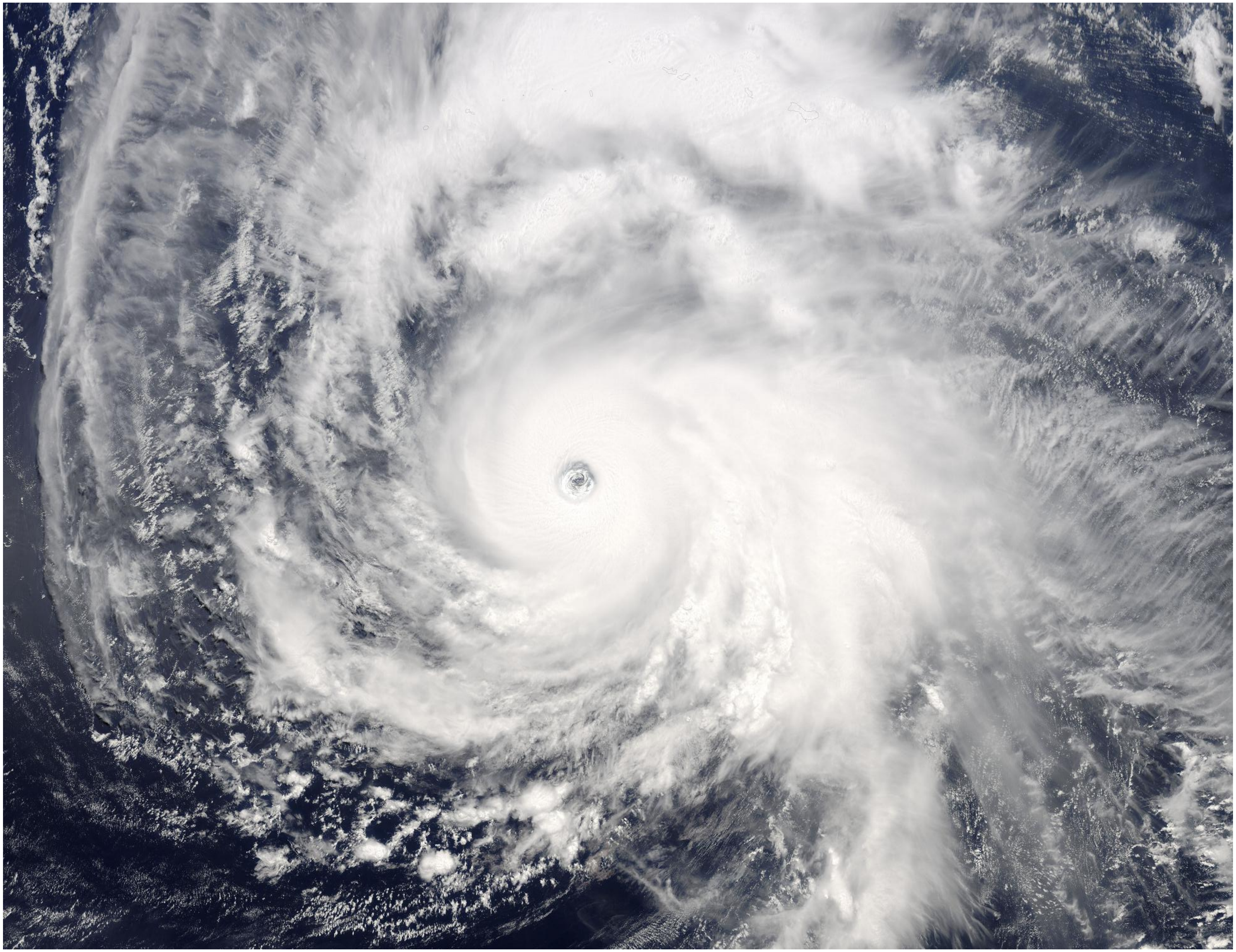
U.S. NAVAL RESEARCH LABORATORY

State Estimation for a Filtered Representation of a Chaotic Field

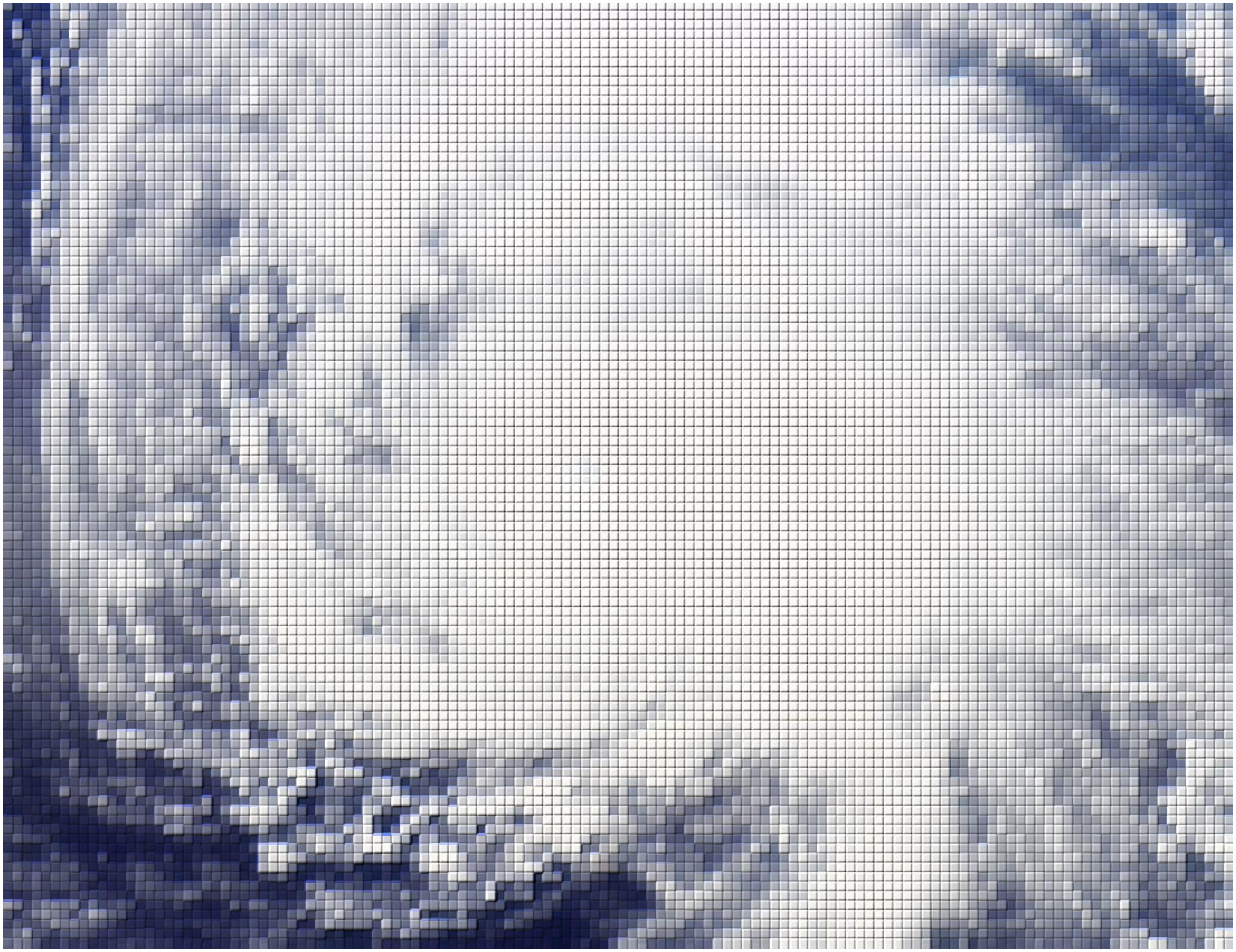
Dan Hodyss[1] and Peter Schwartz[2]

1. Marine Meteorology Division, Naval Research Laboratory, Monterey, CA
2. Applied Numerical Algorithms Group, Lawrence Berkeley National Laboratory, Berkeley, CA

Reality is detailed and full of structure …



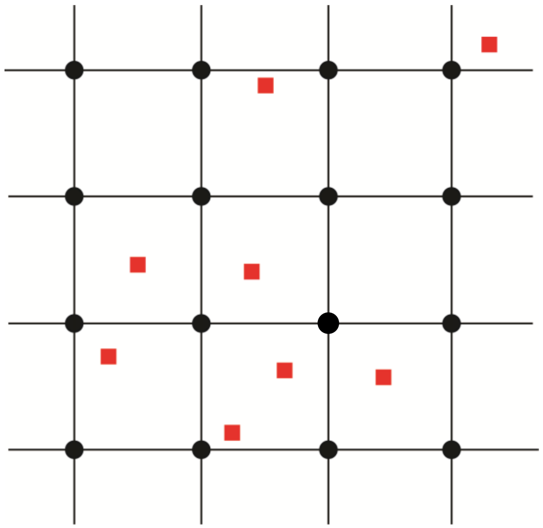
… but our model simulations are coarse and smoothed.



Apples and Oranges

- We have **point** measurements of things like temperature, barometric pressure, wind velocity, etc.
- We have model simulated values of **area-averaged** temperature, barometric pressure, wind velocity, etc.
  - ➢ Cannot run PDE solver at a resolution for which all important physical processes are resolved

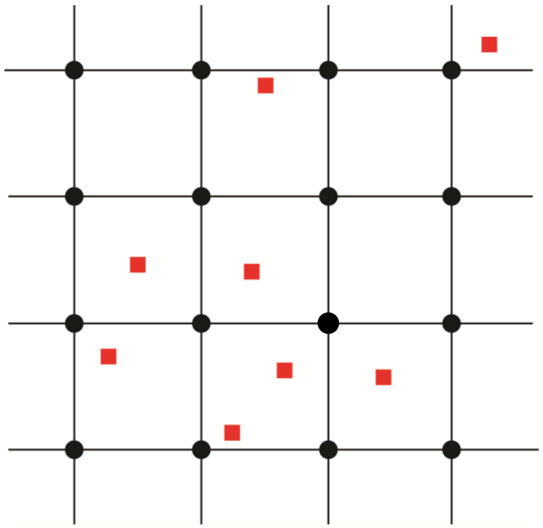
Observations
Model grid points



What's our Goal?

- We have at least two choices:
  1. Search for the best estimate of the pointwise values of temperature, winds, etc. at our model grid points
  2. Search for the best estimate of the area-averaged values of temperature, winds, etc. in each of our grid cells

Observations
Model grid points



High-Resolution Data Assimilation

Step j = 0: $p\left(\mathbf{x}_H^0\right) \longleftarrow$ Climatological PDF

$$\downarrow$$

Step j = 1: $$p\left(\mathbf{x}_H^1 \middle| \mathbf{y}_1\right) = C_1 p\left(\mathbf{y}_1 \middle| \mathbf{x}_H^1\right) p\left(\mathbf{x}_H^1\right)$$

$$\downarrow$$

Integrate Forward $$p\left(\mathbf{x}_H^2 \middle| \mathbf{y}_1\right) = \int_{-\infty}^{\infty} p\left(\mathbf{x}_H^2 \middle| \mathbf{x}_H^1\right) p\left(\mathbf{x}_H^1 \middle| \mathbf{y}_1\right) d\mathbf{x}_H^1$$

$$\downarrow$$

Step j = 2: $$p\left(\mathbf{x}_H^2 \middle| \mathbf{y}_2, \mathbf{y}_1\right) = C_2 p\left(\mathbf{y}_2 \middle| \mathbf{x}_H^2\right) p\left(\mathbf{x}_H^2 \middle| \mathbf{y}_1\right)$$

$$\vdots$$

Step j = N: $$\underbrace{p\left(\mathbf{x}_H^N \middle| \mathbf{Y}\right)}_{\text{Posterior}} = C_N \underbrace{p\left(\mathbf{y}_N \middle| \mathbf{x}_H^N\right)}_{\text{Ob Likelihood}} \underbrace{p\left(\mathbf{x}_H^N \middle| \mathbf{y}_{N-1}, \mathbf{y}_{N-2}, \ldots\right)}_{\text{Prior}}$$

where all the observations are collected together: $\mathbf{Y} = \left[\mathbf{y}_1 \quad \mathbf{y}_2 \quad \cdots \quad \mathbf{y}_N\right]$

Low-Resolution Data Assimilation

Step j = 0: $p\left(\mathbf{x}_L^0\right)$ ← Climatological PDF

Step j = 1: $$p\left(\mathbf{x}_L^1 \middle| \mathbf{y}_1\right) = C_1 p\left(\mathbf{y}_1 \middle| \mathbf{x}_L^1\right) p\left(\mathbf{x}_L^1\right)$$

Integrate Forward $$p\left(\mathbf{x}_L^2 \middle| \mathbf{y}_1\right) = \int_{-\infty}^{\infty} p\left(\mathbf{x}_L^2 \middle| \mathbf{x}_L^1\right) p\left(\mathbf{x}_L^1 \middle| \mathbf{y}_1\right) d\mathbf{x}_L^1$$

Step j = 2: $$p\left(\mathbf{x}_L^2 \middle| \mathbf{y}_2, \mathbf{y}_1\right) = C_2 p\left(\mathbf{y}_2 \middle| \mathbf{x}_L^2\right) p\left(\mathbf{x}_L^2 \middle| \mathbf{y}_1\right)$$

$$\vdots$$

Step j = N: $$\underbrace{p\left(\mathbf{x}_L^N \middle| \mathbf{Y}\right)}_{\text{Posterior}} = C_N \underbrace{p\left(\mathbf{y}_N \middle| \mathbf{x}_L^N\right)}_{\text{Ob Likelihood}} \underbrace{p\left(\mathbf{x}_L^N \middle| \mathbf{y}_{N-1}, \mathbf{y}_{N-2}, \ldots\right)}_{\text{Prior}}$$

where all the observations are collected together: $\mathbf{Y} = \left[\mathbf{y}_1 \quad \mathbf{y}_2 \quad \cdots \quad \mathbf{y}_N\right]$

Low-Resolution Ob Likelihood

A little bit of manipulation with the chain rule of probability finds

$$p\left(\mathbf{y}_1 \middle| \mathbf{x}_L^1\right) = \int_{-\infty}^{\infty} p\left(\mathbf{y}_1 \middle| \mathbf{x}_H^1\right) p\left(\mathbf{x}_H^1 \middle| \mathbf{x}_L^1\right) d\mathbf{x}_H^1$$

High-Resolution Ob Likelihood

Synchronization Density

Synchronization Density

If the high and low-resolution systems are not synchronized then

$$p\left(\mathbf{x}_H^1 \middle| \mathbf{x}_L^1\right) = p\left(\mathbf{x}_H^1\right)$$

which implies

$$p\left(\mathbf{y}_1 \middle| \mathbf{x}_L^1\right) = \int_{-\infty}^{\infty} p\left(\mathbf{y}_1 \middle| \mathbf{x}_H^1\right) p\left(\mathbf{x}_H^1\right) d\mathbf{x}_H^1 = p\left(\mathbf{y}_1\right)$$

Our assimilation of the observation then delivers

$$p\left(\mathbf{x}_L^1 \middle| \mathbf{y}_1\right) = C_1 p\left(\mathbf{y}_1\right) p\left(\mathbf{x}_L^1\right) = p\left(\mathbf{x}_L^1\right)$$

Assume a Structure for the Synchronization

We note that

$$p\left(\mathbf{x}_H^1 \middle| \mathbf{x}_L^1\right) p\left(\mathbf{x}_L^1\right) = p\left(\mathbf{x}_L^1 \middle| \mathbf{x}_H^1\right) p\left(\mathbf{x}_H^1\right)$$

The *converse* synchronization density must satisfy

$$p\left(\mathbf{x}_L^1\right) = \int_{-\infty}^{\infty} p\left(\mathbf{x}_L^1 \middle| \mathbf{x}_H^1\right) p\left(\mathbf{x}_H^1\right) d\mathbf{x}_H^1$$

<u>Assumption</u>: A map exists between high and low-resolution such that

$$\boxed{p\left(\mathbf{x}_L^1 \middle| \mathbf{x}_H^1\right) = \delta\left(\mathbf{x}_L^1 - \mathbf{F}\left(\mathbf{x}_H^1\right)\right)}$$

How do we find F?

Note that **F** is the mean of

$$p\left(\mathbf{x}_L^1 \middle| \mathbf{x}_H^1\right) = \delta\left(\mathbf{x}_L^1 - \mathbf{F}\left(\mathbf{x}_H^1\right)\right)$$

Therefore, standard polynomial regression will find **F**:

$$\mathbf{F}\left(\mathbf{x}_H^1\right) = \int_{-\infty}^{\infty} \mathbf{x}_L^1 \, p\left(\mathbf{x}_L^1 \middle| \mathbf{x}_H^1\right) d\mathbf{x}_L^1 \approx \overline{\mathbf{x}}_L^1 + \mathbf{A}_1\left[\mathbf{x}_H^1 - \overline{\mathbf{x}}_H^1\right] + \ldots$$

When we truncate the expansion we no longer have zero variance:

$$p\left(\mathbf{x}_L^1 \middle| \mathbf{x}_H^1\right) = N \exp\left[-\frac{1}{2}\left(\mathbf{x}_L^1 - \hat{\mathbf{F}}\left(\mathbf{x}_H^1\right)\right)^T \mathbf{B}^{-1}\left(\mathbf{x}_L^1 - \hat{\mathbf{F}}\left(\mathbf{x}_H^1\right)\right)\right]$$

Low-Resolution Ob Likelihood

A little bit of manipulation with the chain rule of probability finds

$$p\left(\mathbf{y}_1 \middle| \mathbf{x}_L^1\right) = \int_{-\infty}^{\infty} p\left(\mathbf{y}_1 \middle| \mathbf{x}_H^1\right) p\left(\mathbf{x}_H^1 \middle| \mathbf{x}_L^1\right) d\mathbf{x}_H^1$$

High-Resolution Ob Likelihood

Synchronization Density

Low-Resolution Data Assimilation

Step j = 0: $$p\left(\mathbf{x}_L^0\right) \longleftarrow \text{Climatological PDF}$$

Step j = 1: $$p\left(\mathbf{x}_L^1 \middle| \mathbf{y}_1\right) = C_1 p\left(\mathbf{y}_1 \middle| \mathbf{x}_L^1\right) p\left(\mathbf{x}_L^1\right)$$

Integrate Forward $$p\left(\mathbf{x}_L^2 \middle| \mathbf{y}_1\right) = \int_{-\infty}^{\infty} p\left(\mathbf{x}_L^2 \middle| \mathbf{x}_L^1\right) p\left(\mathbf{x}_L^1 \middle| \mathbf{y}_1\right) d\mathbf{x}_L^1$$

Step j = 2: $$p\left(\mathbf{x}_L^2 \middle| \mathbf{y}_2, \mathbf{y}_1\right) = C_2 p\left(\mathbf{y}_2 \middle| \mathbf{x}_L^2\right) p\left(\mathbf{x}_L^2 \middle| \mathbf{y}_1\right)$$

$$\vdots$$

Step j = N: $$\underbrace{p\left(\mathbf{x}_L^N \middle| \mathbf{Y}\right)}_{\text{Posterior}} = C_N \underbrace{p\left(\mathbf{y}_N \middle| \mathbf{x}_L^N\right)}_{\text{Ob Likelihood}} \underbrace{p\left(\mathbf{x}_L^N \middle| \mathbf{y}_{N-1}, \mathbf{y}_{N-2}, \ldots\right)}_{\text{Prior}}$$

where all the observations are collected together: $\mathbf{Y} = \left[\mathbf{y}_1 \quad \mathbf{y}_2 \quad \cdots \quad \mathbf{y}_N\right]$

A Test Problem: Solitary Waves in Variable Media

The test problem we will use is a variable-coefficient KdV equation (Hodyss and Nathan 2003, 2006, 2007):

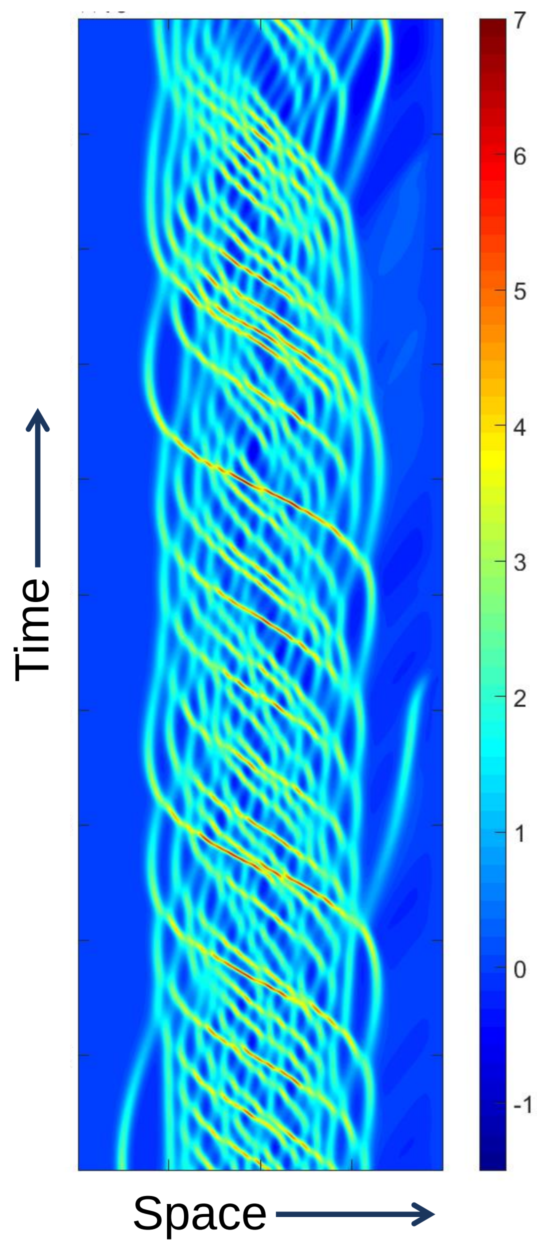
$$\frac{\partial A}{\partial t}+m_d\frac{\partial^3 A}{\partial x^3}+m_p(x)\frac{\partial A}{\partial x}+m_g(x)A+m_n A\frac{\partial A}{\partial x}=0$$

We set the coefficients to:

$$m_d=m_n=-1 \qquad m_p(x)=1-e^{-ax^2} \qquad m_g(x)=-2axe^{-ax^2}$$

Interesting DA problem because:

- Chaotic creation/destruction of solitary waves
- Very large amplitude solitary waves are very narrow
- Large amplitude waves move very fast

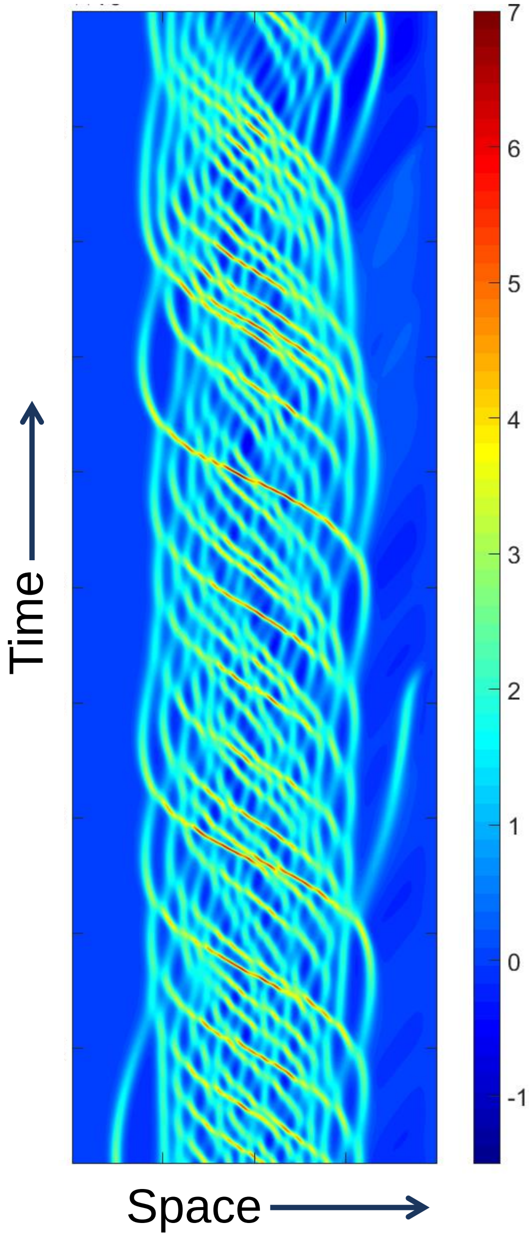


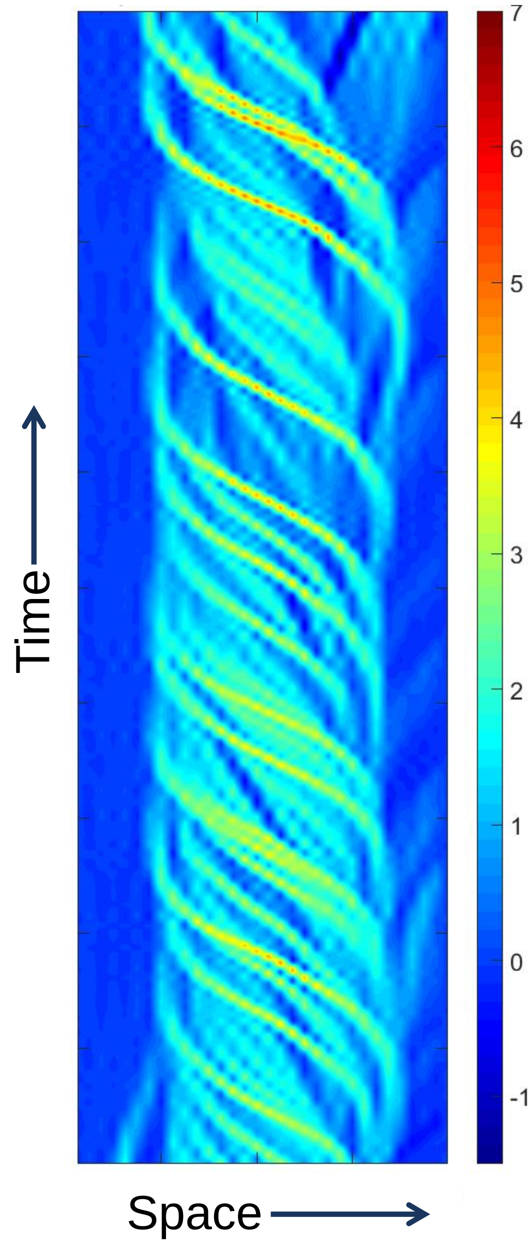
High versus Low Resolution

High-resolution simulation

N = 512

Both simulations will use the same numerical methods





Low-resolution simulation

M = 32

Data Assimilation Problem

- High-resolution – N = 512
- Low-resolution – M = 32

- The locations of the grid points of the low-resolution state vector will be coincident with the high-resolution state-vector subsampled every 16 points.

- Observations will be taken at the location of these overlapping points

- The observation error variance will be R = 0.3, which is approximately 50% of the climatological variance at high-resolution.

- There will be 1 unit of time between observations, which is approximately 1000 (100) time steps at high-resolution (low-resolution).

- We will use 1000 member ensembles

- The contemporary approach is brute-force tuned for best prior and observation inflation parameters

- Note: both methods benefited from some gross localization of the prior covariance matrices

Problem Statement and a Contemporary Approach

- Assume high-resolution reality with state vector of length N.
- Assume low-resolution model state space with state vector of length M.
- We will assume that we can run the model at resolution N, but then must perform our data assimilation at a reduced resolution of length *M.*

- A contemporary (ad hoc) approach using the Ensemble Kalman Filter
- The ensemble update step uses the stochastic observation approach (Evenson 2003)

Low-Resolution Update

$$\overline{\mathbf{x}}_L^c = \overline{\mathbf{x}}_L + \mathbf{G}^c \left[\mathbf{v}_L - \langle \mathbf{v}_L \rangle \right]$$

$$\mathbf{G}^c = \mathbf{P}_L \mathbf{H}_L^T \left[\mathbf{H}_L \mathbf{P}_L \mathbf{H}_L^T + \overline{\mathbf{R}}_c \right]^{-1}$$

$$\overline{\mathbf{R}}_c = \mathbf{R}_{ins} + \mathbf{R}_c$$

$$\mathbf{v}_L = \mathbf{y} - \mathbf{H}_L \overline{\mathbf{x}}_L$$

$$\langle \mathbf{v}_L \rangle = \mathbf{H}_H \overline{\mathbf{x}}_H - \mathbf{H}_L \overline{\mathbf{x}}_L$$

High-Resolution Update

$$\overline{\mathbf{x}}_H^c = \overline{\mathbf{x}}_H + \hat{\mathbf{F}}^\dagger \mathbf{G}^c \left[\mathbf{v}_L - \langle \mathbf{v}_L \rangle \right]$$

$$\mathbf{G}^c = \mathbf{P}_L \mathbf{H}_L^T \left[\mathbf{H}_L \mathbf{P}_L \mathbf{H}_L^T + \overline{\mathbf{R}}_c \right]^{-1}$$

$$\overline{\mathbf{R}}_c = \mathbf{R}_{ins} + \mathbf{R}_c$$

$$\mathbf{v}_L = \mathbf{y} - \mathbf{H}_L \overline{\mathbf{x}}_L$$

$$\langle \mathbf{v}_L \rangle = \mathbf{H}_H \overline{\mathbf{x}}_H - \mathbf{H}_L \overline{\mathbf{x}}_L$$

A Multi-Resolution Kalman Filter

- A multi-resolution Kalman filter approach will make use of the same Ensemble (Monte-Carlo) Kalman Filter framework (Hodyss and Nichols, 2015; Tellus A)
- The ensemble update step uses the stochastic observation approach (Evenson 2003)

Low-Resolution Update

$$\overline{\mathbf{x}}_L^b = \overline{\mathbf{x}}_L + \mathbf{G}\left[\mathbf{v}_L - \langle \mathbf{v}_L \rangle\right]$$

$$\mathbf{G} = \left[\mathbf{P}_L\mathbf{H}_L^T + \mathbf{P}_{LH}\right]\left[\mathbf{H}_L\mathbf{P}_L\mathbf{H}_L^T + \overline{\mathbf{R}}_L^*\right]^{-1}$$

$$\mathbf{P}_{LH} = \hat{\mathbf{F}}\mathbf{P}_H\left(\mathbf{H}_H - \mathbf{H}_L\hat{\mathbf{F}}\right)^T$$

$$\overline{\mathbf{R}}_L^* = \mathbf{R}_{ins} + \mathbf{H}_H\mathbf{P}_H\mathbf{H}_H^T - \mathbf{H}_L\mathbf{P}_L\mathbf{H}_L^T$$

$$\mathbf{v}_L = \mathbf{y} - \mathbf{H}_L\overline{\mathbf{x}}_L$$

$$\langle \mathbf{v}_L \rangle = \mathbf{H}_H\overline{\mathbf{x}}_H - \mathbf{H}_L\overline{\mathbf{x}}_L$$

High-Resolution Update

$$\overline{\mathbf{x}}_H^c = \overline{\mathbf{x}}_H + \hat{\mathbf{F}}^\dagger\mathbf{G}\left[\mathbf{v}_L - \langle \mathbf{v}_L \rangle\right]$$

$$\mathbf{G} = \left[\mathbf{P}_L\mathbf{H}_L^T + \mathbf{P}_{LH}\right]\left[\mathbf{H}_L\mathbf{P}_L\mathbf{H}_L^T + \overline{\mathbf{R}}_L^*\right]^{-1}$$

$$\mathbf{P}_{LH} = \hat{\mathbf{F}}\mathbf{P}_H\left(\mathbf{H}_H - \mathbf{H}_L\hat{\mathbf{F}}\right)^T$$

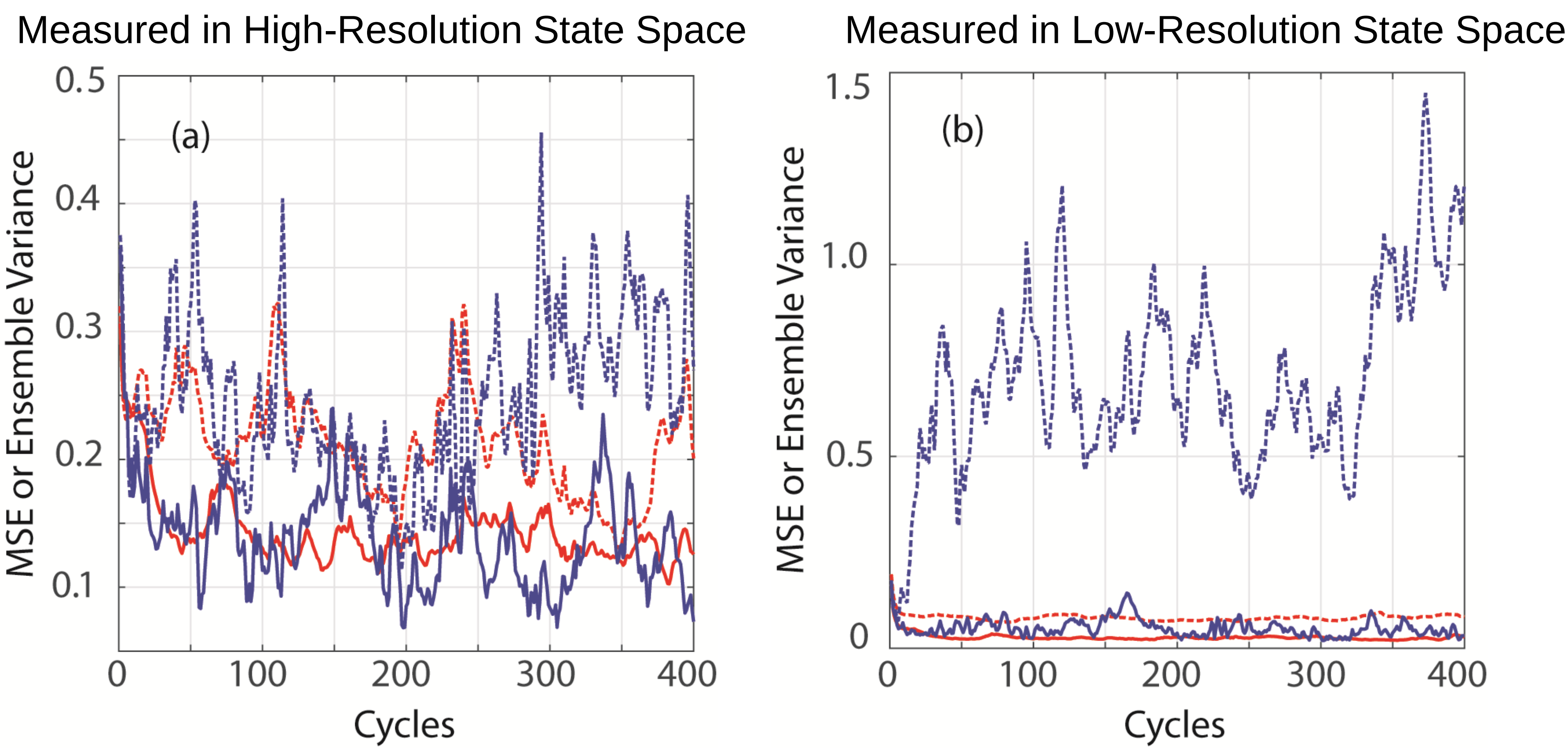
$$\overline{\mathbf{R}}_L^* = \mathbf{R}_{ins} + \mathbf{H}_H\mathbf{P}_H\mathbf{H}_H^T - \mathbf{H}_L\mathbf{P}_L\mathbf{H}_L^T$$

$$\mathbf{v}_L = \mathbf{y} - \mathbf{H}_L\overline{\mathbf{x}}_L$$

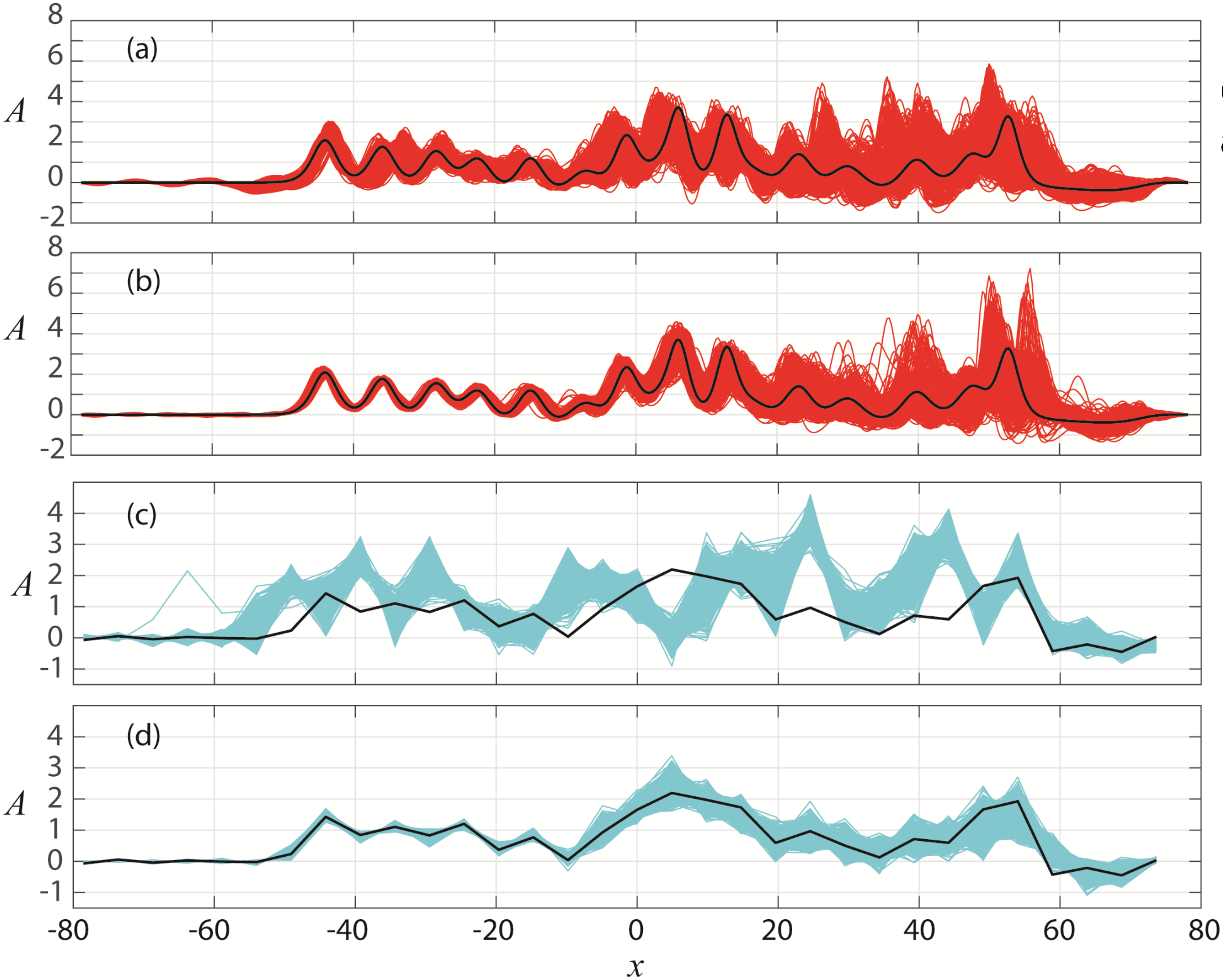
$$\langle \mathbf{v}_L \rangle = \mathbf{H}_H\overline{\mathbf{x}}_H - \mathbf{H}_L\overline{\mathbf{x}}_L$$

Mean-Squared Error and Ensemble Variance

Solid – Multi-Resolution Kalman Technique
Dashed – Contemporary (ad hoc) approach
Blue – Mean-Squared Error (MSE)
Red – Ensemble Variance



Let's see what's happening on the 400th cycle …



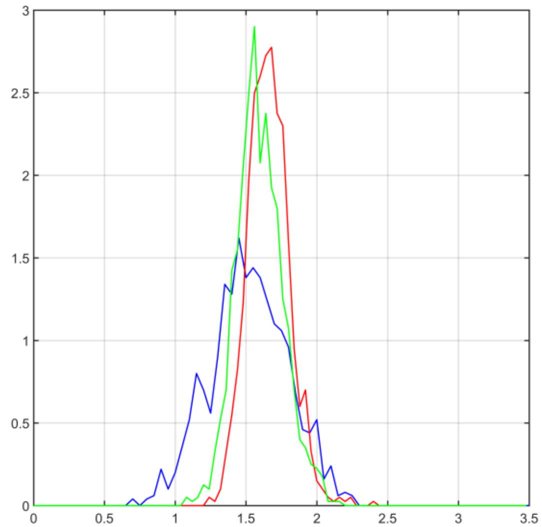
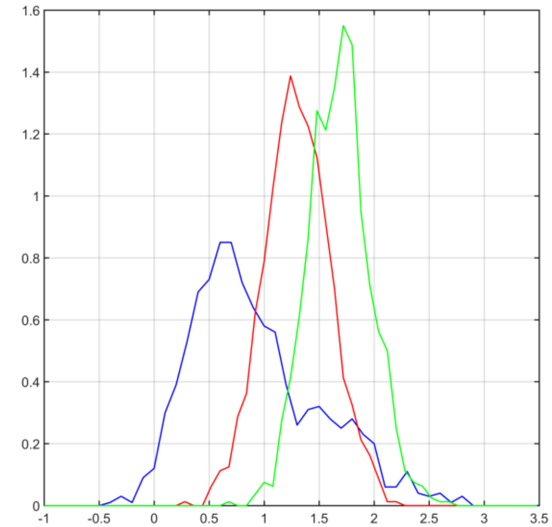
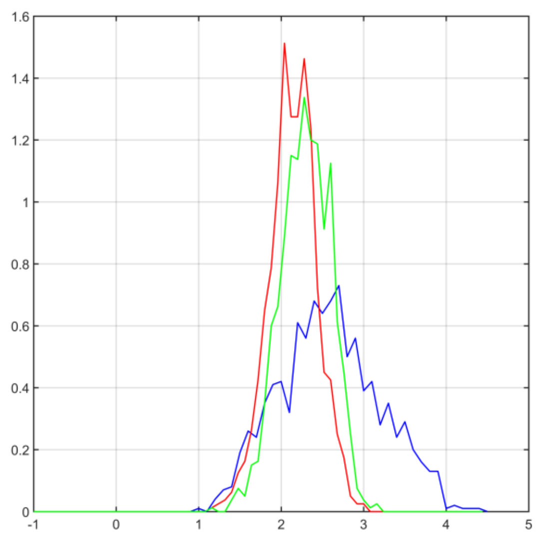
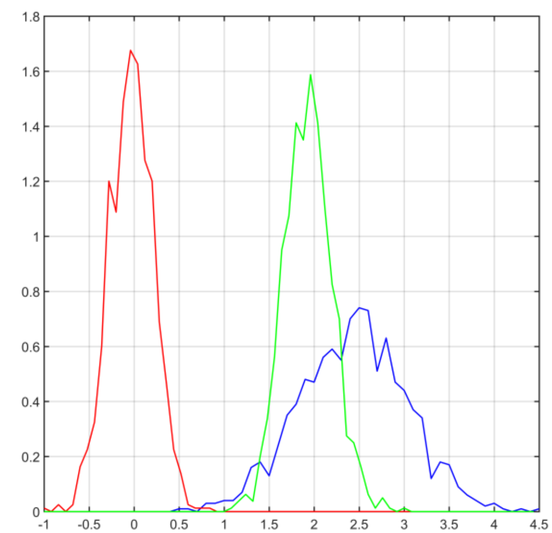
Contemporary Method at High-Resolution

New Method at High-Resolution

Contemporary Method at Low-Resolution

New Method at Low-Resolution

How good was F at t = 400?

	New Method	Ad Hoc Method
x = 0		
x = 5		

Blue – High-Resolution
Red – Low-Resolution
Green – F(High-Res)

Summary and Conclusions

- We described a new framework to understand and account for the coarseness of typical model simulations in the data assimilation process.
- The most important component is the estimation of the correct mapping function from high to low-resolution.
- Presently, we are working on several adaptive methods that update the **F** relationship at each cycle of the data assimilation to account for the new information available.

Hodyss, D. and N. Nichols, 2015: The error of representation: Basic understanding. Tellus, **67A**, 24822.