Autolocalization techniques for ocean modelling using OpenDA

Nils van Velzen, Umer Altaf, Martin Verlaan, Arnold Heemink

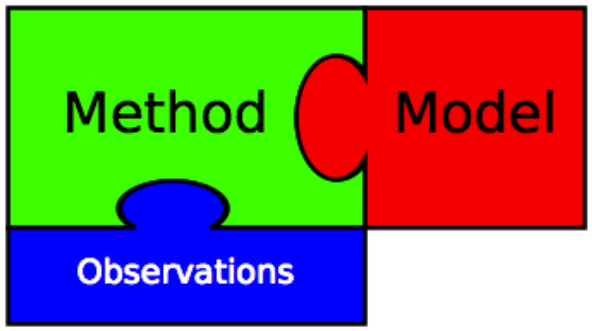
Workshop on Meteorological Sensitivity Analysis and Data Assimilation
1-5 June 2015 Roanoke, West Virginia

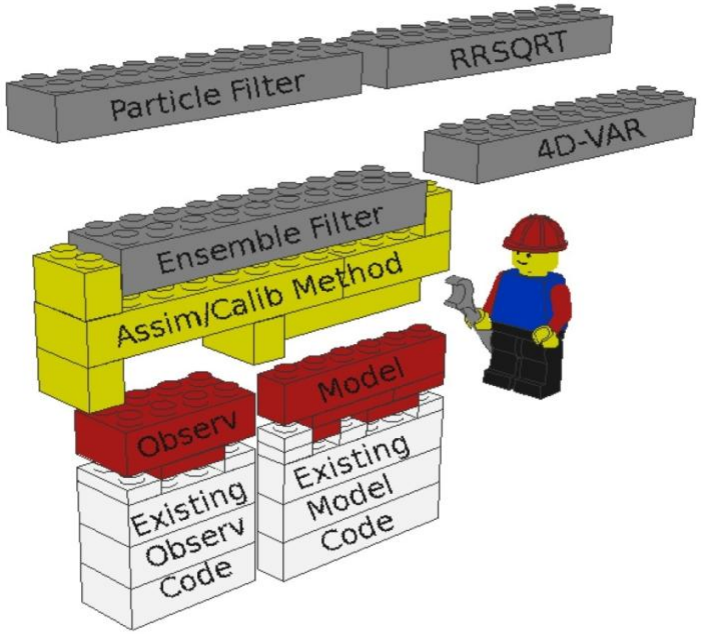
Outline

- OpenDA
- Ensemble methods and localization
- Automatic localization techniques
- Experiments with NEMO model

OpenDA: framework for Data Assimiation

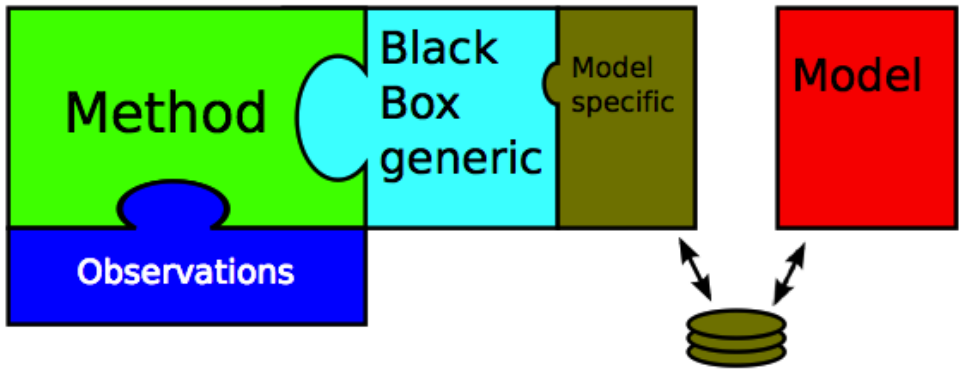
- Content:
  - Set of interfaces that define interactions between components
  - Library of data-assimilation algorithms
  - DA philosophy
  - Building blocks only need to be implemented once





OpenDA: framework for Data Assimiation

- Black box coupling
- Model needs proper restart functionality
- + Easy to implemented
- + No change to model code
- - Restart/file overhead
- - localization might be difficult to implement



Localization

- Ensemble Kalman methods:

$$\hat{\mathbf{P}}_k^f = \frac{1}{n-1}\sum_{i=1}^{n}\left(\mathbf{x}_{k,i}^f - \hat{\mathbf{x}}_k^f\right)\left(\mathbf{x}_{k,i}^f - \hat{\mathbf{x}}_k^f\right)^T$$

$$\mathbf{K}_k = \hat{\mathbf{P}}_k^{xh}\left(\hat{\mathbf{P}}_k^{hh} + \mathbf{R}_k\right)^{-1}$$

- Analyzed ensemble is given by:

$$\mathbf{x}_{k,i}^a = \mathbf{x}_{k,i}^f + \mathbf{K}_k\left(\mathbf{y}_{k,i}^s - \mathcal{H}_k(\mathbf{x}_{k,i}^f)\right), i = 1, 2, ..., n,$$

Localization

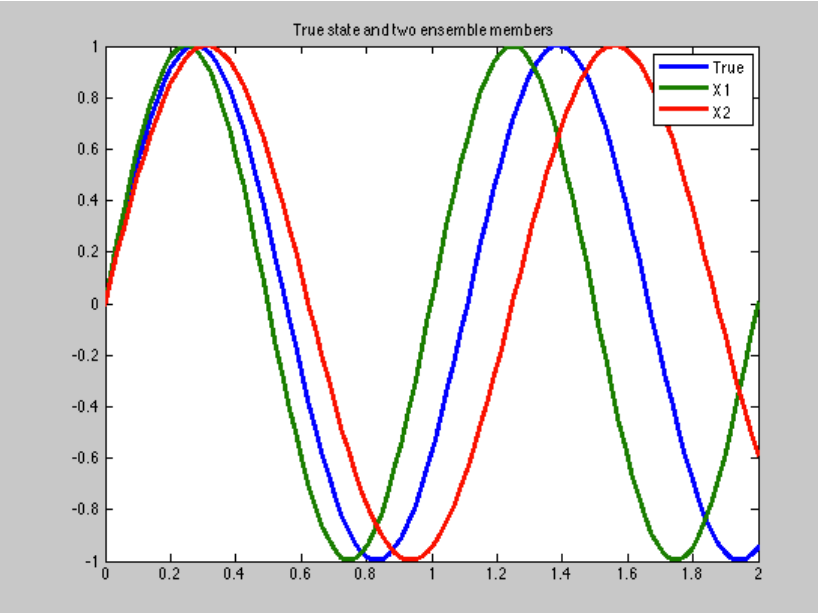
- Covariance: low rank approximation compared to model state:

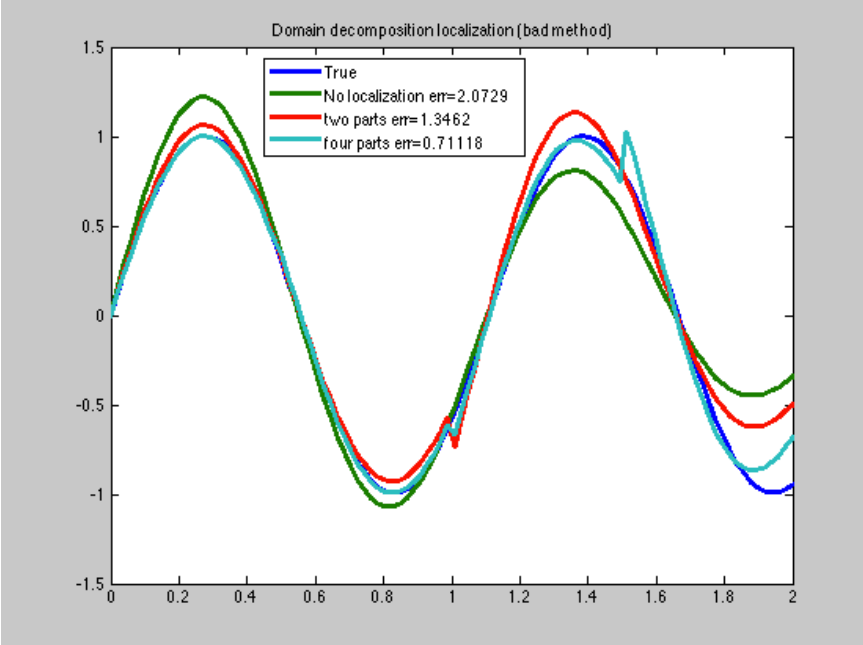
$$\mathcal{O}(10-100)$$

- Structural underestimation of errors
- Spurious correlations

Localization

- Ensemble is set of "basis" vectors for representing errors and update





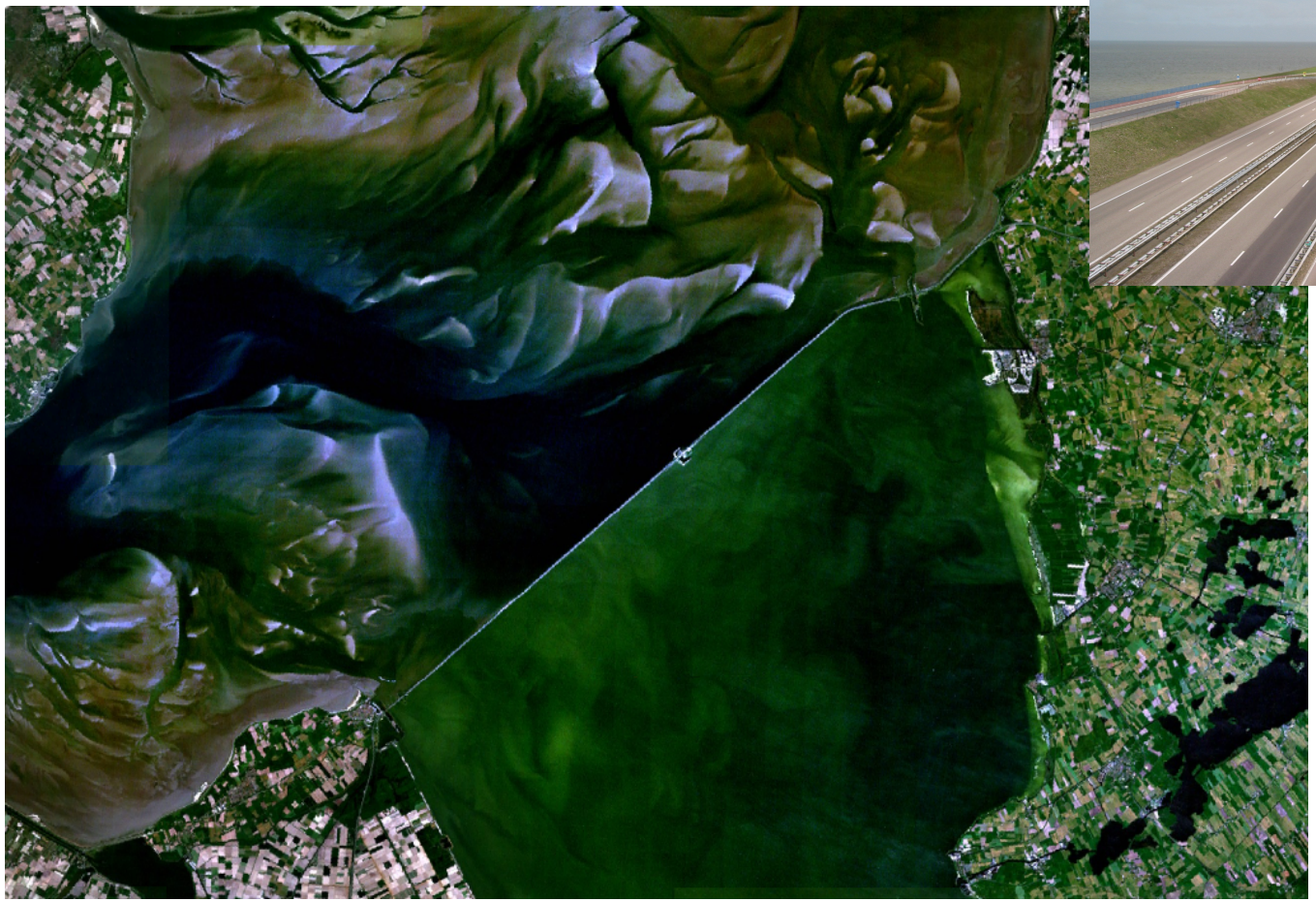
Localization

- Localization of gain matrix

$$\mathbf{K}^{Loc} = \mathbf{K} \circ \beta_{xy}$$

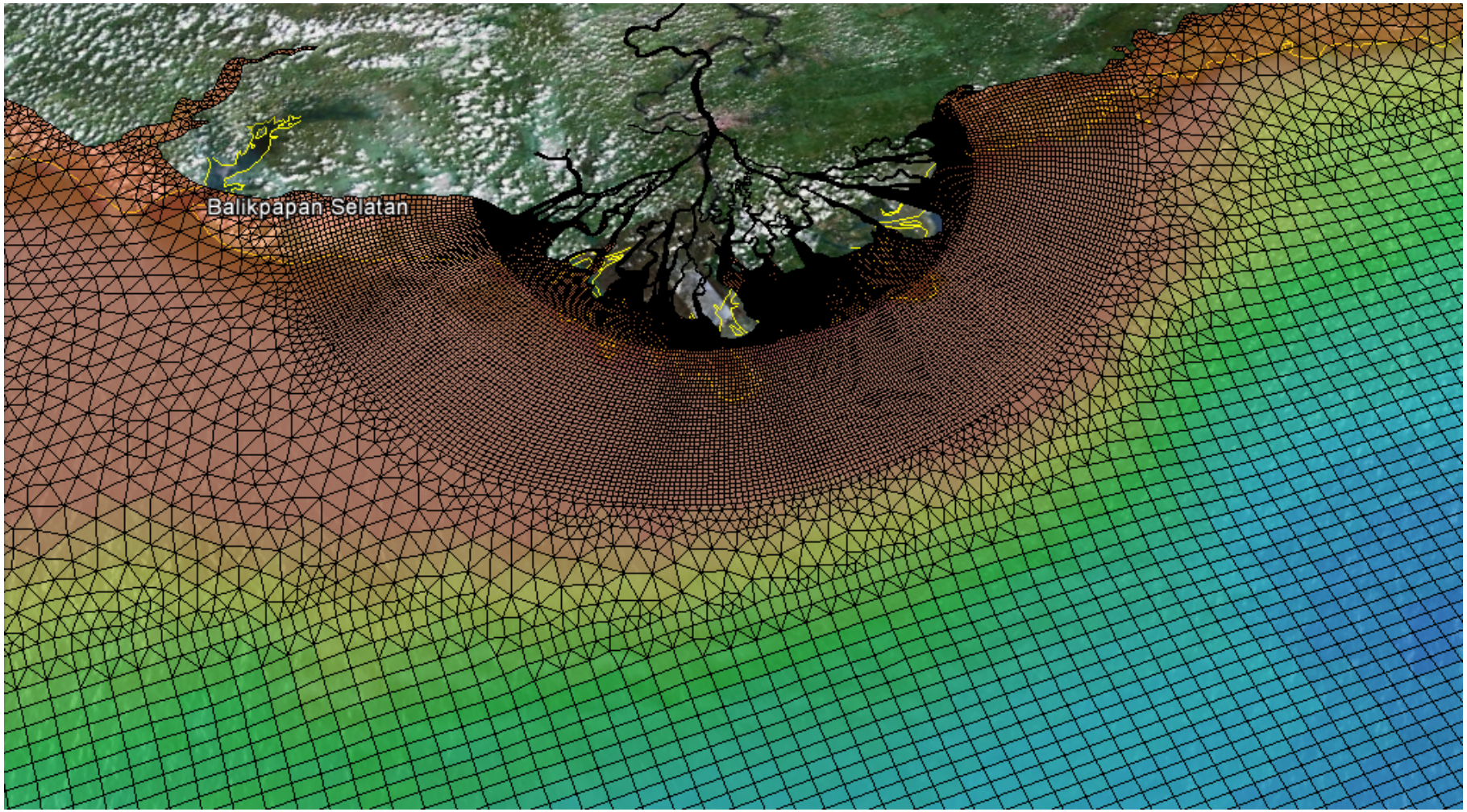
- Increases "dimension" of the update
- How to determine β_{xy} ?
- Is distance a good measure?
- What is the location of state variables

Localization



Source:wikipedia

Localization



Source:deltares

Auto Localization

- Anderson 2004:
  - Define N_g groups of N_e ensembles
  - Each group has its own gain matrix $\mathbf{K}_j, j = 1, .., N_g$

  - Assume elements from gain matrices are drawn from distribution containing the "true" gain matrix.
  - Minimize:

$$\sqrt{\sum_{j=1}^{N_g} \sum_{k=1, k \neq j}^{N_g} (\alpha_i k_{i,k} - k_{i,j}).}$$

  - Too much computational work for real time application
  - Good for investigating good selections of weights

Auto Localization

- Zhang and Oliver 2011:
  - Create ensemble (N_B) of gain matrices based on the initial ensemble using bootstrapping
  - Estimation of variance of each element of the gain matrix

$$\hat{\sigma}^2_{k_{i,j}} = \frac{\sum_{m=1}^{N_B} \left(\hat{k}^*_{i,j,m} - \bar{k}_{i,j} \right)^2}{N_B}$$

  - Ratio between mean and variance

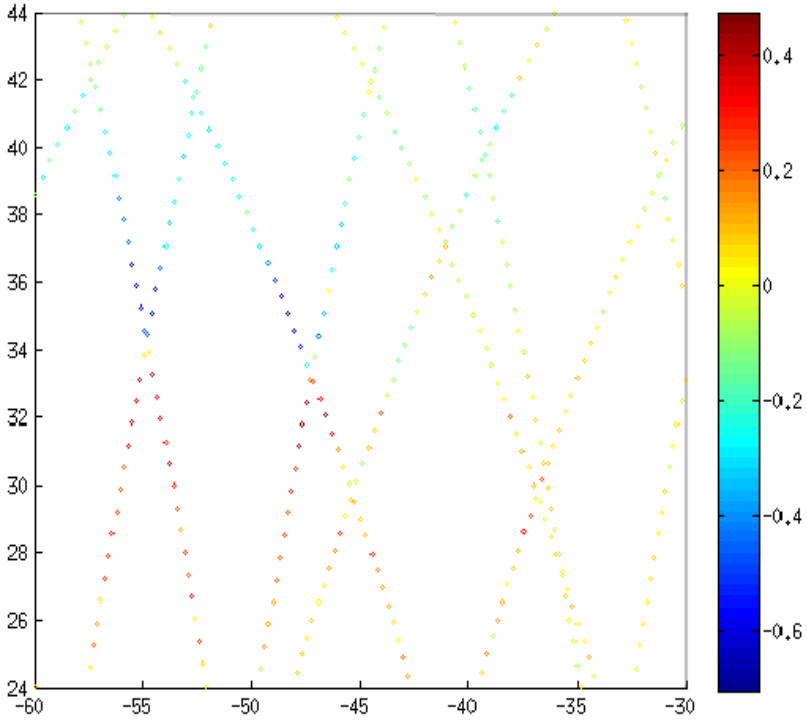
$$\hat{C}^2_{v_{i,j}} = \frac{\hat{\sigma}^2_{\theta_{i,j}}}{\bar{k}^2_{i,j}}$$

  - Localization weights

$$\beta_{i,j} = \frac{1}{1 + \left(1 + 1/\sigma^2_\alpha \hat{C}^2_{v_{i,j}}\right)}$$

  - Balance parameter (Zhang and Oliver 2010) $\sigma^2_\alpha = 0.36$

Experiment Setup



Medium Case Benchmark

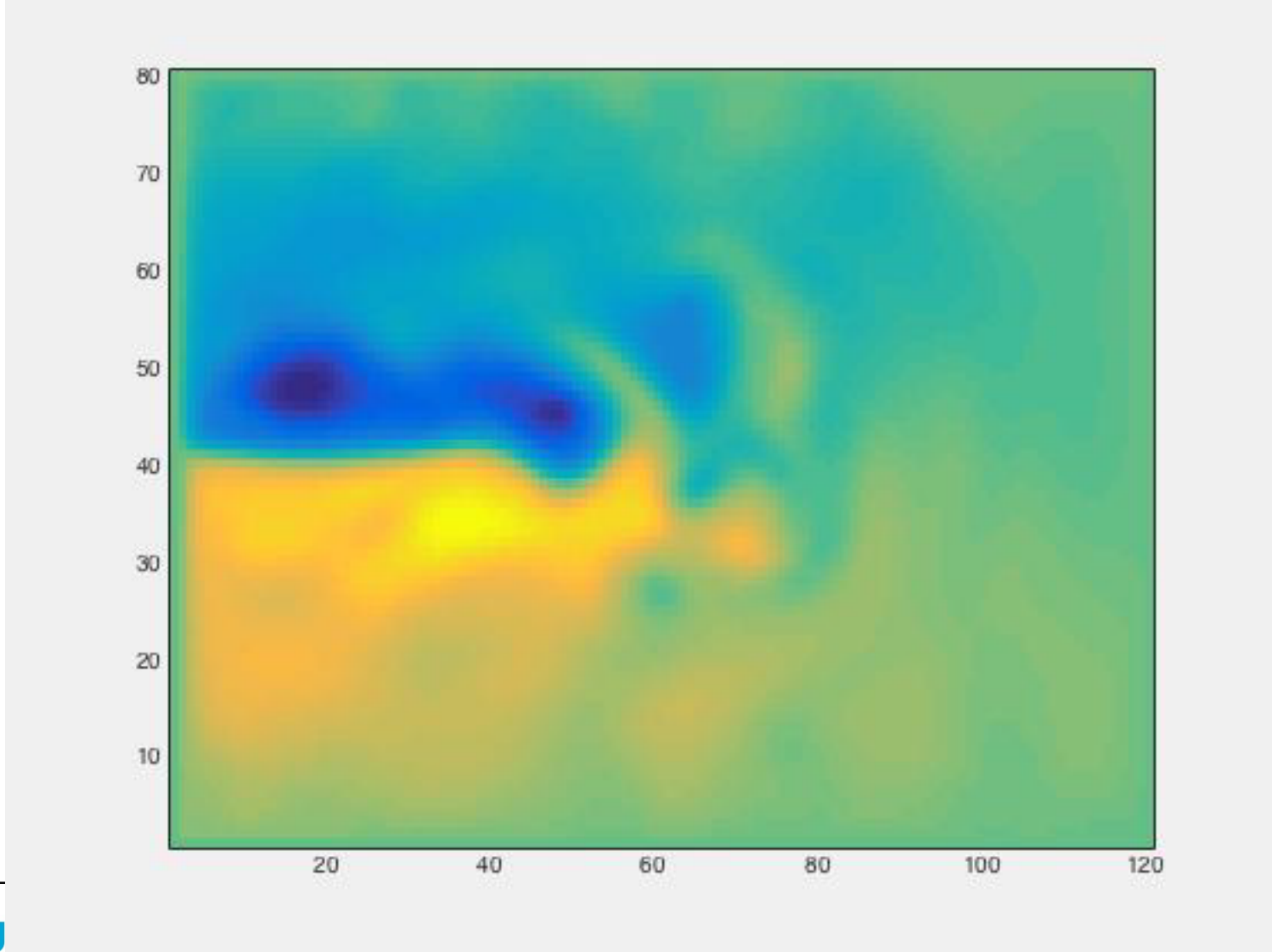
- Free run 40 years.

- State vector includes:
  (ub,vb,tb,sshb,un,vn,tn,sshn)

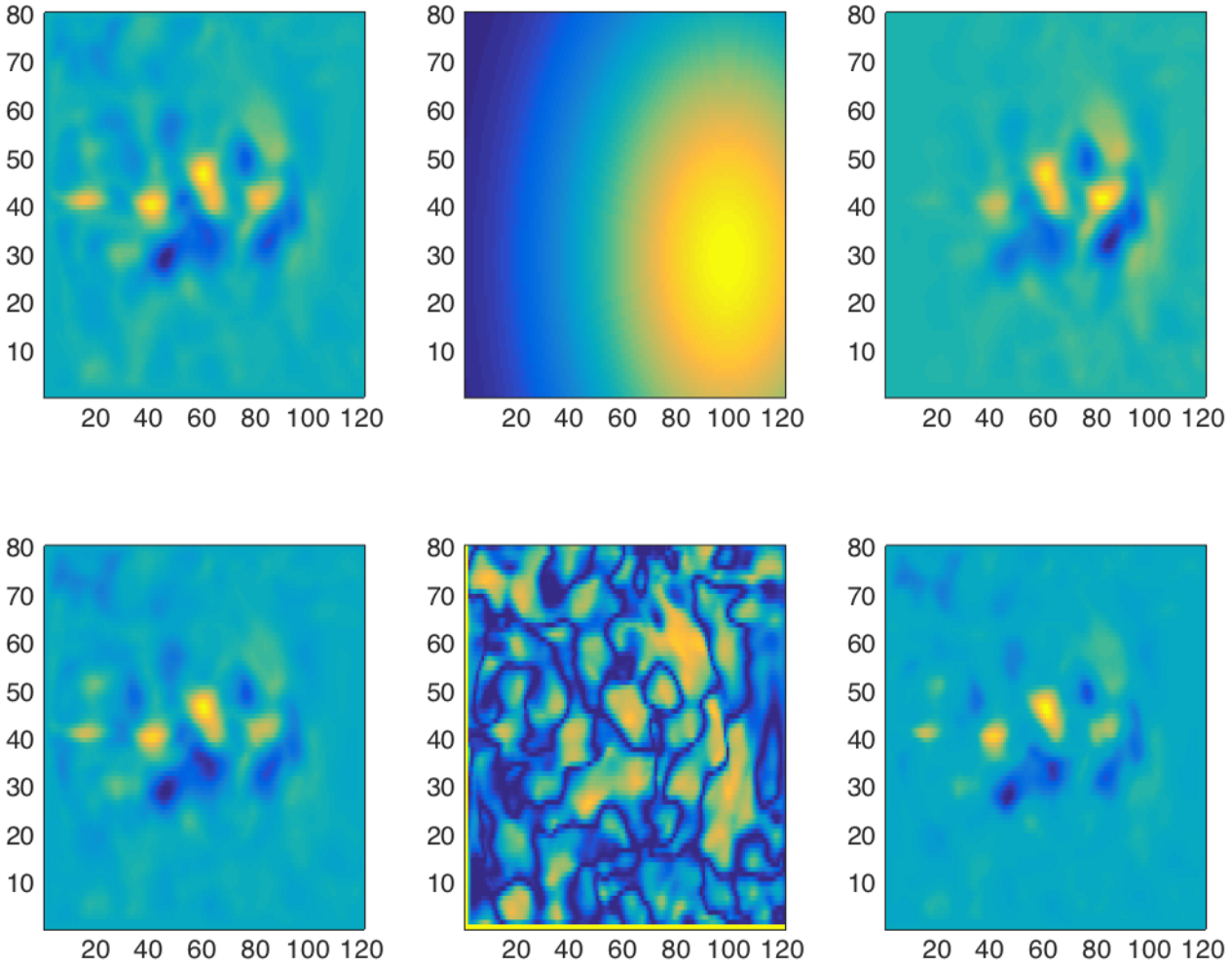
Ensemble Kalman filter

- 30/100 Ensemble members
- SSH observations. ENVISAT, Jason-1
- Analysis: 2 days

Experiment Setup



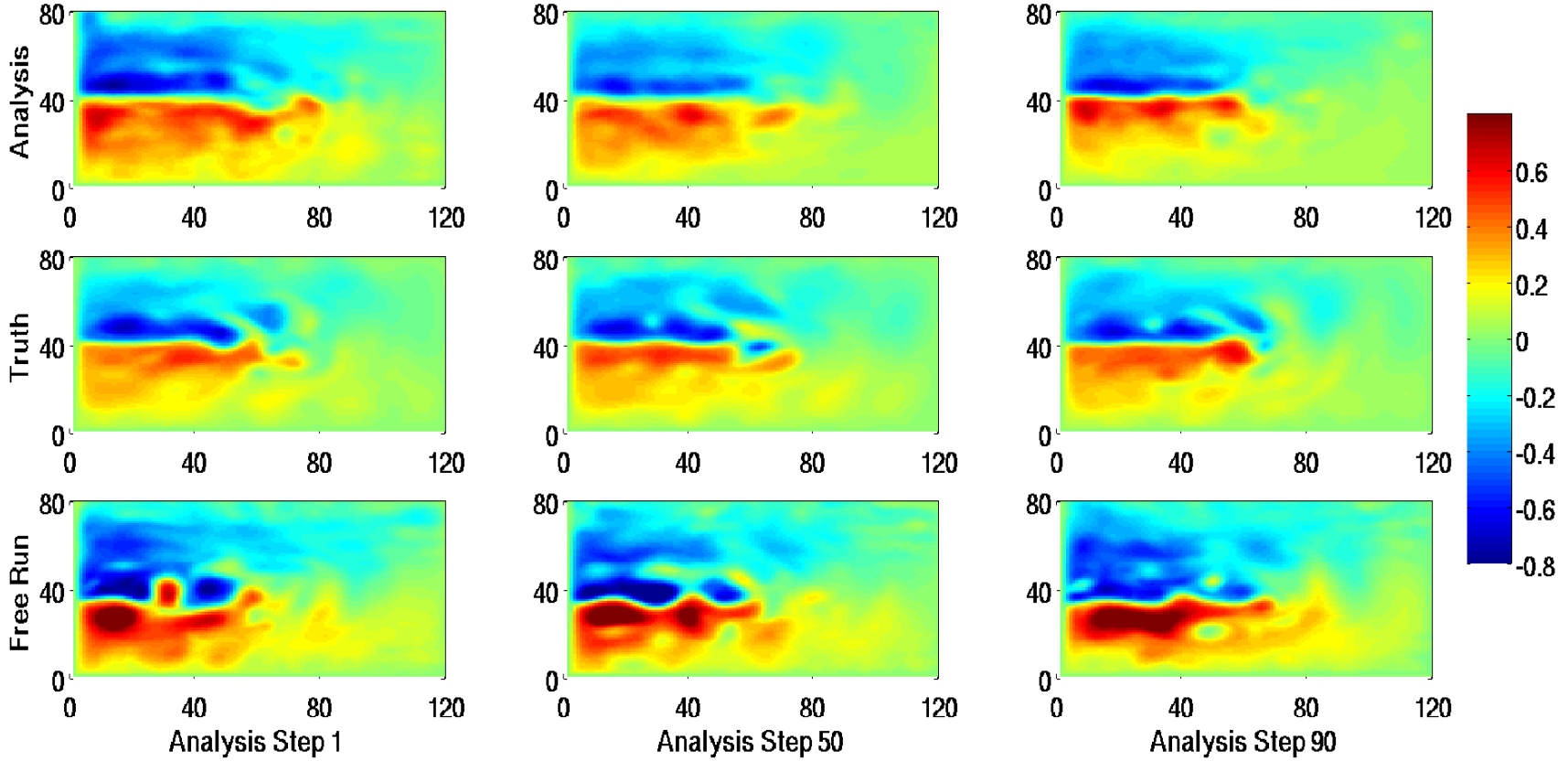
Results

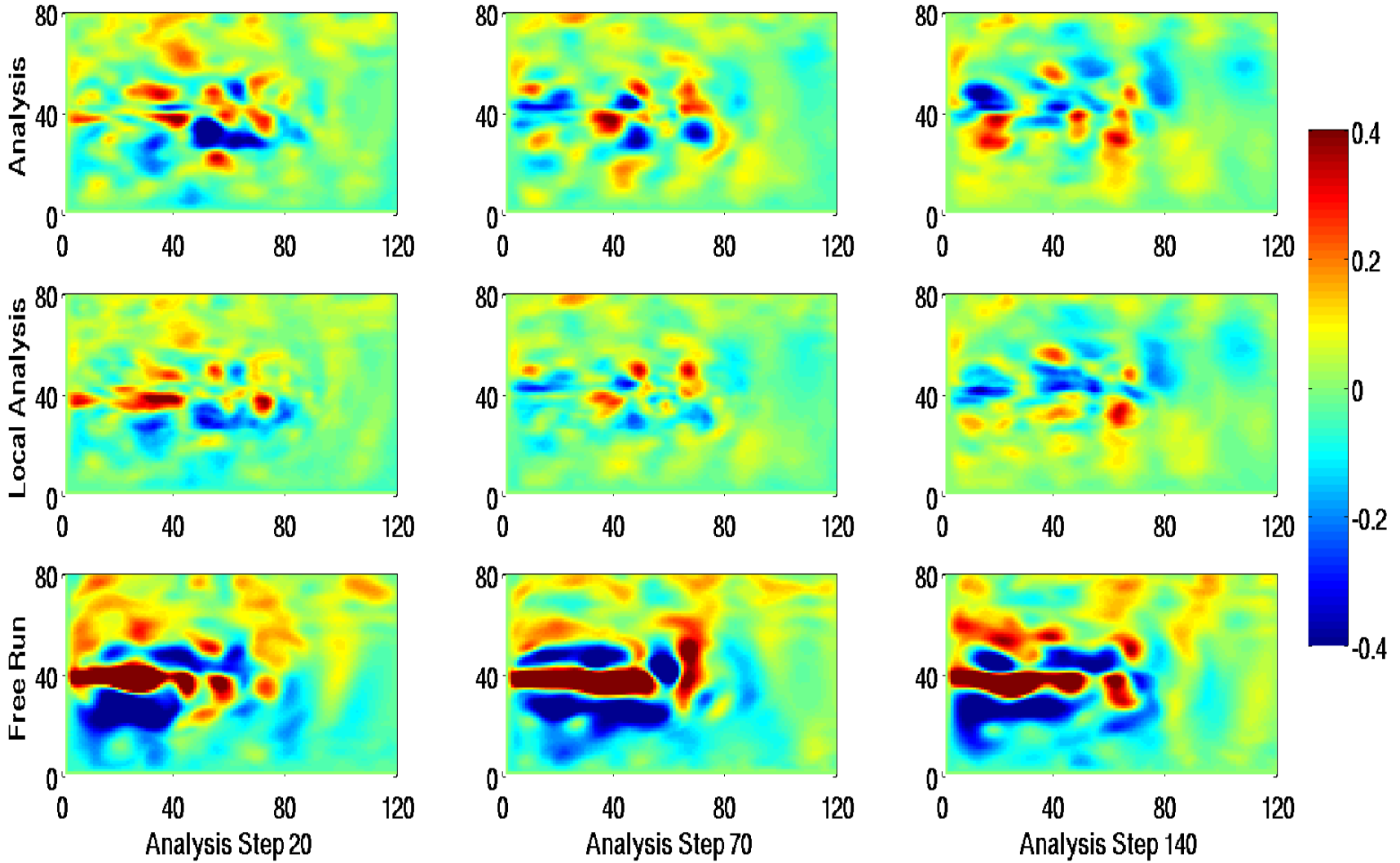


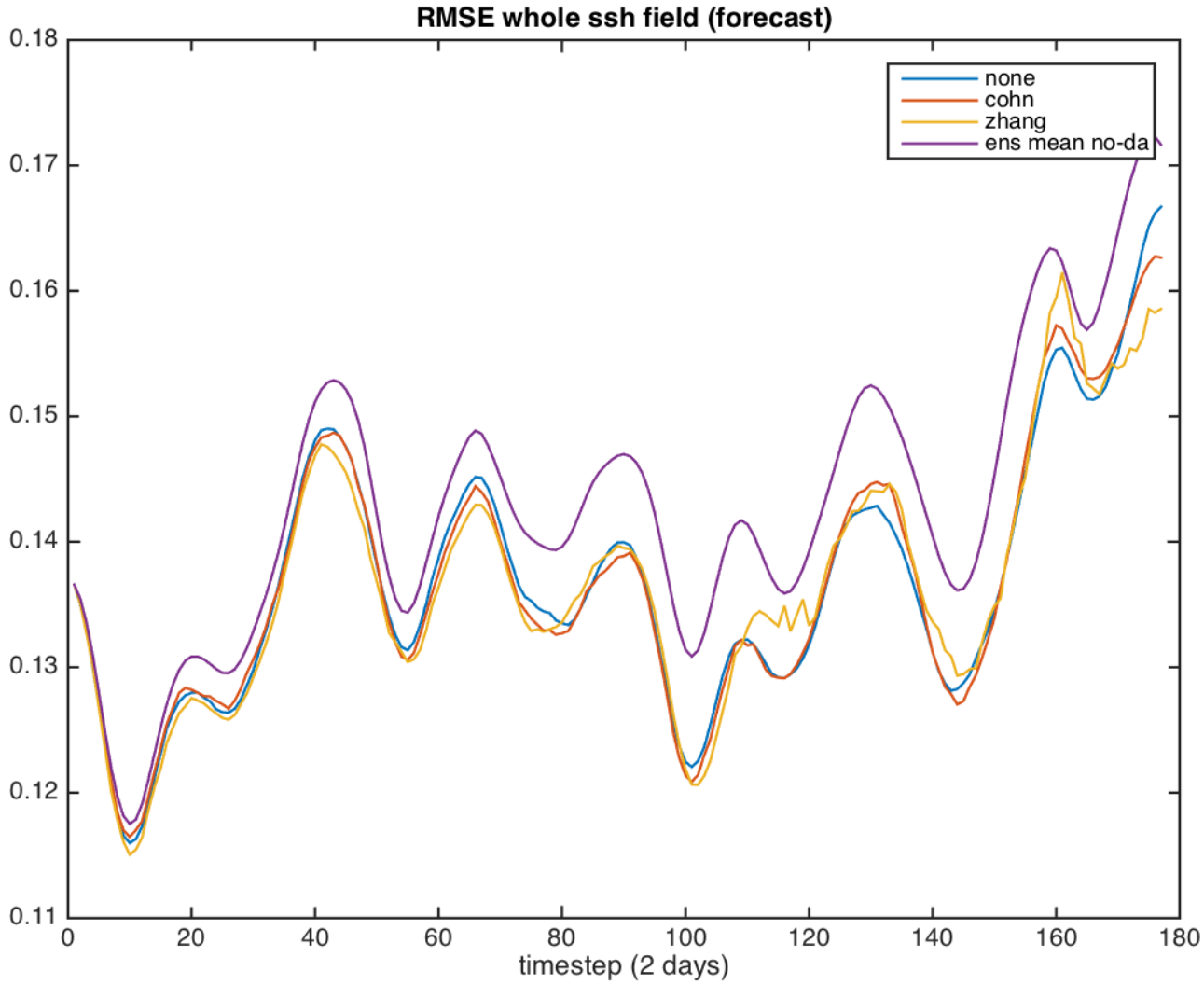
Results

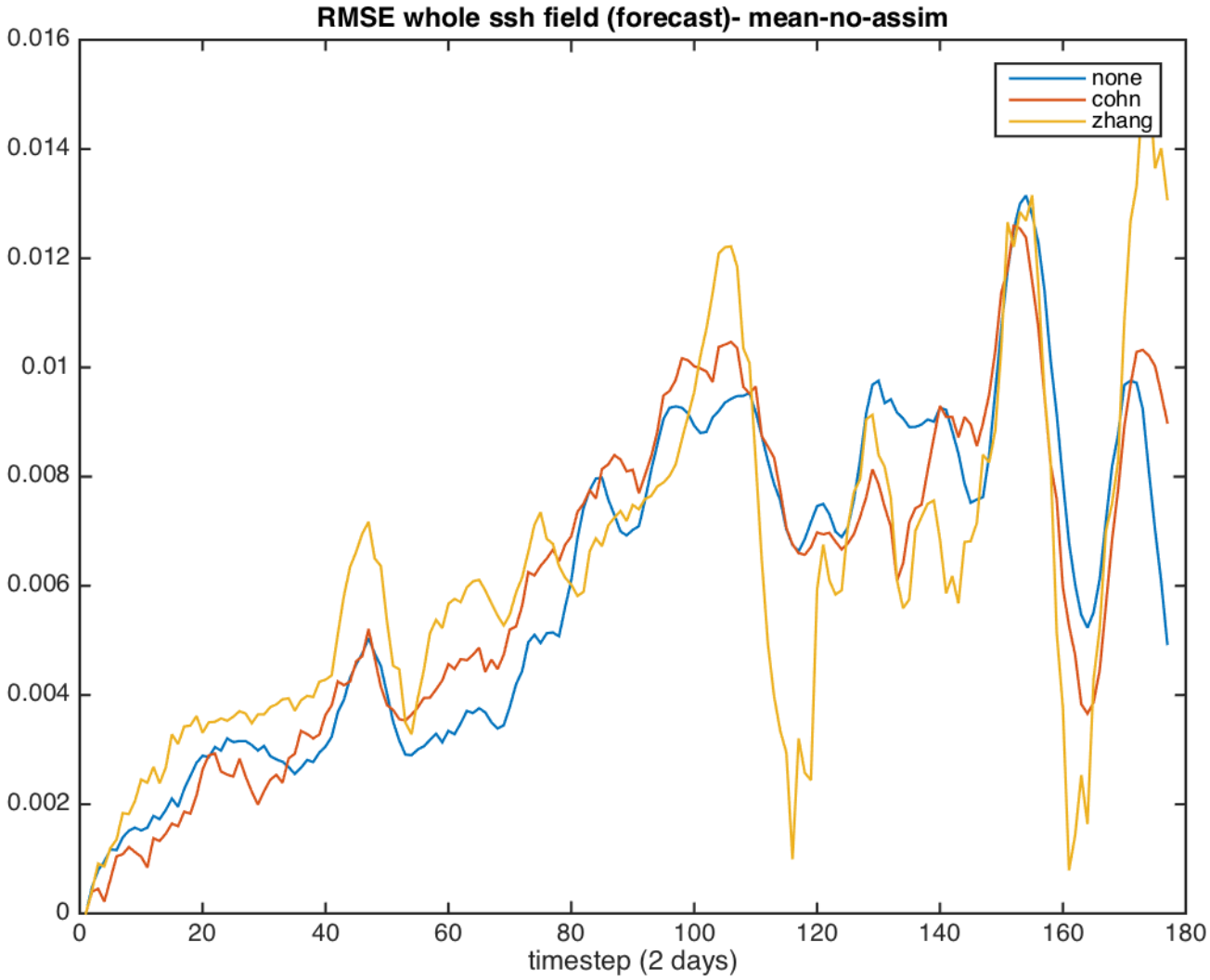
Medium Case Benchmark

- Assimilating SSH every 2 days
- Assimilation period: 1 year



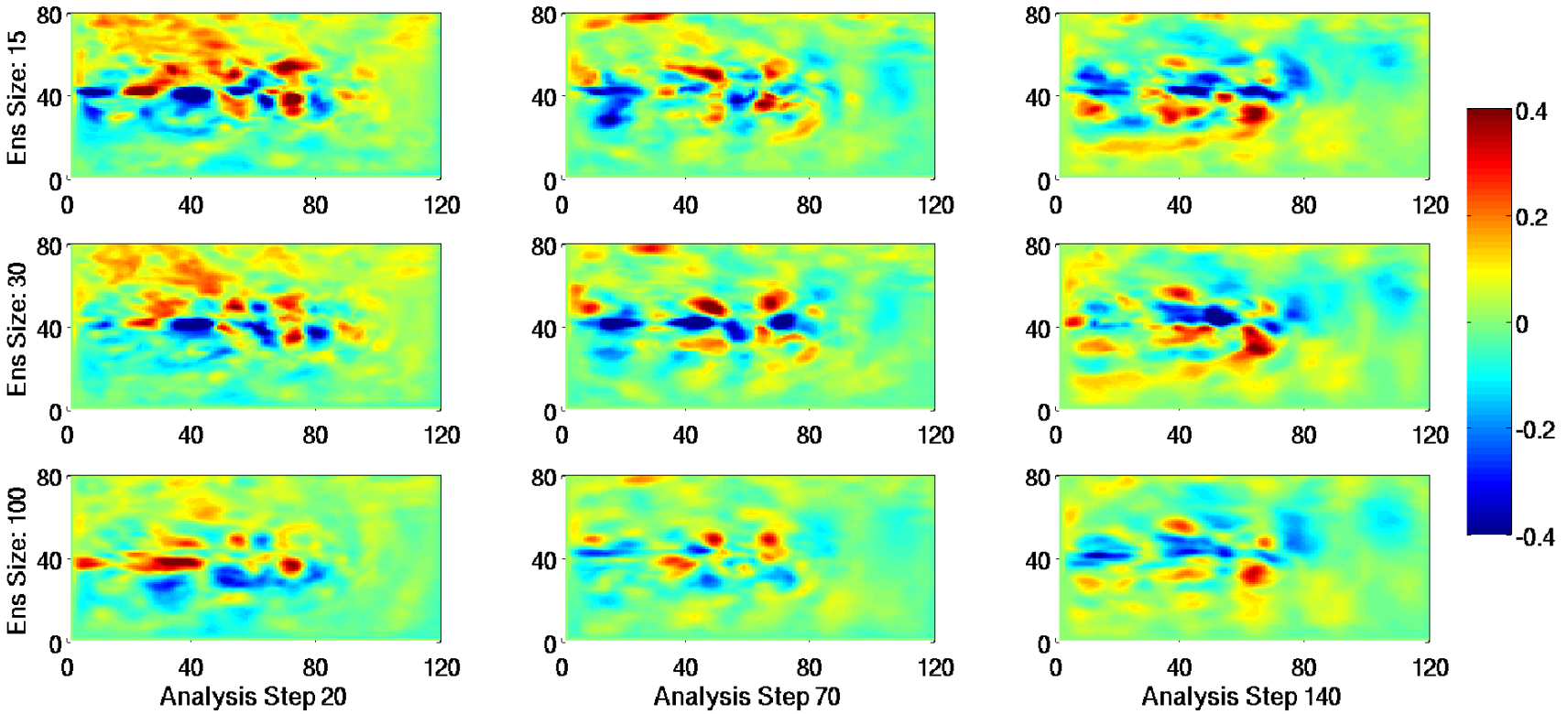

Analysis
Local Analysis
Free Run
Analysis Step 20
Analysis Step 70
Analysis Step 140
0.4
0.2
0
-0.2
-0.4


RMSE whole ssh field (forecast)
0.18
0.17
0.16
0.15
0.14
0.13
0.12
0.11
0
20
40
60
80
100
120
140
160
180
timestep (2 days)
none
cohn
zhang
ens mean no-da


RMSE whole ssh field (forecast)- mean-no-assim
0.016
0.014
0.012
0.01
0.008
0.006
0.004
0.002
0
0
20
40
60
80
100
120
140
160
180
timestep (2 days)
none
cohn
zhang

Medium Case Benchmark Results

- Localization improves results
- Limitation: No. of observation



Summary

- NEMO-OpenDA framework is established.
- Localization is not trivial to implement
- Auto localization seems promising
- Next steps:
  - More experimenting
  - Combine auto localization with normal localization

Acknowledgements:

This work is supported by SANGOMA a European FP7-SPACE-2011 project, Grant 283580.

This work was carried out with the support of the Danish Council for Strategic Research as part of the project "HydroCast e Hydrological Forecasting and Data Assimilation", Contract No. 11- 116880 (http://hydrocast.dhigroup.com/).