Some developments in Data Assimilation at Delft

Arnold Heemink

Delft University of Technology

Joint work with Mohammad Umar Altaf, Nils van Velzen and Martin Verlaan

Overview

- The Mathematical Physics group in Delft
- Adjoint-free variational data assimilation:
  An ensemble approach to variational data assimilation

Delft Institute of Applied Mathematics

In total 60 fte staff, 45 PhD students, 6 Sections:

- Analysis
- Probability
- Statistics
- Mathematical Physics
- Numerical Analysis
- Optimization

Section Mathematical Physics

15 fte staff, 18 PhD students, research topics:

- Partial differential equations
- Inverse modeling and data assimilation
- High performance computing

Research topic
Inverse modeling and Data assimilation

- Staff members: Arnold Heemink, Martin Verlaan, Remus Hanea
- Post-docs: Nils van Velzen, Muhammed Umer Altaf
- 6 PhD students
- Main applications in reservoir modeling, coastal sea modeling and air pollution
- OpenDA is used as programming environment
- Corporation with Deltares, VORTech, TNO, Shell

Model reduction with application to variational data assimilation: An ensemble approach to variational data assimilation

(based on M.U. Altaf, M Verlaan, A.W. Heemink, International Journal on Multi Scale Computational Engineering, 2009)

Strong constraint variational data assimilation

State space model

The (non linear) physics:

$$X_{k+1} = f(X_k, p, k)$$

where X is the state, p is vector of uncertain parameters, f represents the (numerical) model

The measurements:

$$Z_k = M(k) X_k$$

where M is the measurement matrix

Strong constraint variational data assimilation

If we solve the uncoupled system:

$$X_{k+1} = f(X_k, p, k)$$
$$v_k = F(k)^T v_{k+1} + M(k)^T R^{-1}(Z_k - M(k)X_k)$$
$$X_0 = x_0, \quad v_K = 0$$

where F(k) is the tangent linear model,
the gradient of the criterion can be computed by:

$$\frac{\partial J}{\partial p} = -\sum_{k=0}^{k=K} v_k^T \frac{\partial f}{\partial p}$$

Very efficient in combination with a gradient-based optimization scheme.
BUT: we need the adjoint implementation!

A POD model reduction approach to data assimilation: The linear case

Consider the q dimensional sub space:

$$P = [\ldots p_j \ldots]$$

And project the original model onto this sub space:

$$r_{k+1} = [P^T F(k) P] r_k$$
$$Z_k = [M(k) P] r_k + v_k$$

We now have an explicit low dimensional (approximate) system description of the model variations including its adjoint

The sub space can be determined by computing the EOF (Empirical Orthogonal Functions) of an ensemble of model simulations

The nonlinear case

We now have to determine:

$$r_{k+1} = [P^T \left[\frac{\partial f_i}{\partial x_j} \right] P] r_k$$

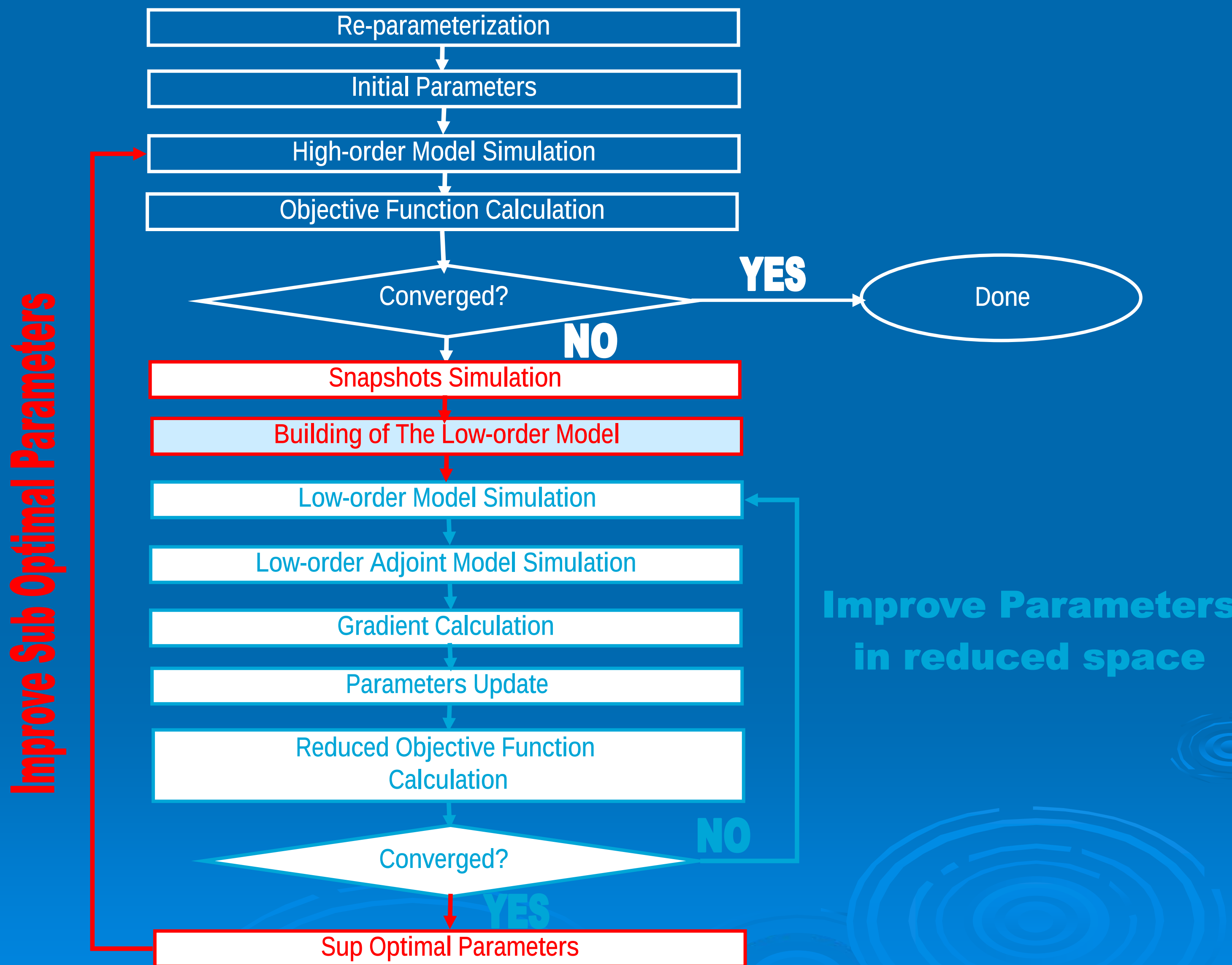
For every column l of P we have

$$\left[\frac{\partial f_i}{\partial x_j} \right] p_l \approx \tfrac{1}{\varepsilon} (f(X_{k-1}^a + \varepsilon p_l) - f(X_{k-1}^a))$$

We do not need the tangent linear approximation!

Parameter estimation problems

- Parameters are included in the original state vector
- The parameter space is not reduced so the reduced state has dimension q+N, where N is the number of parameters



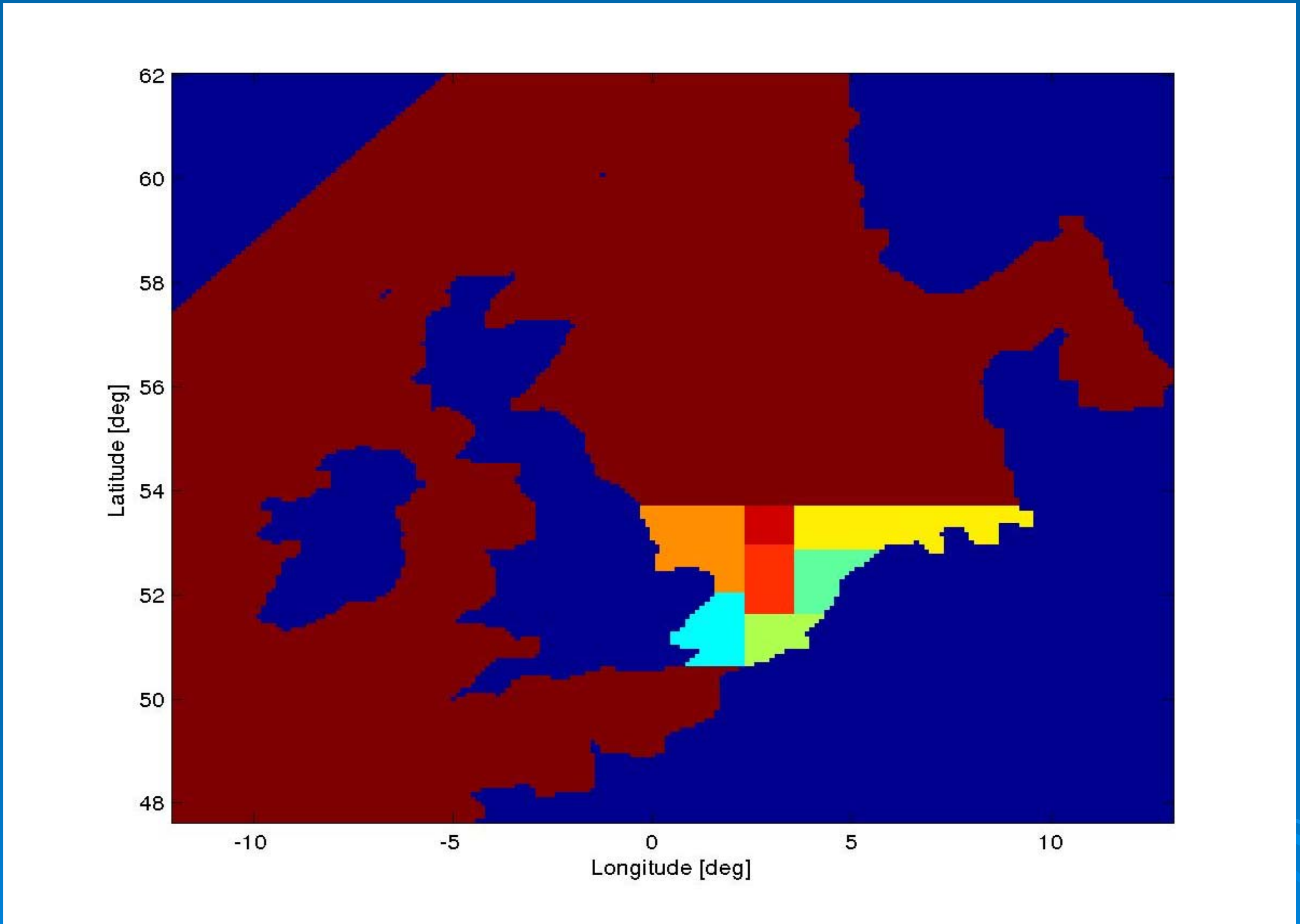
Re-parameterization
Initial Parameters
High-order Model Simulation
Objective Function Calculation
Converged?
YES
Done
NO
Snapshots Simulation
Building of The Low-order Model
Low-order Model Simulation
Low-order Adjoint Model Simulation
Gradient Calculation
Parameters Update
Reduced Objective Function Calculation
Converged?
NO
YES
Sup Optimal Parameters
Improve Sub Optimal Parameters
Improve Parameters in reduced space

Some remarks

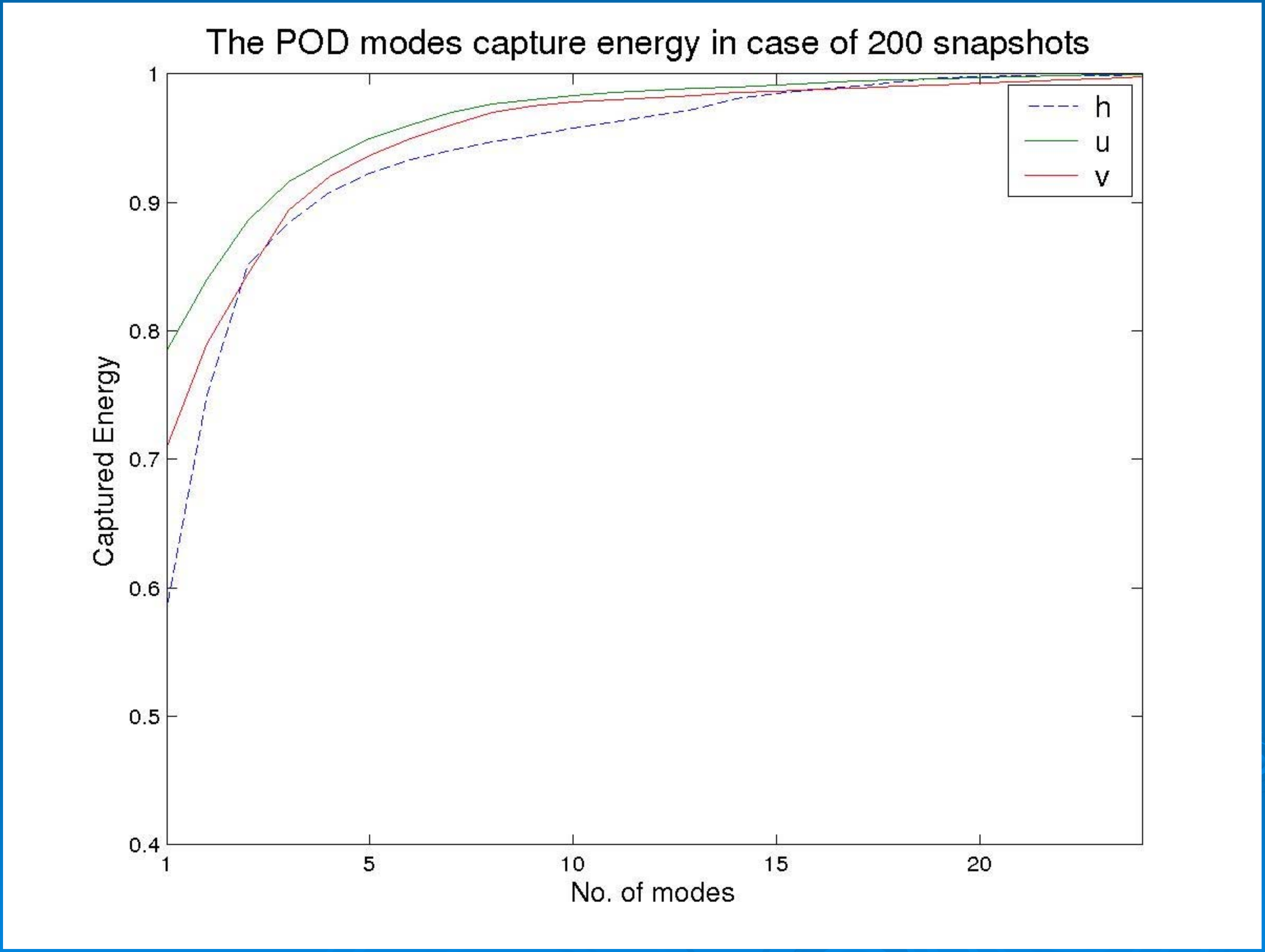
- Very efficient in case the simulation period of the ensemble of model simulation is very small compared to the calibration period
- For some iterations the reduced model can be the same and only the residuals have to be updated
- Not very sensitive to local minima
- Will not find the exact minimum of the original problem
- Balanced truncation takes in account the amount of measurement information that is available: It is not efficient to include model in the model that are not observed. However this requires the adjoint.

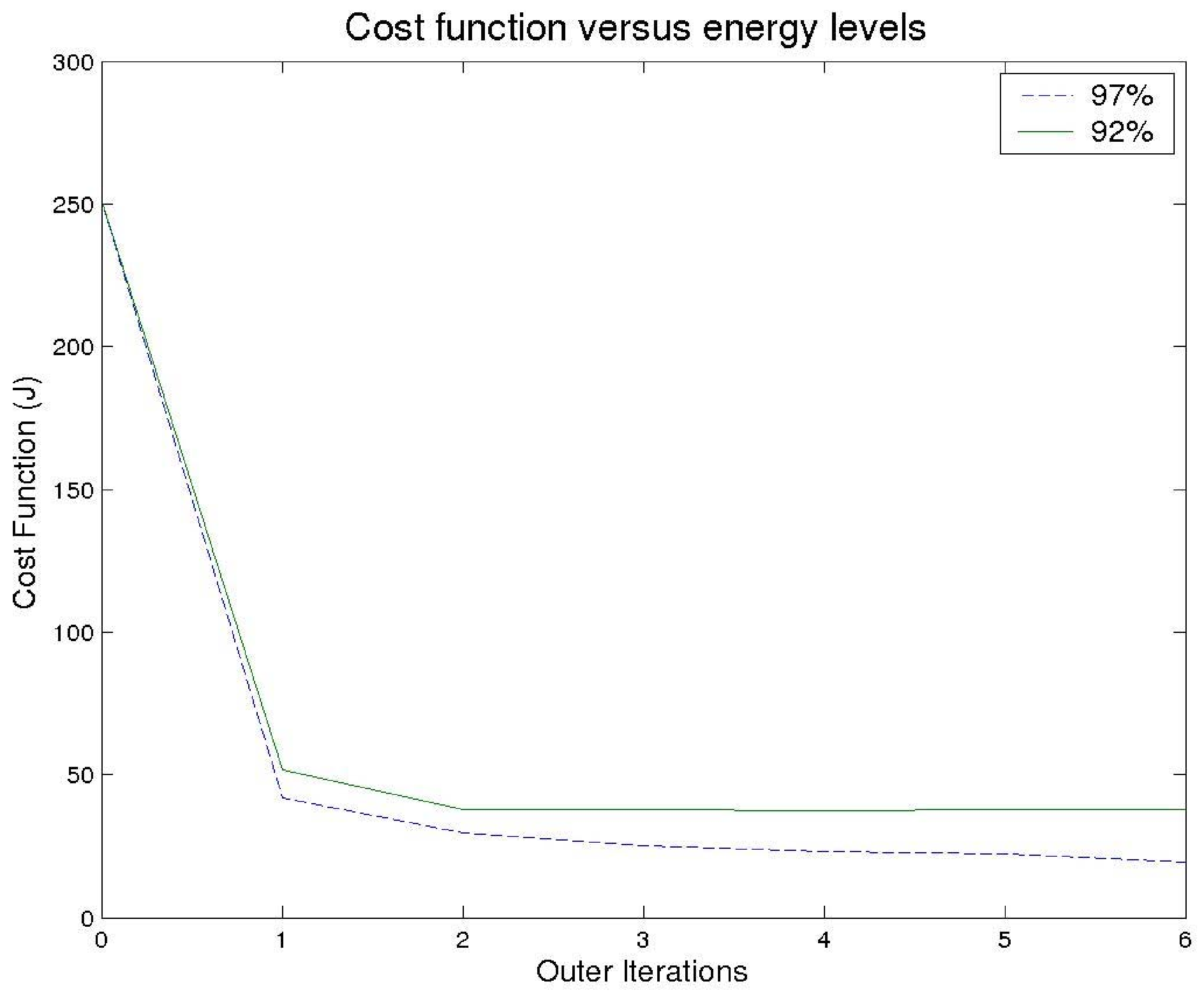
Application to the calibration of a numerical tidal model

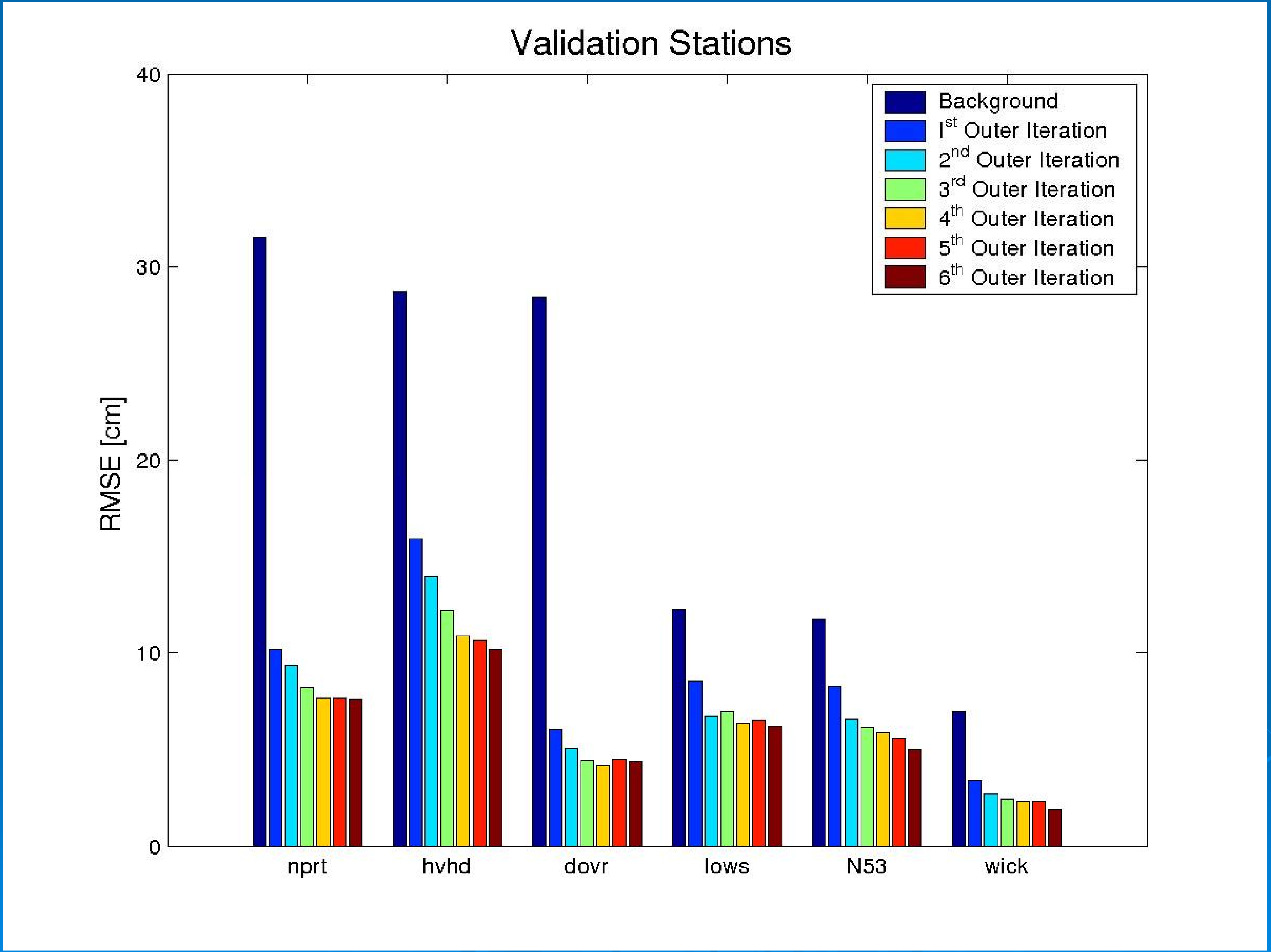
First some experiments with generated data (noise free)

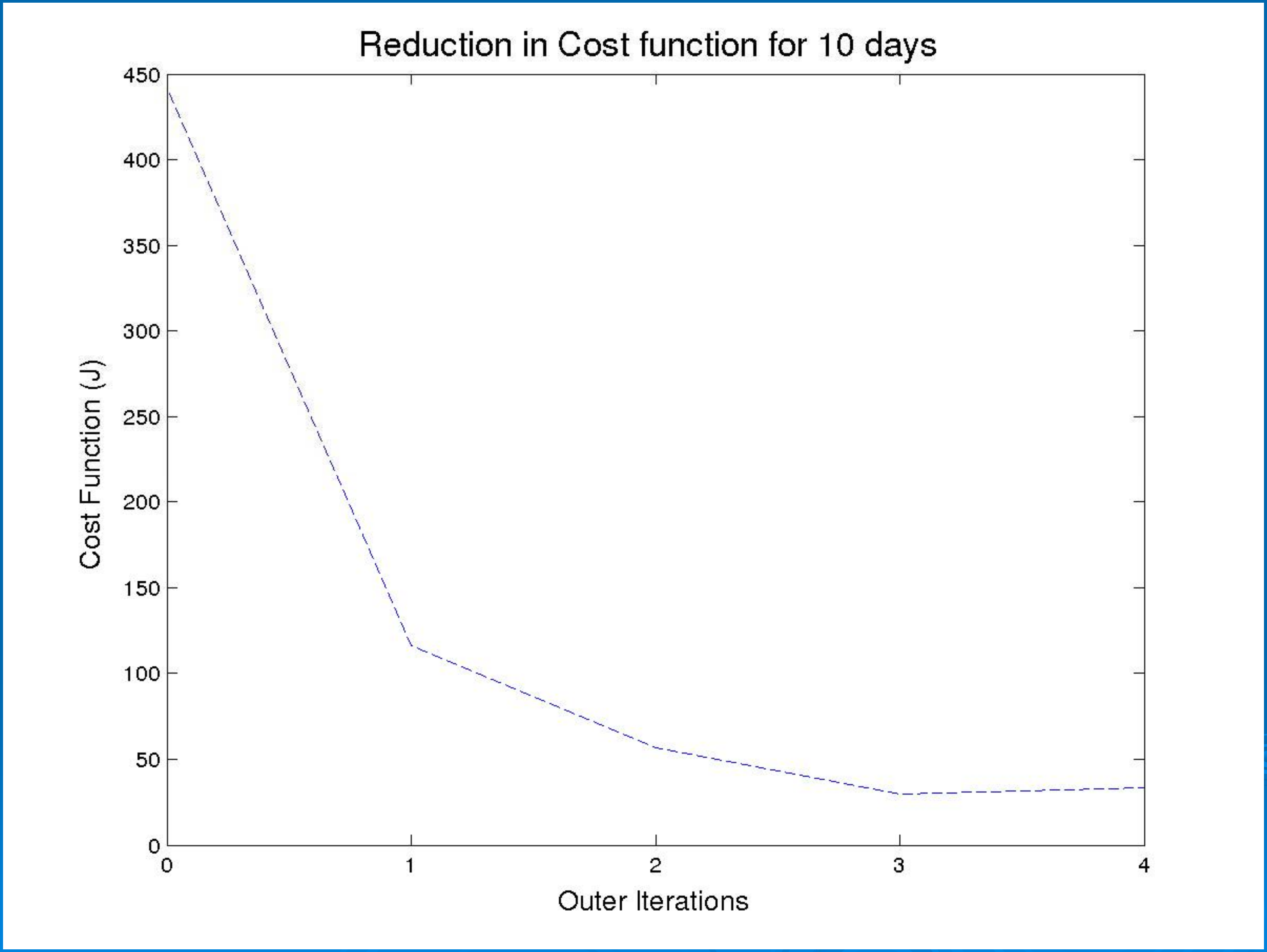


Colors indicate depth parameters

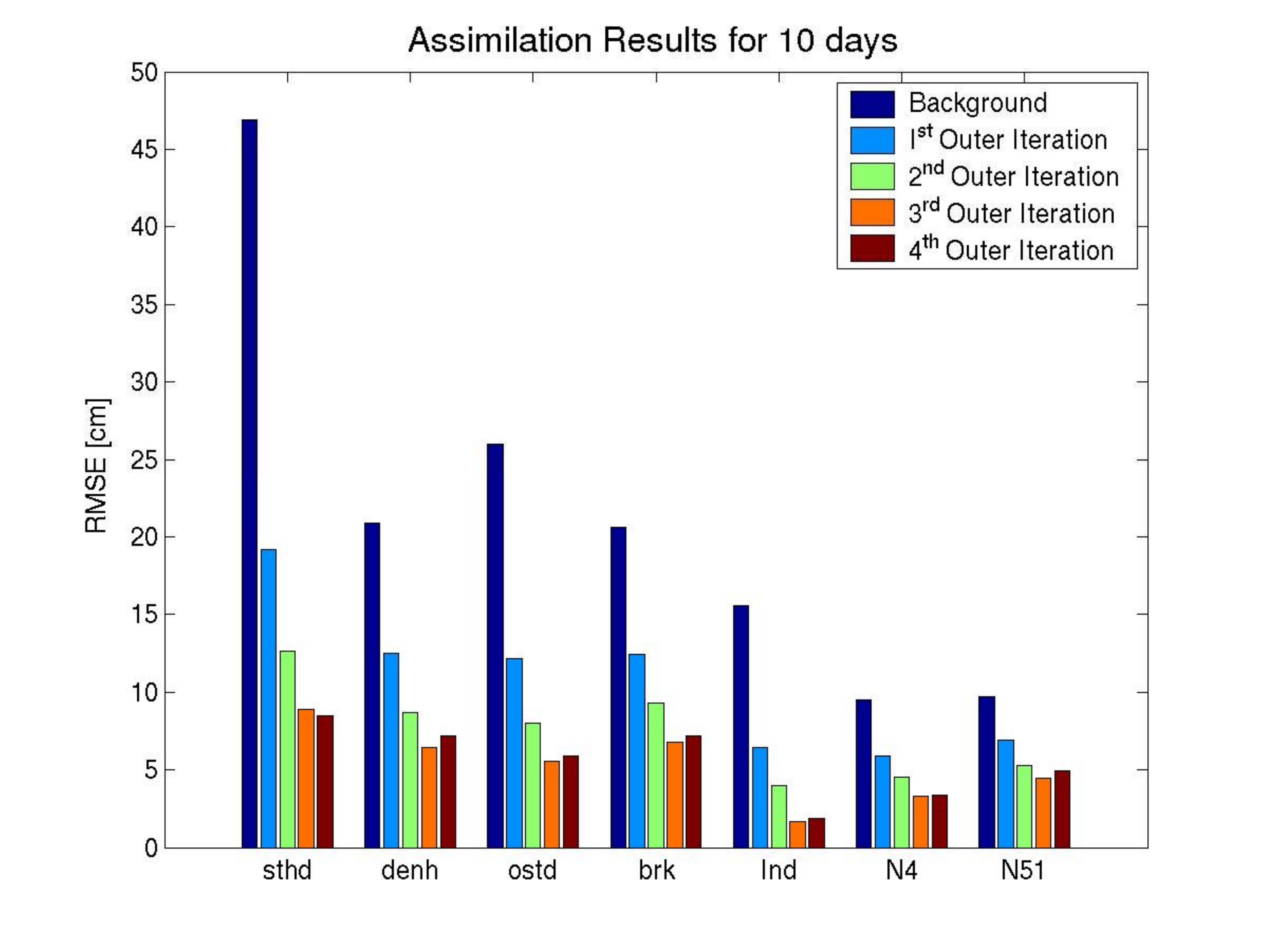

The POD modes capture energy in case of 200 snapshots
Captured Energy
No. of modes
1
0.9
0.8
0.7
0.6
0.5
0.4
1
5
10
15
20
h
u
v

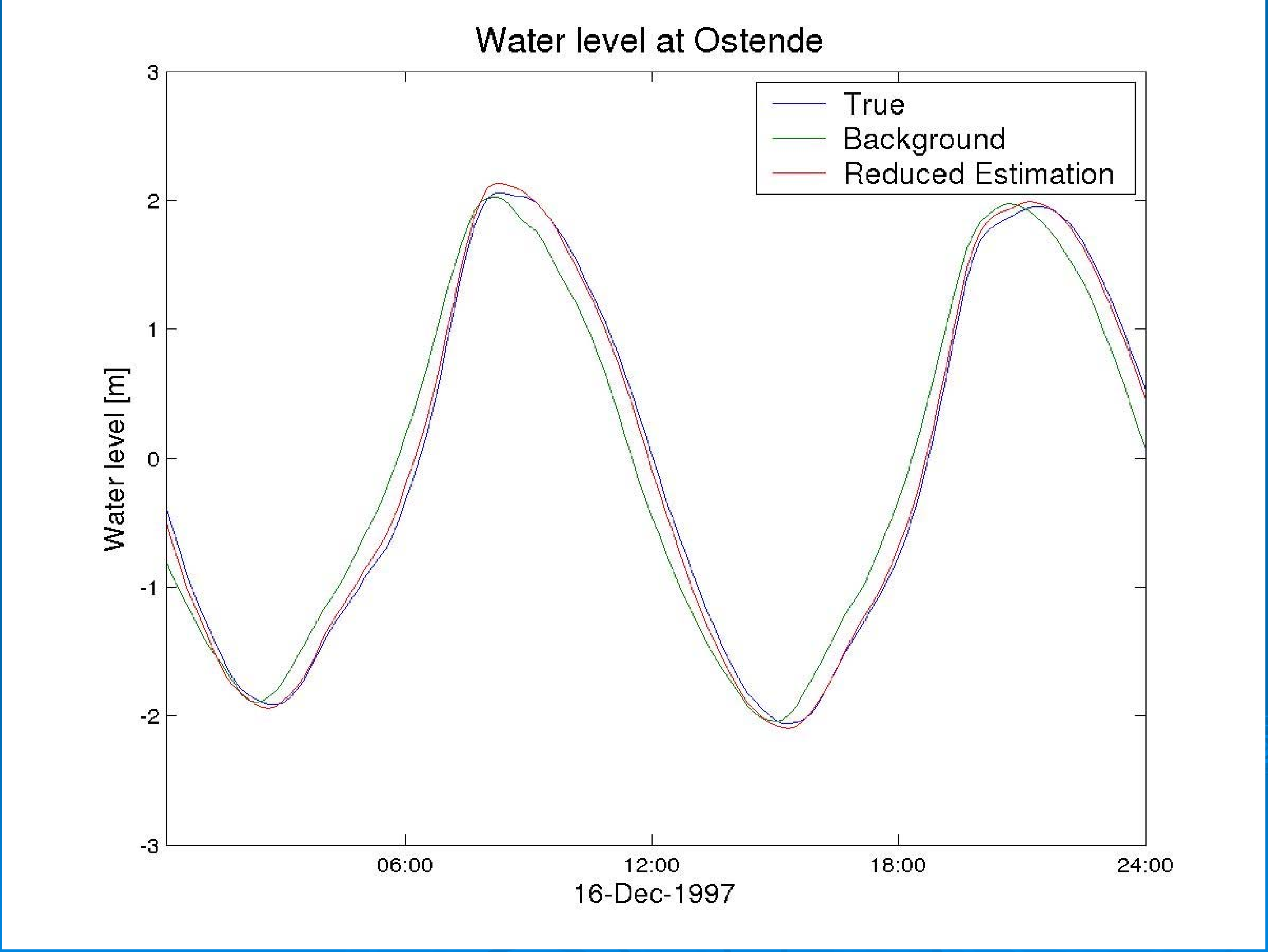

Cost function versus energy levels
97%
92%
Cost Function (J)
Outer Iterations
0
50
100
150
200
250
300
0
1
2
3
4
5
6

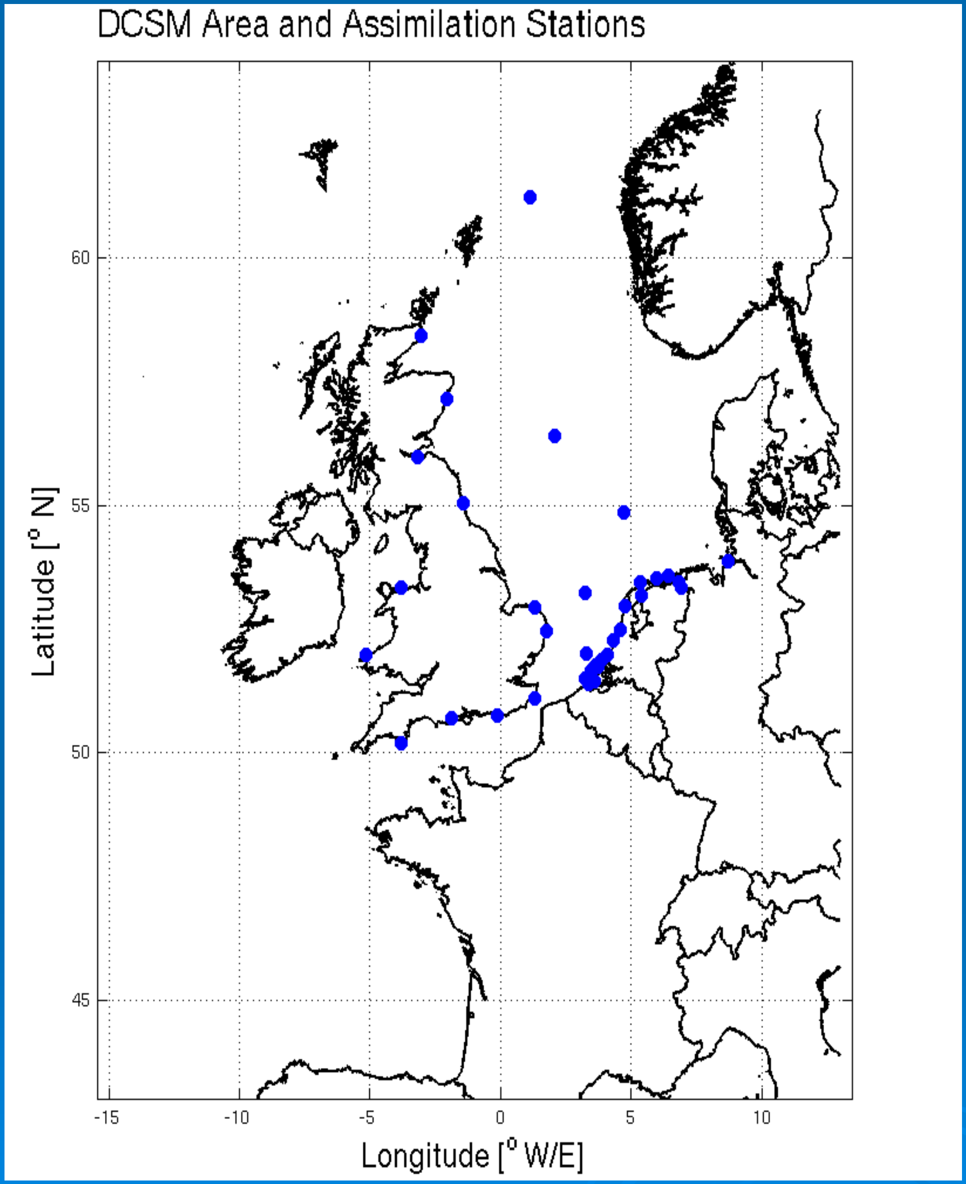

Validation Stations
RMSE [cm]
40
30
20
10
0
Background
I st Outer Iteration
2 nd Outer Iteration
3 rd Outer Iteration
4 th Outer Iteration
5 th Outer Iteration
6 th Outer Iteration
nprt
hvhd
dovr
lows
N53
wick



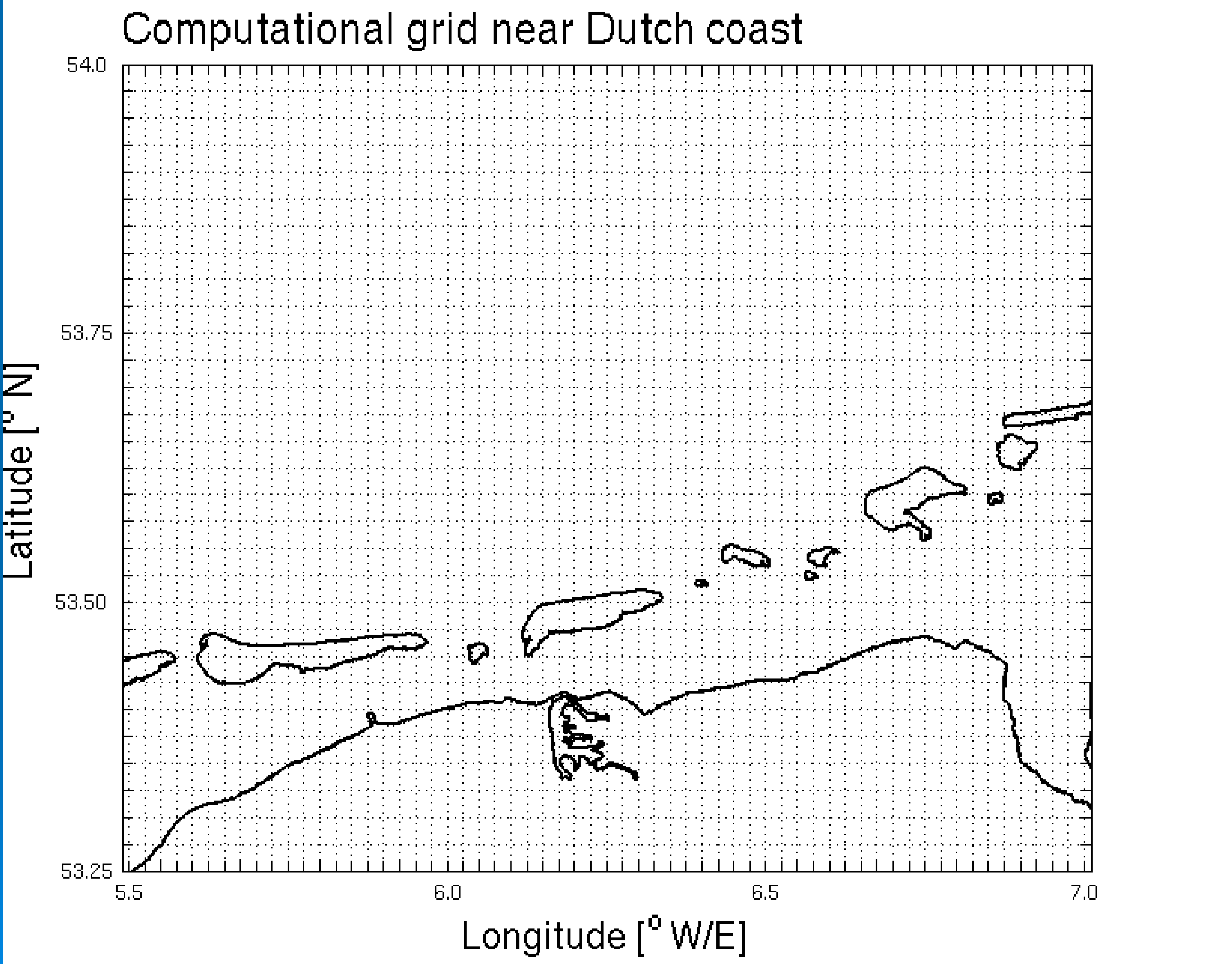
Reduction in Cost function for 10 days
Cost Function (J)
Outer Iterations
450
400
350
300
250
200
150
100
50
0
0
1
2
3
4


Assimilation Results for 10 days
RMSE [cm]
50
45
40
35
30
25
20
15
10
5
0
sthd
denh
ostd
brk
Ind
N4
N51
Background
Ist Outer Iteration
2nd Outer Iteration
3rd Outer Iteration
4th Outer Iteration


Water level at Ostende
True
Background
Reduced Estimation
Water level [m]
3
2
1
0
-1
-2
-3
06:00
12:00
18:00
24:00
16-Dec-1997

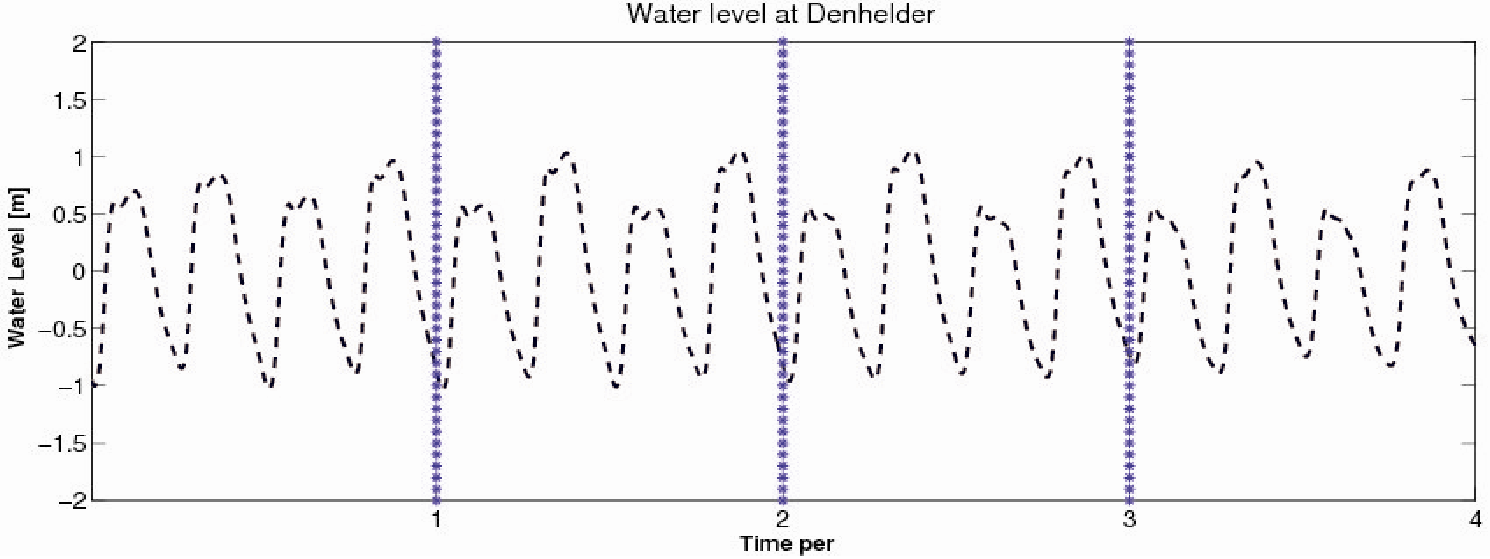


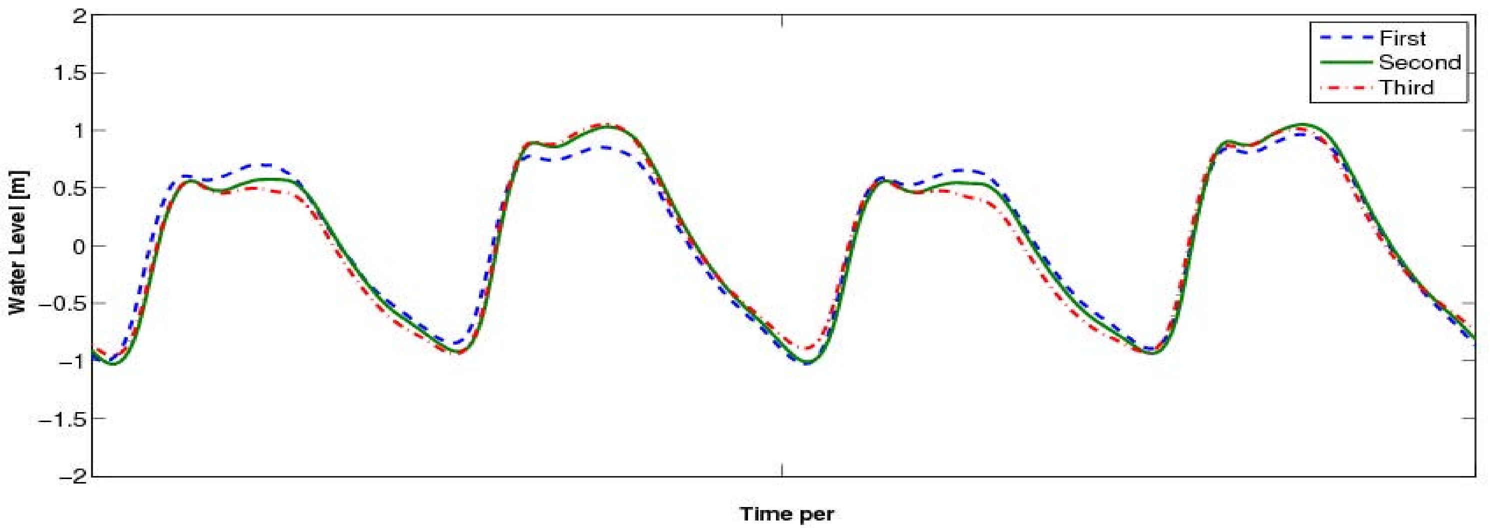
- Based on shallow water equations
- Grid size: 1.5’ by 1.0’ (~2 km)
- Grid dimensions: 1120 x 1260 cells
- Active Grid Points: 869544
- Time step: 2 minutes
- 8 main constituents

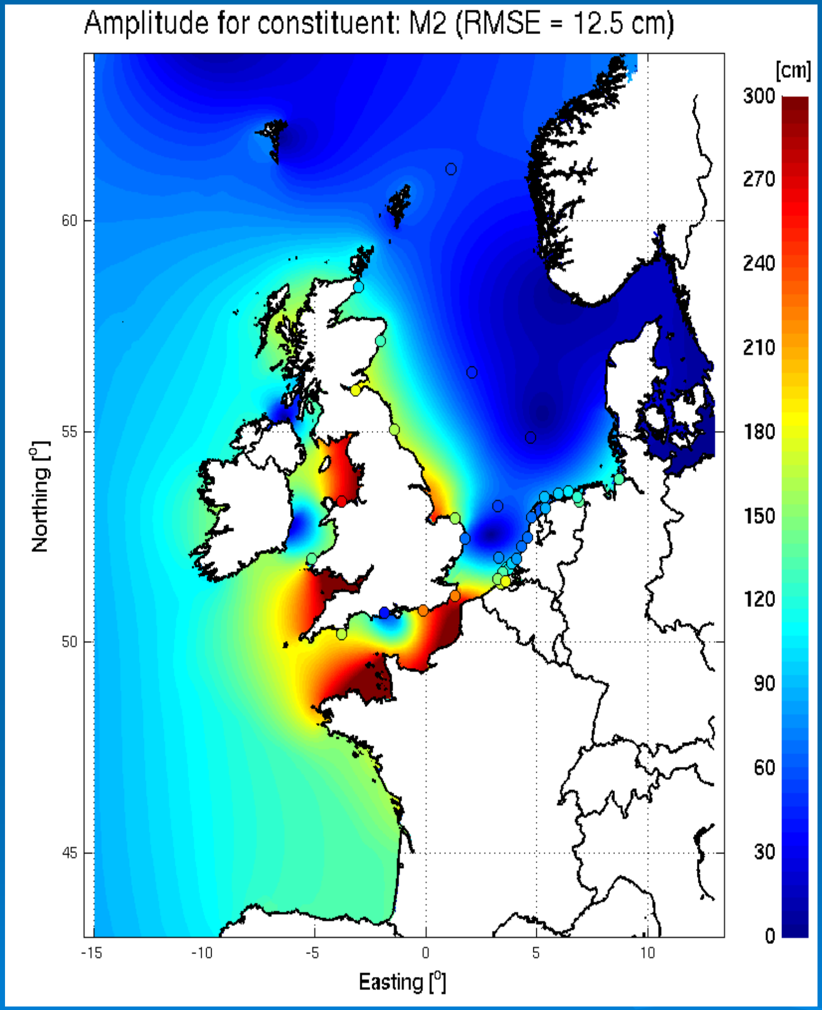


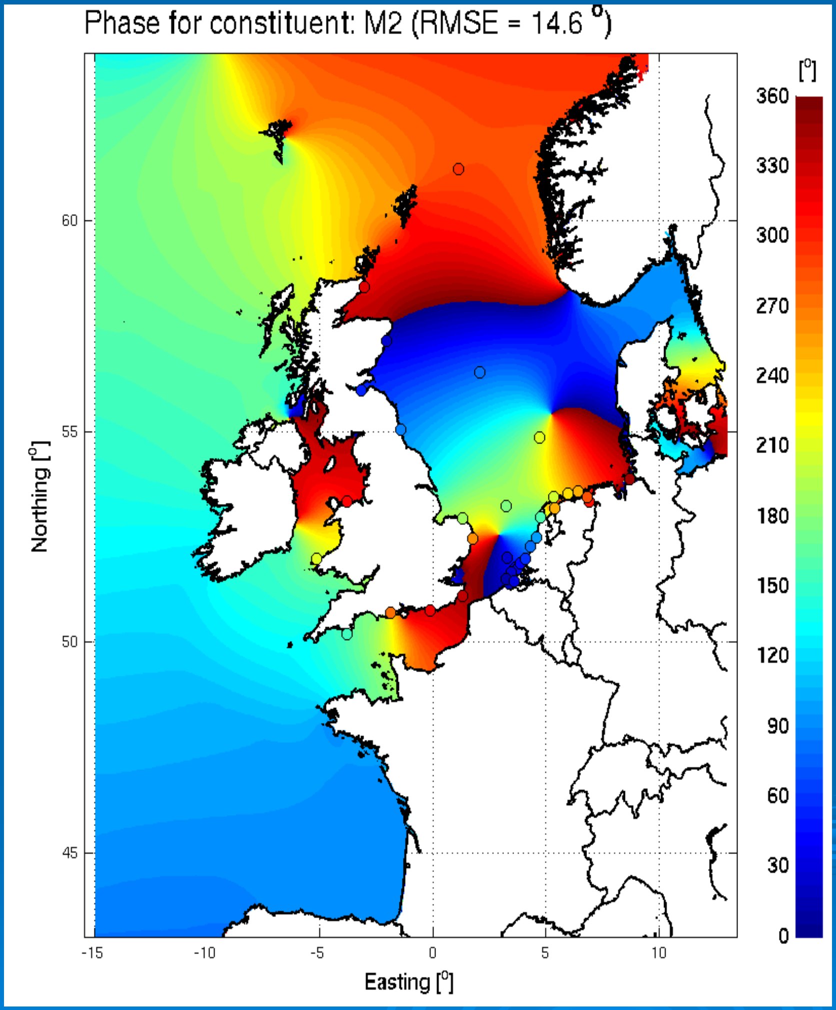
Computational grid near Dutch coast
54.0
53.75
53.50
53.25
Latitude [° N]
5.5
6.0
6.5
7.0
Longitude [° W/E]

DCSM(Water level time series)







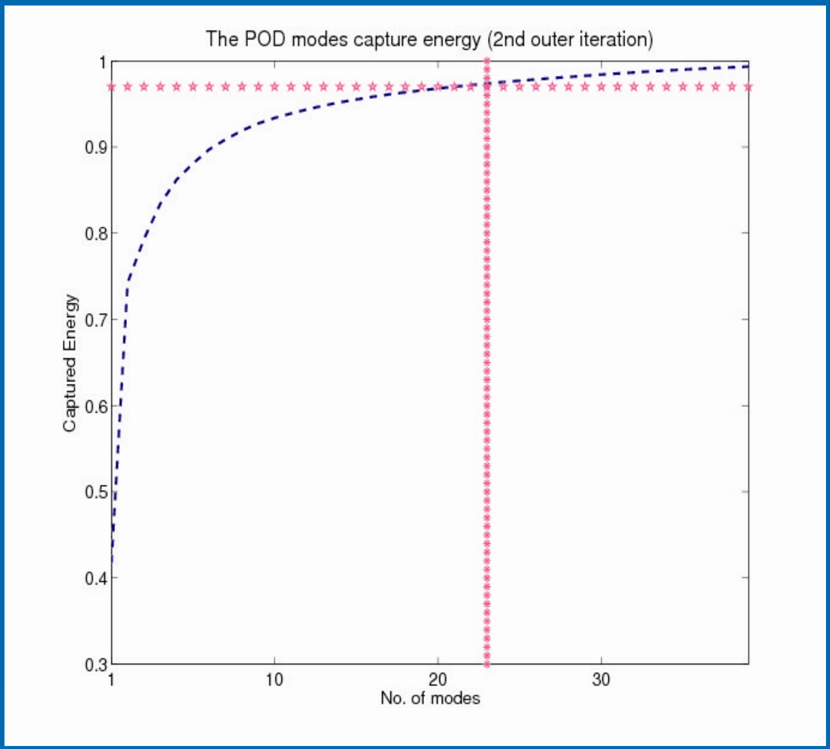


Amplitude and Phase of harmonically analyzed constituent M2

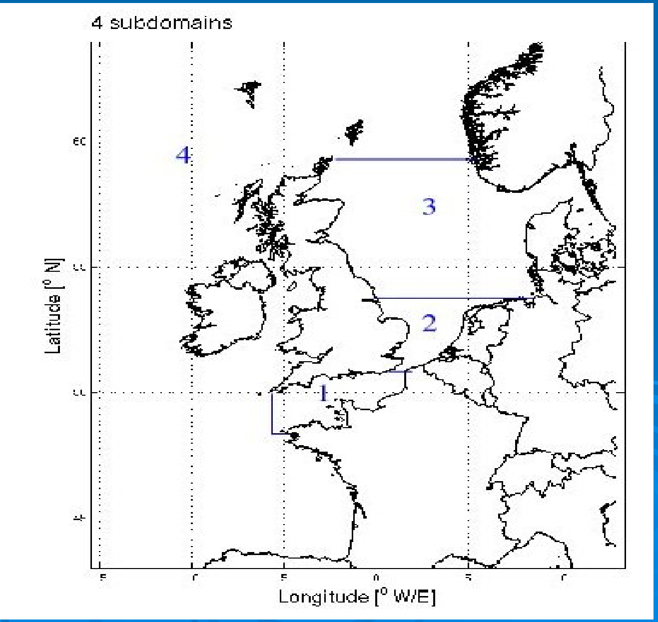
Experiment with field data

- Parameter: Depth
- Calibration run: 28 Dec 2006 to 30 Jan 2007
- Measurement data: 01 Jan 2007 to 30 Jan 2007
- Includes two spring-neap cycles
- Assimilation Stations: 35
- Validation Stations: 15
- Ensemble of forward model simulations for a period of four days (01 Jan 2007 to 04 Jan 2007)

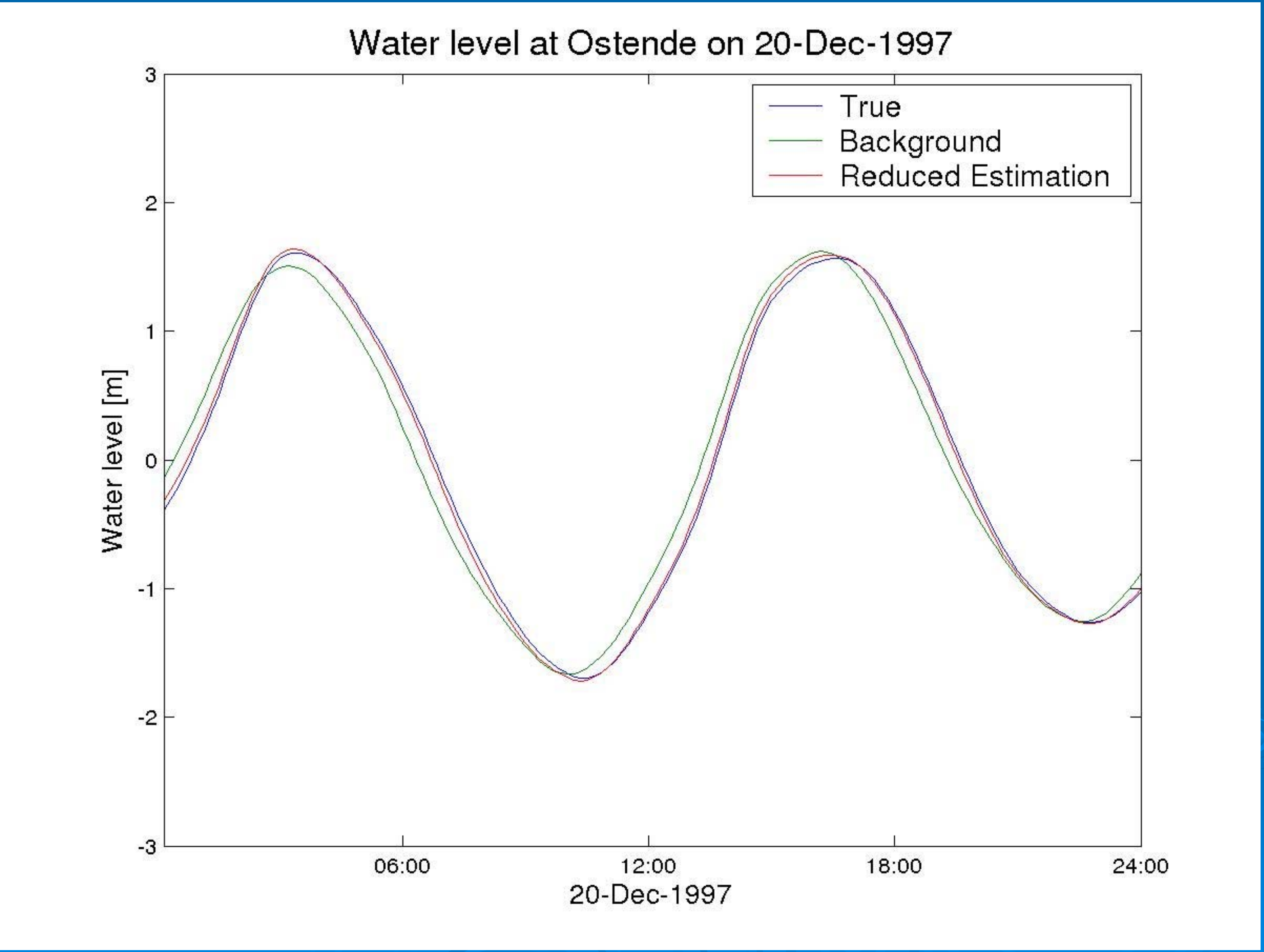
DCSM



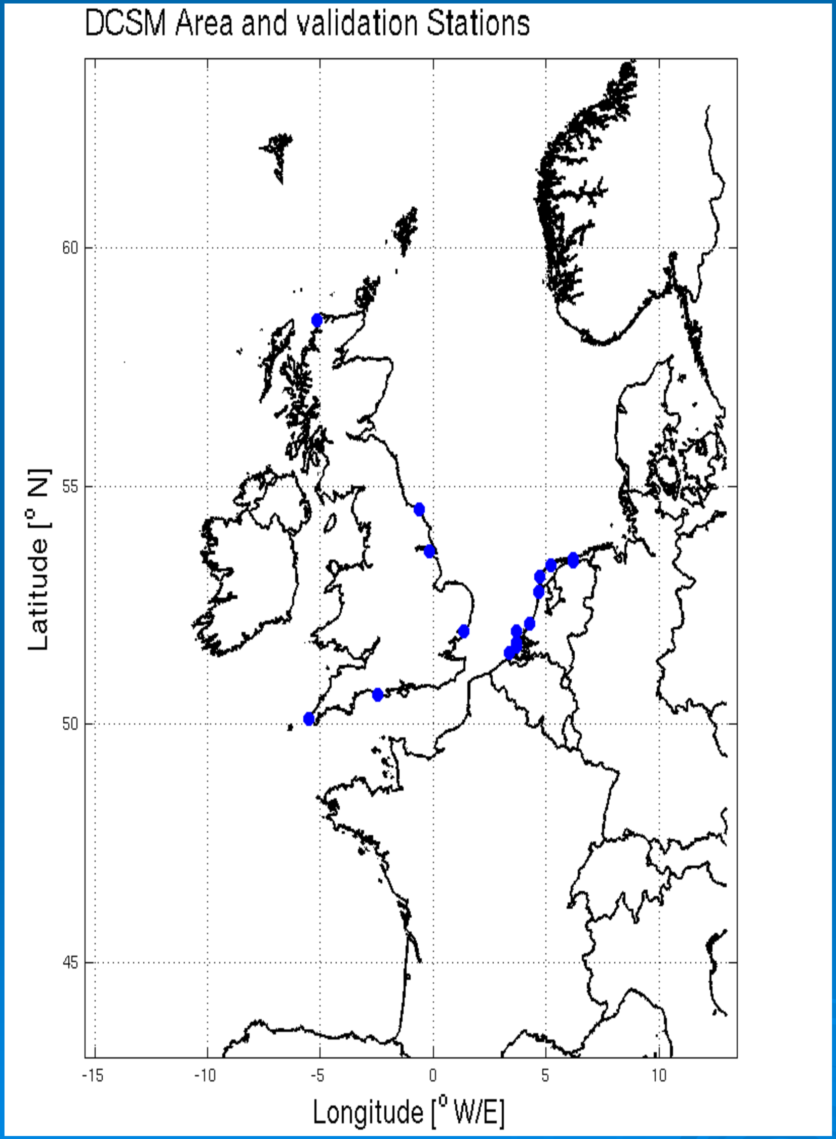
- Divide model area in 4 sub domains + 1 overall parameter
- No. of snapshots: 132 (Every three hours)
- 24 POD modes are required to capture 97% energy
- Same POD modes are used in 2rd iteration

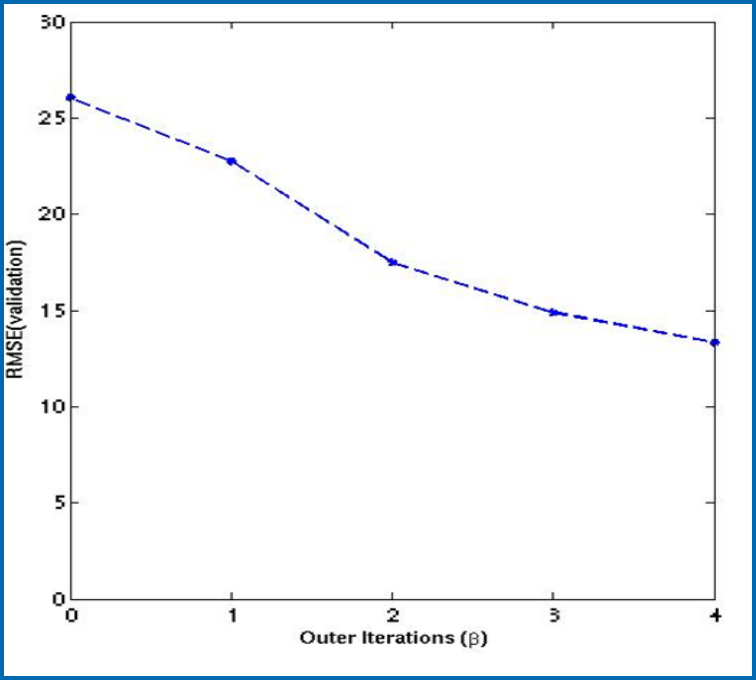


- Initial RMS: 25.7 cm
- After 2rd iteration: 14.9 cm
- Improvement : 42%


Water level at Ostende on 20-Dec-1997
True
Background
Reduced Estimation
Water level [m]
3
2
1
0
-1
-2
-3
06:00
12:00
18:00
24:00
20-Dec-1997

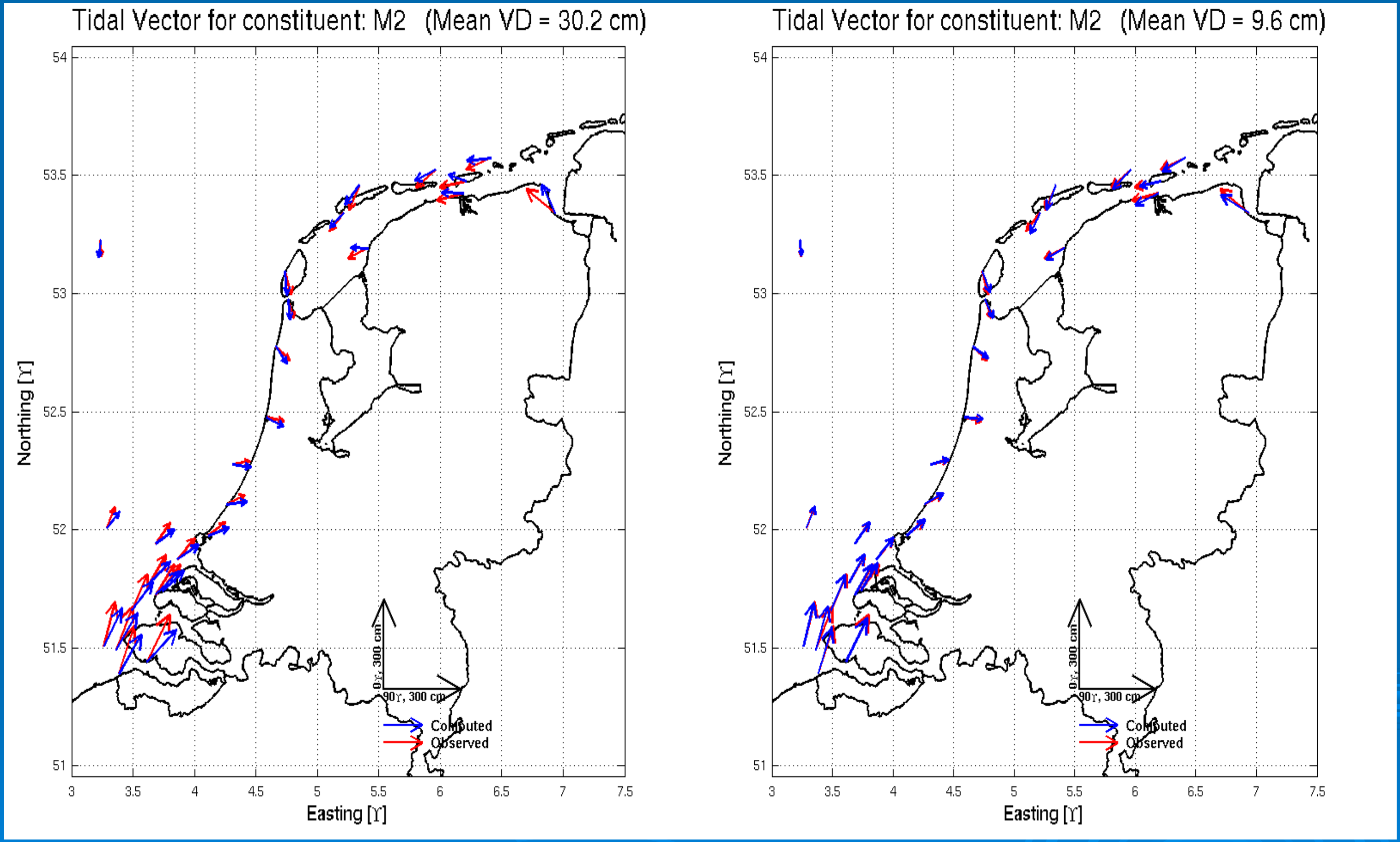
DCSM(Validation results)





- Similar improvement as in the case of assimilation stations

Tidal vector of constituent M2 near Dutch coast



With initial depth

After calibration

Computational Cost

Estimation 5 parameters, calibration period 1 month:
Number of simulations of 1 month 4.7, reduction criterion 42% (2 iterations, no model update in second iteration)

Conclusions

- The adjoint implementation can be avoided using model reduction
- The algorithm is of the ensemble type
- Efficiency MRVAR is very problem dependent, but extremely good for tidal calibration problems
- More research is needed