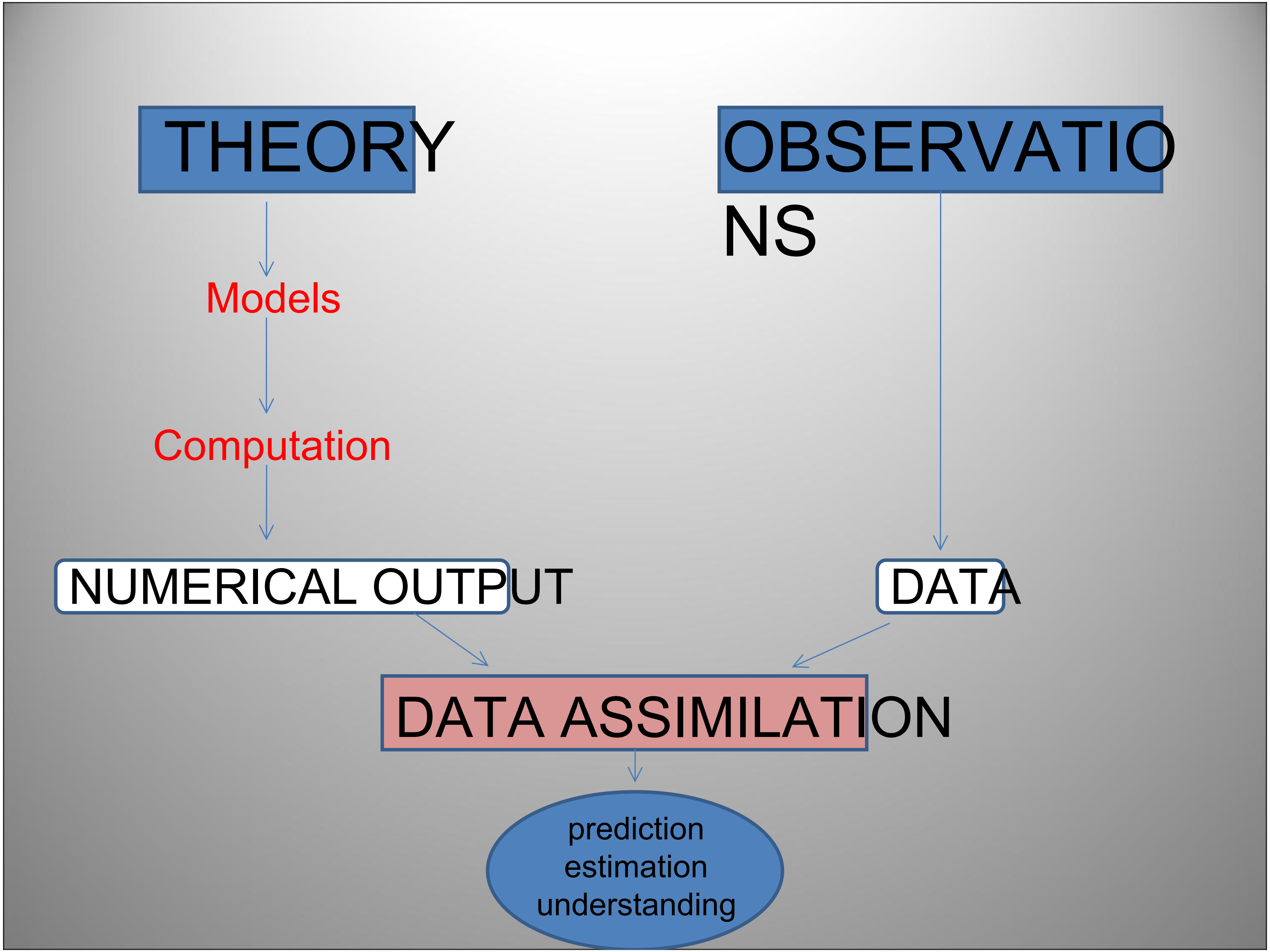
Assimilating Lagrangian Type Data

Chris Jones, UNC-Chapel Hill and University of Warwick

c.k.r.t.jones@warwick.ac.uk

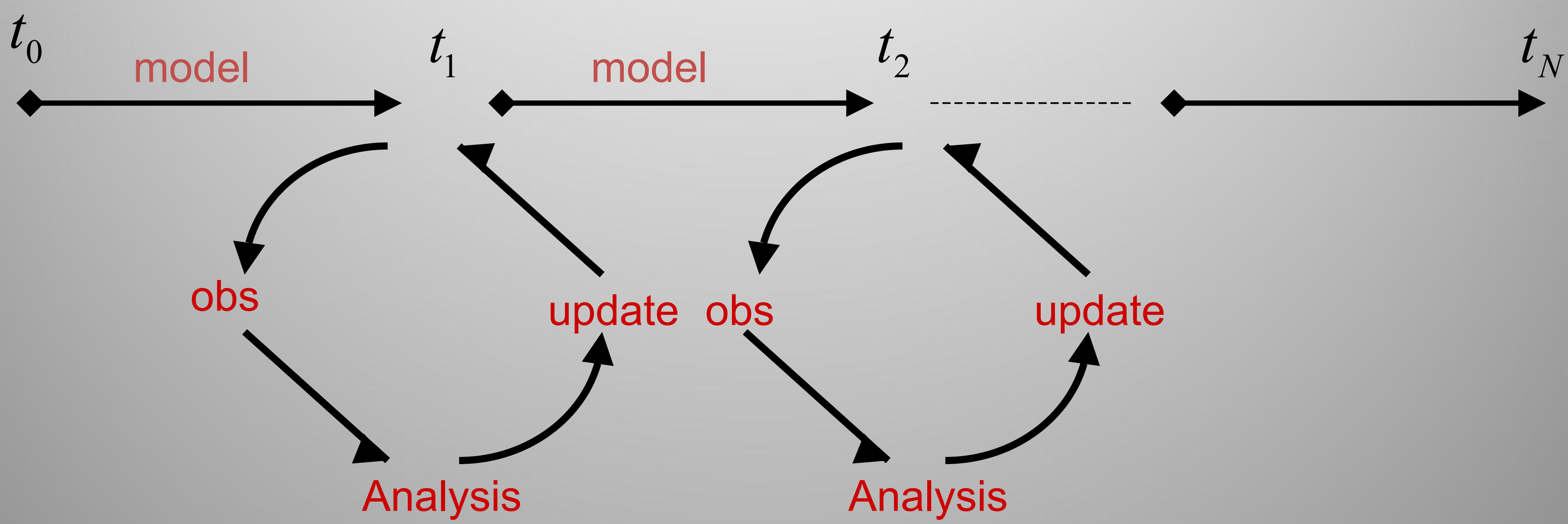
Collaborators: Kayo Ide (Atmos Sci, Maryland) and
Guillaume Vernieres (NASA-Goddard)

Supported by the US Office of Naval Research

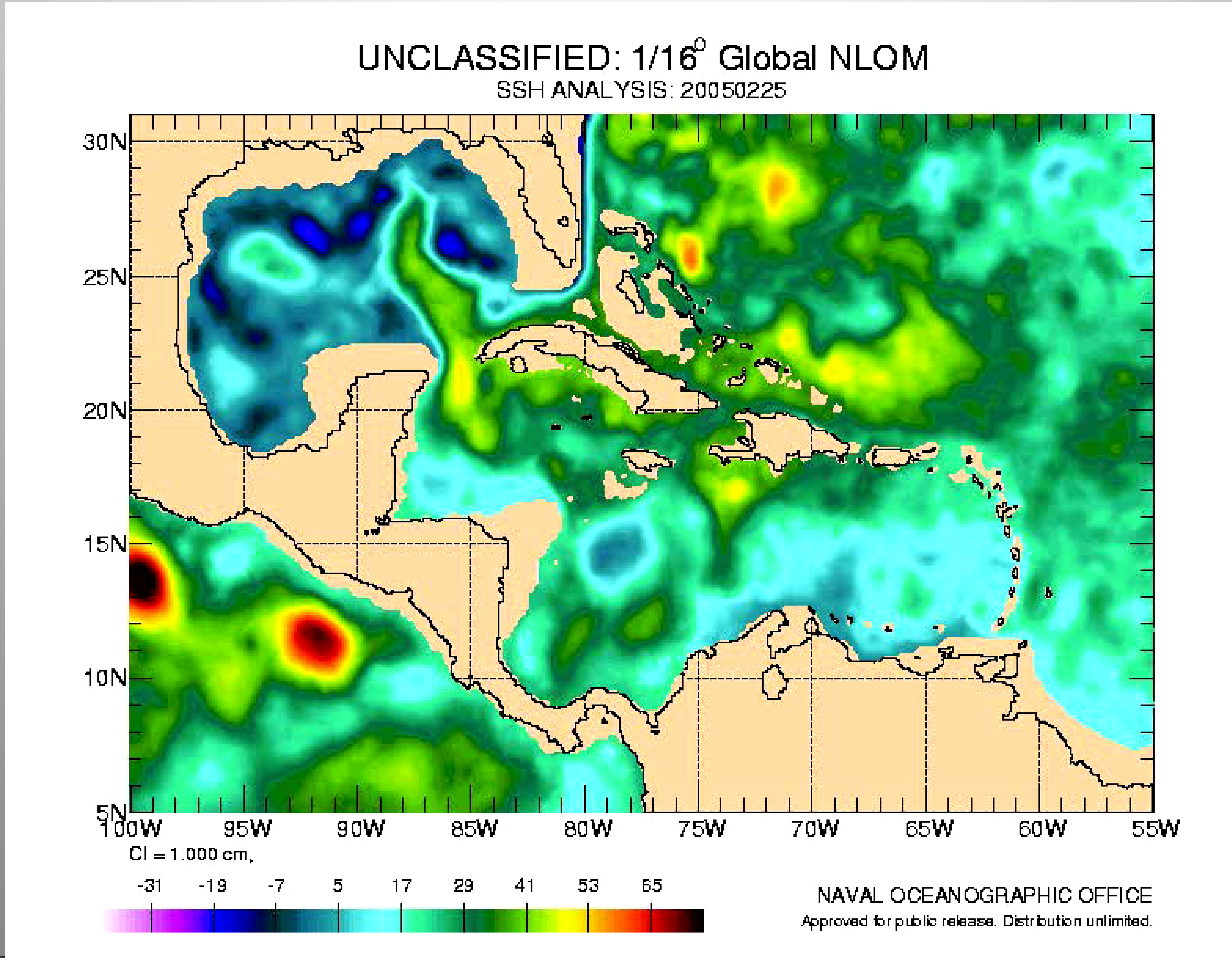

THEORY
OBSERVATIONS
Models
Computation
NUMERICAL OUTPUT
DATA
DATA ASSIMILATION
prediction
estimation
understanding

Sequential Data Assimilation/Filtering

Model + observations ⟶ prediction

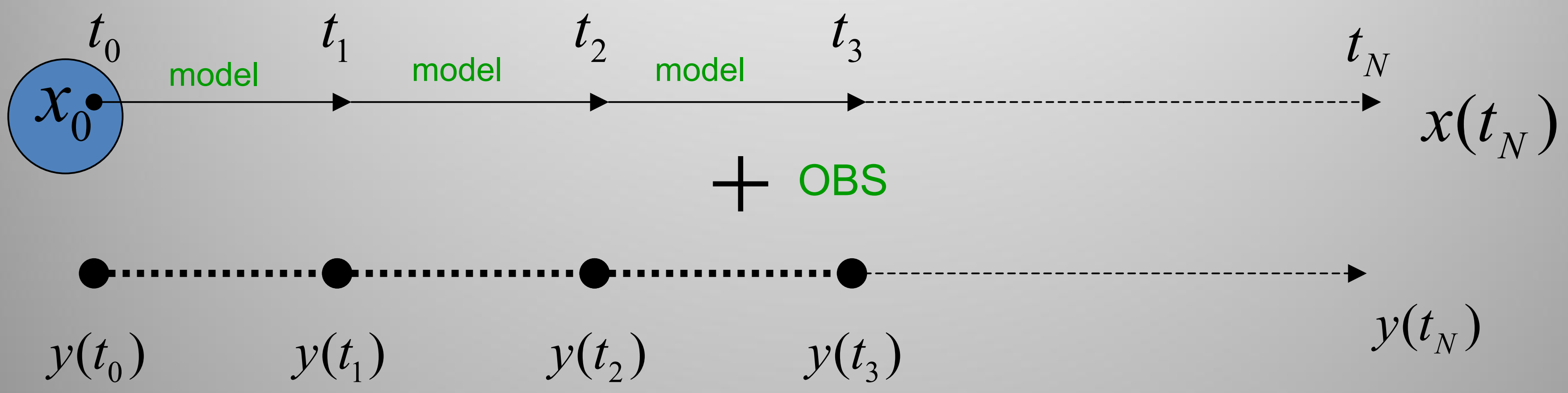


Gulf of Mexico/Carribean



State Estimation/Smoothing

Model + observations ⟶ state estimate



Analysis gives "better" estimate of initial value in light of observations

Ocean model:

$\mathbf{x} \in \mathbf{R}^N$ – state vector comprising all relevant dynamical variables

(e.g. flow velocity, temperature, salinity, etc. at each grid point)

$d\mathbf{x}^{\mathrm{f}} = M(\mathbf{x}^{\mathrm{f}}, t)dt$ prognostic model

$d\mathbf{x}^{\mathrm{t}} = M(\mathbf{x}^{\mathrm{t}}, t)dt + \boldsymbol{\eta}(t)dt$ actual evolution

$E[\boldsymbol{\eta}(t)\boldsymbol{\eta}^T(t')] = \delta(t-t')\mathbf{Q}(t)$ covariance of the model residual

Observations:

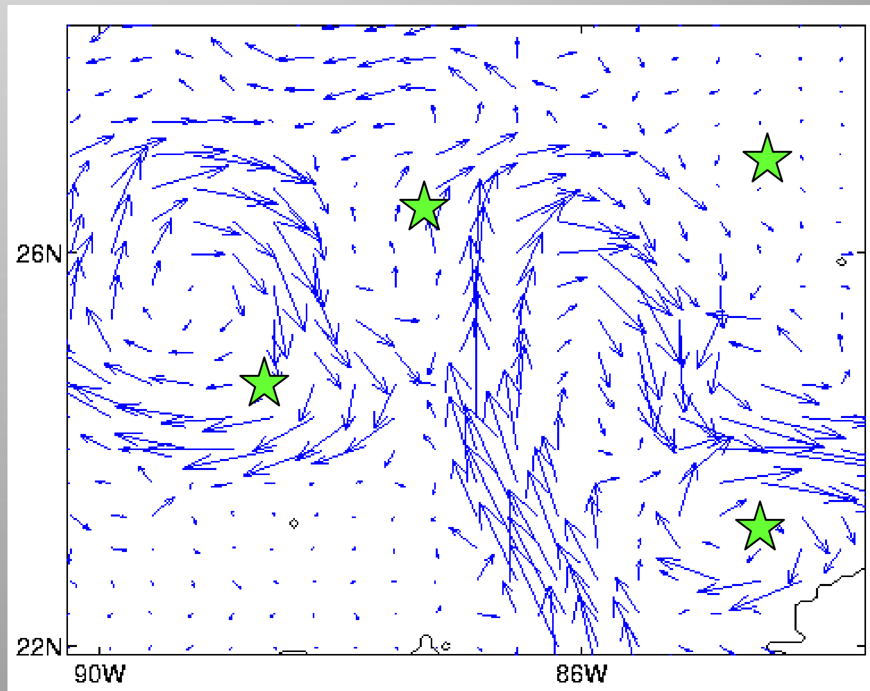
$\mathrm{y}_i^{\mathrm{o}} = H_i[\mathbf{x}_i^{\mathrm{t}}] + \boldsymbol{\varepsilon}_i$

H_i – observation operator

$\boldsymbol{\varepsilon}_i$ – observation error

$\mathrm{y}_i^{\mathrm{o}} \in \mathbf{R}^L$, typically $L = N$

$E[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_m^T] = \delta_{im}\mathbf{R}_i$ – observation error covariance



Fishkill in Lake Kinneret

Vernieres et al. (2006)

Conjecture: due to “lifting” of lower layer of oxygen-free water

- Occasional “fishkill”
- Feeding of 5,000??

Model:

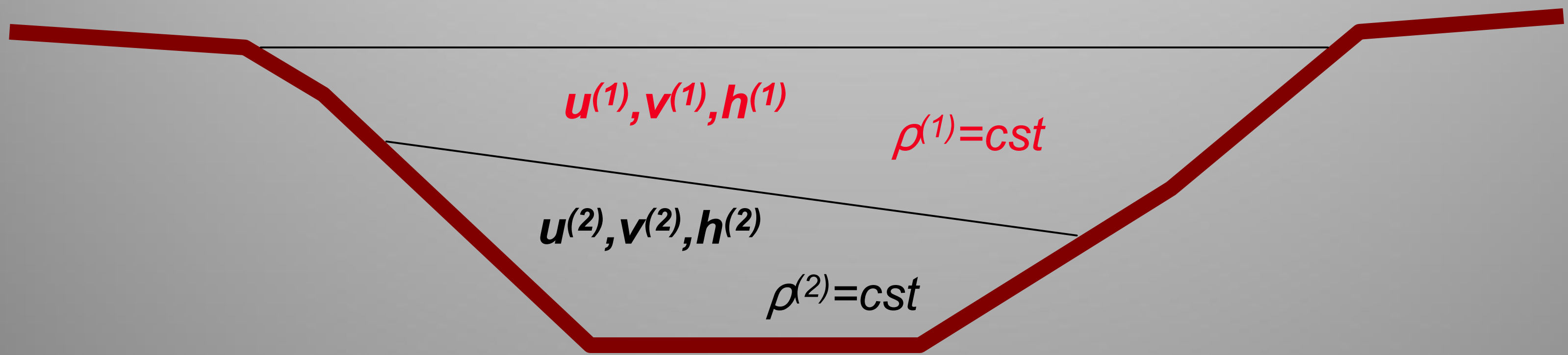
- Stably stratified during summer
- Strong westerly sea breeze

$$\left.\begin{aligned}
&\frac{Du^{(1)}}{Dt}-fv^{(1)}+g\frac{\partial}{\partial x}\left(h^{(1)}+h^{(2)}+D\right)-A_h\nabla^2u^{(1)}-F_u=0\\
&\frac{Dv^{(1)}}{Dt}+fu^{(1)}+g\frac{\partial}{\partial y}\left(h^{(1)}+h^{(2)}+D\right)-A_h\nabla^2v^{(1)}-F_u=0
\end{aligned}\right\}\text{Top layer momentum}$$

$$\frac{Dh^{(1)}}{Dt}=0$$

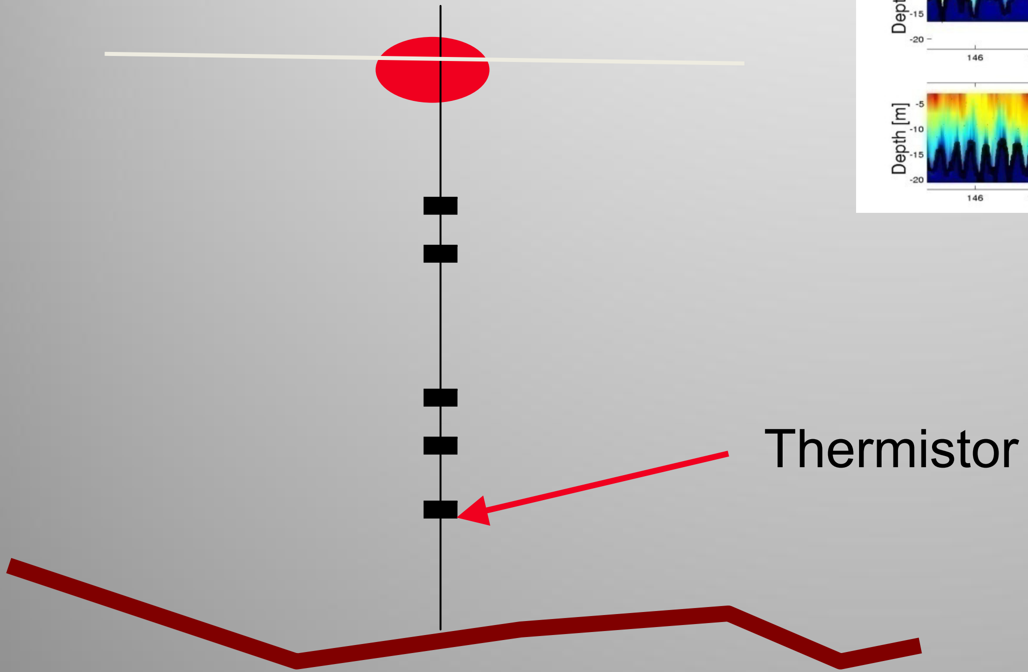
$$\left.\begin{aligned}
&\frac{Du^{(2)}}{Dt}-fv^{(1)}+g\frac{\partial}{\partial x}\left(h^{(1)}+h^{(2)}+D\right)+g'\frac{\partial}{\partial x}\left(h^{(2)}+D\right)-A_h\nabla^2u^{(1)}=0\\
&\frac{Dv^{(2)}}{Dt}+fu^{(1)}+g\frac{\partial}{\partial y}\left(h^{(1)}+h^{(2)}+D\right)+g'\frac{\partial}{\partial y}\left(h^{(2)}+D\right)-A_h\nabla^2v^{(1)}=0
\end{aligned}\right\}\text{Bottom layer momentum}$$

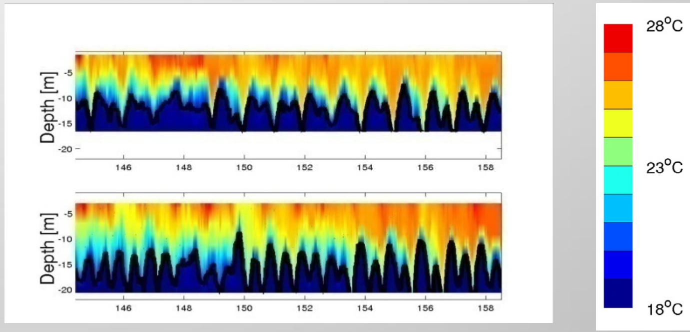
$$\frac{Dh^{(2)}}{Dt}=0$$



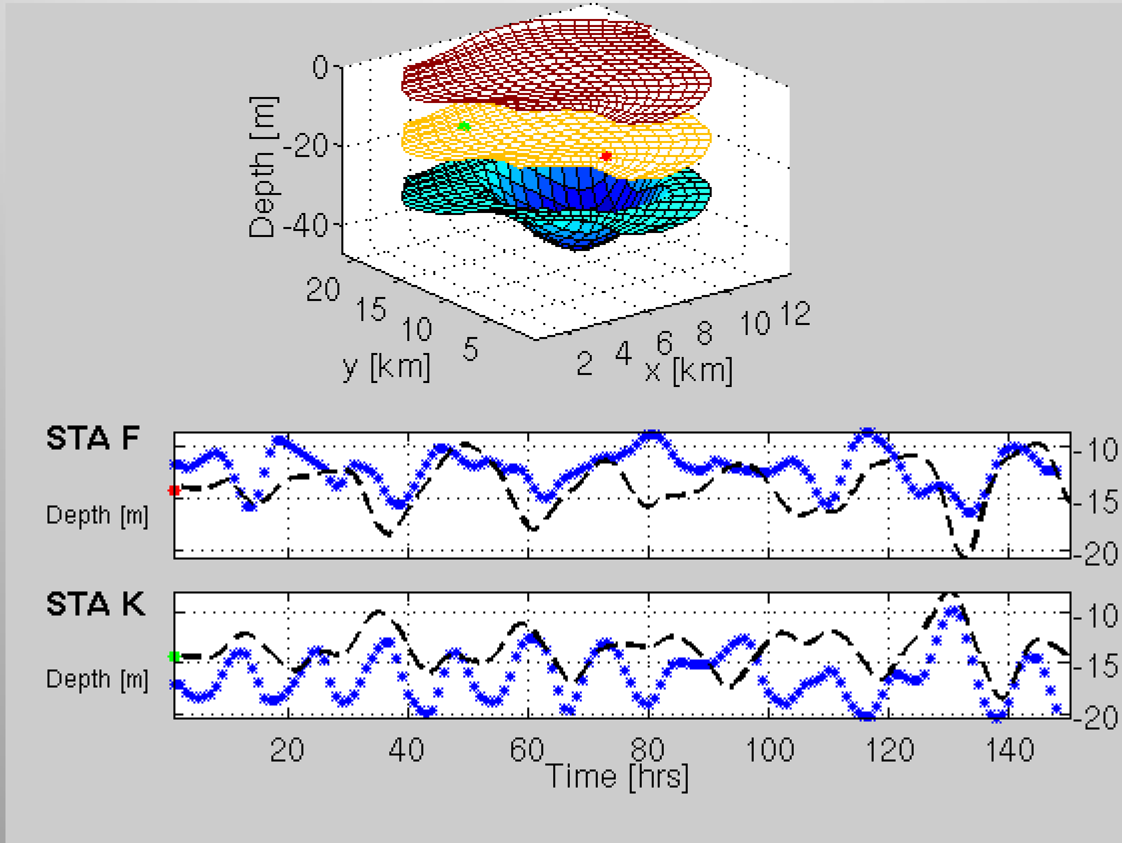
Data

- Thermistor chains

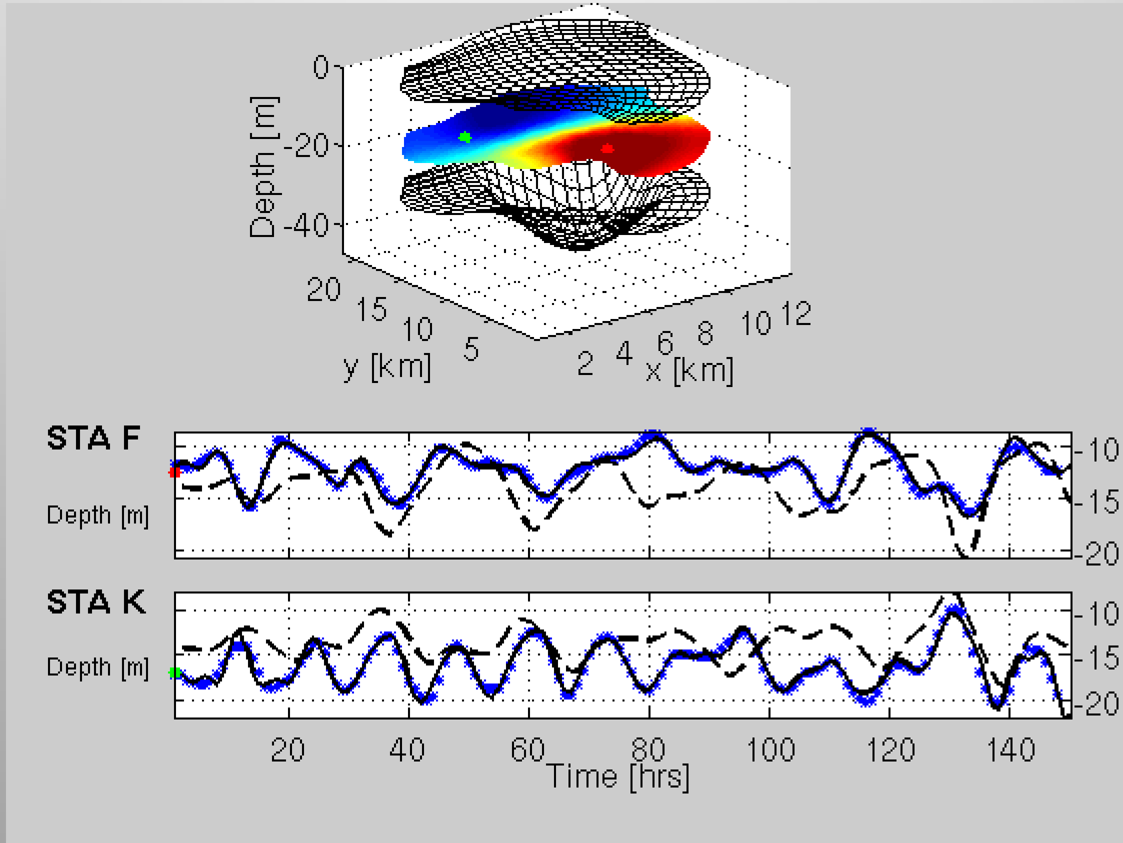




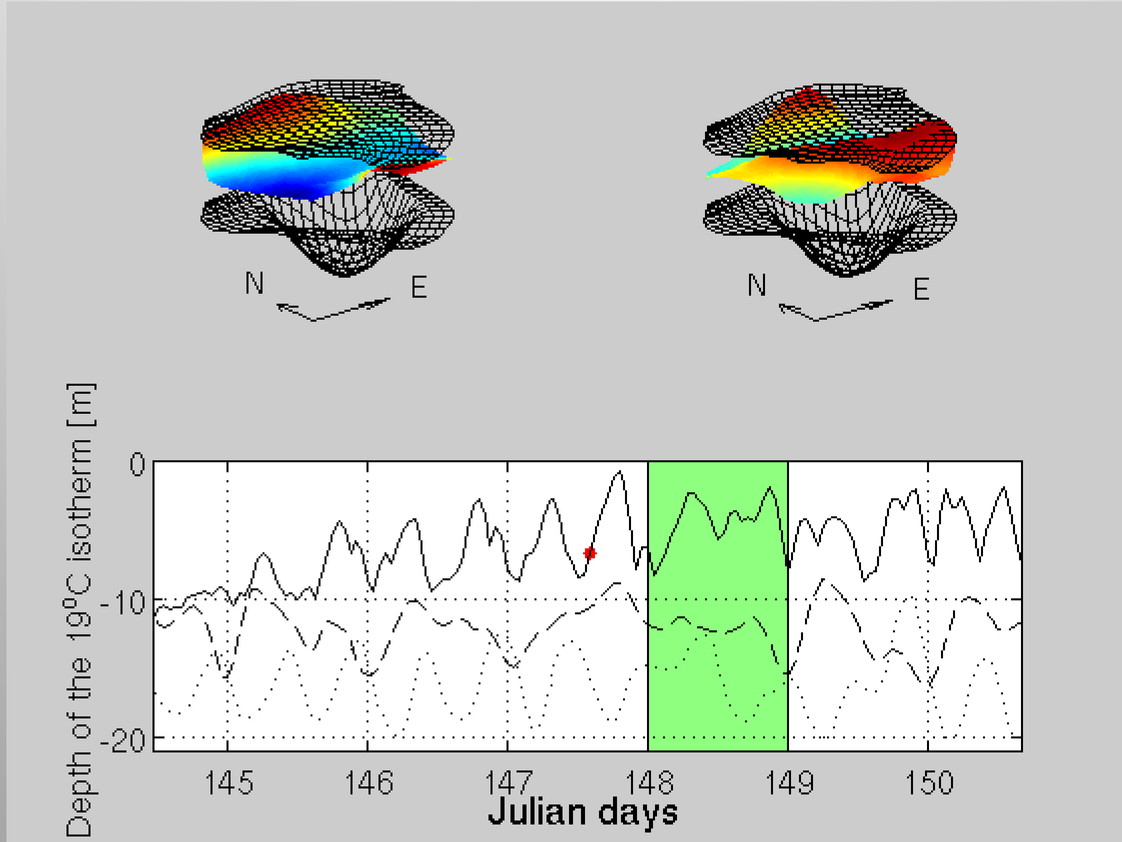
Model running on its own…



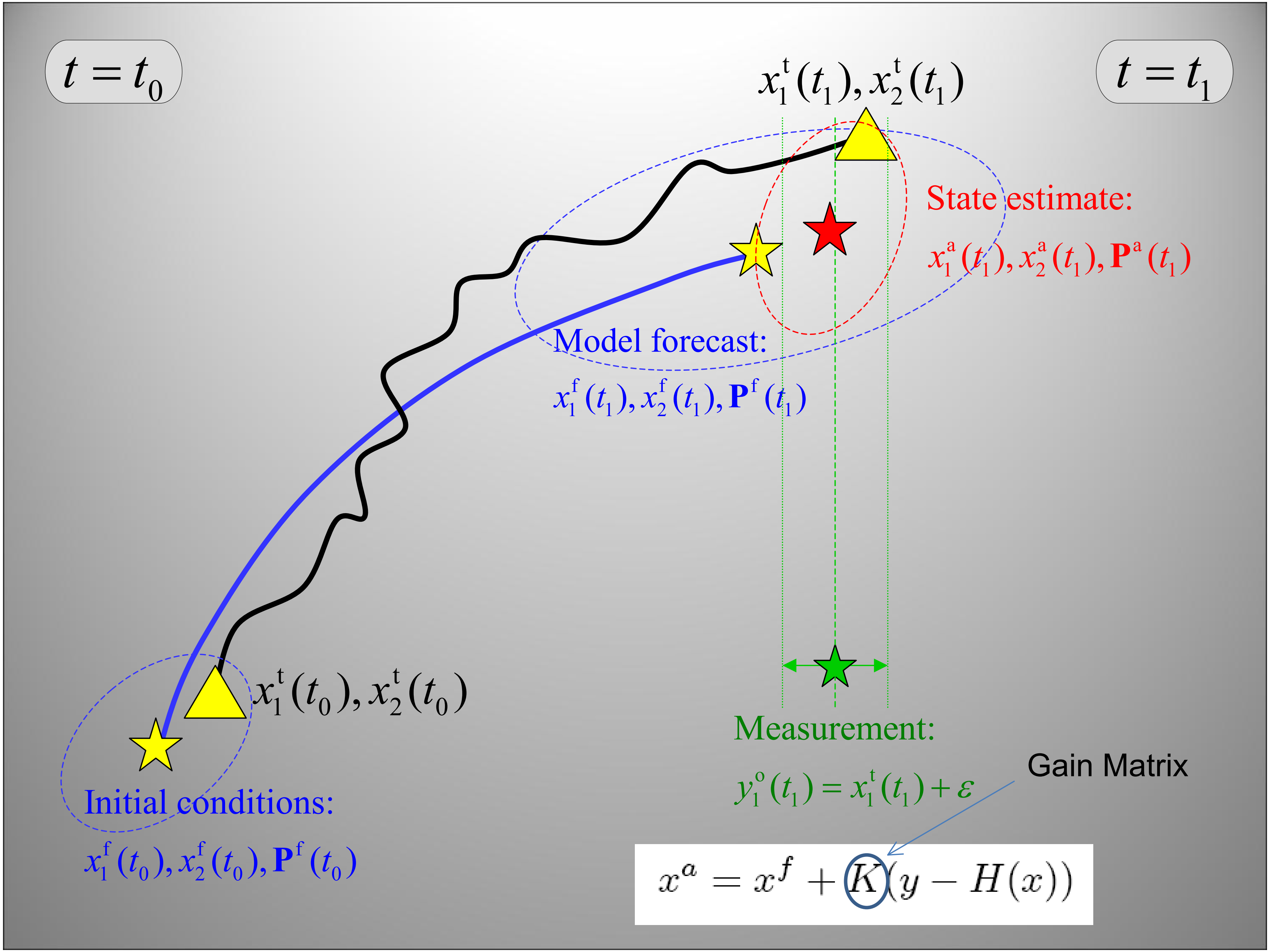
With thermistor data assimilated…



A comparison



Neither model nor data on their own show “fishkill,”
But, together, they do!



$t = t_0$
$t = t_1$
$x_1^{\mathrm{t}}(t_1), x_2^{\mathrm{t}}(t_1)$
State estimate:
$x_1^{\mathrm{a}}(t_1), x_2^{\mathrm{a}}(t_1), \mathbf{P}^{\mathrm{a}}(t_1)$
Model forecast:
$x_1^{\mathrm{f}}(t_1), x_2^{\mathrm{f}}(t_1), \mathbf{P}^{\mathrm{f}}(t_1)$
$x_1^{\mathrm{t}}(t_0), x_2^{\mathrm{t}}(t_0)$
Measurement:
$y_1^{\mathrm{o}}(t_1) = x_1^{\mathrm{t}}(t_1) + \varepsilon$
Gain Matrix
Initial conditions:
$x_1^{\mathrm{f}}(t_0), x_2^{\mathrm{f}}(t_0), \mathbf{P}^{\mathrm{f}}(t_0)$
$x^a = x^f + K(y - H(x))$

Kalman Filter

- Distributions are Gaussian
- Model is linear-TLM (EKF)
- or fit to Gaussian (EnKF)

Extended Kalman Filter

Forecast model error covariance using tangent linear model:

$$\mathbf{P}^{\mathrm{f}} = E[\Delta\mathbf{x}\Delta\mathbf{x}^T]; \qquad \Delta\mathbf{x} \equiv \mathbf{x}^{\mathrm{f}} - \mathbf{x}^{\mathrm{t}}$$

$$\frac{d\mathbf{P}^{\mathrm{f}}}{dt} = \mathbf{M}\mathbf{P}^{\mathrm{f}} + \mathbf{P}^{\mathrm{f}}\mathbf{M}^T + \mathbf{Q}(t)$$

$\mathbf{M}_i \equiv \partial M(\mathbf{x}^{\mathrm{f}}, t)/\partial\mathbf{x}$: linearized model operator

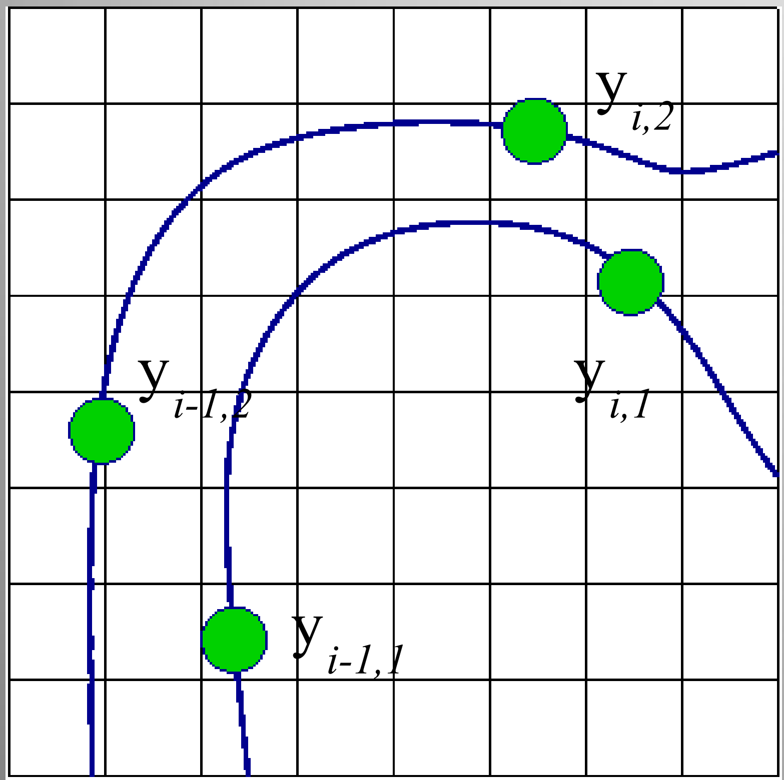
Combine model and observations into a new state $\mathbf{x}^{\mathrm{a}}$ minimizing $\mathrm{tr}\,\mathbf{P}_j^{\mathrm{a}}$

$$\mathbf{x}_j^{\mathrm{a}} = \mathbf{x}_j^{\mathrm{f}} + \mathbf{K}_j\mathbf{d}_j \qquad \mathbf{d}_j = \mathbf{y}_j^{\mathrm{o}} - H_j(\mathbf{x}_j^{\mathrm{f}})$$

$$\mathbf{K}_j = \mathbf{P}_j^{\mathrm{f}}\mathbf{H}_j^T\left(\mathbf{H}_j\mathbf{P}_j^{\mathrm{f}}\mathbf{H}_j^T + \mathbf{R}_j\right)^{-1} \qquad \mathbf{P}_j^{\mathrm{a}} = \left(\mathbf{I} - \mathbf{K}_j\mathbf{H}_j\right)\mathbf{P}_j^{\mathrm{f}}$$

$\mathbf{H}_i \equiv \partial H_i / \partial\mathbf{x}$: linearized observation function

Lagrangian Data Assimilation



Lagrangian observations from drifters and floats do not give the data in terms of model variables.

Solution: Include drifter coordinates into the model

Ide, Kuznetsov and J.
Other methods: Molcard et al.

Augmented system

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

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

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

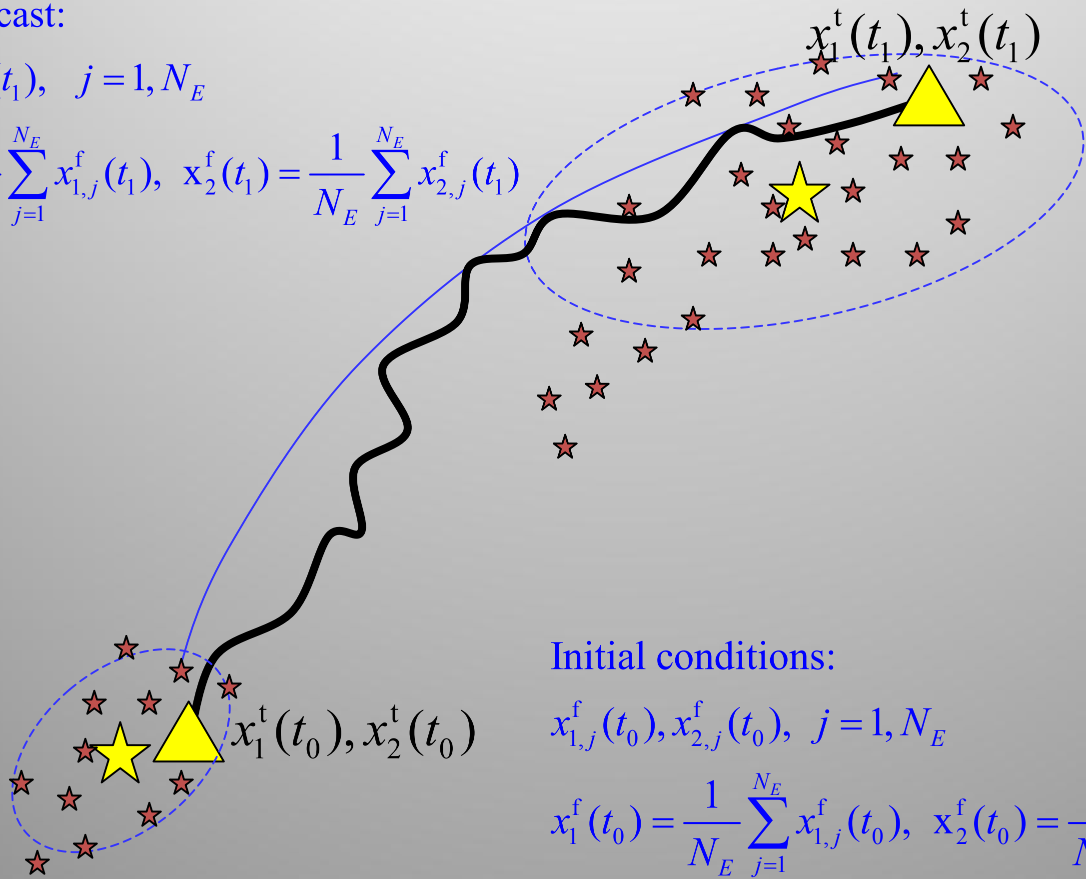
Ensemble Kalman Filter (EnKF)

Error covariance is predicted via solution of full nonlinear system for a Monte-Carlo ensemble of states

Model forecast:

$x^{\mathrm{f}}_{1,j}(t_1), x^{\mathrm{f}}_{2,j}(t_1), \quad j=1, N_E$

$$x^{\mathrm{f}}_1(t_1) = \frac{1}{N_E}\sum_{j=1}^{N_E} x^{\mathrm{f}}_{1,j}(t_1), \;\; \mathrm{x}^{\mathrm{f}}_2(t_1) = \frac{1}{N_E}\sum_{j=1}^{N_E} x^{\mathrm{f}}_{2,j}(t_1)$$



Initial conditions:

$x^{\mathrm{f}}_{1,j}(t_0), x^{\mathrm{f}}_{2,j}(t_0), \;\; j=1, N_E$

$$x^{\mathrm{f}}_1(t_0) = \frac{1}{N_E}\sum_{j=1}^{N_E} x^{\mathrm{f}}_{1,j}(t_0), \;\; \mathrm{x}^{\mathrm{f}}_2(t_0) = \frac{1}{N_E}\sum_{j=1}^{N_E} x^{\mathrm{f}}_{2,j}(t_0)$$

Update step in EnKF

Kalman gain matrix is computed using error covariance matrix derived from the ensemble.
Ensemble members are updated with noisy observations

$$\overline{\mathbf{x}}^{\mathrm{f}} = \frac{1}{N_E}\sum_{j=1}^{N_E}\mathbf{x}_j^{\mathrm{f}} \qquad \mathbf{P}^{\mathrm{f}} = \frac{1}{N_E - 1}\sum_{j=1}^{N_E}\left(\mathbf{x}_j^{\mathrm{f}} - \overline{\mathbf{x}}^{\mathrm{f}}\right)\left(\mathbf{x}_j^{\mathrm{f}} - \overline{\mathbf{x}}^{\mathrm{f}}\right)^{\mathrm{T}}$$

Ensemble of observations: $\mathbf{d}_j = \mathbf{y}^{\mathrm{o}} + \varepsilon_j^{\mathrm{o}} - H(\mathbf{x}_j^{\mathrm{f}}) \qquad E[\varepsilon_j^{\mathrm{o}}\varepsilon_j^{\mathrm{o}T}] = \mathbf{R}$

Update ensemble members:

$$\mathbf{x}_j^{\mathrm{a}} = \mathbf{x}_j^{\mathrm{f}} + \mathbf{K}\mathbf{d}_j \qquad \mathbf{K} = \mathbf{P}^{\mathrm{f}}\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^{\mathrm{f}}\mathbf{H}^T + \mathbf{R}\right)^{-1}$$

Techniques of Data Assimilation

Deterministic techniques

• Variational methods (3DVAR, 4DVAR)
• Kalman filter
• Ensemble Kalman filter

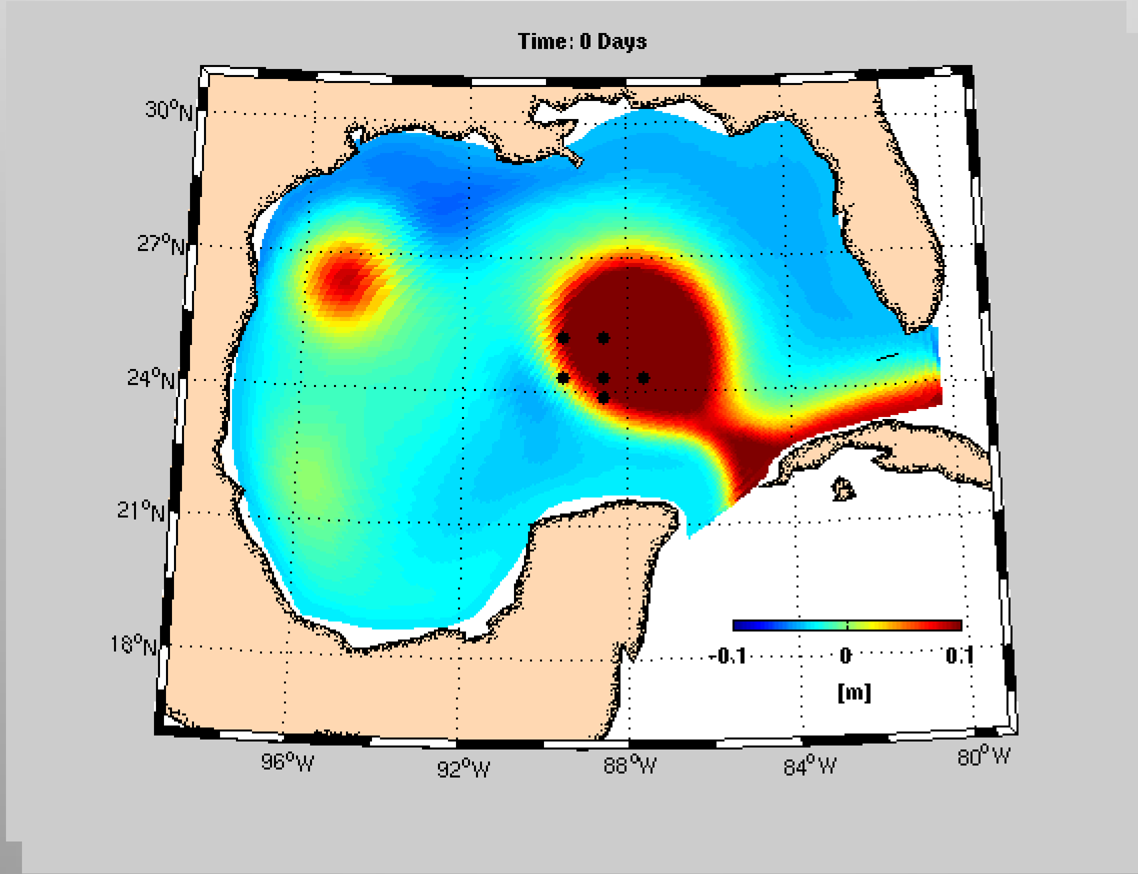
Requirements:
1. Gaussian
2. Close to linear

Statistical techniques

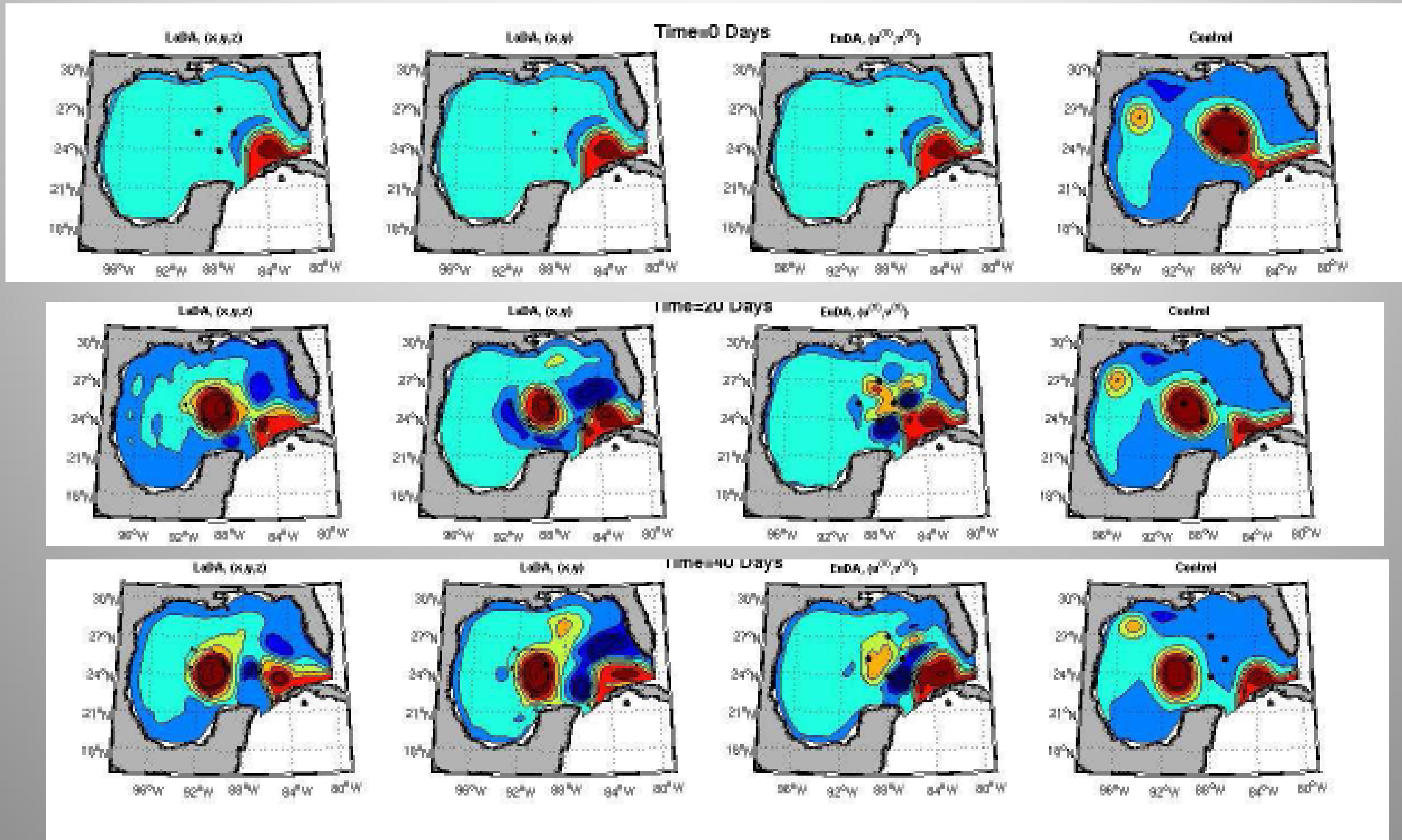
• Particle filtering
• Dynamic Monte-Carlo
• Sampling strategies

Requirement:
Low dimension

Capturing the eddy



Eddies in GoM



Work with Guillaume Vernieres (NASA) and Kayo Ide (Maryland)

Network of Gliders

Virtual leader (VL) + 4 gliders

Glider positions: $x^{(i)}, i = 1, \ldots, 4$

Equations of motion:

$$\dot{x}^{(i)} = u_F(x) + u_g^{(i)} = u(x)$$

flow speed

relative glider speed

Glider controls

$$\dot{u}^{(i)} = -\sum_{j=i}^{N_D} f_I\left(\left\| x^{(ij)} \right\|\right) \frac{x^{(ij)}}{\left\| x^{(ij)} \right\|} - f_h\left(\left\| h^{(i)} \right\|\right) \frac{h^{(i)}}{\left\| h^{(i)} \right\|} - K u^{(i)}$$

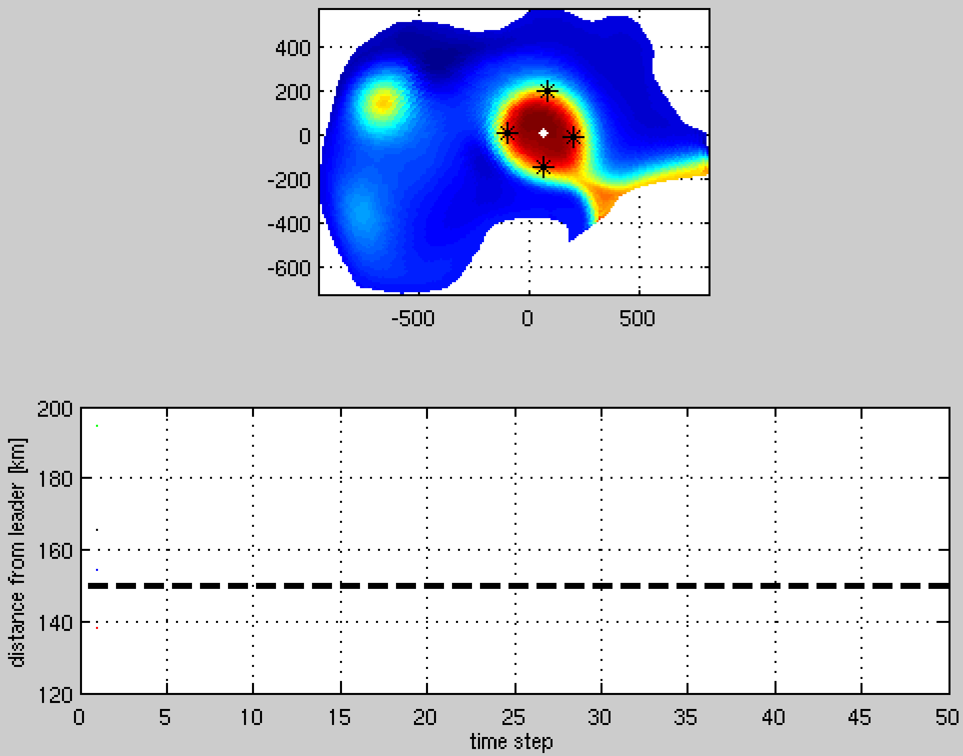
$$x^{(ij)} = x^{(i)} - x^{(j)}$$

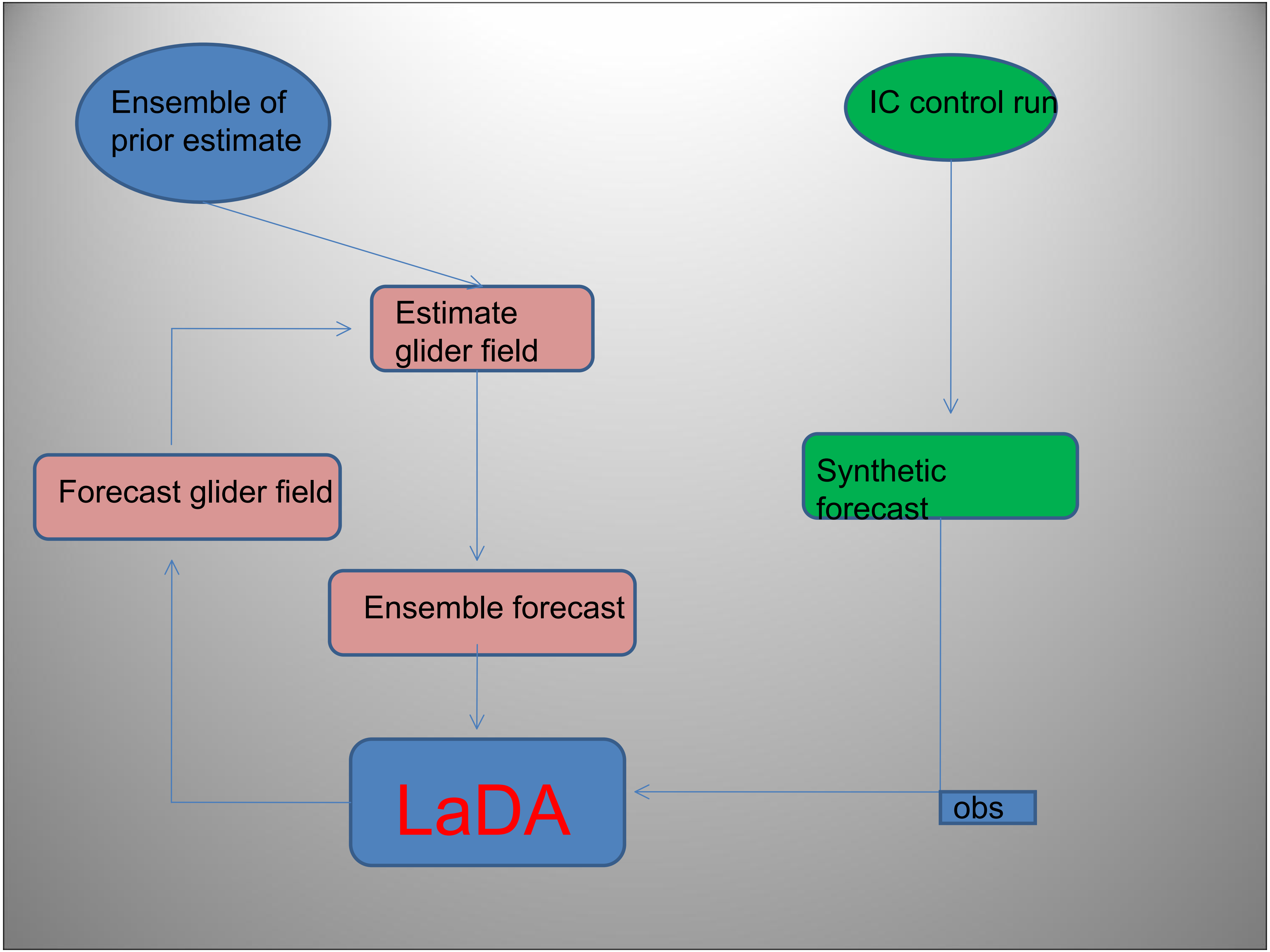
$$h^{(i)} = x^{(i)} - h_0$$

h_0 position of virtual leader

f_I f_h control forces derived from artificial potential

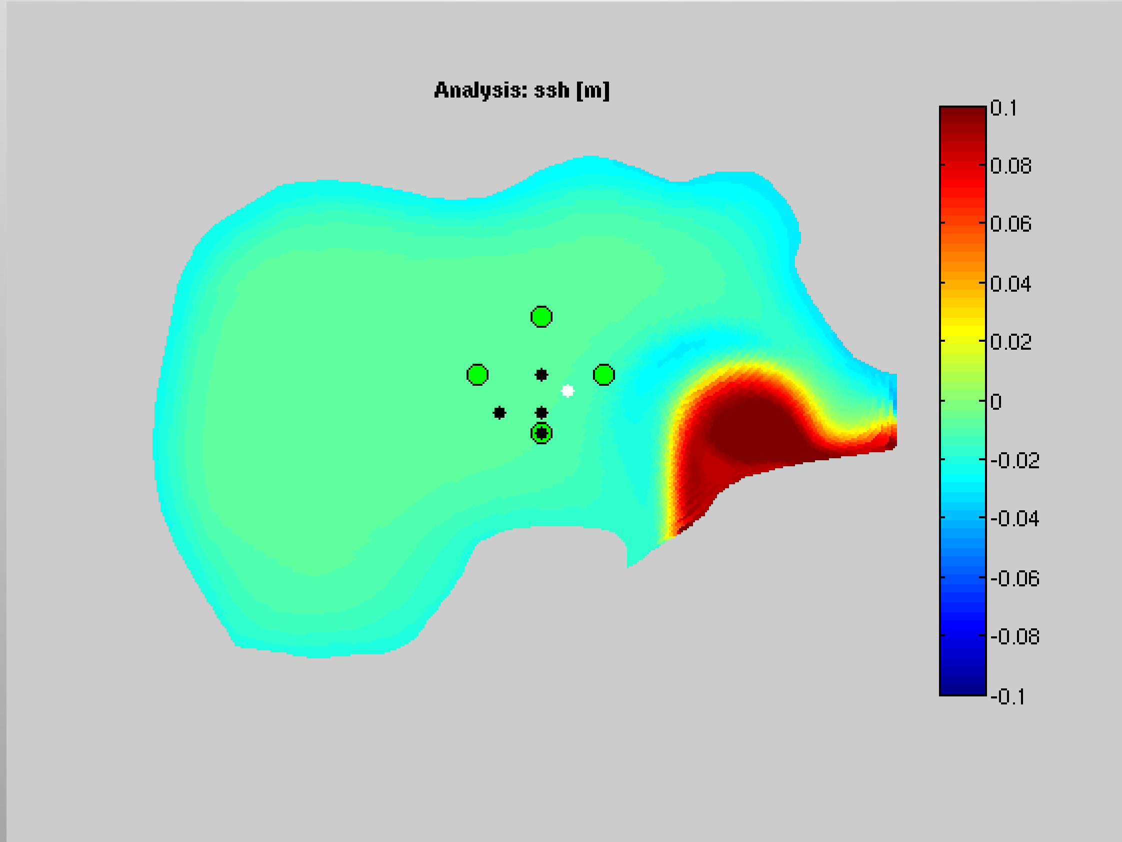
Gliders following the leader





Ensemble of prior estimate
IC control run
Estimate glider field
Forecast glider field
Synthetic forecast
Ensemble forecast
LaDA
obs

Capturing the eddy



Conclusions

- Lagrangian DA scheme is effective
- High information content in drifter/glider data
- Important to develop effective methods for assimilating such quasi-Lagrangian data
- Potential interplay between DA and data collection