Hybrid EnKF and Particle Filter:

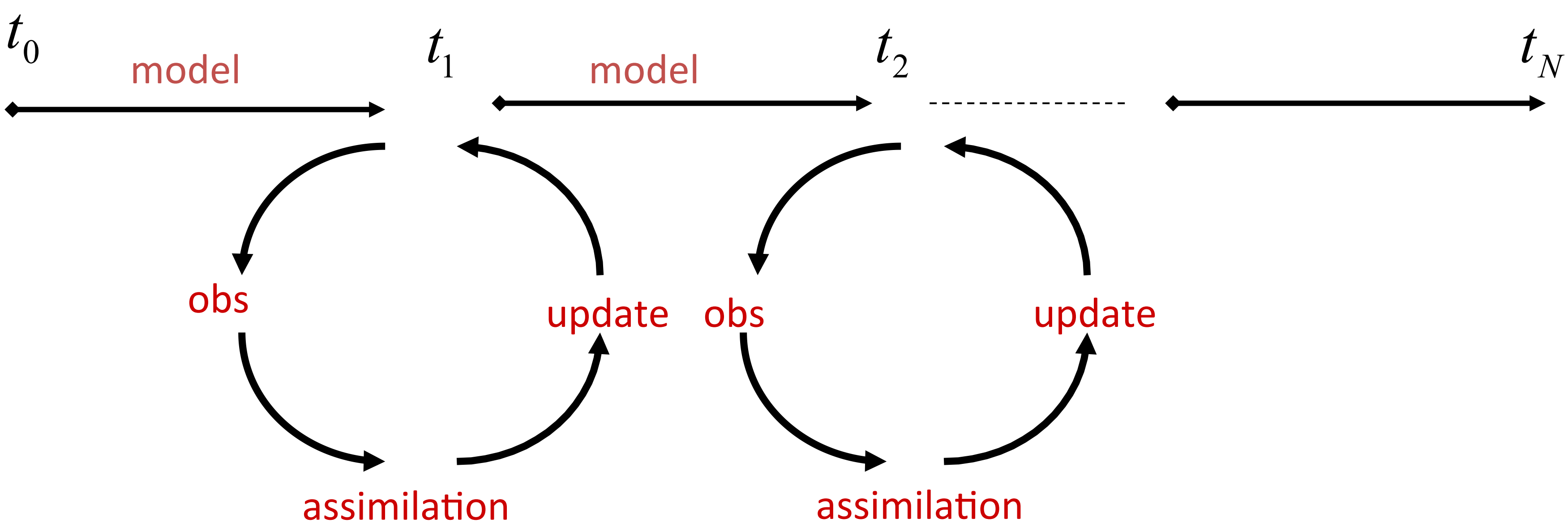
Lagrangian DA and Parameter Estimation

Chris Jones and
Naratip Santitissadeekorn*
RENCI and Mathematics,
UNC-Chapel HIll

*University of Surrey

Data Assimilation in Sequential Mode

Model + observations ⟶ prediction



Assimilation at: $t = t_k$

$$x_k^a = x_k^f + K_k\left(\eta_k - H(x_k^f)\right)$$

$$P^{\text{posterior}}(x_k \mid y_k) \propto P^{\text{obs}}(y_k \mid x_k)\, P^{\text{prior}}\left(x_k\right)$$

Skew-Product Structure of Dynamics

$$x_k = m_k\left(x_{k-1}\right) + \varepsilon_k$$

$$y_k = h\left(x_k\right) + \delta_k$$

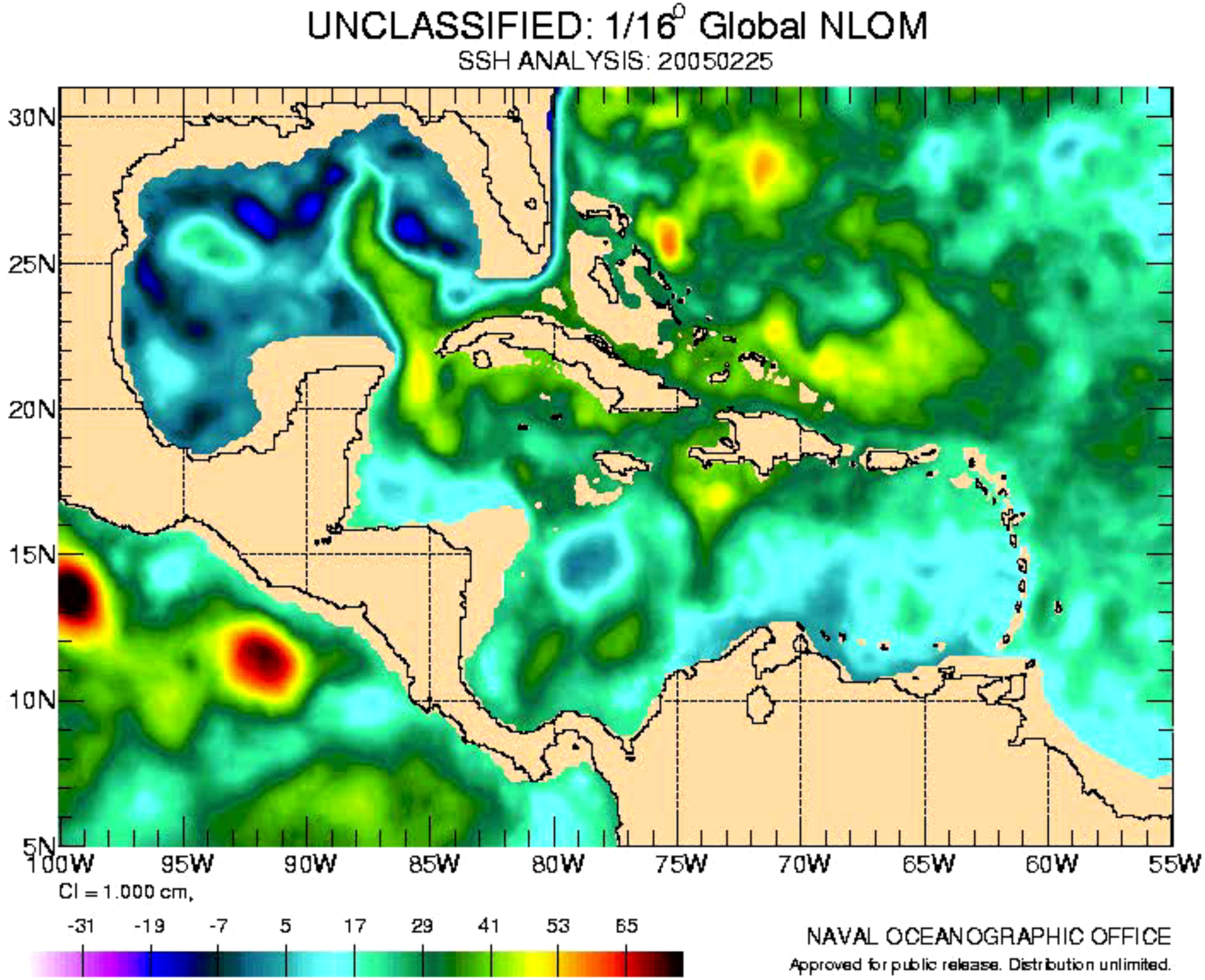
$$x_k = \left(x_k^1, x_k^2\right)$$

$$x_k^1 = m_k^1\left(x_{k-1}^1\right) + \varepsilon_k^1$$

$$x_k^2 = m_k^2\left(x_{k-1}^1, x_{k-1}^2\right) + \varepsilon_k^2$$

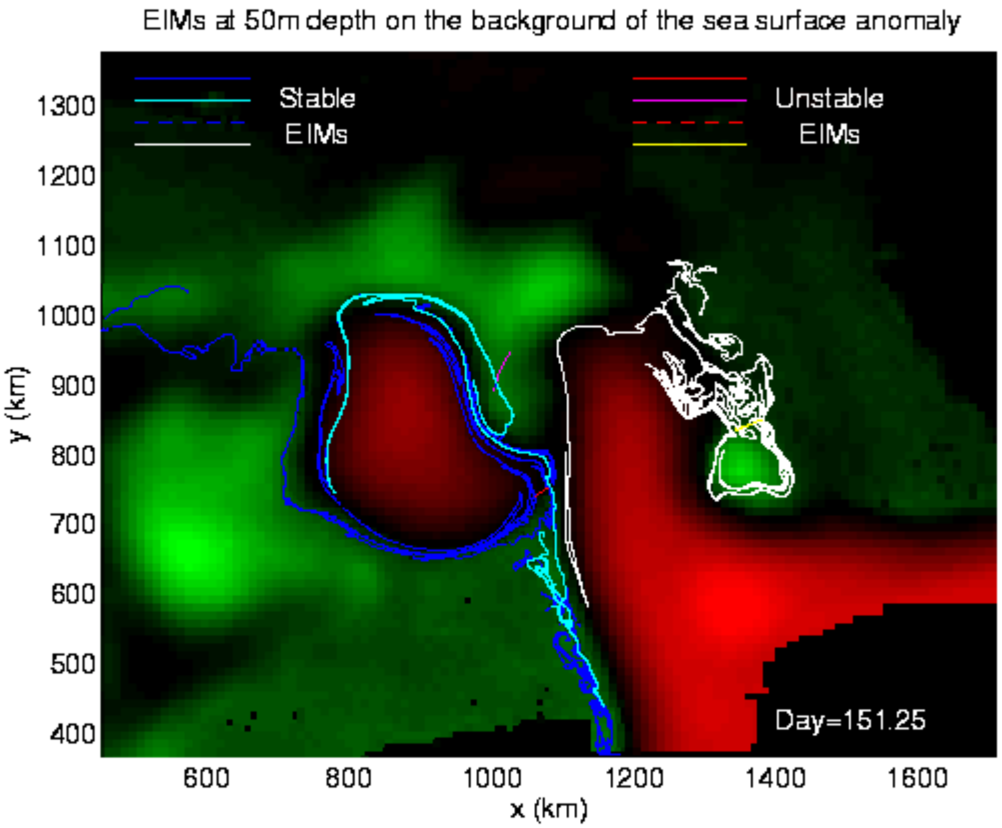
- One of $x_k^i, i = 1,2$ is low-dimensional
- Idea: EnKF on *high-dimensional* part
  PF on *low-dimensional* part

Gulf of Mexico/Carribean



Dynamics in GoM

Key structures: elliptic points (trajectories)
hyperbolic points (trajectories)



From: Kuznetsov et al. JMR 2002

MISSION

CARTHE brings together over 50 of the nation's top ocean modelers and air-sea interaction experts to share knowledge and explore the fate of the hydrocarbons released as a result of the Deepwater Horizon oil spill. Funded through the Gulf of Mexico Research Institute, this collaborative effort will produce the first-ever comprehensive modeling hierarchy that offers a combined space and time (3D+1) description of the oil and dispersants' fate and transport. From their current studies of oceanic and atmospheric turbulence

Above: CARTHE researchers releasing the first custom drifter from the R/V Walton Smith near the site of the Deepwater Horizon oil spill.

and mixing, tropical cyclones, and coastal and nearshore observations, the team members will:

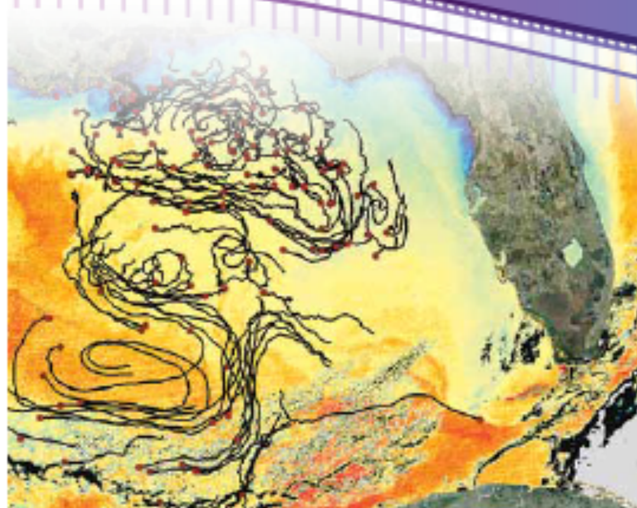
- **Develop a multi-scale modeling tool that incorporates state-of-the-art knowledge.**
- **Conduct *in-situ* observations and laboratory experiments specifically designed to quantifying and follow dispersion.**
- **Create a robust set of tools to assess model performance and quantify predictive uncertainty.**
- **Establish sampling strategies and investigations that may be used in other petroleum release scenarios.**

PARTNERS IN THIS ACTIVITY INCLUDE SCIENTISTS FROM:

City University of New York — Staten Island
Florida International University
Florida State University
Naval Postgraduate School
Naval Research Laboratory
Nova Southeastern University
Texas A&M University-Corpus Christi
Tulane University
University of Arizona
University of Delaware
University of Miami
University of Texas at Austin

www.carthe.org

Contact:
Dr. Tamay Özgökmen
CARTHE Director
University of Miami
Rosenstiel School of Marine & Atmospheric Science
4600 Rickenbacker Causeway, Miami, FL 33149
tozgokmen@rsmas.miami.edu
+1-305-613-2851



CARTHE

CONSORTIUM for ADVANCED RESEARCH on TRANSPORT of HYDROCARBON in the ENVIRONMENT

A Project of the Gulf of Mexico Research Initiative

GULF of MEXICO RESEARCH INITIATIVE

www.carthe.org

Augmented system

Append equations for drifters (floats)

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_F \\ \mathbf{x}_D \end{pmatrix} \quad \text{-- augmented state vector}$$

$$x^F \leftrightarrow (u, v, w)$$

$$x^D \leftrightarrow (x, y, z)$$

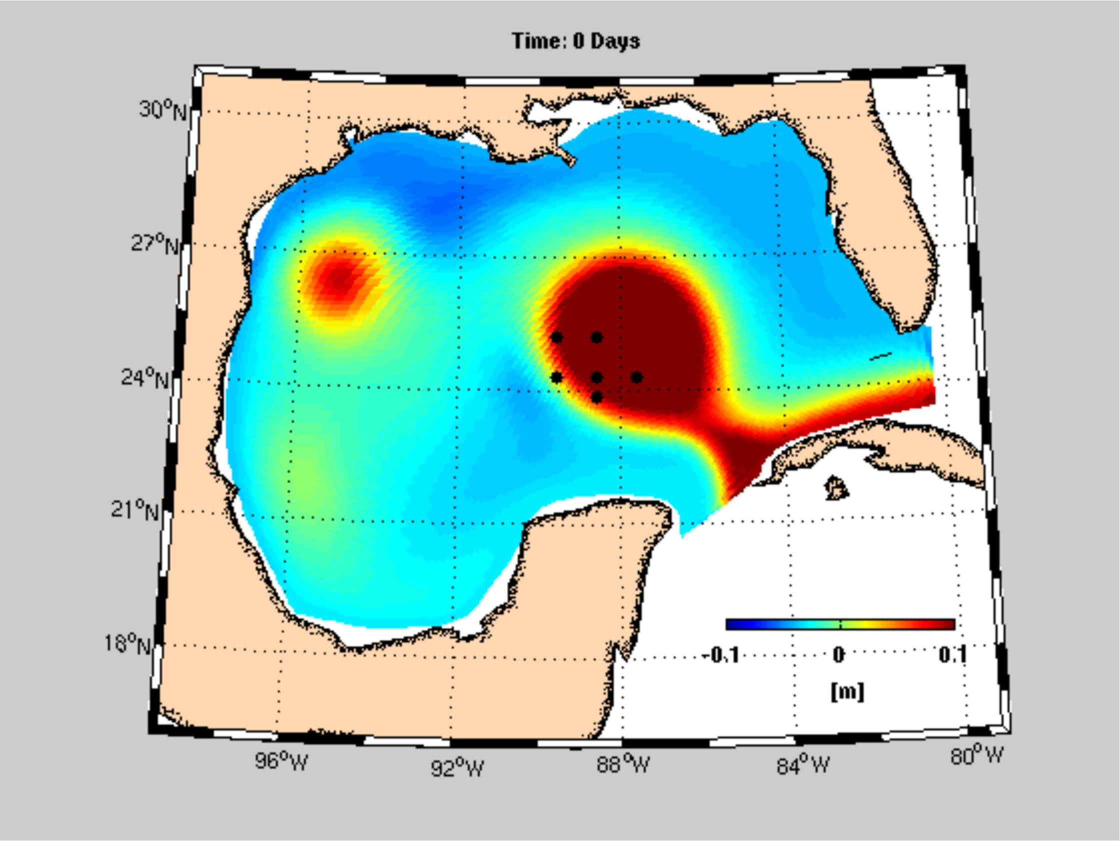
$$\frac{d\mathbf{x}_F^{\mathrm{f}}}{dt} = M_F(\mathbf{x}_F^{\mathrm{f}}, t) \quad \text{-- flow equations}$$

$$\frac{d\mathbf{x}_D^{\mathrm{f}}}{dt} = M_D(\mathbf{x}_D^{\mathrm{f}}, \mathbf{x}_F^{\mathrm{f}}, t) \quad \text{-- tracer advection equation}$$

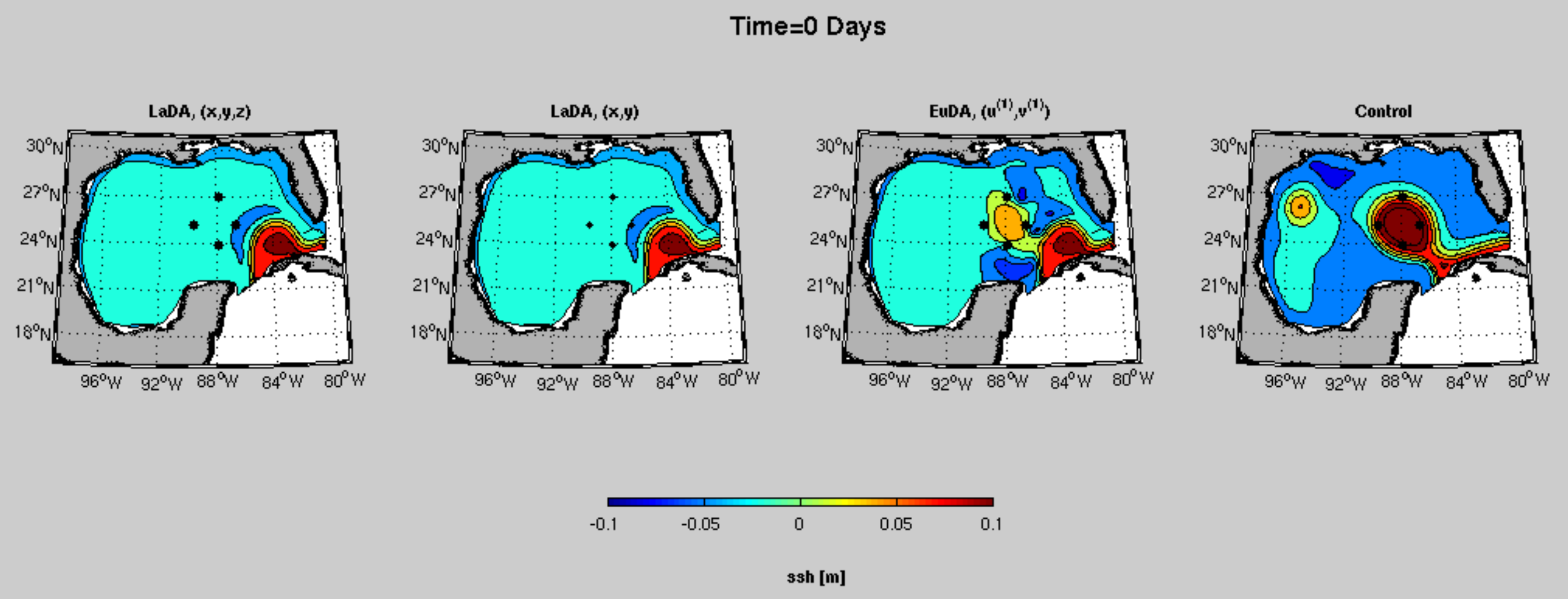
Apply DA to augmented system

Ide, Jones and Kuznetsov 2002

Recapturing an eddy



Eddies in GoM



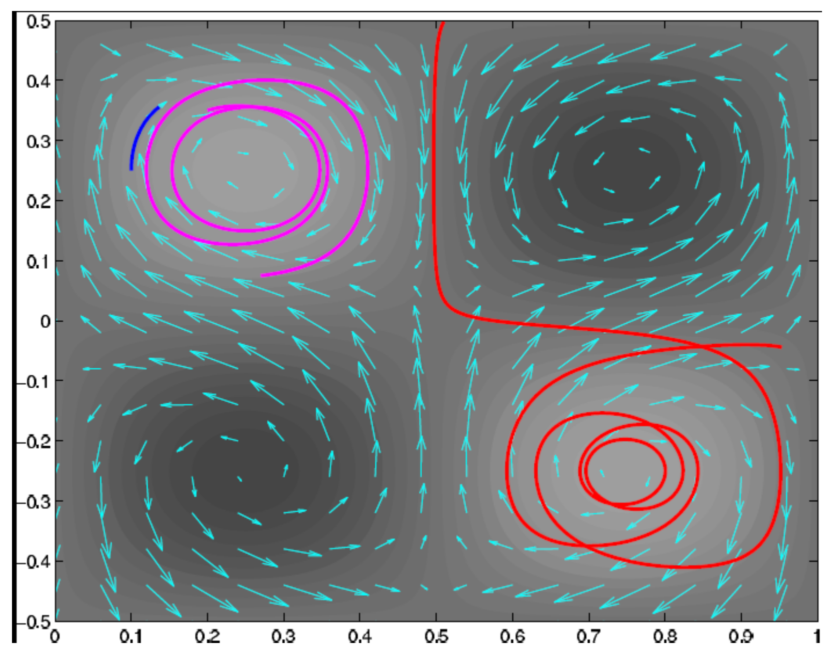
Work with Guillaume Vernières (NASA) and Kayo Ide (MD) -Physica D, 2011

Perturbed Cellular Flow Field

$$\frac{\partial u}{\partial t} = v - \frac{\partial h}{\partial x},$$
$$\frac{\partial v}{\partial t} = -u - \frac{\partial h}{\partial y},$$
$$\frac{\partial h}{\partial t} = -\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y},$$

$$u(x,y,t) = -2\pi l \sin(2\pi kx)\cos(2\pi ly)u_0 + \cos(2\pi my)u_1(t),$$
$$v(x,y,t) = 2\pi k \cos(2\pi kx)\sin(2\pi ly)u_0 + \cos(2\pi my)v_1(t),$$
$$h(x,y,t) = \sin(2\pi kx)\sin(2\pi ly)u_0 + \sin(2\pi my)h_1(t),$$

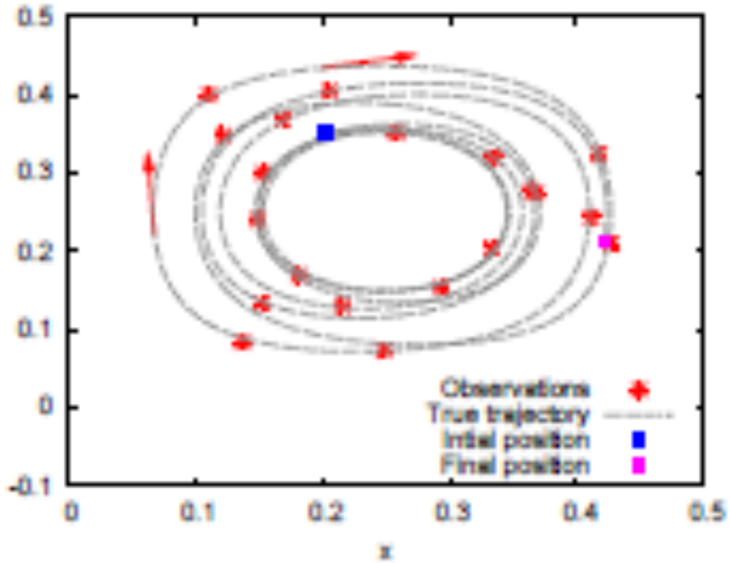
$$\dot{u}_0 = 0,$$
$$\dot{u}_1 = v_1,$$
$$\dot{v}_1 = -u_1 - 2\pi m h_1,$$
$$\dot{h}_1 = 2\pi m v_1,$$

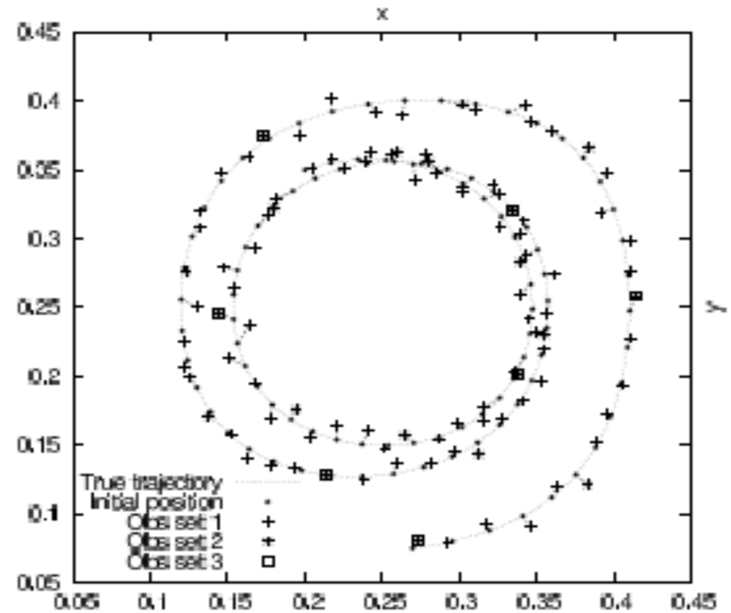


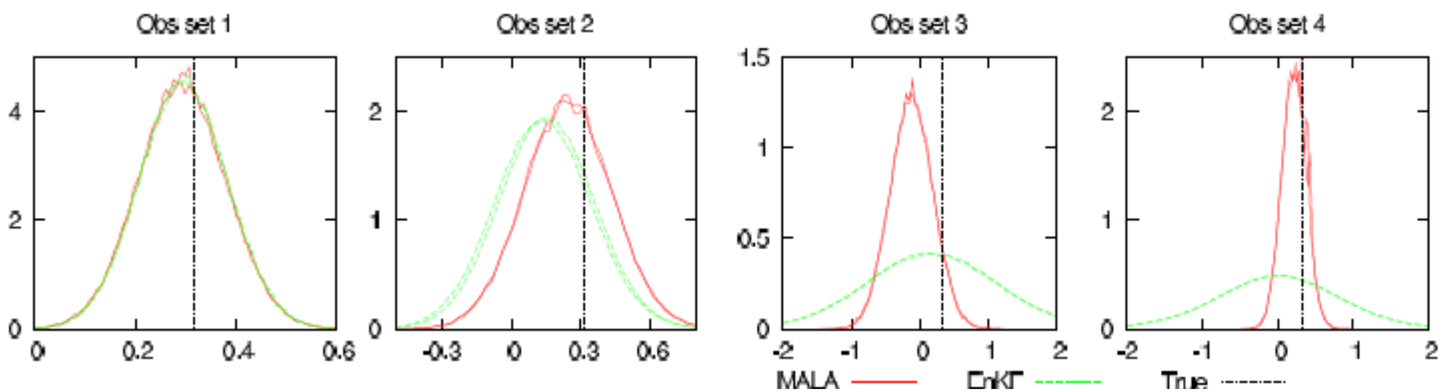
Apte, J and Stuart Tellus A 2008
Apte and J. 2014

Assimilating from trajectory staying in one cell

Expt: estimate i.c. from observations of trajectory

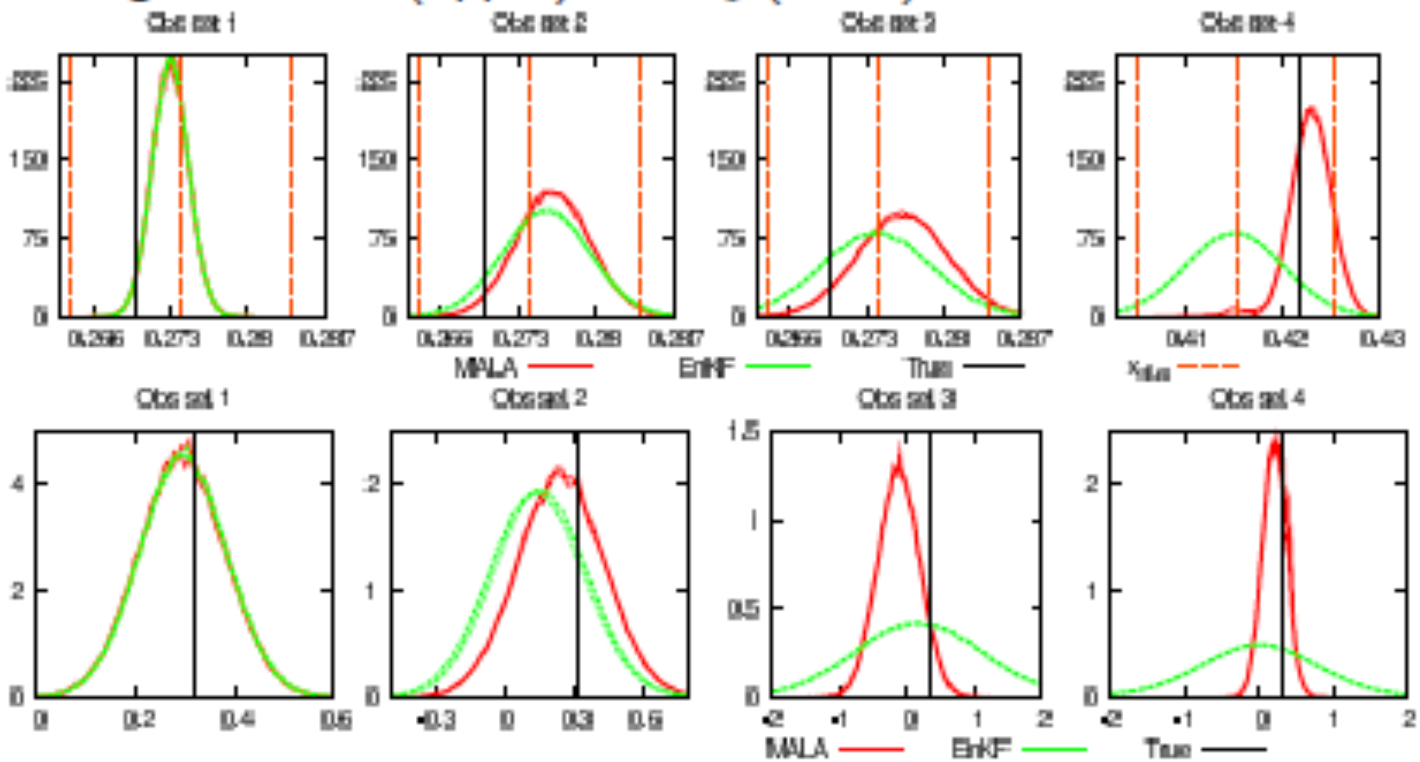






Varying observational frequency – around the center

Histograms for x (upper) and u_1 (lower)



- Frequent observations: EnKF works well – the PDF does not develop significant non-Gaussianity between observations
- For "large" time interval between observations (set 3 and 4), increasing the number of observations does not make EnKF "recover"

Lagrangian Data Assimilation (LaDA)

$$\frac{dx^F}{dt} = f^F(x^F, t)$$

$$\frac{dx^D}{dt} = f^D(x^F, x^D, t)$$

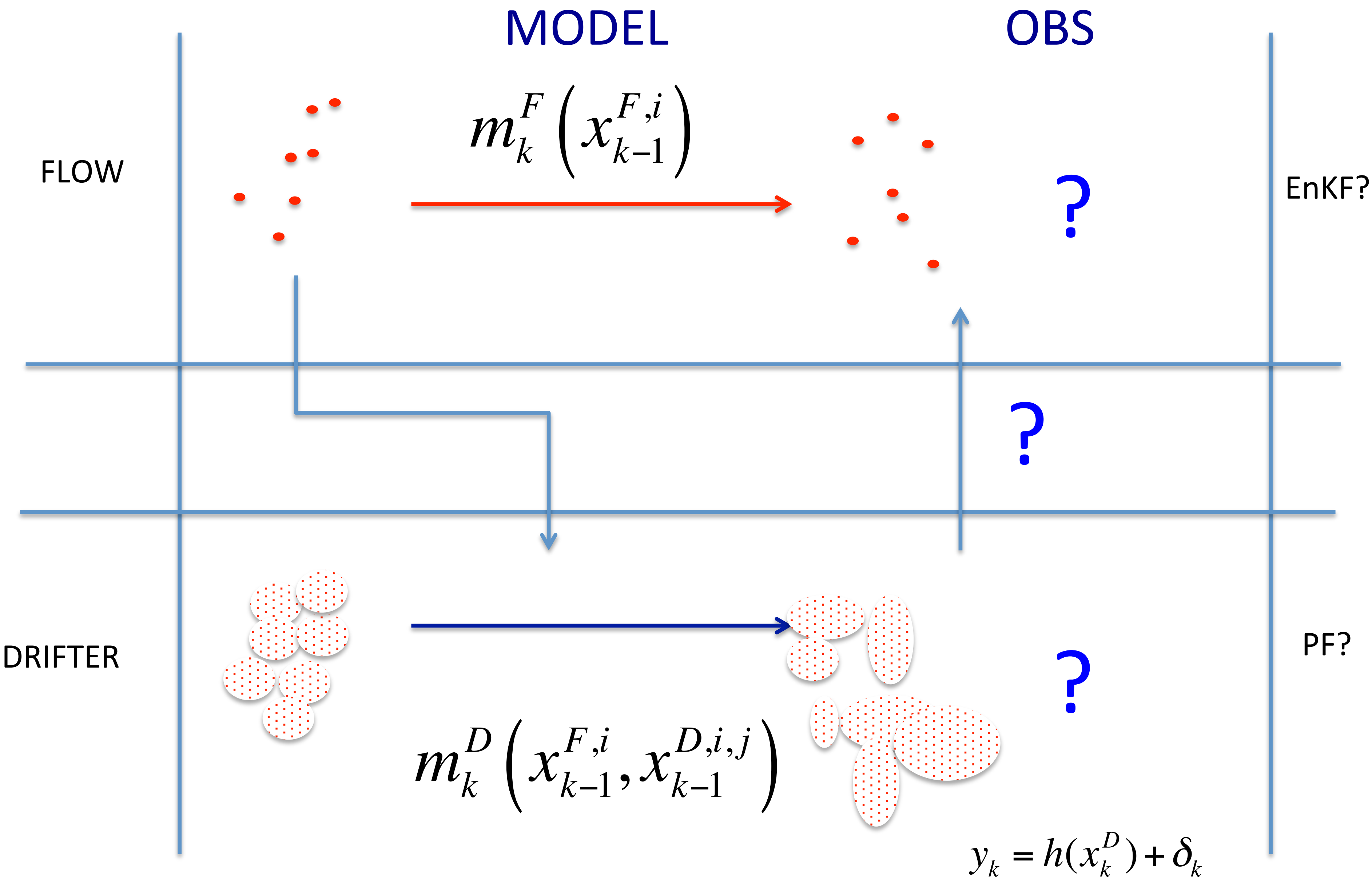
$$x_k^F = m_k(x_{k-1}^F)$$

Estimate: high

$$x_k^D = m_{k-1}^D\left(x_{k-1}^F, x_{k-1}^D\right)$$

Observe: low

Slivinskii, Spiller, Apte and Sandstede *Monthly Weather Review* 2014


MODEL
OBS
FLOW
DRIFTER
EnKF?
PF?
$m_k^F\left(x_{k-1}^{F,i}\right)$
$m_k^D\left(x_{k-1}^{F,i}, x_{k-1}^{D,i,j}\right)$
$y_k = h(x_k^D) + \delta_k$
?
?
?

Update Step

No resampling:

(according to some criterion on “paucity” of particle ensemble)

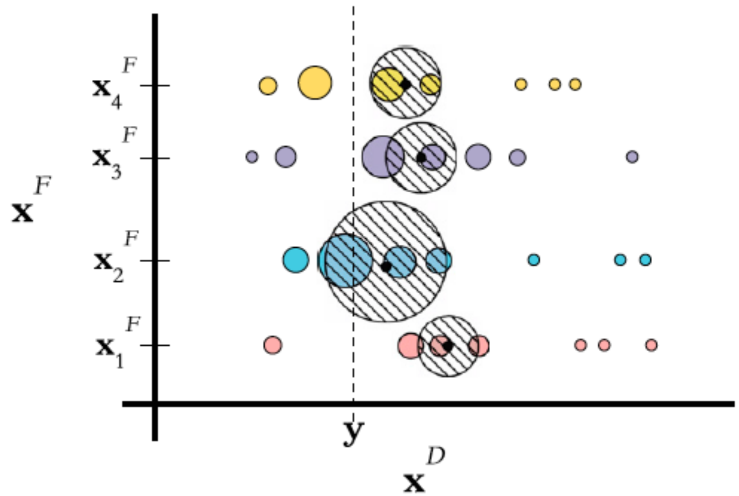
Drifter only

$$\sum_j w_{ij}^k \delta\left(x^{D,k} - x_{ij}^{D,k}\right)$$

Computed from obs

Joint PDF

$$\sum_{i,j} w_{ij}^k \delta\left(x^{D,k} - x_{ij}^{D,k}\right)\delta\left(x^{F,k} - x_i^{F,k}\right)$$



Update Step

With resampling: EnKF on flow variables

Step 1: Move flow states w/ EnKF:

Average over set of drifter ensembles

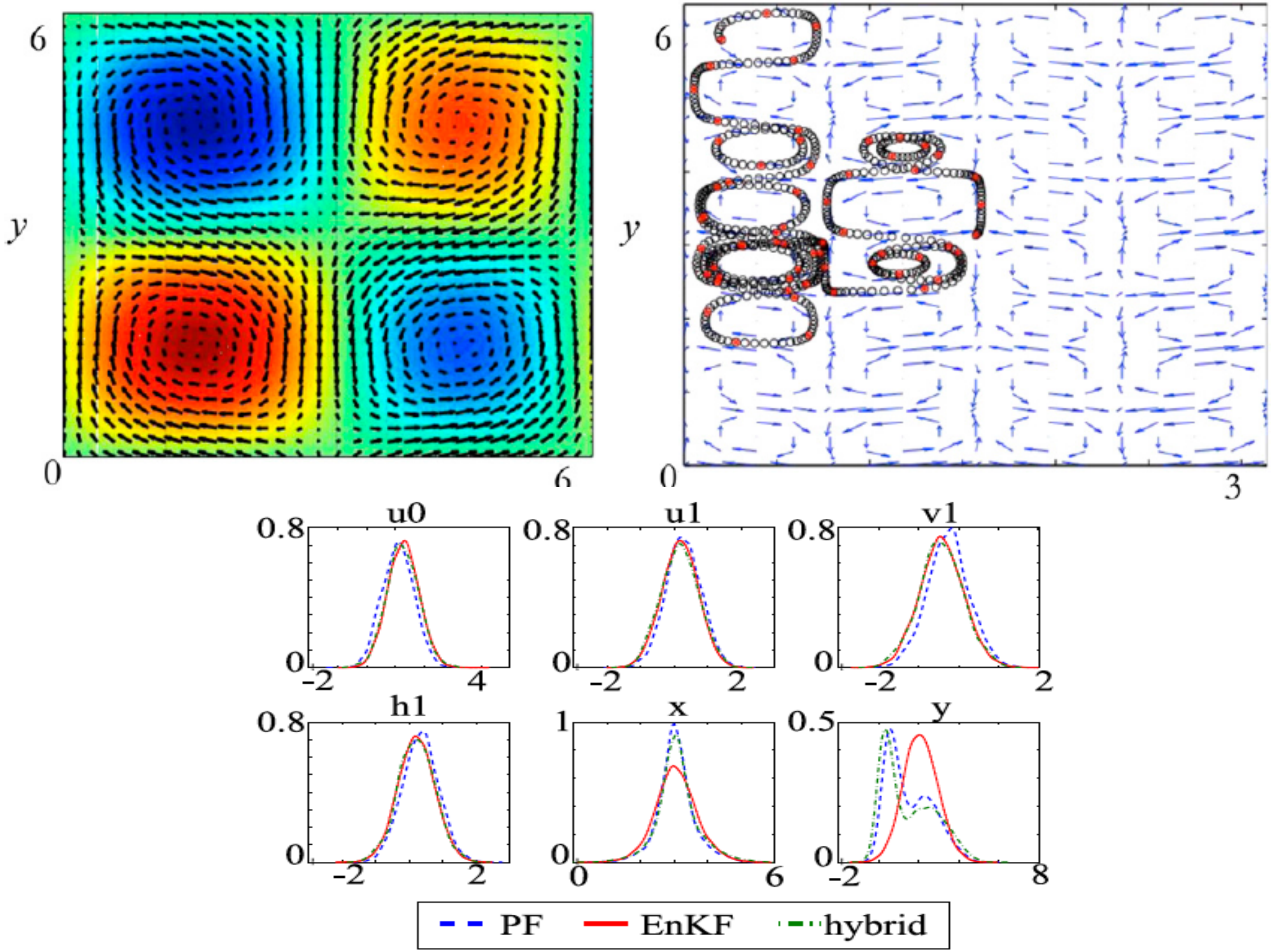
$$x_i^{F,k} = x_i^{F,f} + P_{FD}^f \left(P_{DD}^f + R\right)^{-1} \left(Y - \bar{x}_i^{F,D}\right)$$

$$P = \begin{bmatrix} P_{FF} & P_{FD} \\ P_{DF} & P_{DD} \end{bmatrix}$$

Forecast error covariance

Step 2: Form joint posterior PDF $\left\{x_i^{F,k}, \tilde{w}_i^k\right\}\left\{x_{ij}^D, w_{ij}^k\right\} \quad \tilde{w}_i^k = \sum_j w_{ij}^k$

Step 3: Resample, reset weights and proceed



6
y
0
6
6
y
0
3
u0
0.8
0
-2
4
u1
0.8
0
-2
2
v1
0.8
0
-2
2
h1
0.8
0
-2
2
x
1
0
0
6
y
0.5
0
-2
8
PF
EnKF
hybrid

Joint State-Parameter Estimation

Lorenz 96

$$\dot{x}_i = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F_i, \qquad i = 1,\ldots,40$$

+ cyclic BC

$$F_i = 8 + \theta_1 \sin\left(\frac{2\pi}{\theta_2} i\right)$$

Identical twin expt:

$$\theta^* = [\theta_1, \theta_2] = [2, 40] \qquad \text{"Truth"}$$

Goal: Estimate both state and parameters from obs

Filtering Options

EnKF on augmented system:

Update based on linear regression. Fails if correlation is not linear (Yang and DelSole, 2009)

PF on augmented system:

Computationally expensive….

Rao-Blackwellized Particle Filter:

Computationally more expensive!
But basis for an approach

Parameter Estimation

$$\frac{dx}{dt} = F(x, \theta)$$

$$x_{k+1} = f(x_k, \theta_k)$$

Observe: high

$$\frac{d\theta}{dt} = 0$$

$$\theta_{k+1} = \theta_k$$

Estimate: low

MIXED FILTER

RBPF

$$w = [x, \theta] \qquad\qquad y = h(x)$$

$$p(w_{1:k} | y_{1:k}) \propto p(x_{1:k} | \theta_{1:k}, y_{1:k}) \, p(\theta_{1:k} | y_{1:k})$$

$$p(w_{1:k}|y_{1:k}) \approx \sum_{i=1}^{N} \omega_k^{(i)} p(x_{1:k}|\theta_{1:k}^{(i)}, y_{1:k}) \delta(\theta_{1:k} - \theta_{1:k}^{(i)})$$

$$\omega_k^{(i)} \propto p(y_k | y_{1:k-1}, \theta_k^{(i)}) \omega_{k-1}^{(i)}$$

If model is linear, then Gaussian

Parameter Models

Persistence model

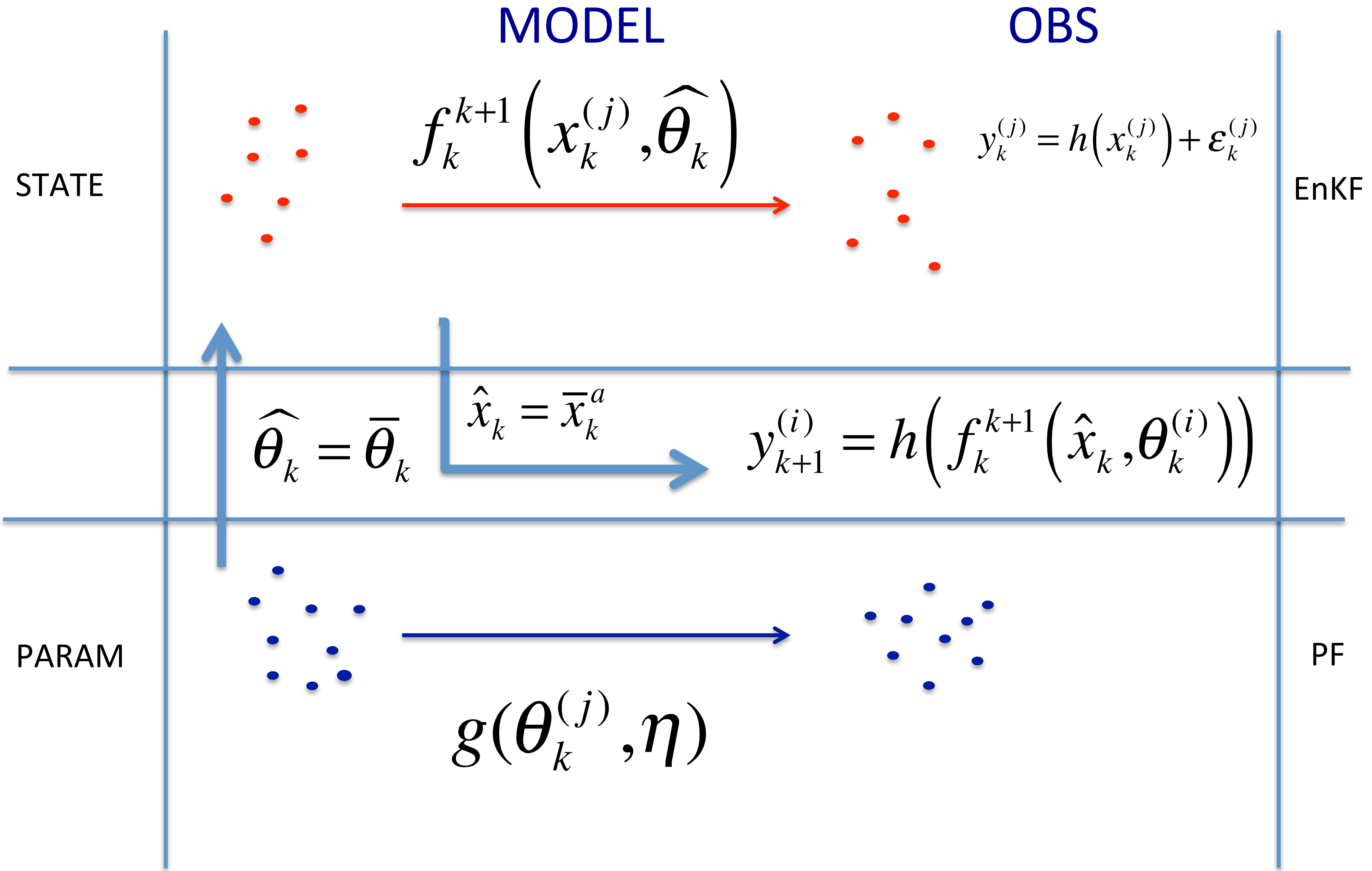
$$\theta_{k+1} = \theta_k$$

Random walk model

$$\theta_{k+1} = \theta_k + \eta_k, \quad \eta_k \sim N(0, W_k)$$

Liu-West model

$$\theta_{k+1} = \alpha\theta_k + (1-\alpha)\bar{\theta}_k + \eta_k$$


MODEL
OBS
STATE
EnKF
PARAM
PF
$f_k^{k+1}\left(x_k^{(j)},\widehat{\theta}_k\right)$
$y_k^{(j)} = h\left(x_k^{(j)}\right) + \varepsilon_k^{(j)}$
$\hat{x}_k = \overline{x}_k^a$
$\widehat{\theta}_k = \overline{\theta}_k$
$y_{k+1}^{(i)} = h\left(f_k^{k+1}\left(\hat{x}_k,\theta_k^{(i)}\right)\right)$
$g(\theta_k^{(j)},\eta)$

PROBLEM

Lorenz 96

$$\dot{x}_i = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F_i, \qquad i = 1,\ldots,40$$

+ cyclic BC

$$F_i = 8 + \theta_1 \sin(\frac{2\pi}{\theta_2} i)$$

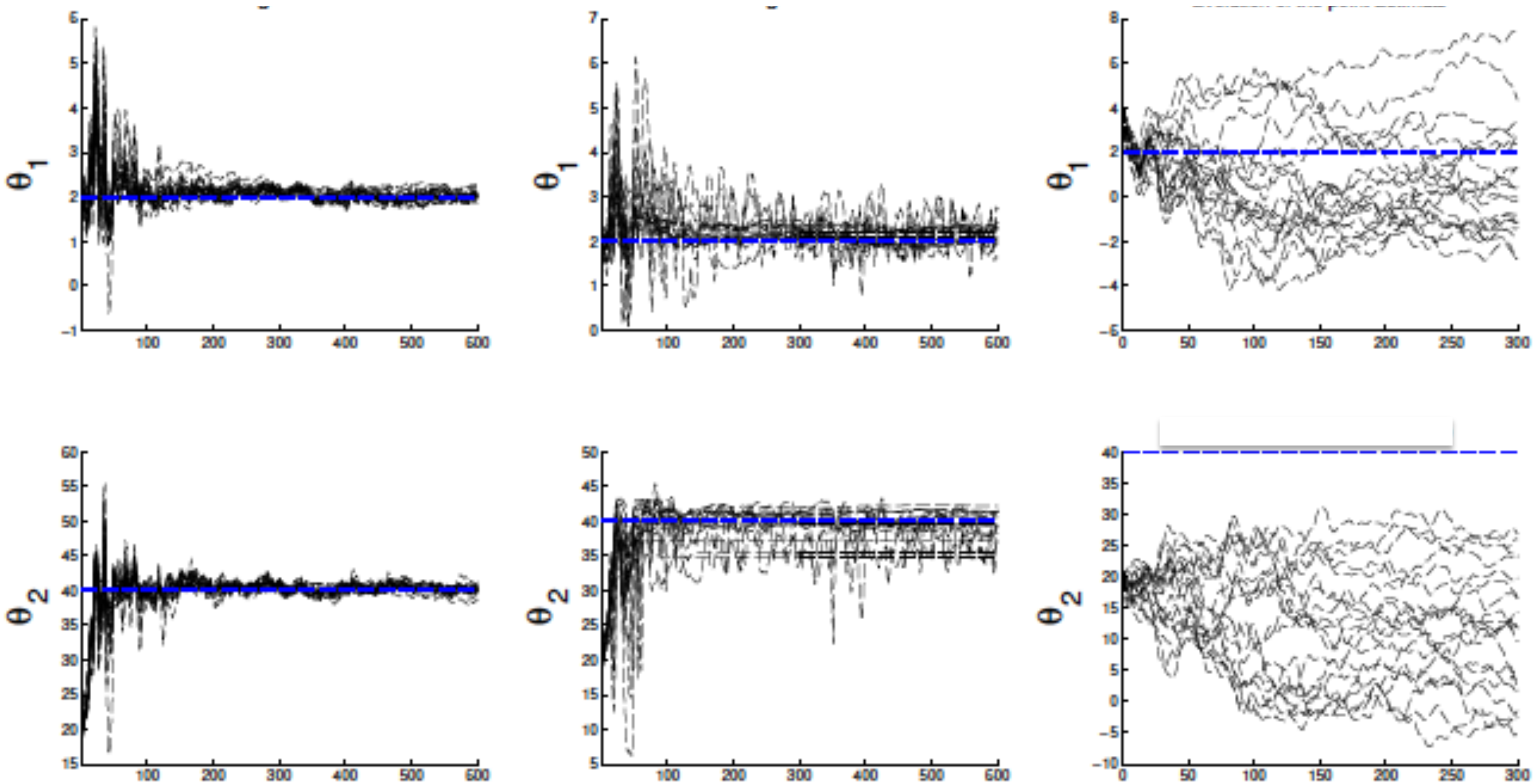
Identical twin expt:

$$\theta^* = [\theta_1, \theta_2] = [2, 40]$$

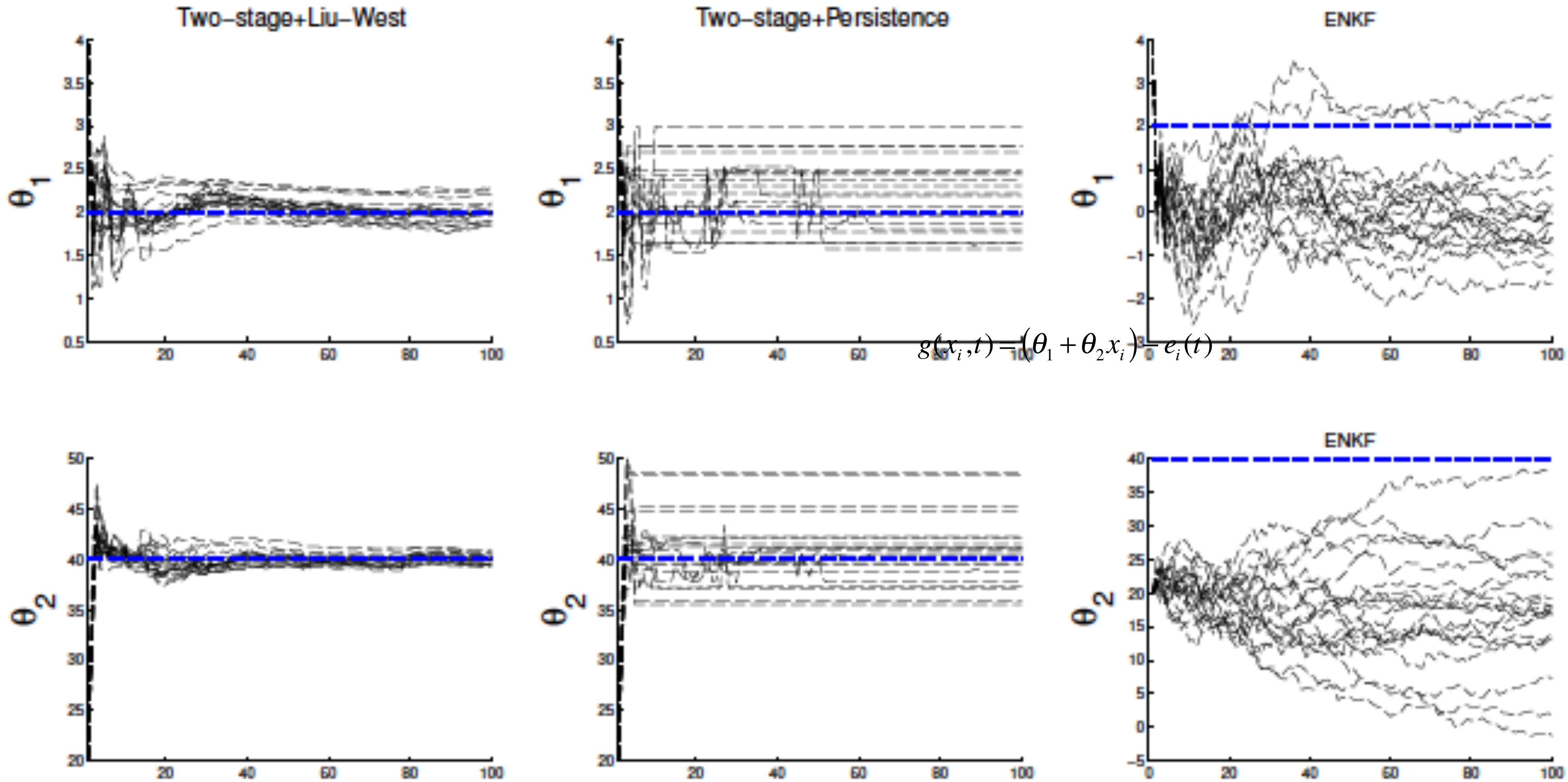
"Truth"

Goal: Estimate both state and parameters from obs

Two-Stage +Liu-West | Two-Stage +Persistence | EnKF



$\delta t = \Delta t$ 200/50 vs. 250



$$g(x_i,t) = (\theta_1 + \theta_2 x_i) + e_i(t)$$

$$\delta t = 10\,\Delta t$$

2-layer QG:

$$\psi_1 = \epsilon_1 A_1 \sin(k_1 x - c_1 t)\sin(l_1 y) + \epsilon_2 A_2 \sin(k_2 x - c_2 t)\sin(l_2 y), \qquad \text{L1}$$

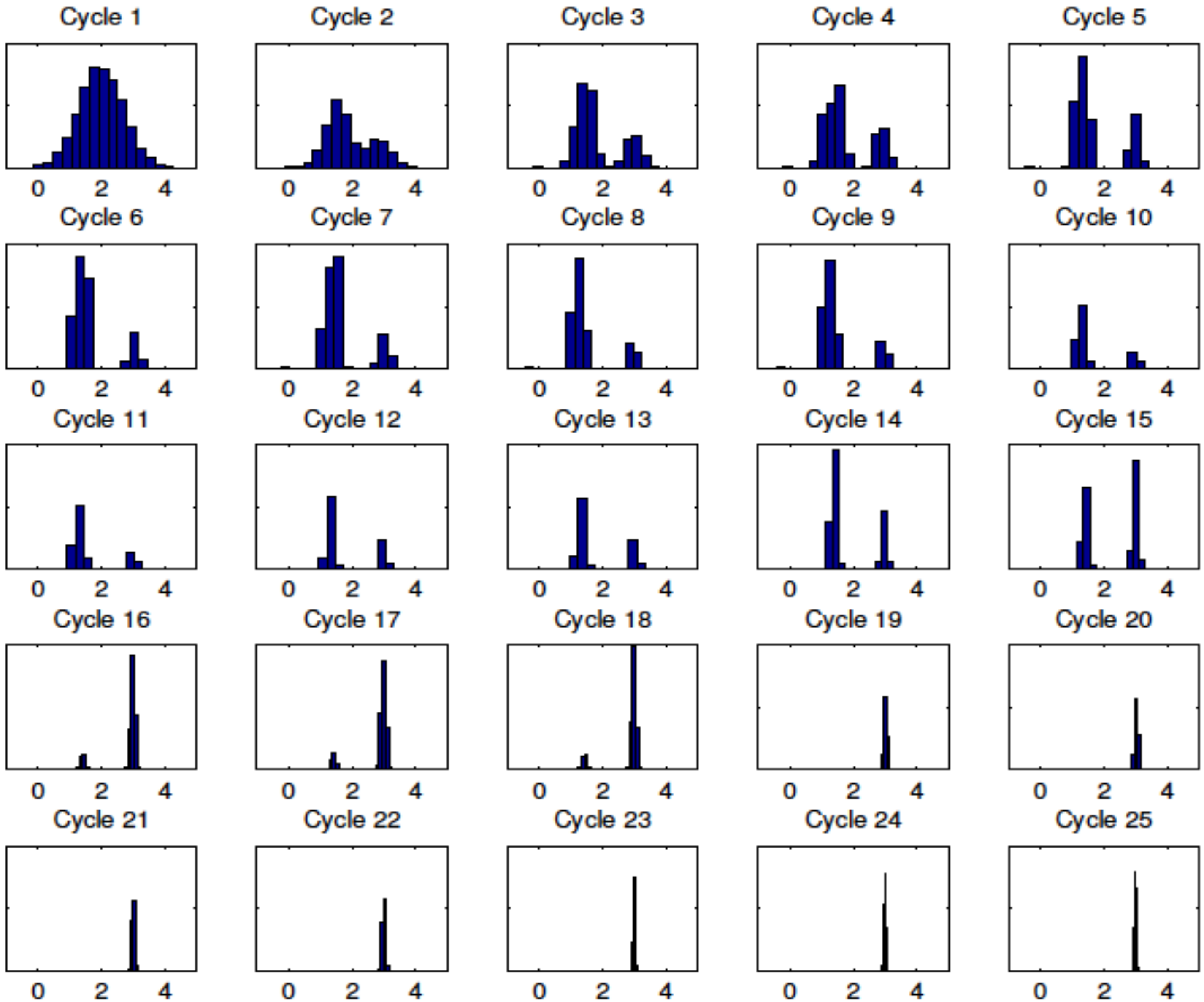
$$\psi_2 = \epsilon_1 A_1 \sin(k_1 x - c_1 t)\sin(l_1 y) - \epsilon_2 A_2 \sin(k_2 x - c_2 t)\sin(l_2 y). \qquad \text{L2}$$

$$\dot{x}_i = \frac{\partial \psi_i}{\partial y_i}$$

$$\dot{y}_i = -\frac{\partial \psi_i}{\partial x_i} \qquad i = 1,2$$

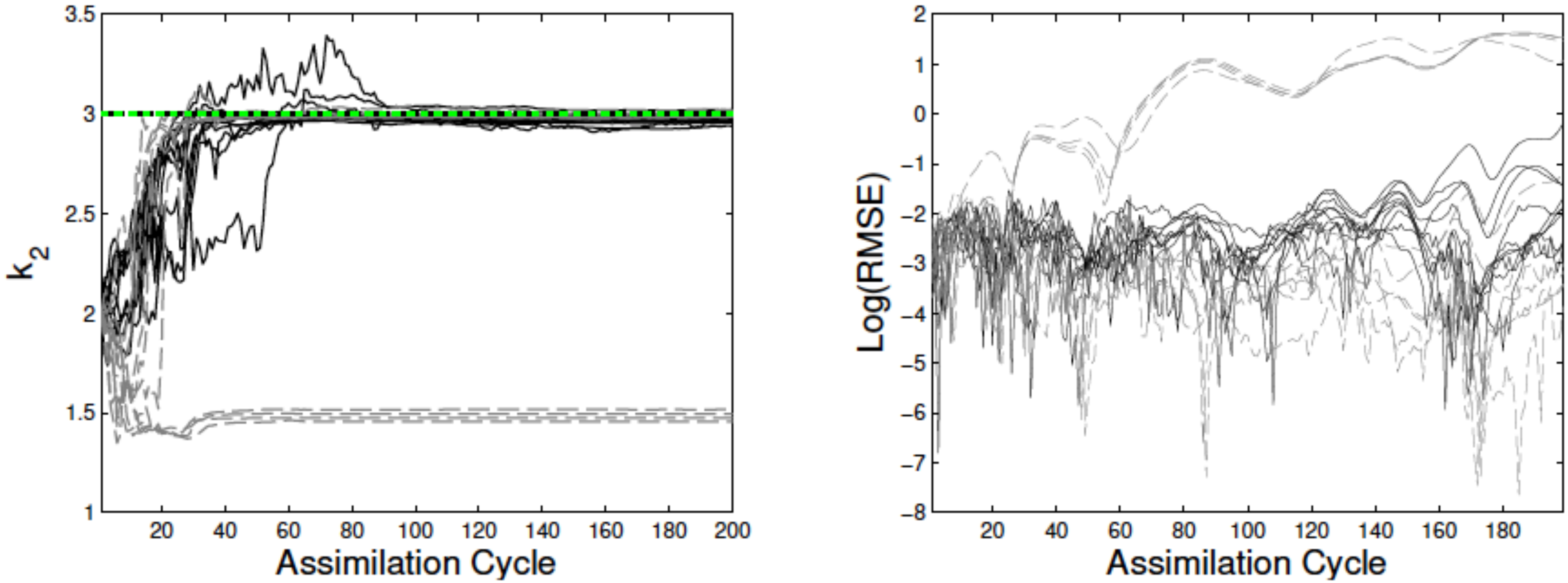
Problem: observe trajectory in layer 1 and estimate k_2

Bimodality trap



Horizontal axis: $p = k_2$

Results from Identical Twin Expt



Mean estimate (Left) and RMSE for the tracer position (Right) for 10 different experiments and $N = 200$. The results for the PF with Liu-West model are plotted in the gray dash lines and for the two-stage filtering $(M = 0.75N)$ in black solid lines.

Conclusions

- Skew-product structure of problem can be exploited to create new filtering approaches
- Two examples: LaDA and JSP estimation
- Issues are different in each case (reverse of dimensional issues)
- Basic idea: Use EnKF on high-dimensional part
- Issues with nonlinearity focused into low-dimensional part
- Key decision in implementation is in *crosstalk*