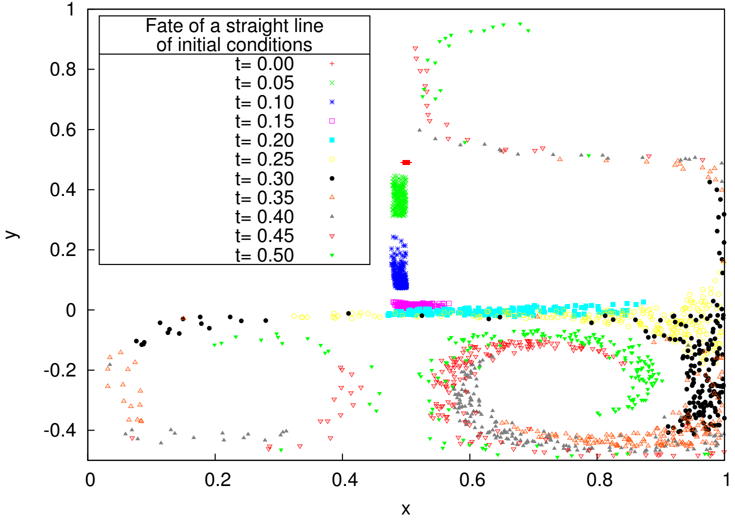
Near the saddle, the effect of perturbation is dramatic

Positions of drifters up to $t = 0.05$ (recall period of the flow $= 1$)



The perturbations of course lead to chaos

Positions of drifters up to $t = 0.5$ (recall period of the flow $= 1$)



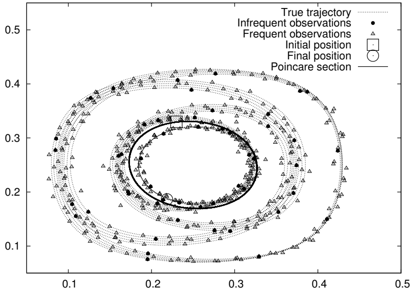
Implications for data assimilation

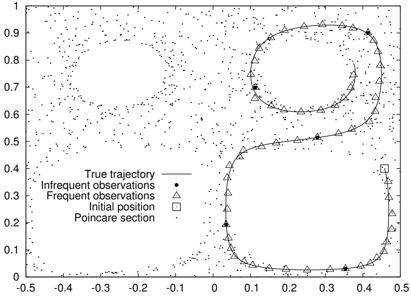
- Initial Gaussian distributions near the centre are distorted into "annulus" shaped non-Gaussian distributions
- Trajectories near the centre contain information mostly about the first Fourier mode u_0 and much less about the other modes (e.g. u_1, v_1, h_1)
- Trajectories near the saddle do contain information about u_1, v_1, h_1

Two notes:

- Such *a priori* conclusions about the model dynamics have implications for "adaptive observations."
- Dynamically important features affect the filter performance (e.g. eddy separation point).

Two true trajectories used for this study

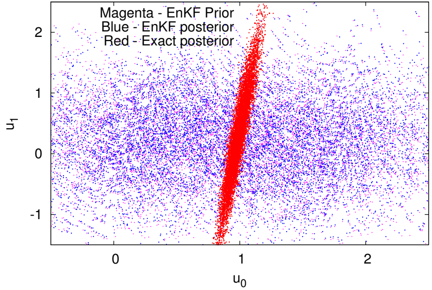


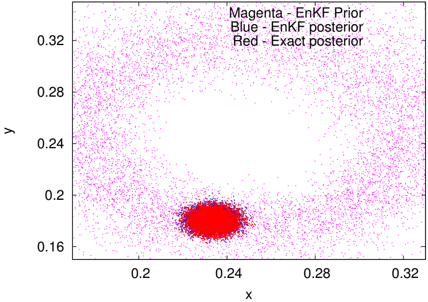


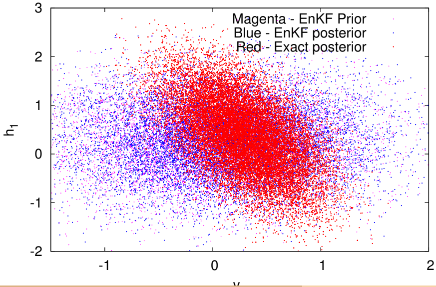
For each of the above trajectories, we chose

- frequent ($T_{obs} = 0.01$) or infrequent ($T_{obs} = 0.1$) observations
- prior distribution at $t = 0$ with a broad covariance $(1.0, 0.7, 0.7, 0.7, 0.005, 0.005)$ or peaked covariance $(0.1, 0.07, 0.07, 0.07, 0.005, 0.005)$ for the velocity variables

The prior and posterior at time of first observation

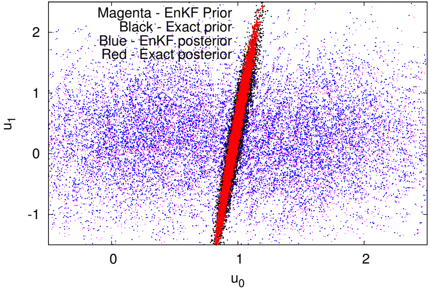


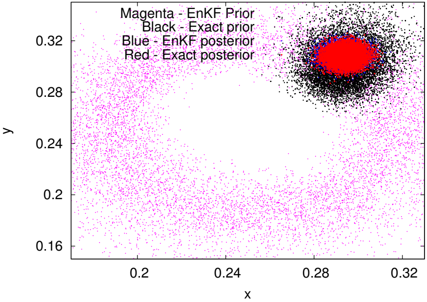


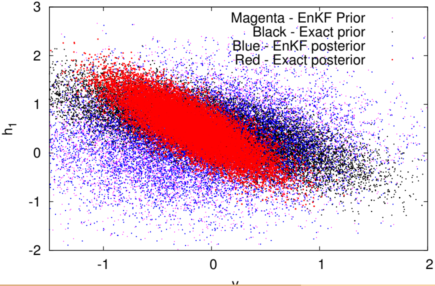


- "Annulus" shaped non-Gaussian distribution for drifter position leads to
- "wrong" posterior for velocity variables
- The correlation between different variables leads to a narrow exact

The prior and posterior at time of second observation







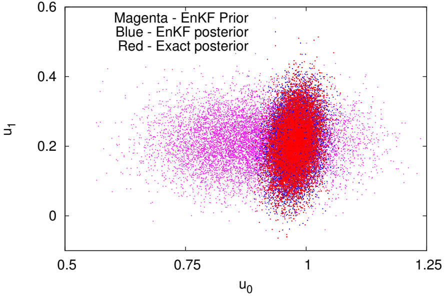
- The EnKF is unable to recover the "lost" information, and hence continues to perform poorly.
- As noted earlier, there is very little information about u_1, v_1, h_1

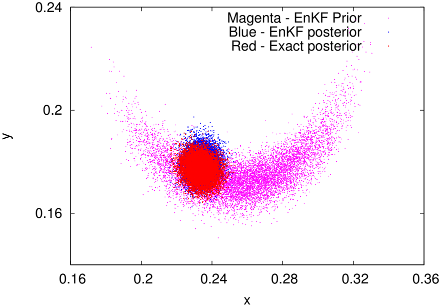
Decreasing the prior covariance or time interval between the observations fixes this problem

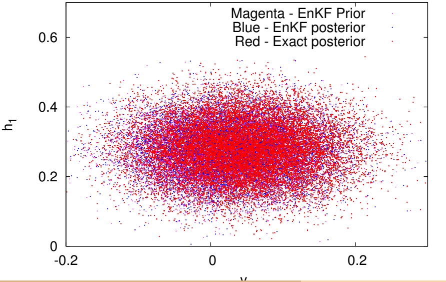
Two ways that the effects of nonlinearity can be made less pronounced:

- Reduce the prior covariance at $t = 0$ from $(1.0, 0.7, 0.7, 0.7, 0.005, 0.005)$ (previous figures) to $(0.1, 0.07, 0.07, 0.07, 0.005, 0.005)$ for the next few set of figures.
- Make more frequent observations: instead of $T_{pbs} = 0.1$ (previous figures) to $T_{obs} = 0.01$ (subsequent figures - the posterior will be indistinguishable from the case of narrow prior covariance)

The prior and posterior at time of first observation

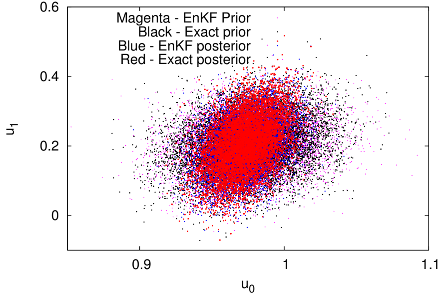


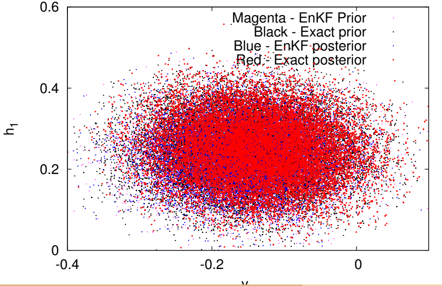


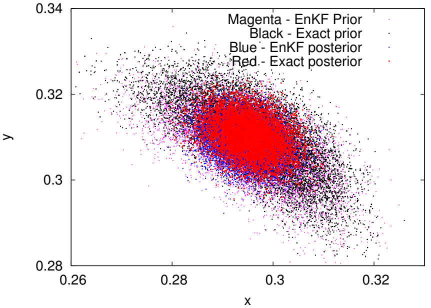


- The effect of shear is mitigated by the added information in the prior covariance.

The prior and posterior at time of second observation

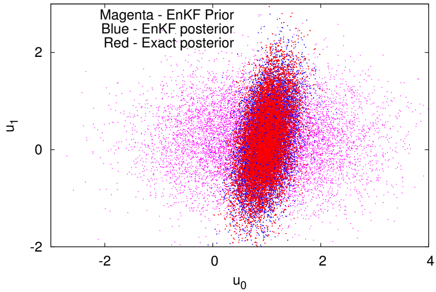


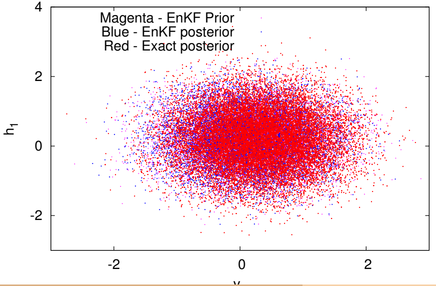


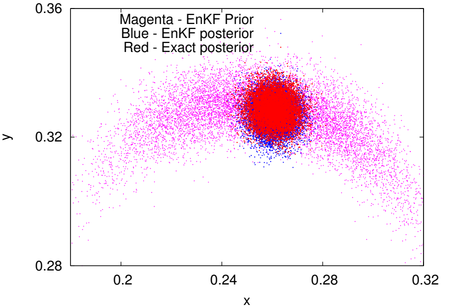


- The EnKF is able to capture the exact posterior well.

The prior and posterior at time of first observation

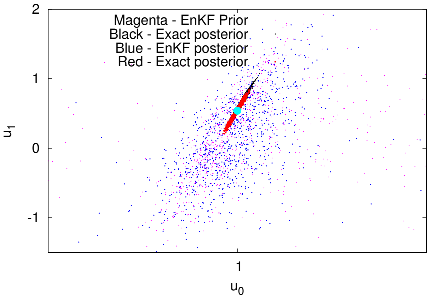


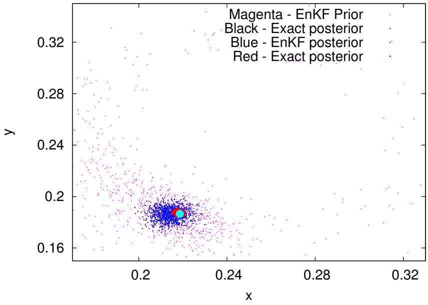


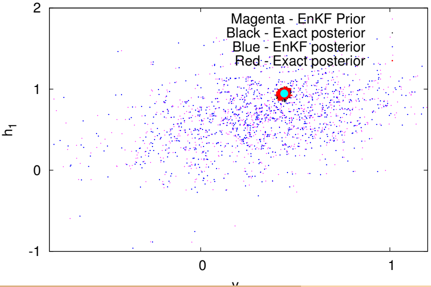


- The effect of shear is also mitigated by frequent observations.

The prior and posterior at the time of 50th obs ($t = 5.0$)

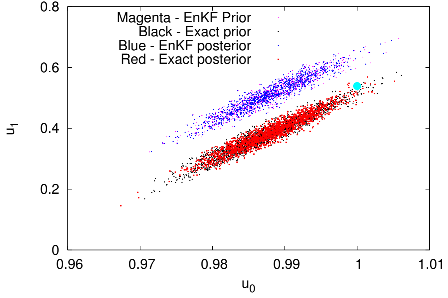


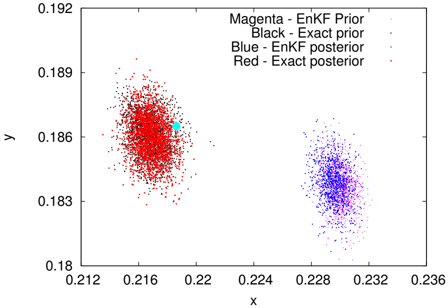


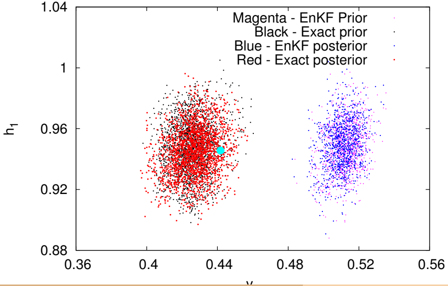


- The exact posterior in the case of data from the trajectory near the centre seems to converge to the true value.
- EnKF does not converge (as may be expected because of the non-Gaussianity of priors at each stage).

The prior and posterior at the time of 50th obs ($t = 5.0$)

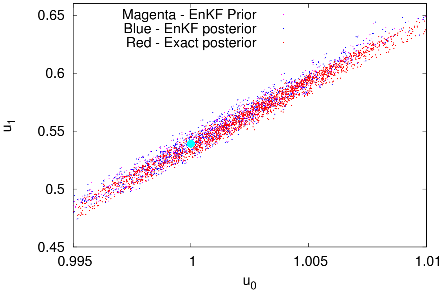


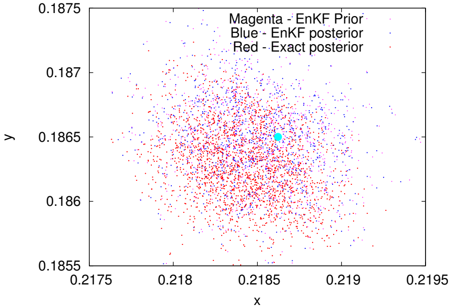


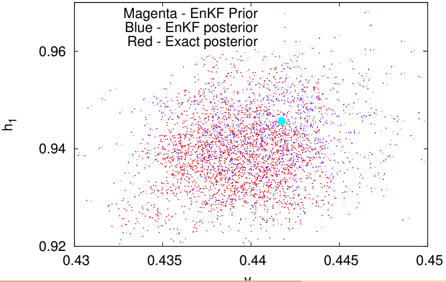


- For the narrow prior, the exact posterior converges, but
- EnKF seems to converge to a mean away from the truth.

The prior and posterior at the time of 500th obs ($t = 5.0$)

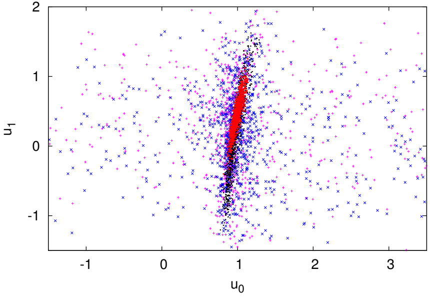


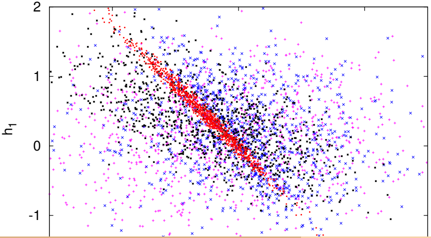


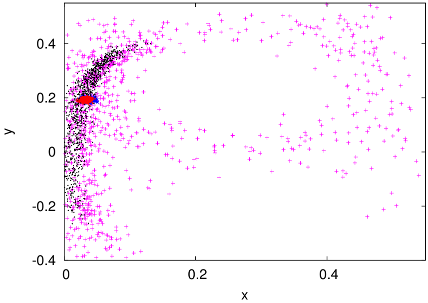


- As earlier, the exact posterior converges to the true value.
- EnKF converges as well, when the observations are frequent.

The prior and posterior at time of second observation: observations near the saddle







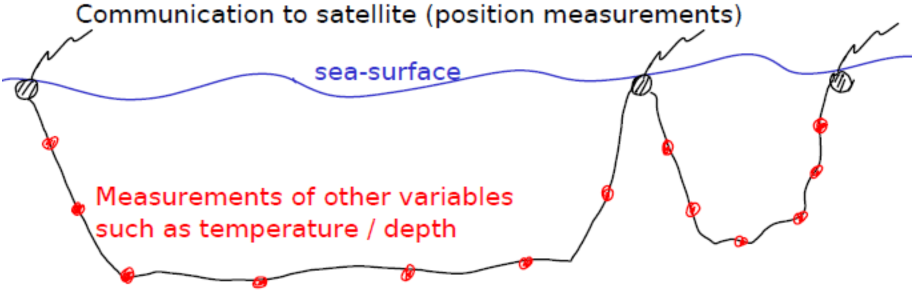
- The EnKF is unable to recover the "lost" information, and hence continues to perform poorly.
- As noted earlier, there is very little information about

Outline

1. Quick recap of the probabilistic formulation of data assimilation
2. What is Lagrangian data assimilation?
3. Data assimilation for skew-product systems
4. Effects of nonlinearity
5. En-route data assimilation

Lagrangian instruments also collect other data

A cartoon of the path of a glider / float (such as ARGO float):



- Subsurface drifter path is unknown, except possibly depth (no communication below surface)
- The time at which the observations are collected is available

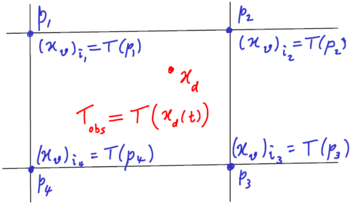
En-route data assimilation

Use of observations for which location information is unavailable

We need a new observation operator

Suppose the en-route data is temperature, and temperature is one of the variables coupled to velocity.

- Temperature observed at time t at the position of the drifter $x_d(t)$
- Consider state vector x_v consisting of, e.g., velocity and temperature at grid points. Let $x_{v,i_1}, x_{v,i_2}, x_{v,i_3}, x_{v,i_4}$ be the components corresponding to temperature at grid points surrounding the drifter, i.e.,
  $x_{v,i_j} = T(p_j), \; j = 1, 2, 3, 4$



Linear interpolation of $T(p_j)$ gives the temperature at drifter location, thus giving us the required observation operator:

$$T_{obs} = T(x_d) = \sum_{j=1}^{4} \alpha_j T(p_j) = \sum_{j=1}^{4} \alpha_j(x_d) x_{v,i_j} =: H(x_v, x_d)$$

where α_j depend on x_d as well as on p_1, p_2, p_3, p_4.

Example in the context of linear shallow water equations

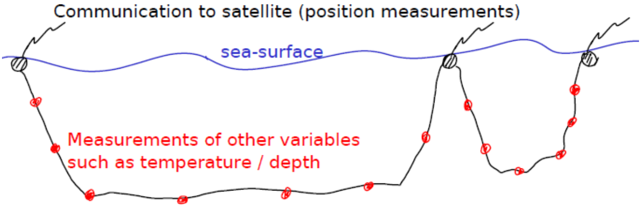
- Recall the state vector is $x = (x_v, x_d) = (u_0, u_1, v_1, h_1, s_1, s_2)$
- We will consider the "subsurface" observations to be the height field at the drifter location:

$$
\begin{aligned}
h_{obs} &= h(s_1(t), s_2(t), t) \\
&= \sin(2\pi k s_1(t)) \sin(2\pi l s_2(t)) u_0(t) + \sin(2\pi m s_2(t)) h_1(t) \\
&=: H(x)
\end{aligned}
$$

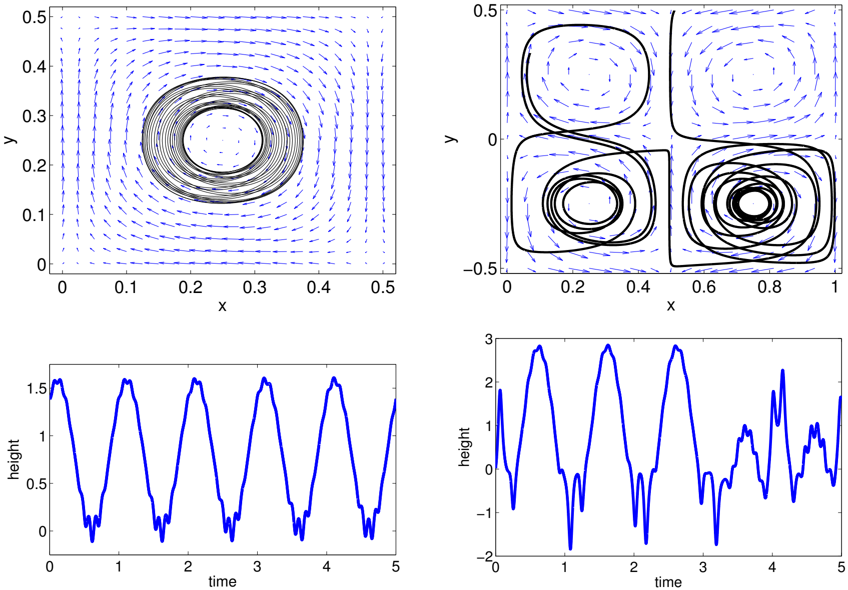
- In contrast with Lagrangian DA, the observation operator H is a nonlinear function of the state x.

Identical twin experiments with en-route data

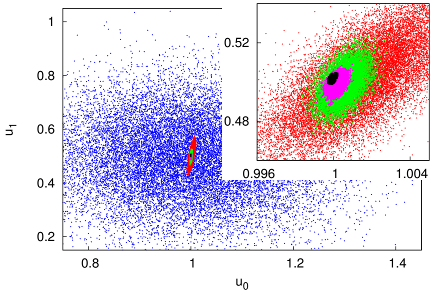
- The efficacy of using the subsurface observations is assessed by comparison with the case of only assimilating position information.
- Four cases: only position information used = Lagrangian DA, 9 height observations between each "surfacing", 99 height obs, **999 height obs**
- We use two assimilation methods:
  - MCMC to sample the posterior (smoother)
  - Particle filter (in this case, the subsurface observations are only assimilated at the time when the drifter surfaces)

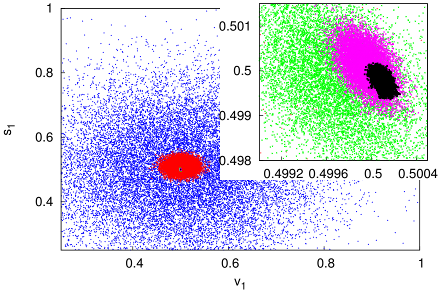


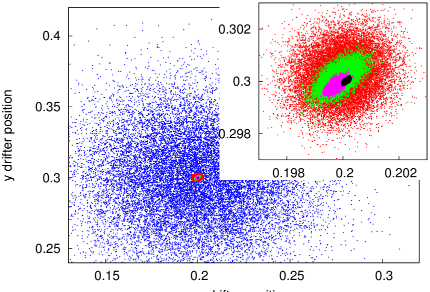
Some trajectories used for these comparisons



Subsurface observations are highly informative







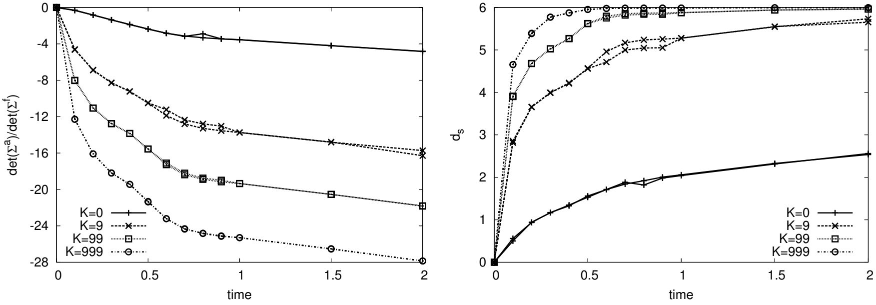
- Prior = blue dots
- Posteriors for K = 0, 9, 99, **999**
- "True" values: $u_0 = 1, u_1 = v_1 = h_1 = 0.5, x = 0.2, y = 0.3$

Quantifying the gain of information

Left: $r(t) = \det(\Sigma_F^a(t))/\det(\Sigma_F^f)$

Right: degree of freedom (DOF) for signal (Zupanski 2007):

$$d_s(t) = tr\left[\mathbf{I} - \Sigma_F^a(t)(\Sigma_F^f)^{-1}\right]$$



Some main advantages:

- Significant information about unobserved variables (velocity)
- Accurate estimates of subsurface drifter paths - useful for planning glider paths using "adaptive" controls

Summary

1. Two main aspects of nonlinearity
   - shear around a center and
   - "nonlinear" divergence around a saddle

strongly affect assimilation because

- they lead to non-Gaussian distribution functions, which manifest themselves in
- failures of "linear" assimilation methods (such as Kalman filter variants) and can aid the search for
- approximate methods such as Gaussian mixture models, etc.

The main factors that control the strength of these effects is

- time between observations for development of nonlinearity, and
- covariance of the prior.

2. Glider / subsurface observations can be highly informative for ocean state estimation. (Caveat: we have not looked at happens in the case of large model error, such as systematic bias.)