# Data Assimilation at ULg

Alexander Barth, Aida Alvera Azcárate, Charles Troupin, Ouberdous Mohamed, Jean-Marie Beckers

GHER, University of Liege, Belgium

Sangoma Kick-Off meeting, Liege, 25 November 2011

# Outline

- Weakly constrained ensemble perturbations
- German Bight experiments (estimation of tidal boundary conditions and wind forcing using HF radar observations)
- Ligurian Sea experiments

# Improved parametrization of error covariance

- error covariance is crucial for data assimilation
- the model error covariance defines the vector space of possible model states
- ensemble method relies on perturbing model initial condition, forcing, ... within the limit of their uncertainty
- error covariance matrix $\mathbf{P} \Longleftrightarrow$ method for creating perturbations $\mathbf{x}^{(k)}$

$$\mathbf{x}^{(k)} = \mathbf{P}^{1/2}\mathbf{z}^{(k)} \qquad k \text{ ensemble index}$$

$$\mathbf{P} = E\left[\left(\mathbf{x} - E\left[\mathbf{x}\right]\right)\left(\mathbf{x} - E\left[\mathbf{x}\right]\right)^T\right]$$

- Instead of enforcing a dynamical balance as a post-processing step after the analysis, it is preferable to choose only dynamically balanced ensemble members (if possible)

# Weakly constrained ensemble perturbations

- By validation of the model with observations one can obtain an estimate of the magnitude of the perturbation.
- But which spatial structure?
- Method to create ensemble perturbation that satisfy *a priori* linear constraints
- Example of constraints:
    - geostrophic equilibrium
    - zero horizontal divergence of surface winds
    - stationary solution to the advection-diffusion equation
    - the linear shallow water equations
    - perturbations should be close to a subspace defined by *e.g.* empirical orthogonal functions (EOFs).
    - ...

# Probability of a perturbation

- To describe our *a priori* knowledge of what a realistic perturbation is, we introduce a cost function $J$, similar to the cost function used in variational analysis techniques:

$$J(\mathbf{x}) = \text{“linear balance”}^2 + \text{“smooth”}^2 + \text{“limited amplitude”}^2$$

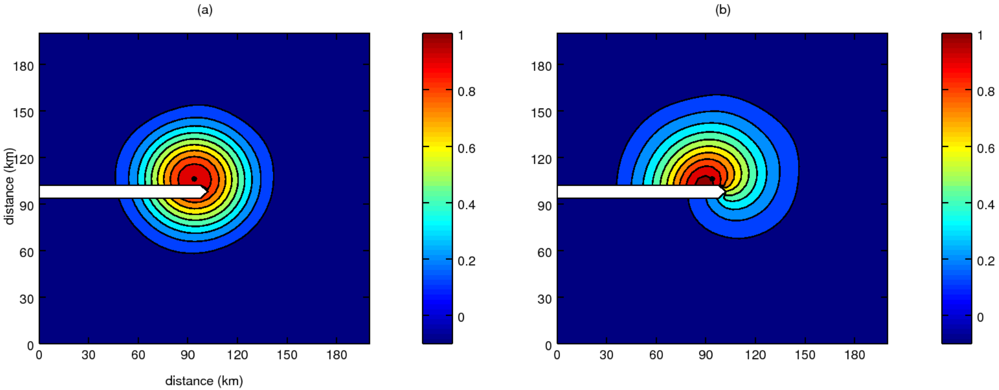
- The cost function can be used to define the probability of a perturbation $\mathbf{x}$ (*e.g.* Kalnay, 2002):

$$p(\mathbf{x}) = \alpha \exp\left(-J(\mathbf{x})\right) \tag{1}$$

- Perturbations are derived from the Hessian matrix of $J$.
- Article and source code (for MATLAB and GNU Octave) is available at `http://modb.oce.ulg.ac.be/mediawiki/index.php/WCE`

# Impact of barriers

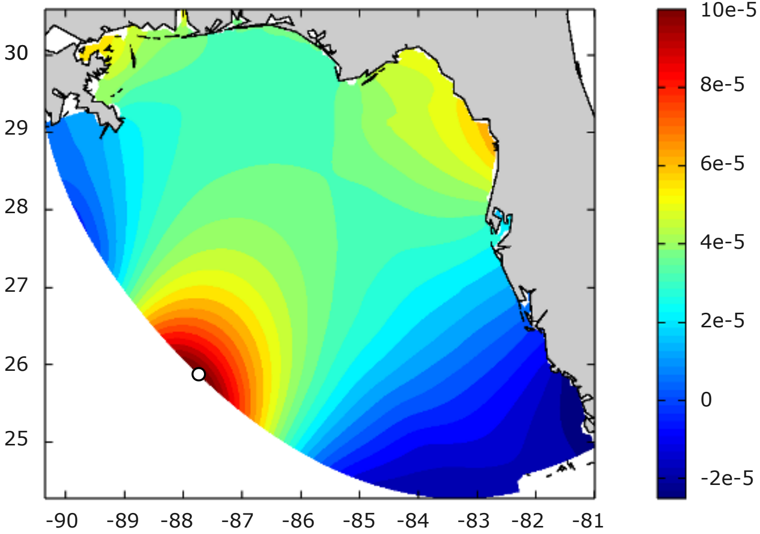
- The "smoothness" constraint is implemented through a diffusion operator (laplacian), it takes thus the land-sea mask into account




- Ensemble covariance using "classical" Fourier modes (a) and constrained perturbations based on the land-sea mask (b).

# Harmonic shallow water equations

- For tidal models, perturbations should be approximately a harmonic solution to the shallow water equations

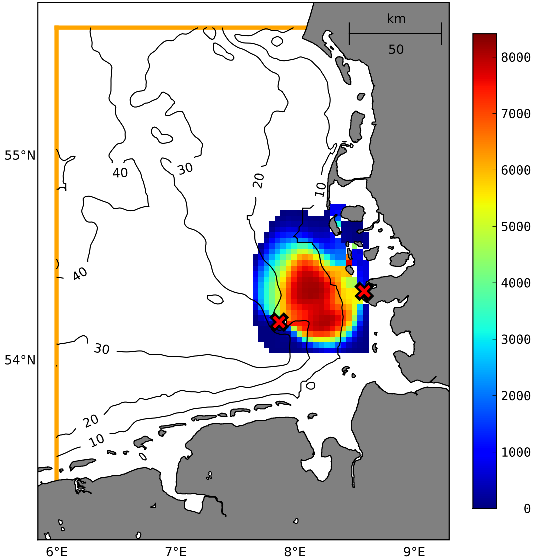



- Horizontal covariance of the constrained perturbations between the point near the open boundary marked by a black dot and all other grid points.

# German Bight model

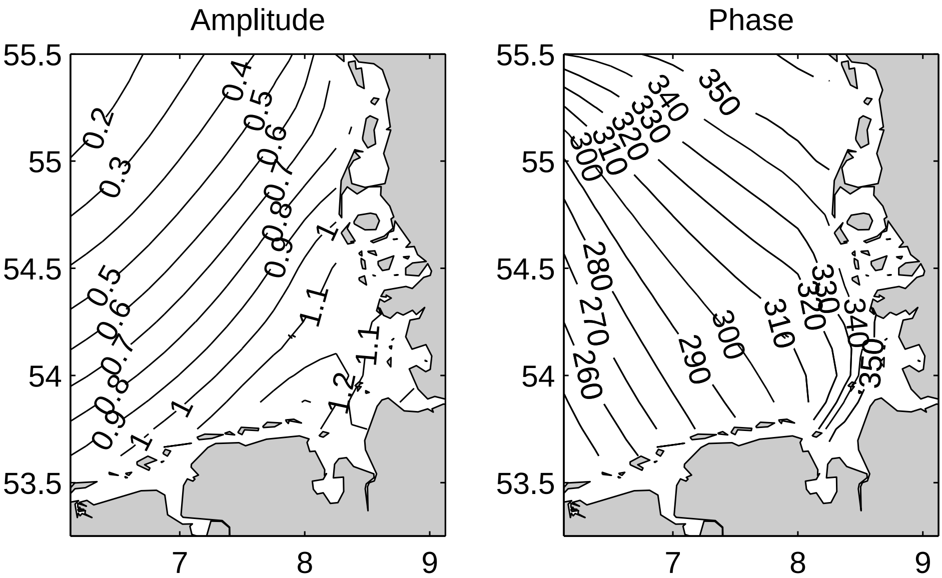
- General Estuarine Ocean Model (GETM Burchard and Bolding, 2002)
- 3-D primitive equations with a free-surface
- 21 $\sigma$ levels, resolution of about 0.9 km.
- nested in a 5-km resolution North Sea-Baltic Sea model
- ETOPO-1 topography with observations from BSH
- Atmospheric fluxes are estimated by the bulk formulation using 6-hourly ECMWF re-analysis
- Implementation by GKSS (Staneva *et al.*, 2009).

# HF radar observations

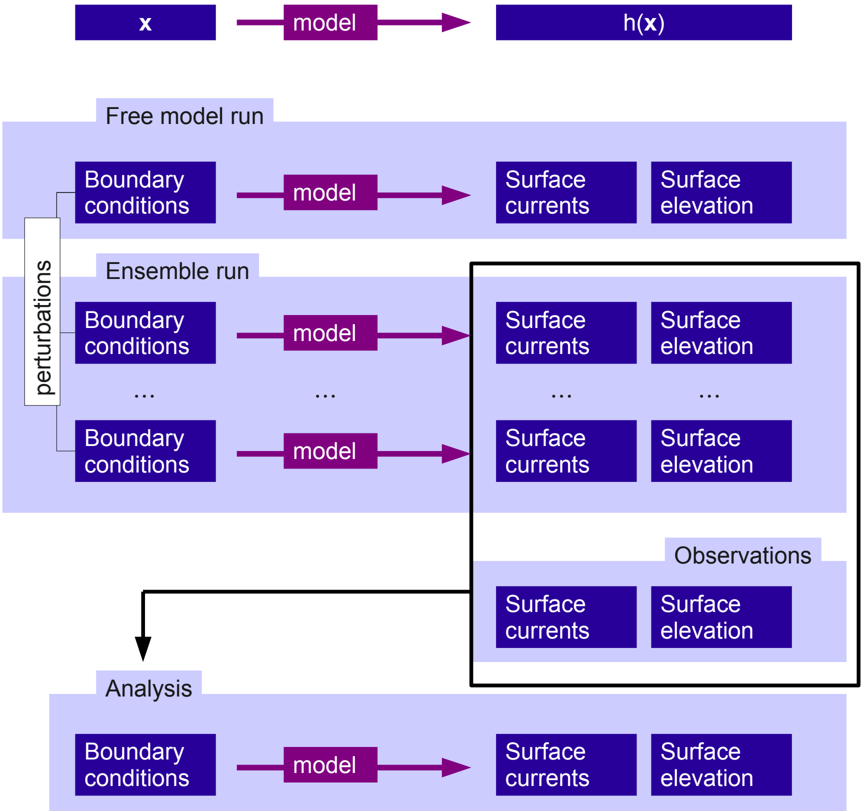



- Spatial coverage of the HF radar zonal and meridional surface velocity observations
- The number of samples available at each observation grid point is color–coded according to the color-bar.
- The crosses show the location of HF radar antennas.
- The operating frequency: 29.85 MHz (coupling to 5.02 m long ocean waves).
- HF Radar measurements from University of Hamburg (PRISMA project)

# Empirical Ocean Tides (EOT08a)




- M2 amplitude (in m) and phase (in degrees) of EOT08a for the German Bight based on altimetry.
- complex tidal parameters are assimilated


x
model
h(x)
Free model run
Boundary conditions
model
Surface currents
Surface elevation
perturbations
Ensemble run
Boundary conditions
model
Surface currents
Surface elevation
...
...
...
...
Boundary conditions
model
Surface currents
Surface elevation
Observations
Surface currents
Surface elevation
Analysis
Boundary conditions
model
Surface currents
Surface elevation

# Smoother scheme

- M2 tidal boundary conditions are perturbed within the range of their uncertainty to create a ensemble with 51 members. Perturbations are constrained by the linear shallow water equations.
- The GETM model is run for 40 days with each of those perturbed boundary values.
- All HF radar observations at any time instance within the integration period and the EOT parameters are grouped in the observation vector (vector $\mathbf{y}^o$) with their corresponding error covariance (matrix $\mathbf{R}$) estimated by cross-validation.
- The observations are extracted from every ensemble member (vector $h(\mathbf{x}^{(k)})$).
- Schematically, the non-linear operator $h(\cdot)$ performs the following operations:

$$h(\cdot) = \text{Interpolation to obs. location} \circ \text{Model integration with perturbed forcing} \tag{2}$$

# Smoother scheme

- The optimal perturbation is given the Kalman analysis (using non-linear observation operators as in Chen and Snyder (2007)):

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{A}\left(\mathbf{B}+\mathbf{R}\right)^{-1}\left(\mathbf{y}^o - h(\mathbf{x}^b)\right) \tag{3}$$

- where the matrices $\mathbf{A}$ and $\mathbf{B}$ are covariances estimated from the ensemble.

$$\mathbf{A} = \mathrm{cov}(\mathbf{x}^b, h(\mathbf{x}^b)) = \left\langle (\mathbf{x} - \langle\mathbf{x}\rangle)\left(h(\mathbf{x}) - \langle h(\mathbf{x})\rangle\right)^T \right\rangle \tag{4}$$

$$\mathbf{B} = \mathrm{cov}(h(\mathbf{x}^b), h(\mathbf{x}^b)) = \left\langle (h(\mathbf{x}) - \langle h(\mathbf{x})\rangle)\left(h(\mathbf{x}) - \langle h(\mathbf{x})\rangle\right)^T \right\rangle \tag{5}$$

where $\langle\cdot\rangle$ is the ensemble average.

- But covariance matrices do not need to be formed explicitly. Analysis is performed in the subspace defined by the ensemble members.

# Smoother scheme

- For a linear model and an infinite large ensemble, equation (14) minimizes,

$$J(x) = (\mathbf{x} - \mathbf{x}^b)^T \mathbf{P}^{b^{-1}} (\mathbf{x} - \mathbf{x}^b) + (\mathbf{y}^o - h(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y}^o - h(\mathbf{x})) \quad (6)$$

  or

$$J(x) = (\mathbf{x} - \mathbf{x}^b)^T \mathbf{P}^{b^{-1}} (\mathbf{x} - \mathbf{x}^b) + \sum_n (\mathbf{y}_n^o - (h(\mathbf{x})_n))^T \mathbf{R}_n{}^{-1} (\mathbf{y}_n^o - (h(\mathbf{x})_n)) \quad (7)$$

  where $n$ references to the indexed quantifies at time $n$. This is the cost function from which 4D-Var and Kalman Smoother can be derived.

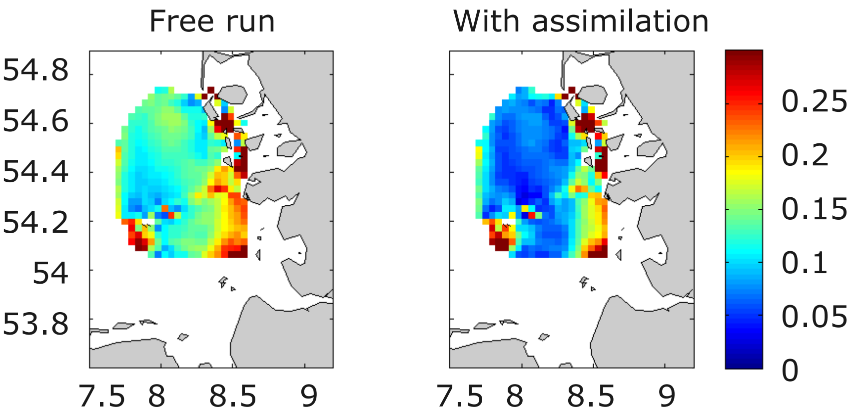
- Approach is closely related to Ensemble Smoother (van Leeuwen, 2001), 4D-EnKF (Hunt *et al.*, 2007) and AEnKF (Sakov *et al.*, 2010) where model trajectories instead of model states are optimized and to the Green's method with stochastic "search directions"

- The model is rerun with the optimized boundary values for 60 days.

# RMS difference

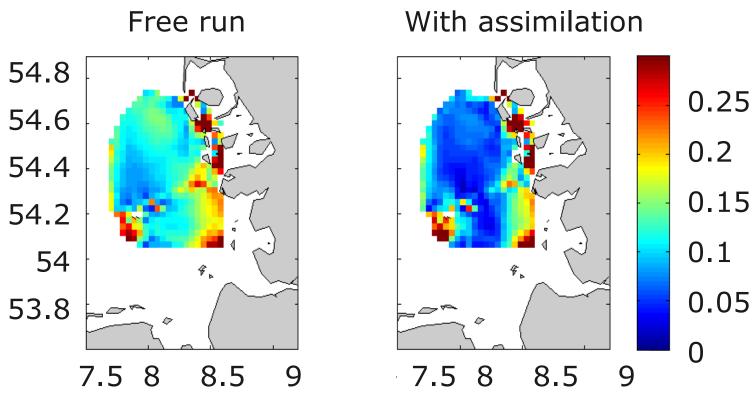
$$\text{RMS}^2 = \lim_{T\to\infty} \frac{1}{T} \int_0^T (A\cos(\omega t - \phi) - A' \cos(\omega t - \phi'))^2 dt \quad (8)$$

$$= \frac{A^2 + A'^2}{2} - AA' \cos(\phi - \phi') \quad (9)$$




RMS difference between surface current observations due to the M2 tides and the corresponding model results without (left panel) and with assimilation (right panel).

# Comparison with un-assimilated observations (M2)




- RMS difference between surface current observations (not used in the assimilation) due to the M2 tides and the corresponding model results without (left panel) and with assimilation (right panel).
- Analysis RMS compared to unassimilated data is only 0.002 m/s larger than compared to assimilated data

# Tide gage observations

| | Helgoland | | | Cuxhaven | | |
|---|---|---|---|---|---|---|
| | amplitude | phase | RMS | amplitude | phase | RMS |
| Observations | 1.13 | 304 | | 1.36 | 334 | |
| Free | 0.81 | 318 | 0.28 | 0.95 | 15 | 0.63 |
| Assimilation | 0.97 | 302 | 0.12 | 1.08 | 2 | 0.46 |

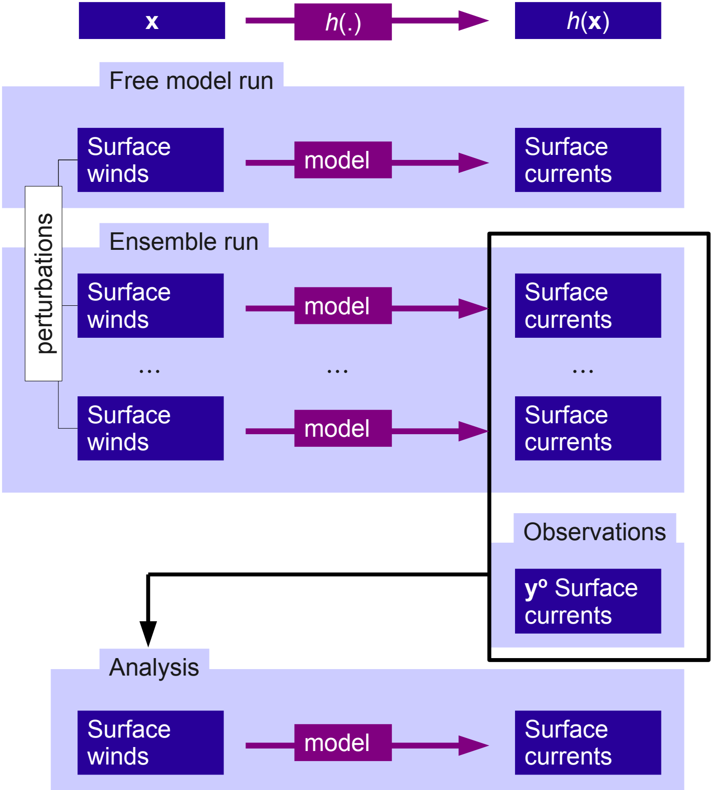
Table 1: Comparison with tide gage observations. Amplitude is in m and phase in degrees.

- Tide gage observations from different time period $\rightarrow$ only comparison of tidal parameters
- Helgoland within the area covered by radar, but not Cuxhaven
- The assimilation reduces the RMS error by a factor of 2 for Helgoland and by a factor of 1.4 for Cuxhaven.
- Ocean Science, 6, 161–178, 2010 http://www.ocean-sci.net/6/161/2010/os-6-161-2010.pdf.

# Wind estimation from HF radar observations

- Ensemble of 100 wind forcings are created (by using a Fourier decomposition)
- estimation vector $\mathbf{x}$: u- and v- component of wind forcing
- observations: $\mathbf{y}^o$: surface currents
- "observation operator" $h(\cdot)$:

$$h(\cdot) = \text{Interpolation to obs. location} \circ \text{Model integration with perturbed wind} \tag{10}$$



x
h(.)
h(x)
Free model run
Surface winds
model
Surface currents
perturbations
Ensemble run
Surface winds
model
Surface currents
...
...
...
Surface winds
model
Surface currents
Observations
yº Surface currents
Analysis
Surface winds
model
Surface currents

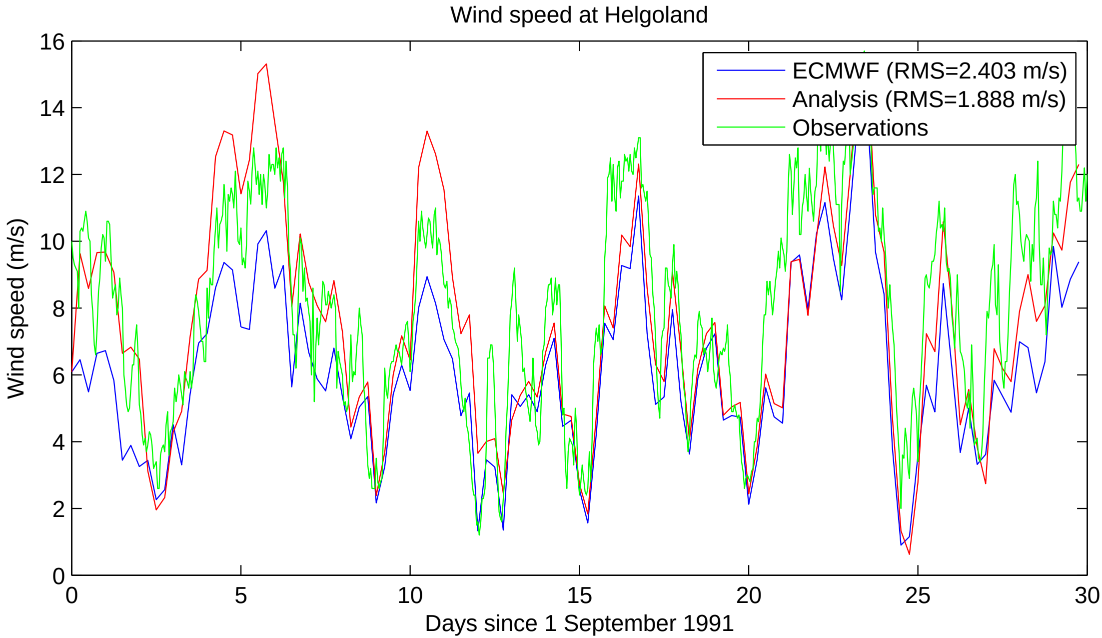



Figure 1: Measured wind speed, wind speed from ECMWF and analyzed wind speed at Helgoland. Units are m/s.

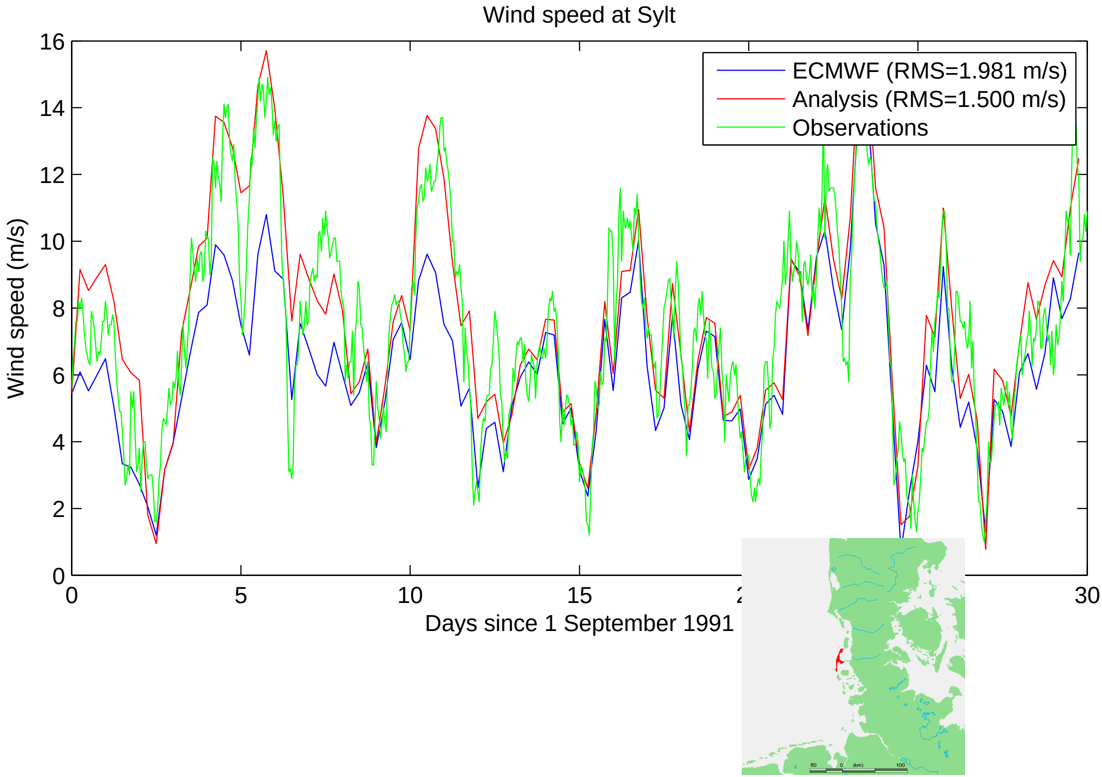



Figure 2: Measured wind speed, wind speed from ECMWF and analyzed wind speed at Sylt. Units are m/s.

# Comparison with satellite SST

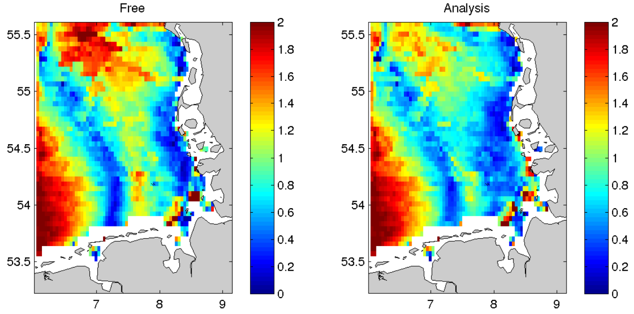


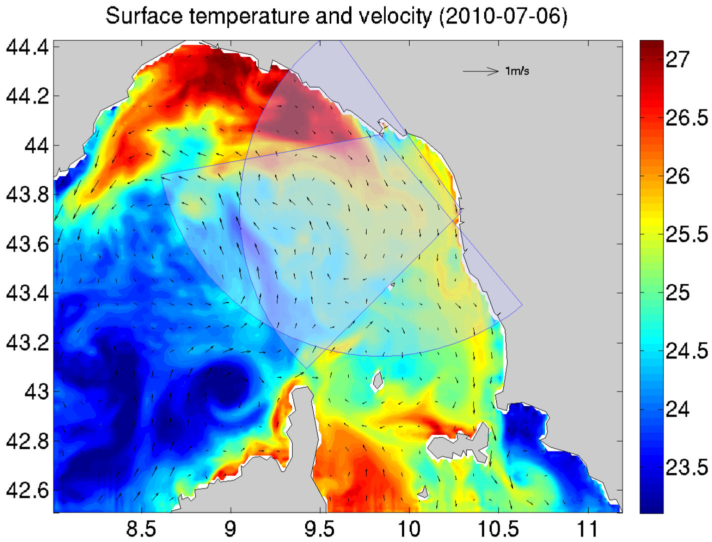

Figure 3: RMS difference between AVHRR SST and model SST without assimilation (left panel) and with assimilation (right panel)

| | $S_{\mathrm{HF}}$ | RMS | skill score |
|---|---|---|---|
| Free | – | 1.21 | 0.00 |
| Analysis | 0.5 | 1.09 | 0.19 |
| | 1.0 | 1.09 | 0.19 |
| | 1.5 | 1.10 | 0.18 |
| | 2.0 | 1.11 | 0.16 |
| | 2.5 | 1.12 | 0.14 |
| | 5.0 | 1.16 | 0.08 |

RMS is expressed in °C and $S_{\mathrm{HF}}$ in m/s.

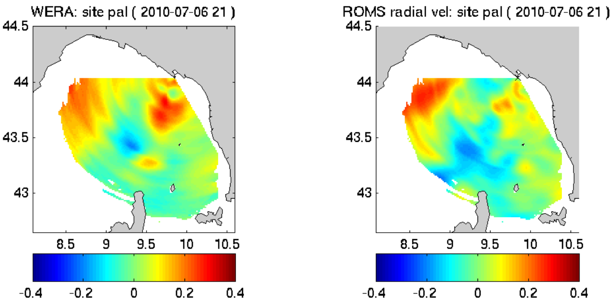
# Ligurian Sea Model (WP5)

- ROMS nested in Mediterranean Ocean Forecasting System
- 1/60 degree resolution and 32 vertical levels
- Currents: Western & Eastern Corsican Current, Northern Current, inertial oscillation
- Two WERA HF radar systems (Palmaria, San Rossore) by NATO Undersea Research Centre (NURC)



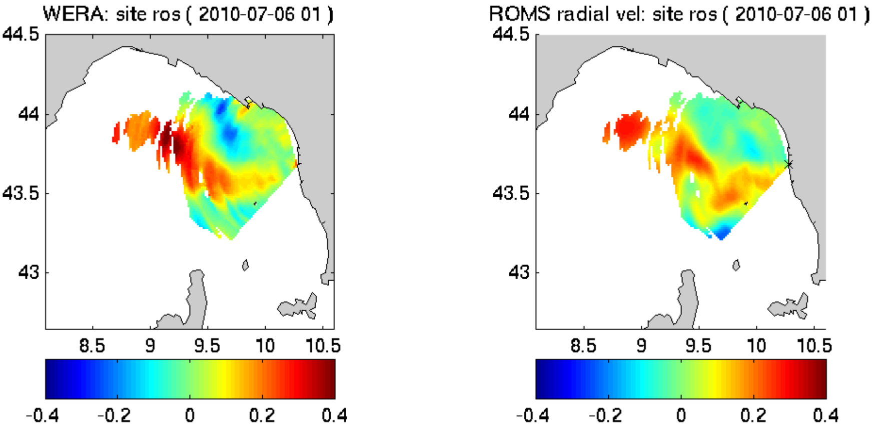

# Observations

- Frequency of $\nu = 12.359$ MHz and coupled to a wave length of $\lambda_b = 12.13$ m,
- Radial currents are used for the assimilation
- Azimuthal resolution of 6 degrees
- Currents are averaged over 1 h




Radial currents on 2010-07-06 21:30 relative to the Palmaria site: left panel shows WERA measurements and right panel shows ROMS results without assimilation.

# Observations




Radial currents on 2010-07-06 01:30 relative to the San Rossore site: left panel shows WERA measurements and right panel shows ROMS results without assimilation.

# Observation operator

- Radial currents are extracted by:

$$u_{\mathrm{HF}} = \frac{k_b}{1 - \exp(-k_b h)} \int_{-h}^{0} \mathbf{u}(z) \cdot \mathbf{e}_r \exp(k_b z) dz \tag{11}$$

  - $k_b = \frac{2\pi}{\lambda_b}$
  - $\mathbf{e}_r$ is the unit vector pointing in the direction opposite to the location of the HF radar site
  - positive values: current away from the system
  - essentially represent an average over the upper meters.
- Smoothed in the azimuthal direction by a diffusion operator to filter scales smaller than 6 degrees

# Model errors covariance

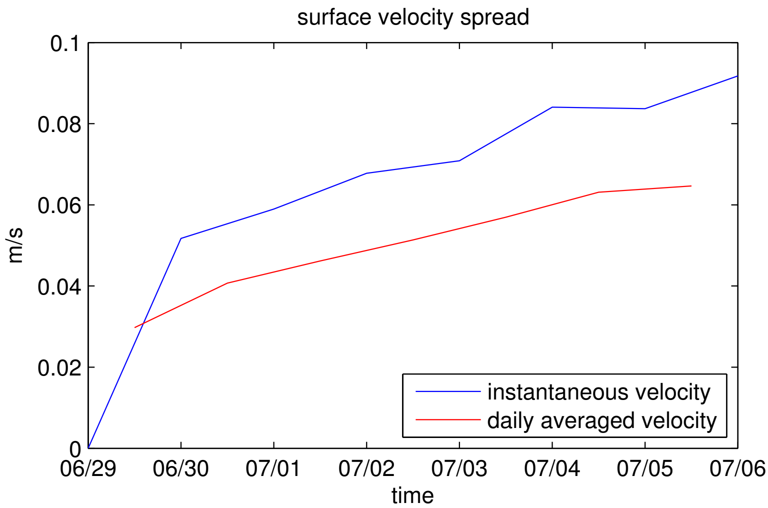
- Estimated by ensemble simulation where uncertain aspect of the model are perturbed
- Perturbed zonal and meridional wind forcing
- Perturbed boundary conditions (elevation, velocity, temperature and salinity)
- Perturbed momentum equation ($\varepsilon$)

$$\frac{d\mathbf{u}}{dt} + \mathbf{\Omega} \wedge \mathbf{u} = -\frac{1}{\rho_0}\nabla_h p + \frac{1}{\rho_0}\nabla \cdot \mathbf{F}^{\mathbf{u}} + \nabla_h \wedge \ \varepsilon \, \mathbf{e}_z \tag{12}$$

$$\tag{13}$$

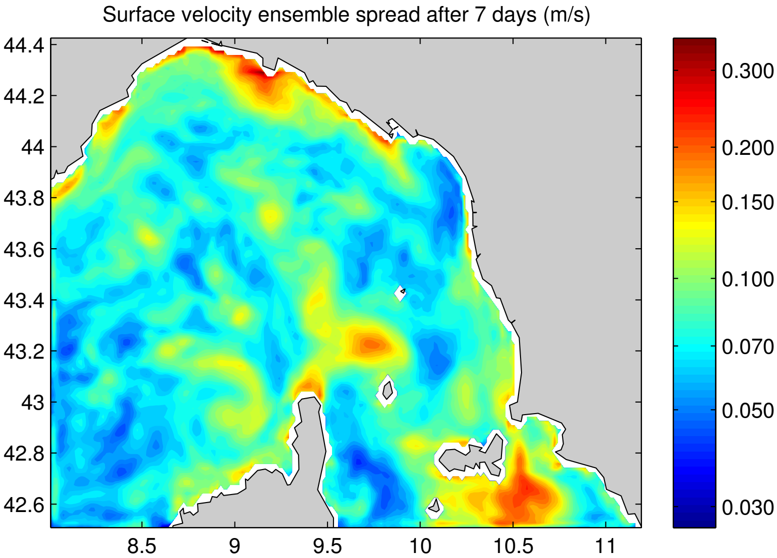
- where $\nabla_h = \mathbf{e}_x \frac{\partial}{\partial x} + \mathbf{e}_y \frac{\partial}{\partial y}$
- does not create horizontal convergence or divergence (linked to barotropic waves)
- can create mesoscale flow structures (absent or misplaced)

# Ensemble spin-up



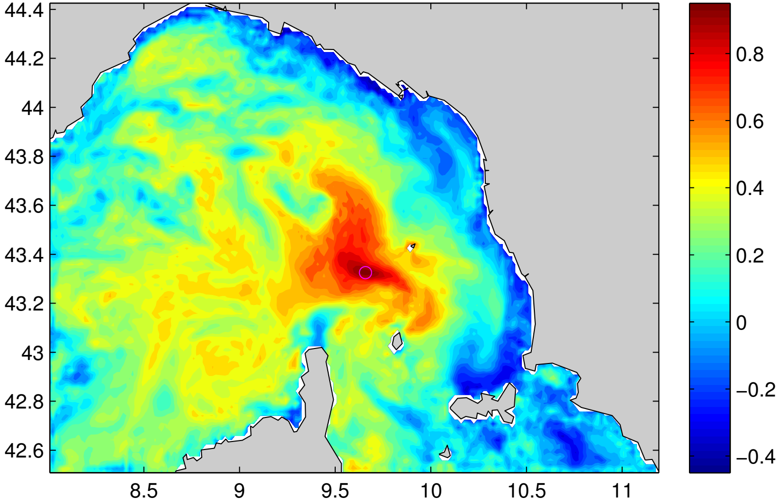

- Ensemble of IC is created by a 7 day ensemble integration starting from the same IC but with perturbed forcing (ensemble spin-up)
- Spin-up should create mesoscale circulation features

# Velocity spread



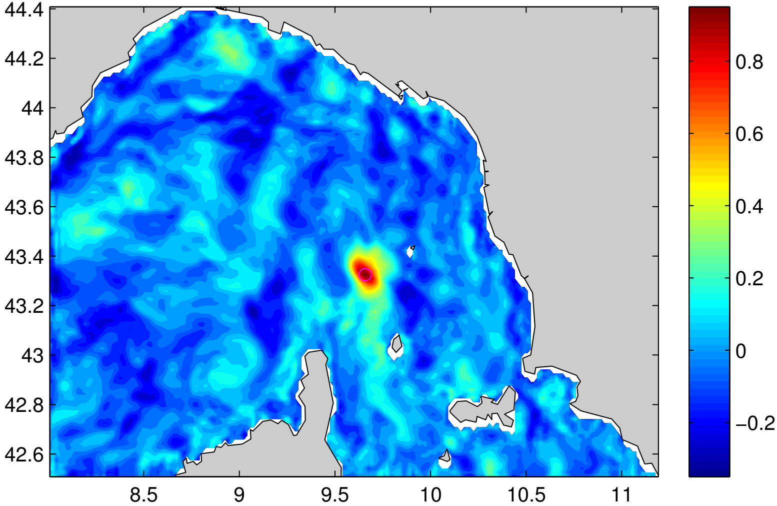

- Velocity spread after 7 days
- Largest uncertainties near eddies

# Spatial correlation



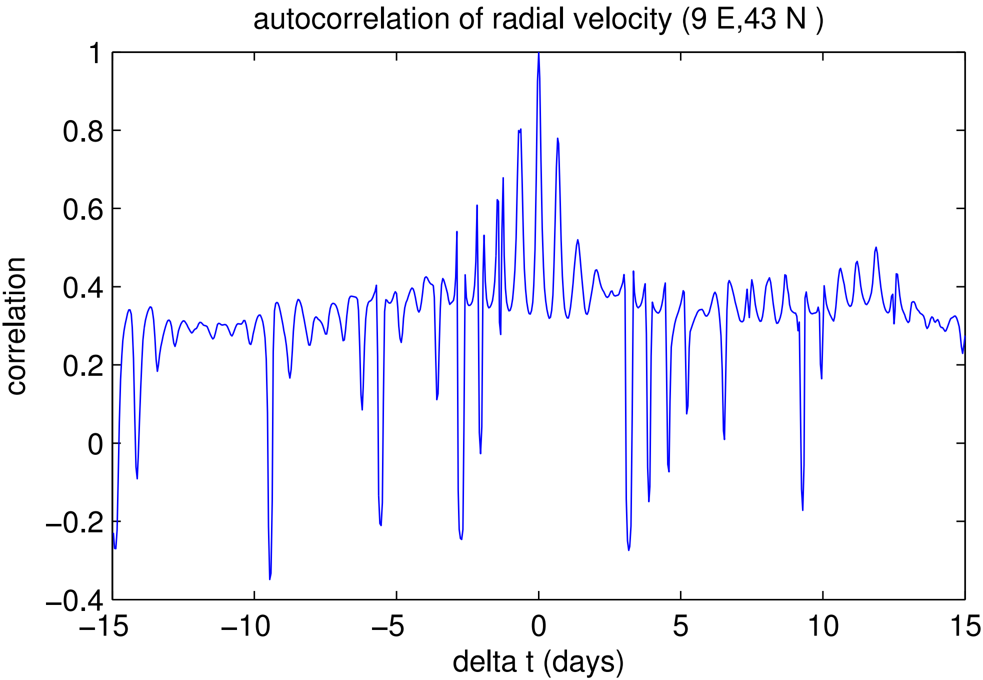

- Correlation of temperature at a specific point (magenta circle) and other surface grid points
- Resulting length-scale is about 50 km

# Spatial correlation




- Correlation of zonal velocity at a specific point (magenta circle) and other surface grid points
- Resulting length-scale is about 10 km
- Adequately observing surface velocity would require measurements with higher spatial resolution that the resolution of temperature measurements

## Temporal correlation




periodicity of 16 h (period of inertial oscillations is 17.6 h)

# Data assimilation scheme

- Time dimension embedded in estimation vector $\mathbf{x}$
- Different definitions of estimation vector are possible:
  - $\mathbf{x} =$ (model trajectory), *i.e.* model state at all time instances
  - $\mathbf{x} =$ (uncertain forcing fields), here IC, BC, wind and stochastic error term at all time instances
  - $\mathbf{x} =$ (model trajectory, uncertain forcing fields)
- The optimal $\mathbf{x}$ is given by the Kalman analysis (using non-linear observation operators as in Chen and Snyder (2007)):

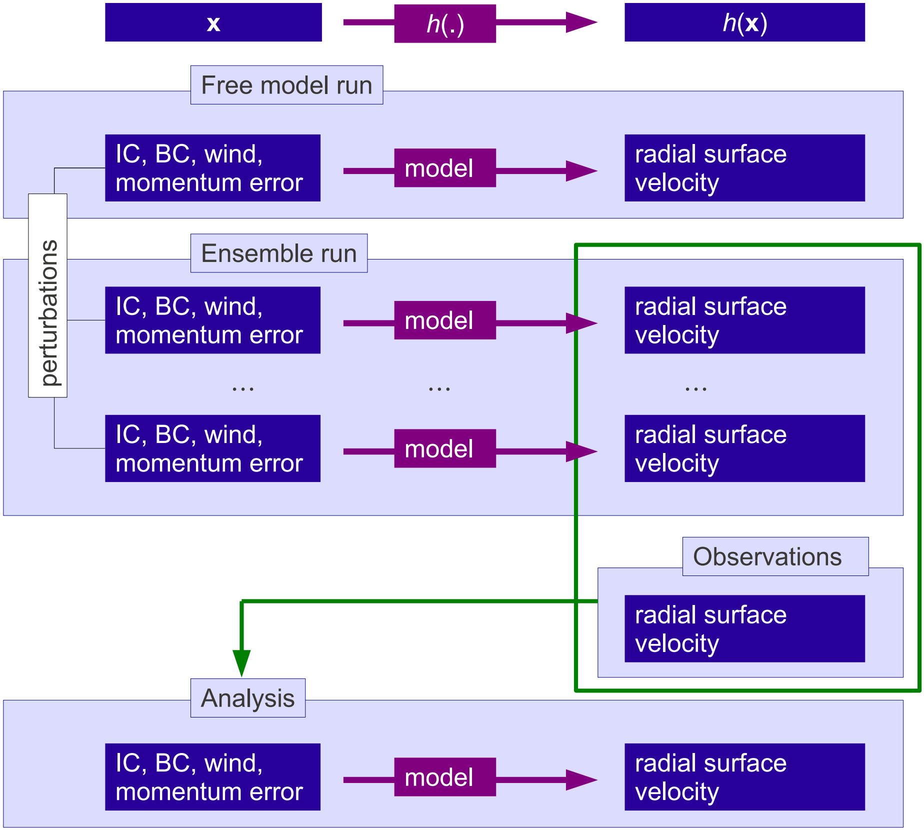
$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{A}\left(\mathbf{B}+\mathbf{R}\right)^{-1}\left(\mathbf{y}^o - h(\mathbf{x}^b)\right) \tag{14}$$

- where the matrices $\mathbf{A}$ and $\mathbf{B}$ are covariances estimated from the ensemble.

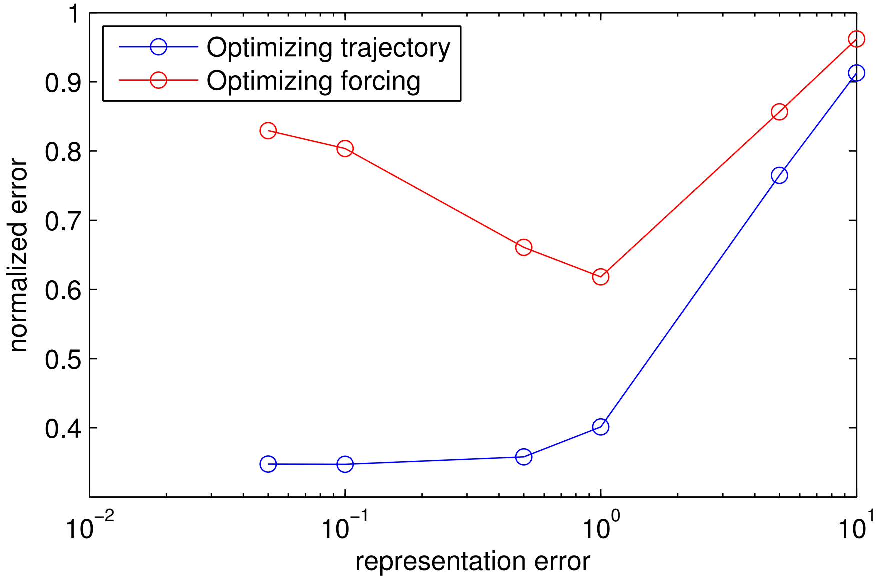
$$\mathbf{A} = \mathrm{cov}(\mathbf{x}^b, h(\mathbf{x}^b)) = \left\langle \left(\mathbf{x} - \langle\mathbf{x}\rangle\right)\left(h(\mathbf{x}) - \langle h(\mathbf{x})\rangle\right)^T \right\rangle \tag{15}$$

$$\mathbf{B} = \mathrm{cov}(h(\mathbf{x}^b), h(\mathbf{x}^b)) = \left\langle \left(h(\mathbf{x}) - \langle h(\mathbf{x})\rangle\right)\left(h(\mathbf{x}) - \langle h(\mathbf{x})\rangle\right)^T \right\rangle \tag{16}$$

where $\langle\cdot\rangle$ is the ensemble average.


x
h(.)
h(x)
Free model run
IC, BC, wind, momentum error
model
radial surface velocity
perturbations
Ensemble run
IC, BC, wind, momentum error
model
radial surface velocity
...
...
...
IC, BC, wind, momentum error
model
radial surface velocity
Observations
radial surface velocity
Analysis
IC, BC, wind, momentum error
model
radial surface velocity

## Estimation of trajectory versus estimation of forcing fields

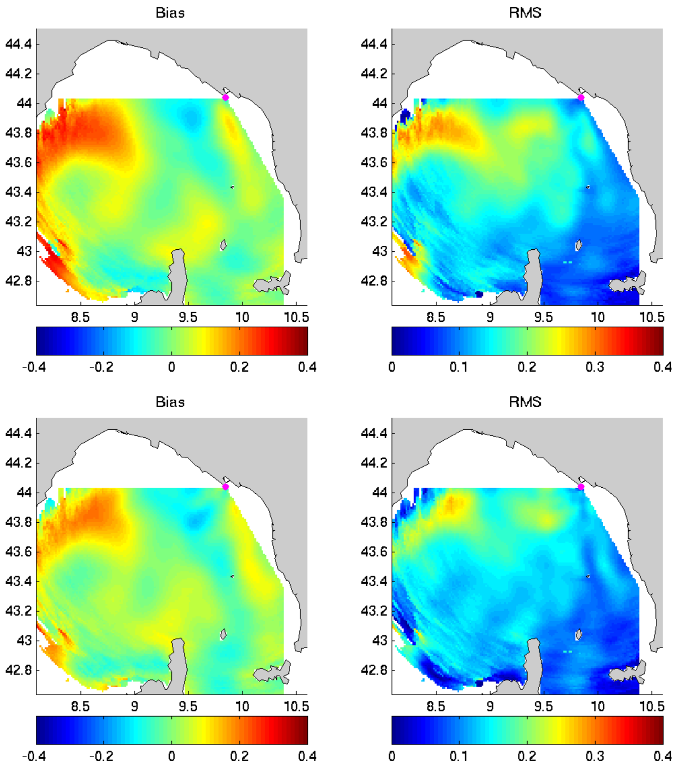



- Both approaches equivalent for linear system (and additive noise)
- Unrealistic "ensemble extrapolation" when too small observation errors are used $\rightarrow$ model trajectory and forcing fields are inconsistent

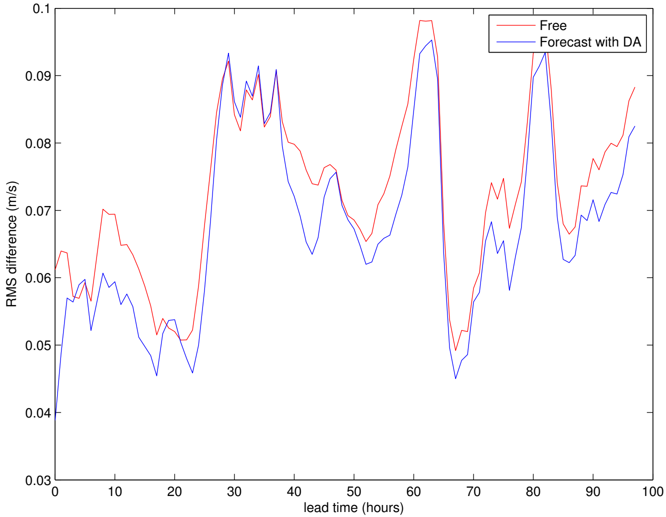
# Error statistics for Palmaria Site

Without assimilation (positive values: current away from the magenta dot)

With assimilation



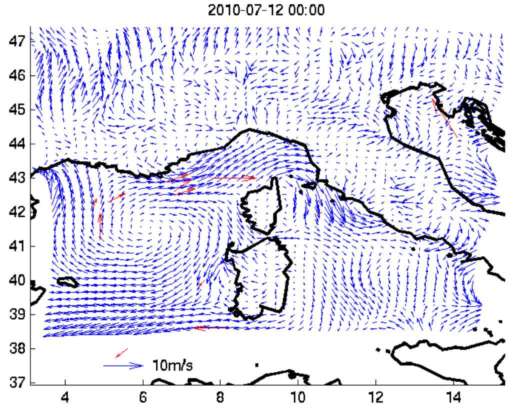
# Forecasts




- Impact of data assimilation on current forecast
- Comparison with surface currents from Palmaria
- HF radar assimilation improves the strength of the Northern Current and this improvement persists for some time.

# Simulation with atmospheric model (WRF)

- blue arrows: WRF 10m wind vectors, red arrows: in situ wind measurements from ICOADS (International Comprehensive Ocean-Atmosphere Data Set).wind_LS2.mp4
- 3 WRF domains at 30, 10, 3.33 km resolution (two-way nesting). The limit of those domains are shown in black.
- 30-km grid model nested (one-way) into the Global Forecast System
- 28 vertical layers



# Model results with different wind forcings

- Total RMS differences (m/s):

| | Pal. | Ros. |
|---|---|---|
| COSMO | 0.14 | 0.11 |
| WRF | 0.13 | 0.14 |

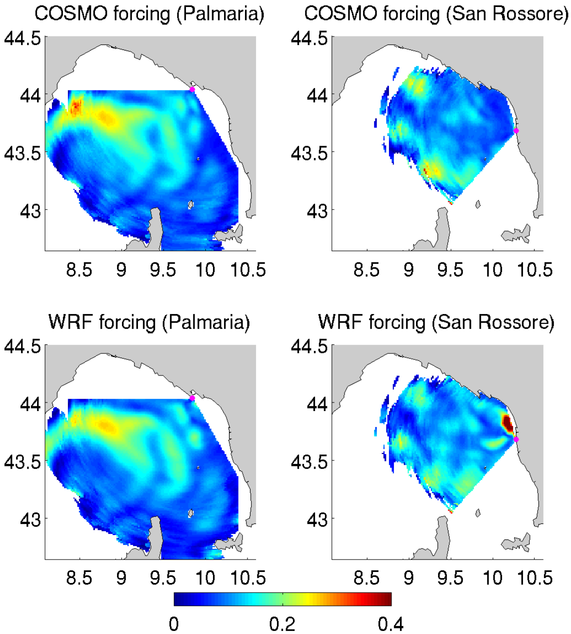



Figure 4: Radial surface current RMS difference

# Ocean Assimilation Kit

- Reduced rank square root analysis
- Global and local algorithm
- Modular Fortran 90 program
- Flexible definition of state vector
- Supports arbitrary curvilinear grid
- Local algorithm parallelized with OpenMP and MPI
- NetCDF or Fortran binary files as input

```
Model.variables = [                  'zeta',          'temp',         'salt']
Model.gridX     = ['domain.nc#lon(:,:,end)','domain.nc#lon','domain.nc#lon']
Model.gridY     = ['domain.nc#lat(:,:,end)','domain.nc#lat','domain.nc#lat']
Model.gridZ     = [  'domain.nc#z(:,:,end)',  'domain.nc#z',  'domain.nc#z']
Model.mask      = [  'domain.nc#z(:,:,end)',  'domain.nc#z',  'domain.nc#z']
Model.path      = '/home/user/Data/'

Syntax:
NetCDF_filename.nc#NetCDF_variable(index list)
```

# Ocean Assimilation Kit - Observations

```
Obs001.time       = '2010-07-06T00:30:00.00'  ! time as YYYY-MM-DDThh:mm:ss
Obs001.path       = 'Obs/'                     ! where the file can be found
Obs001.variables  = [               'TEM']    ! name as in Model.variables
Obs001.names      = [      'temp_profile']    ! descriptive name
Obs001.gridX      = [       'obs1.nc#lon']    ! longitude
Obs001.gridY      = [       'obs1.nc#lat']    ! latitude
Obs001.gridZ      = [         'obs1.nc#z']    ! depth
Obs001.value      = [      'obs1.nc#temp']    ! value of the observations
Obs001.mask       = [      'obs1.nc#temp']    ! mask of the observations
Obs001.rmse       = [ 'obs1.nc#temp_rmse']    ! root mean square error
```

# Implementation

Compact algorithm using Fortran operators:

```
Hxf = H.x.xf
increment = Sf.x.(U.x.(lambda.dx.(U.tx.(HSf.tx.(invsqrtR**2*(yo-Hxf))))))
```

Definition of operators .x., .tx., .dx and sparse matrix type H

# Conclusions

- Ensemble assimilation methods require realistic perturbation schemes (error co-variances) which can be based on dynamical relationships (similar to Variational analysis)
- Tidal boundary conditions can be constrained by HF radar measurements.
- Correcting tidal boundary conditions avoids (or at least reduces) systematic errors in the model solution.
- Similar approach can also be used to adjust wind forcings using HF radar data.
- Embedding the time dimension into the state vector leads to a smoother scheme (which is very simple to implement)
- Smoother schemes can be used to estimate the optimal model trajectory or forcing field
- Both approaches are not equivalent for non-linear systems or multiplicative noise

# References


Barth, A., A. Alvera-Azcárate, K.-W. Gurgel, J. Staneva, A. Port, J.-M. Beckers, and E. V. Stanev, 2010: Ensemble perturbation smoother for optimizing tidal boundary conditions by assimilation of High-Frequency radar surface currents - application to the German Bight. *Ocean Science*, **6**, 161–178, doi:10.5194/os-6-161-2010. URL http://www.ocean-sci.net/6/161/2010/

Barth, A., A. Alvera-Azcárate, J.-M. Beckers, J. Staneva, E. V. Stanev, and J. Schulz-Stellenfleth, 2011: Correcting surface winds by assimilating High-Frequency Radar surface currents in the German Bight. *Ocean Dynamics*, **61**, 599–610, doi:10.1007/s10236-010-0369-0. URL http://hdl.handle.net/2268/83330

Burchard, H. and K. Bolding, 2002: GETM – a general estuarine transport model. Scientific documentation. Technical Report EUR 20253 EN, European Commission.

Chen, Y. and C. Snyder, 2007: Assimilating Vortex Position with an Ensemble Kalman Filter. *Monthly Weather Review*, **135**, 1828–1845.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, **230**, 112–126.

Kalnay, E., 2002: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 1 edition.

Sakov, P., G. Evensen, and L. Bertino, 2010: Asynchronous data assimilation with the EnKF. *Tellus*, **62A**, 24–29.

Staneva, J., E. V. Stanev, J.-O. Wolff, T. H. Badewien, R. Reuter, B. Flemming, A. Bartholomä, and K. Bolding, 2009: Hydrodynamics and sediment dynamics in the German Bight. A focus on observations and numerical modelling in the East Frisian Wadden Sea. *Continental Shelf Research*, **29**, 302–319.

van Leeuwen, P. J., 2001: An Ensemble Smoother with Error Estimates. *Monthly Weather Review*, **129**, 709–728.

# Perturbations scheme

The cost function is a quadratic function in $\mathbf{x}$ and can thus be written as:

$$
\begin{aligned}
2J(\mathbf{x}) &= \mathbf{x}^T(\mathbf{M}^T\mathbf{W}_M\mathbf{M} + \mathbf{D}^T\mathbf{W}_D\mathbf{D} + \mathbf{W}_E)\mathbf{x} \qquad (17)\\
&= \mathbf{x}^T\mathbf{B}^{-1}\mathbf{x} \qquad (18)
\end{aligned}
$$

where the matrix $\mathbf{B}$ (covariance matrix, not formed explicitly) is defined as:

$$\mathbf{B} = (\mathbf{M}^T \mathbf{W}_M \mathbf{M} + \mathbf{D}^T \mathbf{W}_D \mathbf{D} + \mathbf{W}_E)^{-1} \tag{19}$$

To generate an ensemble of perturbations that follows the previous pdf, the matrix $\mathbf{B}$ is decomposed in eigenvectors (rows of $\mathbf{U}$) and eigenvalues (diagonal elements of $\mathbf{\Lambda}$) :

$$\mathbf{B} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \tag{20}$$

The smaller an eigenvalue is, the stronger the corresponding eigenvector violates the dynamical and smoothness constraint.

An ensemble of vectors $\mathbf{z}^{(k)}$ where the subscript $k$ is the ensemble member, is created following a normal distribution.

$$\mathbf{z} \sim N(0, \mathbf{I}_n) \tag{21}$$

An ensemble of perturbations $\mathbf{x}^{(k)}$ following (1) can be obtained by:

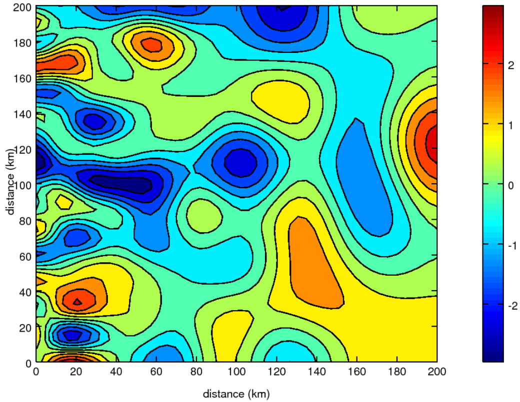
$$\mathbf{x}^{(k)} = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{z}^{(k)} \tag{22}$$

Alternatively, one can use the 2nd order exact re-sampling method (SEIK):

$$\mathbf{x}^{(k)} = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{H}_w(\Omega)_k \tag{23}$$

where columns of $\mathbf{H}_w$ are all perpendicular to the vector $\mathbf{1}_{N\times 1}$ and $(\Omega)_k$ is the $k$-column of a random orthogonal matrix $\Omega$.

- Also perturbations with a spatially varying correlation length can be created.
- Scale of mesoscale variability $\rightarrow$ internal radius of deformation which varies in space:



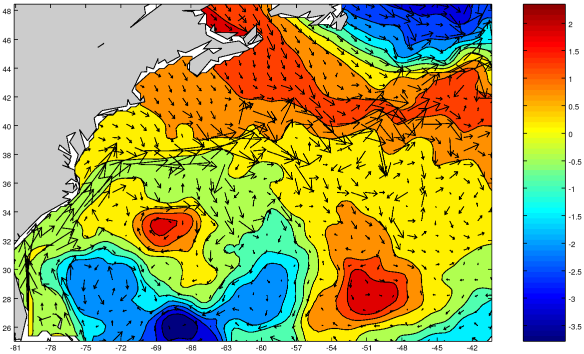

- Illustration of a random field with a variable correlation length.

# Examples for linear constraints

## Advection constraint

- For large-scale models, perturbations should be approximately stationary solutions to the advection equation

$$\mathbf{v} \cdot \nabla \phi = 0 \tag{24}$$



- Example of ensemble perturbations using the advection constraint

# Application to HF Radar assimilation in the German Bight (tidal BC)

- Only M2 tidal boundary conditions are perturbed:

$$\zeta^{(k)} = \zeta^{(b)} + \Re\left(\zeta'(x,y)\exp(i\omega t)\right) \tag{25}$$

where $\omega$ is the M2 angular frequency and $\zeta'(x,y)$ is a random field satisfying approximately the harmonic shallow water equations:

$$i\omega\zeta' + \frac{\partial(hu')}{\partial x} + \frac{\partial(hv')}{\partial y} = 0 \tag{26}$$

$$i\omega u' - fv' + g\frac{\partial \zeta'}{\partial x} = 0 \tag{27}$$

$$i\omega v' + fu' + g\frac{\partial \zeta'}{\partial y} = 0 \tag{28}$$

- The 50 eigenvector with the largest eigenvalues of the matrix $\mathbf{B}$ from (20) are calculated (providing the spatial structure of the perturbation).

- From those 50 eigenvector/eigenvalues an ensemble of 51 members is created with zero mean (2nd order exact re-sampling).
- The GETM model is run for 40 days with each of those perturbed boundary values.
- Observations are assimilated with an expected RMS error of 0.3 m/s (including representativity error and error that cannot be corrected modifying only the boundary conditions) providing an optimal increment of the boundary values.
- The model is rerun with the optimized boundary values for 60 days.