Possible roles of the data assimilation framework towards quantitative ocean science

P. De Mey, LEGOS, Toulouse, France

COST-ESF Workshop on Coastal Model Validation
Brest, France, November 2009

Quantitative ocean science $\cong$ ocean science with error bars

Outline

- Introduction
  - How can DA help validation?
- Built-in tools
  - Short-term prediction (NIV)
  - Ensemble-based statistics
- Are model-data misfits within nominal range?
  - Internal consistency criterion (Jmin)
- Array design for the detection of model errors
  - Array design and model validation are interdependent
  - The RMSpectrum method (Le Hénaff et al., 2009)
- Probabilistic model testing
- Conclusion

How can DA help validation?

- Validation generally understood as “validating a model wrt. observations”
- But observations are not truth!
  - Measurement errors are often not well known (drifts, biases, saturation), and their models are imperfect
  - The transfer function of an instrument can critically depend on environmental factors (e.g. particle traps) or electronics
  - The time response of an instrument can contain transients which can be impossible to identify as such (e.g. altimeter range trackers)
  - Observation operators (viewed as relating observed variables with variables of interest) are sometimes complex engineering models by themselves (most remote sensing data)
  - Measurements do not always observe the same processes as the model (representativity errors: e.g. Lagrangian data/invariants)
  - Observation operators (viewed as relating observed variables with model variables) are imperfect in their handling of subgridscale processes

  --> Need to take into account both model errors and observation errors in order to check whether difference is in acceptable range
- Data assimilation provides a theoretical framework matching imperfect models with imperfect observations (“innovation”)
  - Match through observation-space statistics (GODAE Class 4)

Extended Kalman Filter (e.g. Gelb, 1974)

Discrete dynamical system model

$$\mathbf{x}_k = M_{k-1}\left(\mathbf{x}_{k-1}\right) + \eta_{k-1} \quad \text{with } \eta_k \approx N(0,\mathbf{Q}_k) \qquad \text{(MOD1)}$$

Observation model

$$\mathbf{y}_k^o = H_k\left(\mathbf{x}_k\right) + \varepsilon_k \quad \text{with } \varepsilon_k \approx N(0,\mathbf{R}_k) \qquad \text{(MOD2)}$$

"Optimal" state estimate obtained through EKF:

State forecast $\quad \mathbf{x}_k^f = M_{k-1}\left(\mathbf{x}_{k-1}^a\right) \qquad$ (EKF1)

Error forecast $\quad \mathbf{P}_k^f = \mathbf{M}_{k-1}\mathbf{P}_{k-1}^a\mathbf{M}_{k-1}^T + \mathbf{Q}_{k-1} \qquad$ (EKF2)

State update $\quad \mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k\left(\mathbf{y}_k^o - H_k(\mathbf{x}_k^f)\right) \qquad$ **Innovation** (EKF3)

Error update $\quad \mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^f \qquad$ (EKF4)

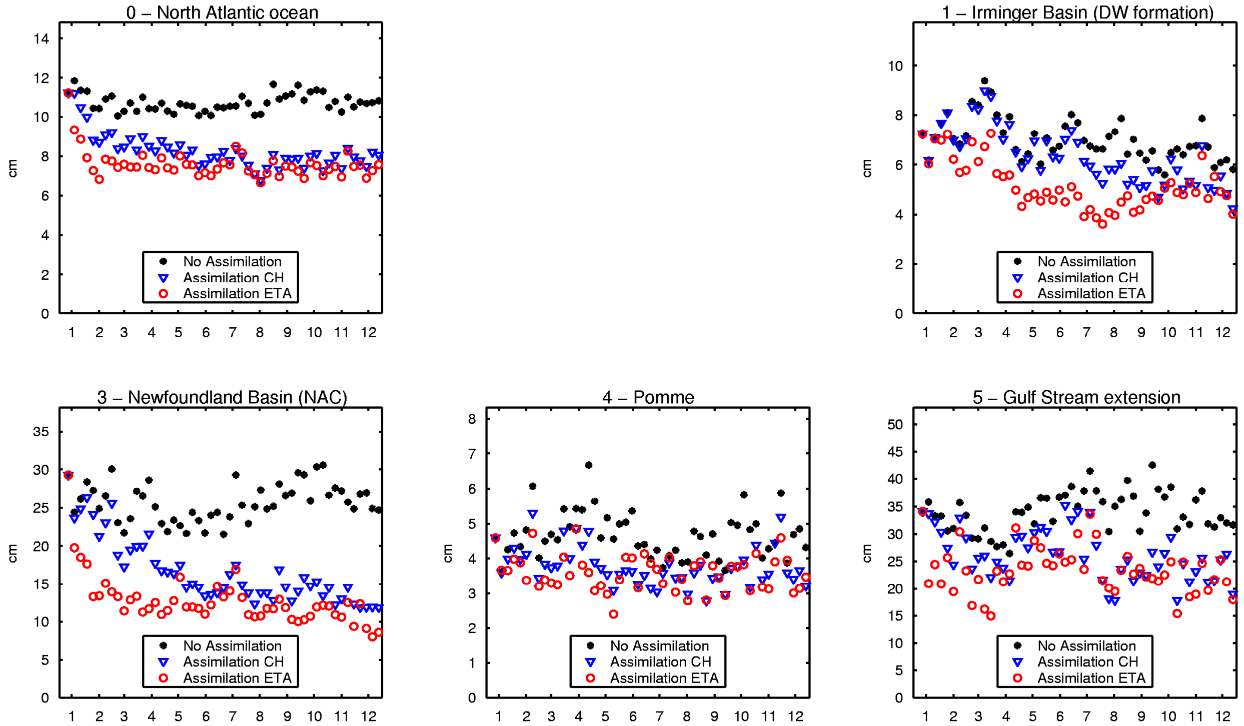
Kalman gain $\quad \mathbf{K}_k = \mathbf{P}_k^f\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k)^{-1} \qquad$ **Innovation covariance** (EKF5)

from e.g. $\quad \min\left\{J_k = \mathrm{trace}(\mathbf{P}_k^a)\right\}$

DA framework can be used even without actually assimilating observations!

Innovation vector statistics (NIV = $||y^o-H(x^f)||$)

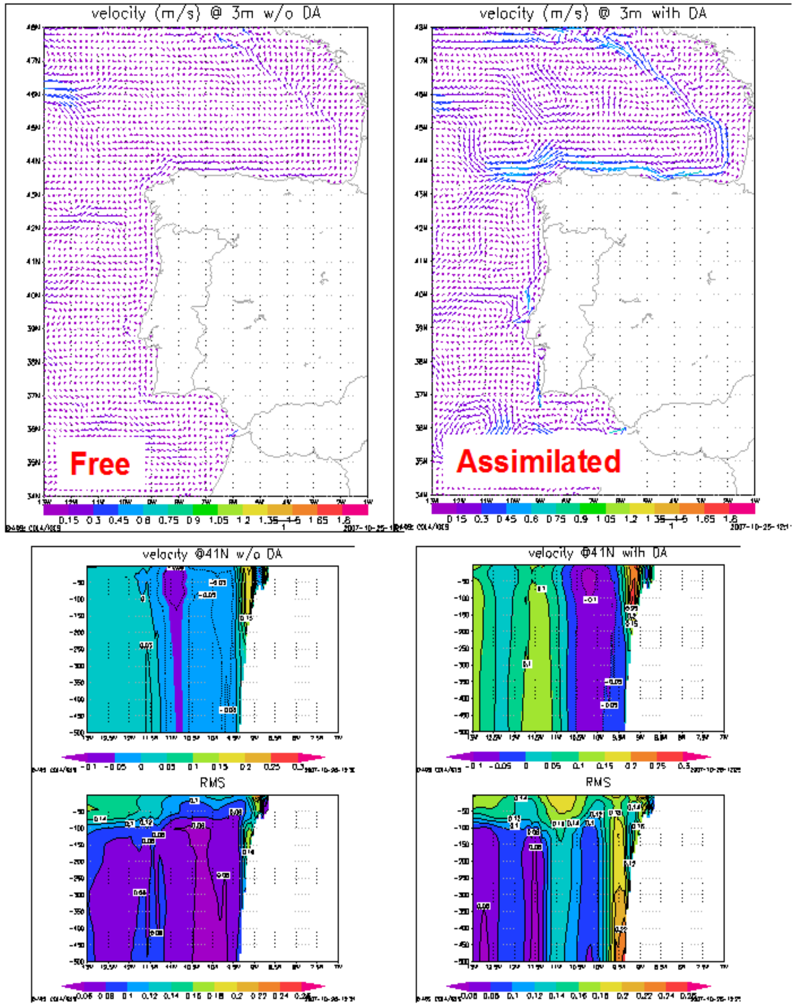
- Isopycnal DA in North Atlantic 3D-OGCM (OPA/NEMO), 1993
- Statistics wrt. not-yet-assimilated (not independent) observations
- Objective: visualize gain in short-term predictive skill in observation space, in particular assessing how the numerical model cooperates



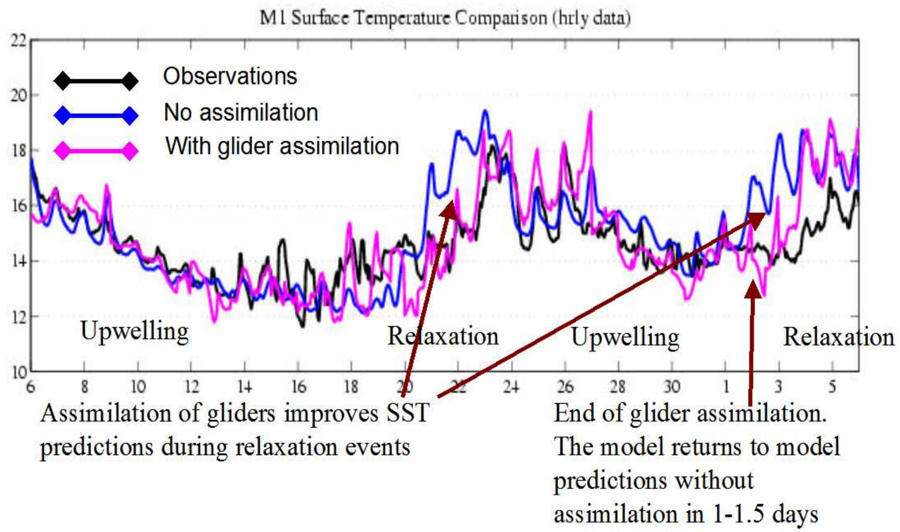
RMS(MISFIT) with MISFIT(t)=DATA(t)–MODEL(t) – YEAR 1993
Assimilated data: along–tracks TP/ERS w/ref. 93–95 – 51 cycles of 7 days
DATA: Topex/Poseidon – MODEL: MNATL–07 ; 20S–70N ; 1/3 degree

(Faucher, Gavart, and De Mey, 2001)

Visual/physical consistency : impact of assimilation



- Use pre-existing knowledge or independent observations
- Performance enhancement by assimilation validates choice of processes to control, and provides an indirect criticism over the original model features

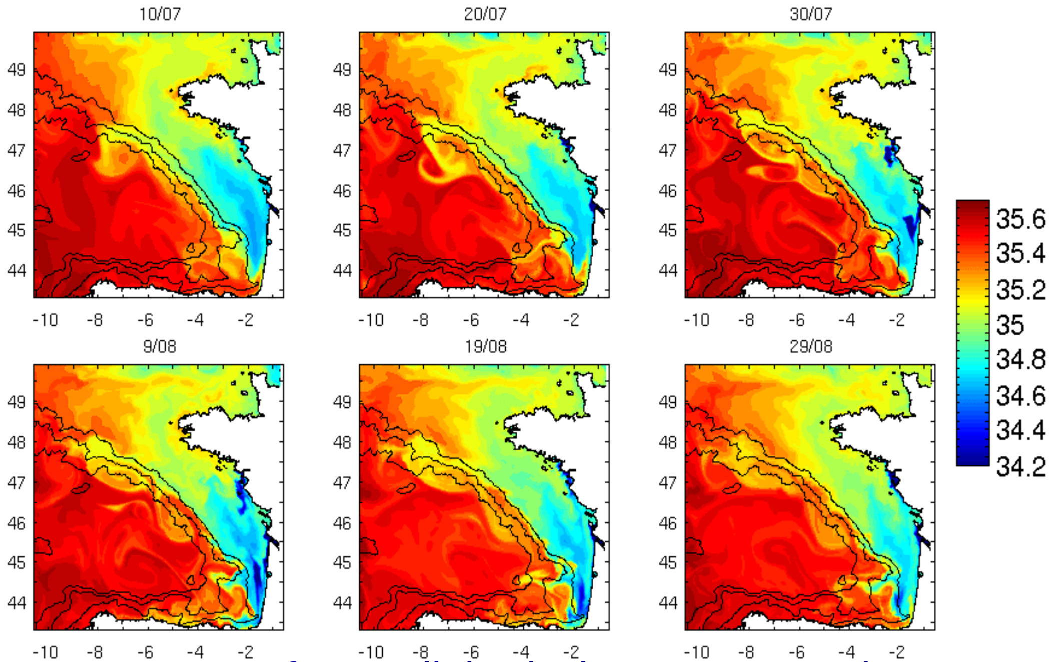


DA enhances Iberian Poleward Current + Bay of Biscay slope circulation (Fernandez and De Mey, ESEOO/LEGOS)

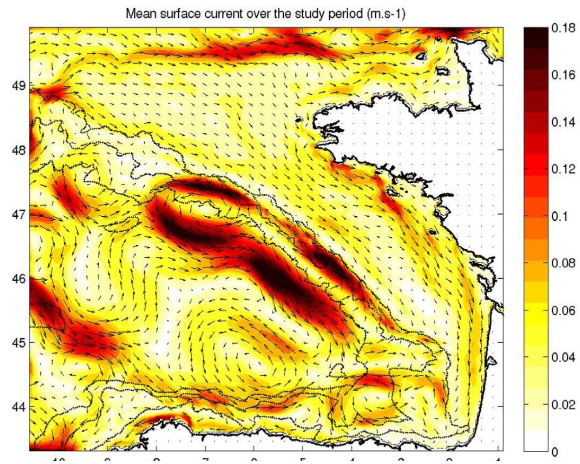
Independent observations: positive impact of assimilation of gliders off Monterey Bay (Kindle, NRL)

Ensemble-based statistics: Bay of Biscay (BoB) configuration

- SYMPHONIE 3DFD, 3-km horizontal resolution, free surface, sigma-step vertical scheme (41 levels max), major river runoff, tidal friction
- Downscaled from MERCATOR PSY2v3 (1/12°)
- Cyclonic slope circulation, anticyclonic recirculation
- Mesoscale activity above abyssal plain
- Integrated in Ensemble Kalman Filter (BELUGA; De Mey, 2008)



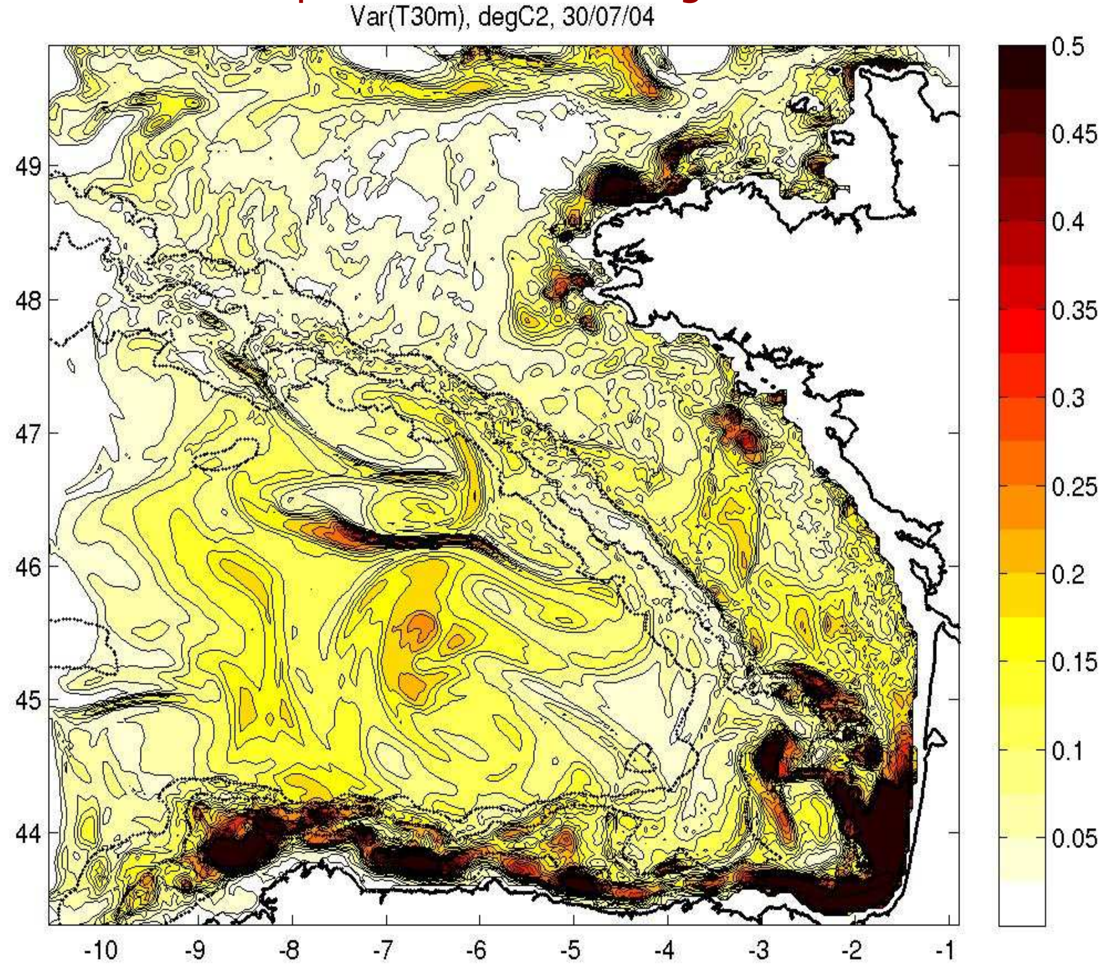
Sea-surface salinity (July-August 2004)



BoB mean surface currents (July-August 2004) & bathymetry

Ensemble variance as a proxy tracer of model state error

BoB 3DFD 30m temperature ensemble variance, July 30, 2004
Response to wind forcing errors



A proxy of state error variance and of its time evolution as a tracer
-> mix of shelf, shelf-break, upwelling and mesoscale responses

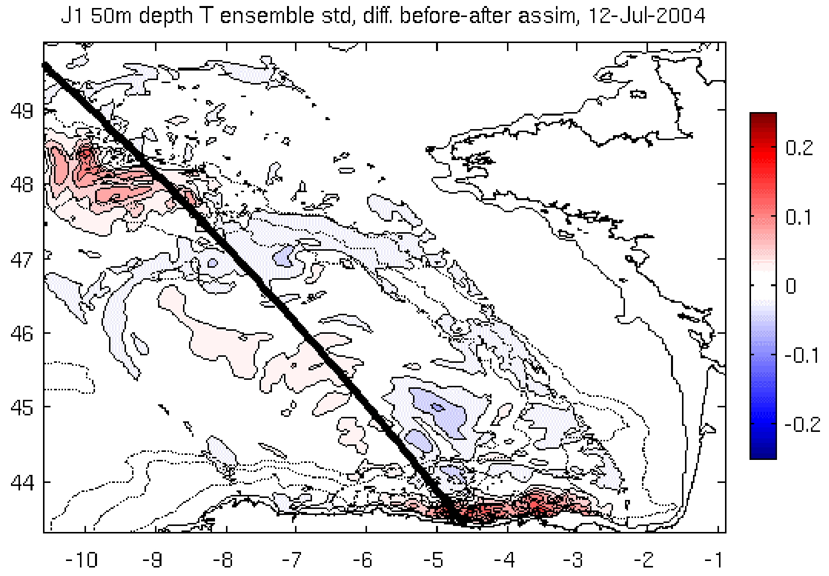
(Le Hénaff & De Mey, 2008)

Posterior ensemble variance:
Impact of assimilating wide-swath vs. nadir altimeter data

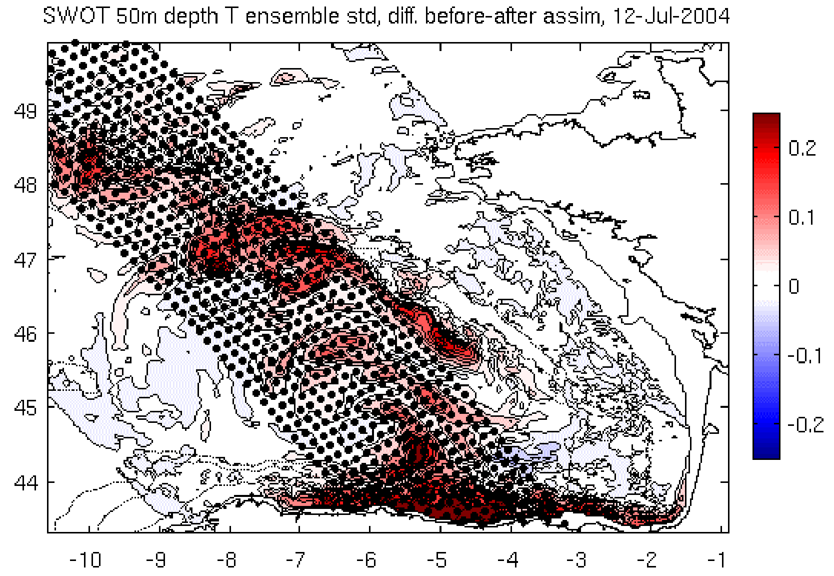
EnKF, ocean-only state vector

Reduction of BoB 50-m temperature ensemble variance at analysis time (˚C rms) in presence of wind forcing errors

Jason-1



SWOT on Jason-type orbit



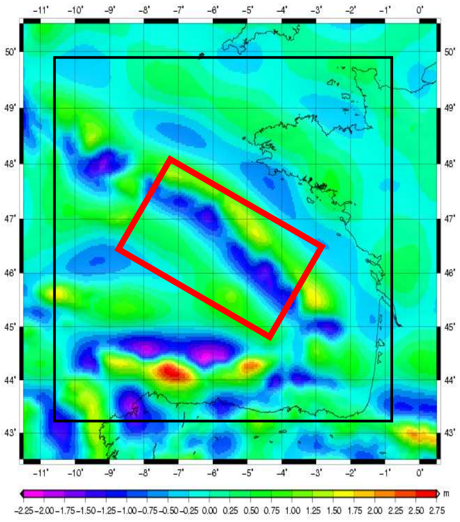
- Performance enhancement by assimilation validates choice of processes to control, and provides an indirect criticism over the original model features

(Le Hénaff & De Mey, 2009)

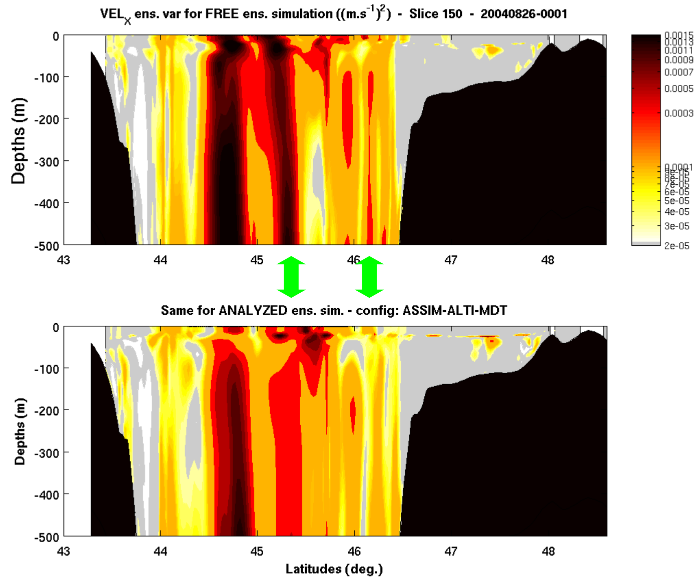
Posterior ensemble variance:
impact of assimilating the GOCE geoid

- The signal associated with the slope current is small and at the limit of the ability of GOCE to detect + the geoid omission errors over the slope are large
  - Use guess of slope current, e.g. model forecast
- Simultaneous assimilation of simulated SLA and GOCE geoid estimate with EnKF
- Reduction of posterior ensemble variance in coastal current system

GOCE omission error estimate



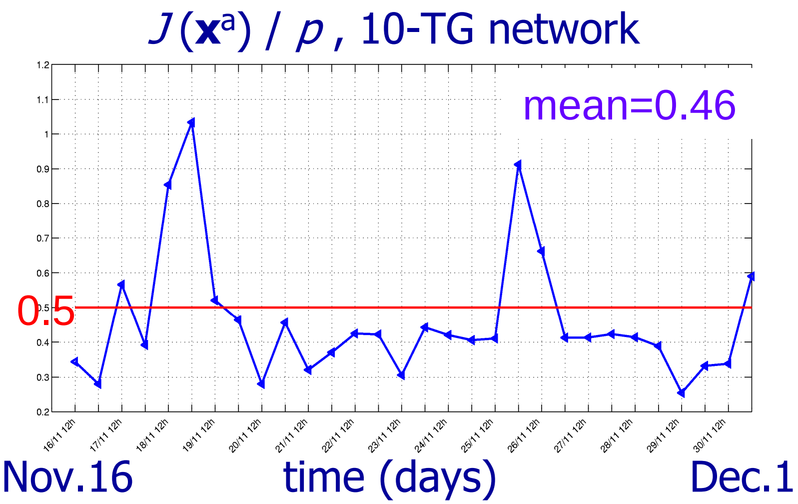
Velocity ensemble variance along N-S section, BoB



(Lamouroux and De Mey, 2009)

Are model-data misfits within expected range?

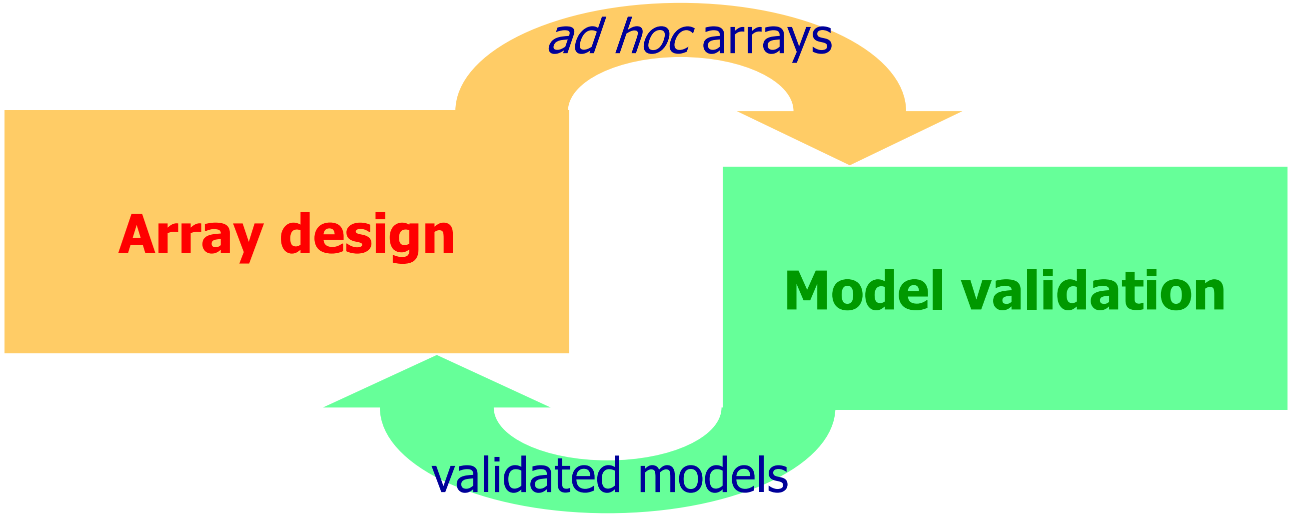
- Ensemble OI configuration in Northeast Atlantic (T-UGOm)
- Parameterize atmosphere-driven system errors using 100 time-independent multivariate ensemble EOFs (“reduced-order” approach)
- Use assimilation to correct state (ocean + surface atm. variables)
- Use Jmin=p/2 internal consistency criterion (Talagrand, 1999, ECMWF rep.):
  $\mathbf{R} + \mathbf{H}\mathbf{P}^{f}\mathbf{H}^{T}$ <-> <$\mathbf{d}\mathbf{d}^{T}$>



- Although overall agreement is good, overfitted data outside of extreme events probably mean that the specification or prior error estimates must be refined

(Lamouroux and De Mey, 2006)

Array design and model validation are interdependent



- One reasonable criterion for array design is to ensure a better detection of model errors, for model validation & assimilation
- In turn, models which are used to derive realistic criteria of array performance must be realistic and therefore validated against data
- Because it is in a (relatively) early stage, coastal ocean forecasting provides an opportunity to develop both components together

Array design & optimization

- What for?
  - For the resolution of processes of interest (unknown/poorly known)
    - Guess: none or climatology
    - Physical sense
  - For validation & assimilation
    - Guess: prior simulation
    - Detection of model errors & impact on analysis
    - Control of model variables for better forecasting (constraint of “null space”)
- How?
  - OSEs/OSSEs
  - Targeted observations (e.g. Langland, QJRMS, 2005), adjoint sensitivity, etc.
  - Adaptive sampling (e.g. Palmer et al., JAS, 1998)
  - Representer-based analysis (Bennett, JGR, 1990; Le Hénaff et al., OD, 2009), sometimes using stochastic (Ensemble) modelling
  - …

Can we optimize arrays for detecting model errors?
(with later objectives of validation and assimilation)

$\mathbf{x}$ "augmented state vector" $(n,1)$

$\mathbf{y}^o$ "observations" $(p,1)$ verifying $\mathbf{y}^o = H(\mathbf{x}^t) + \varepsilon$, with:

$H(\,)$ "observation operator"

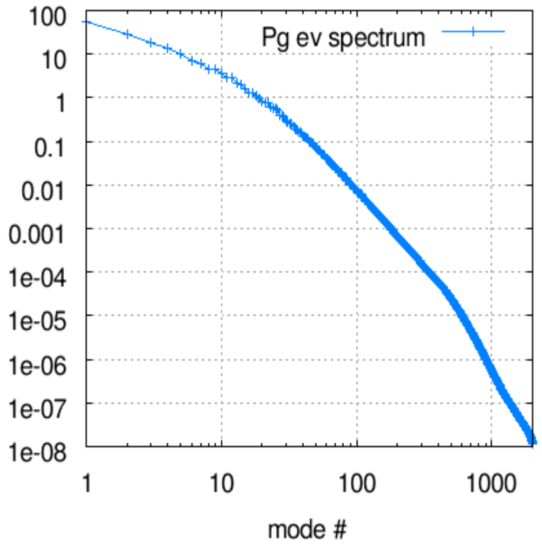
$\varepsilon \in N(0, \mathbf{R})$

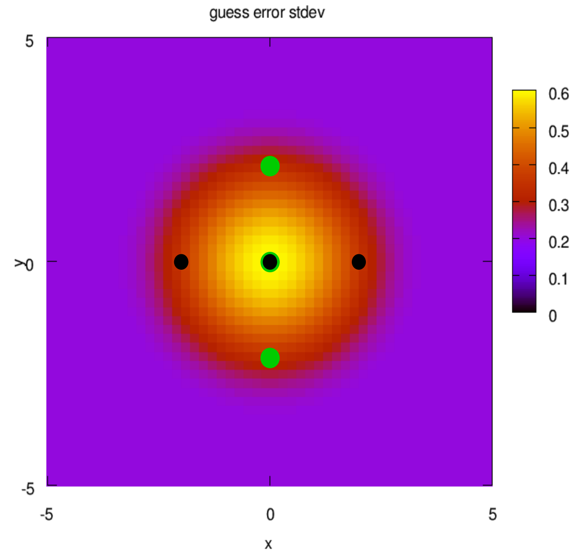
Q: which of $(H_1, \mathbf{R}_1)$ and $(H_2, \mathbf{R}_2)$ is "better"?

Assume we have an *a priori* estimate – a "guess" – of $\mathbf{x}$ and associated error statistics (if not, any observational array will bring valuable information proportionately to its cost):

$\mathbf{x}^g = \mathbf{x}^t + \eta$, with:

$\eta \in N(0, \mathbf{P}^g)$

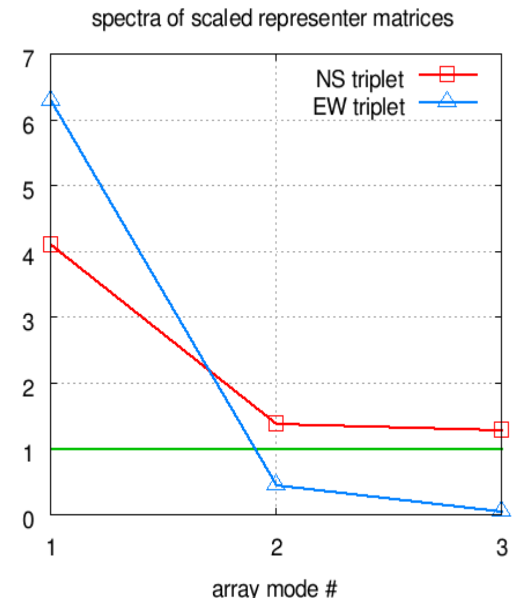




(Le Hénaff & De Mey, ODYN, 2009)

Detection of model state errors with RMSpectrum: Formal criteria

Formalize intuitive order relationship…



"array modes" (e.g. Bennett et al., 1997): orthonormal rotation vectors $\boldsymbol{\mu}$ obtained by diagonalizing the representer matrix:

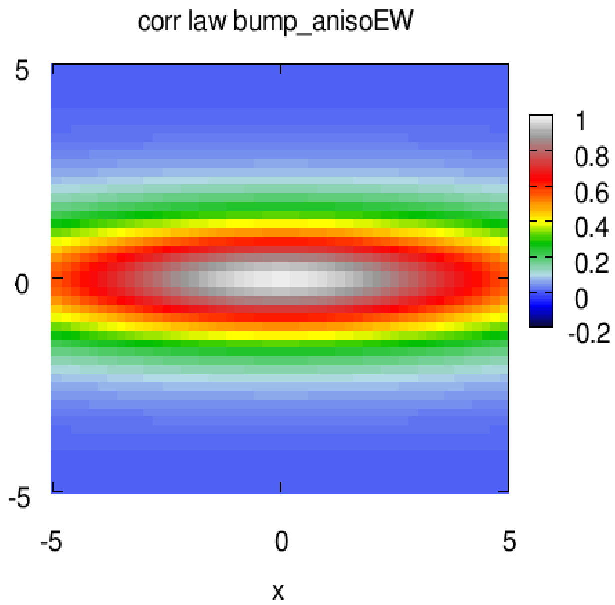
$$\mathbf{HP}^g\mathbf{H}^T = \boldsymbol{\mu\sigma\mu}^T$$

$\boldsymbol{\mu}$: "observable degrees of freedom" of the physical system for that configuration

$\boldsymbol{\sigma}$: spectrum of RM ←→ spectrum of **R**

In general case of non-homogeneous, non-diagonal **R**, use spectrum $\boldsymbol{\sigma}$' and array modes $\boldsymbol{\mu}$' of scaled representer matrix $\boldsymbol{\chi}$:

$$\boldsymbol{\chi} = \mathbf{R}^{-1/2}\mathbf{HP}^g\mathbf{H}^T\mathbf{R}^{-1/2} = \boldsymbol{\mu}'\boldsymbol{\sigma}'\boldsymbol{\mu}'^T$$

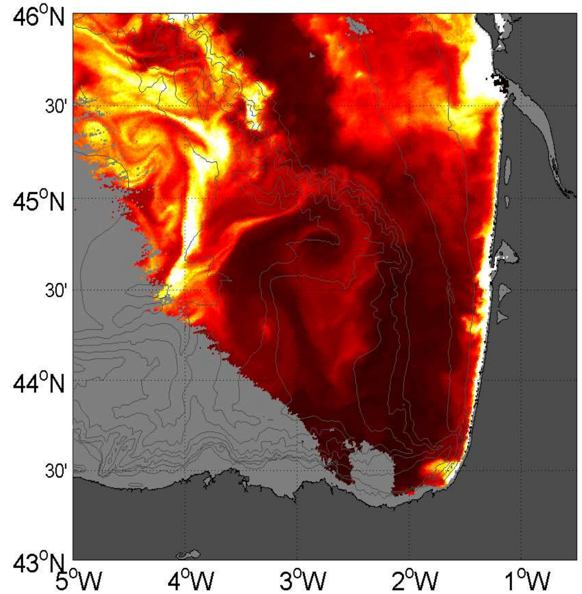


→ Quantitative (albeit complex) criterion of array performance
Takes into account space/time sampling characteristics of array

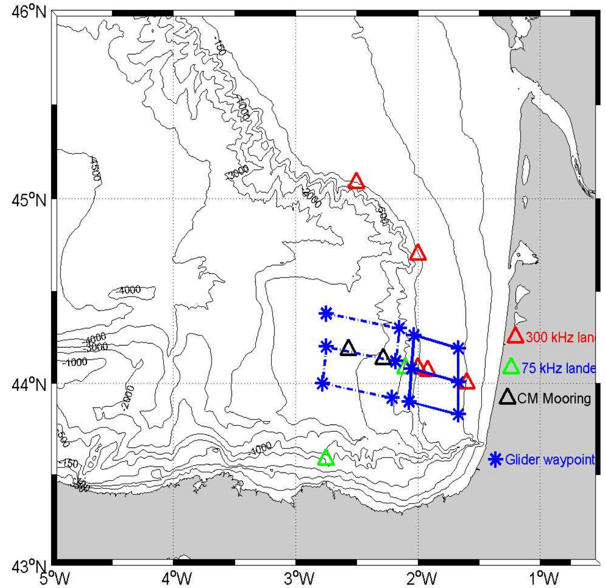
(Le Hénaff & De Mey, ODYN, 2009)

Detection of model state errors with RMSpectrum: Landes plateau example

- A few years ago, the former AQUILAND project (le Cann et al., 2007) aimed at studying mass and property exchanges across continental shelf break
- Secondary objective was the IPC, Navidad event, and associated "swoddies"
- Observing array planned included coastal and deep gliders, fixed-point profilers, a CM/ADCP cross-shelf transect, and more CM moorings to North and West
- Evaluate several array options with RMspectrum



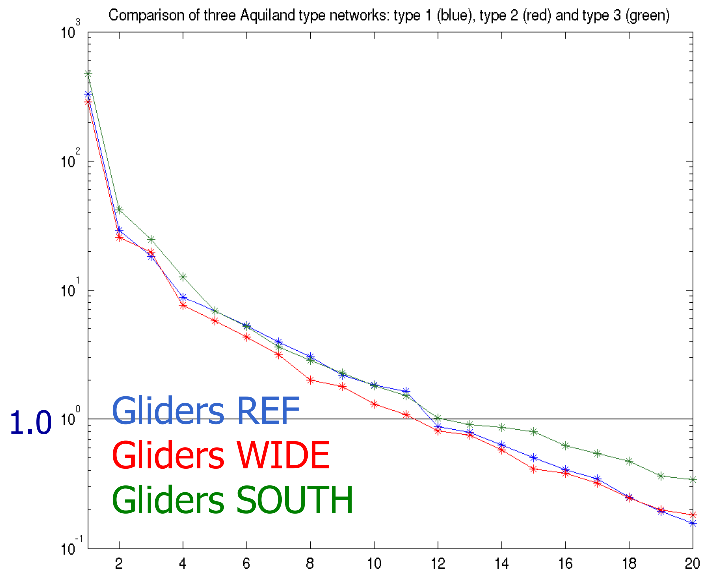
Chl - Apr 29, 2006 - AQUA/MODIS



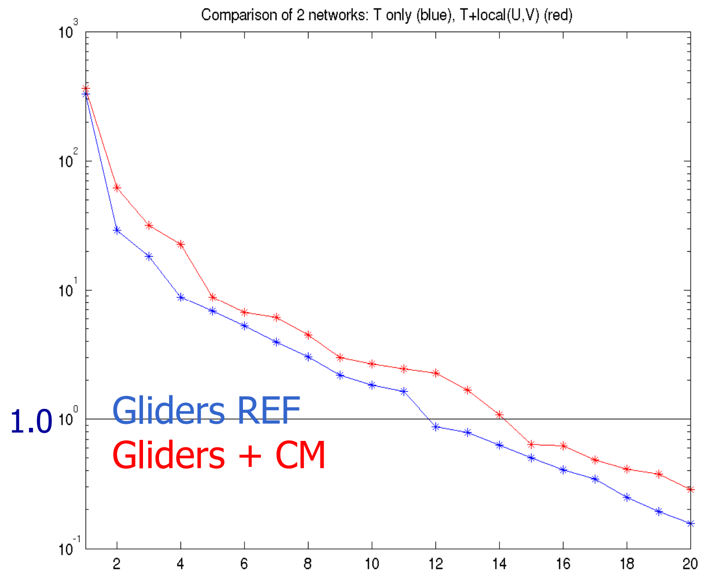
AQUILAND formerly projected array

Detection of model state errors with RMSpectrum: Landes plateau array

Glider array options



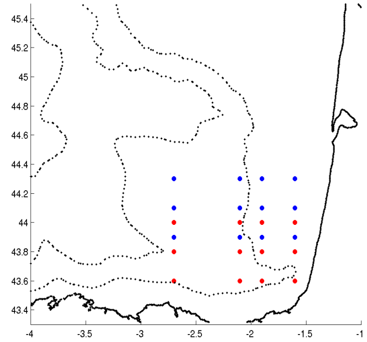
Added benefit of central CM transect



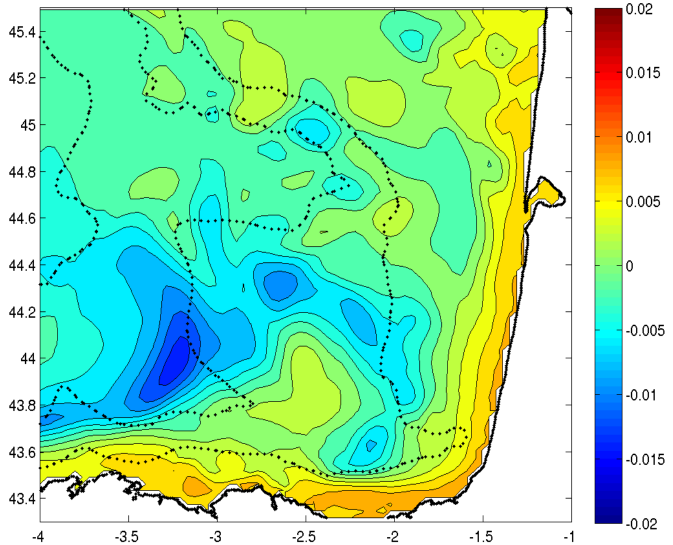
Spectra of scaled RM (log scale)

(Le Hénaff & De Mey, 2007)

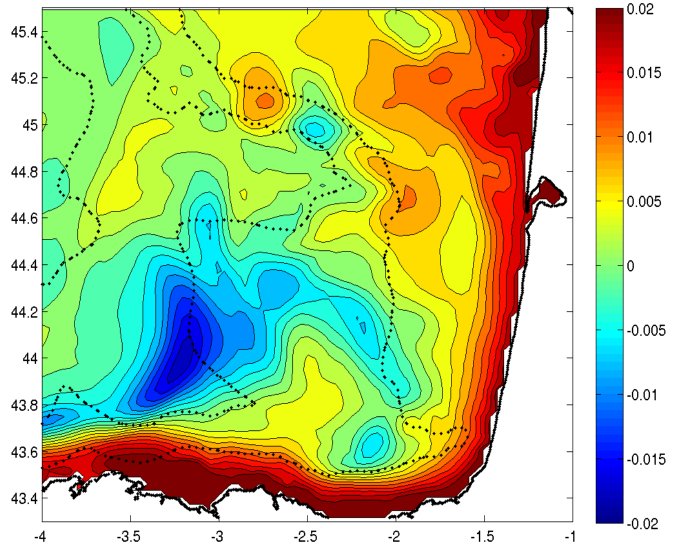
Detection of model state errors with RMSpectrum: Landes plateau array



Gliders REF



Gliders SOUTH



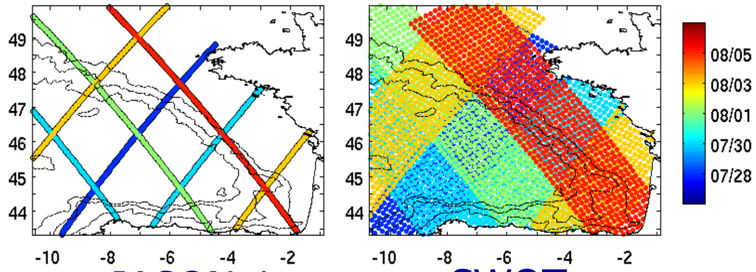
Modal representers for array mode 2 - Surface elevation
Same color scheme

"modal representers" $\boldsymbol{\rho}_{\mu} = \mathbf{P}^{g}\mathbf{H}^{T}\mathbf{R}^{-1/2}\boldsymbol{\mu}' =$ representers for the array modes

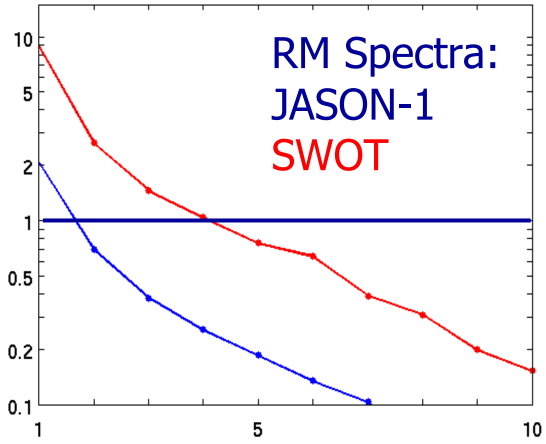
(Le Hénaff & De Mey, 2007)

Detection of model state errors with RMSpectrum: Wide-swath vs. nadir altimeter

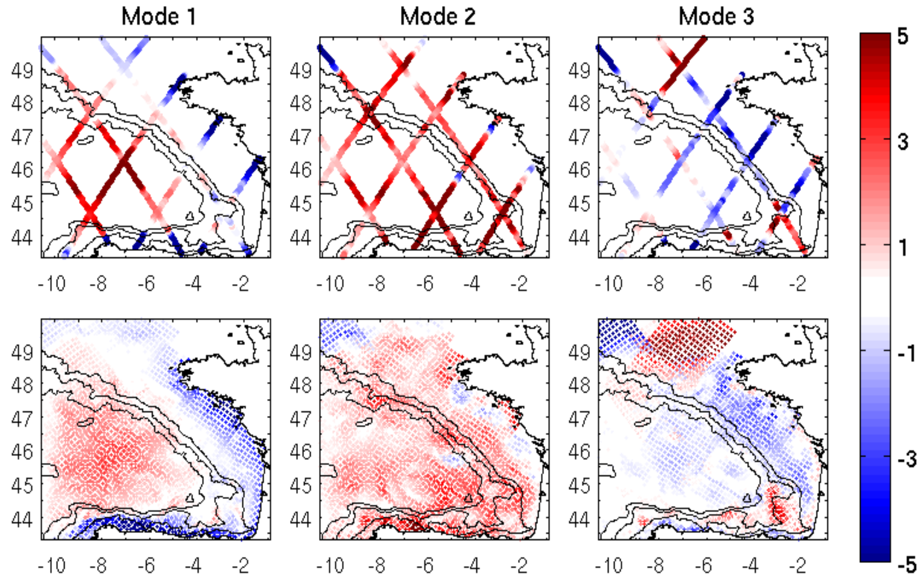
Space-time sampling schemes



JASON-1 SWOT



Array Modes (SLA)



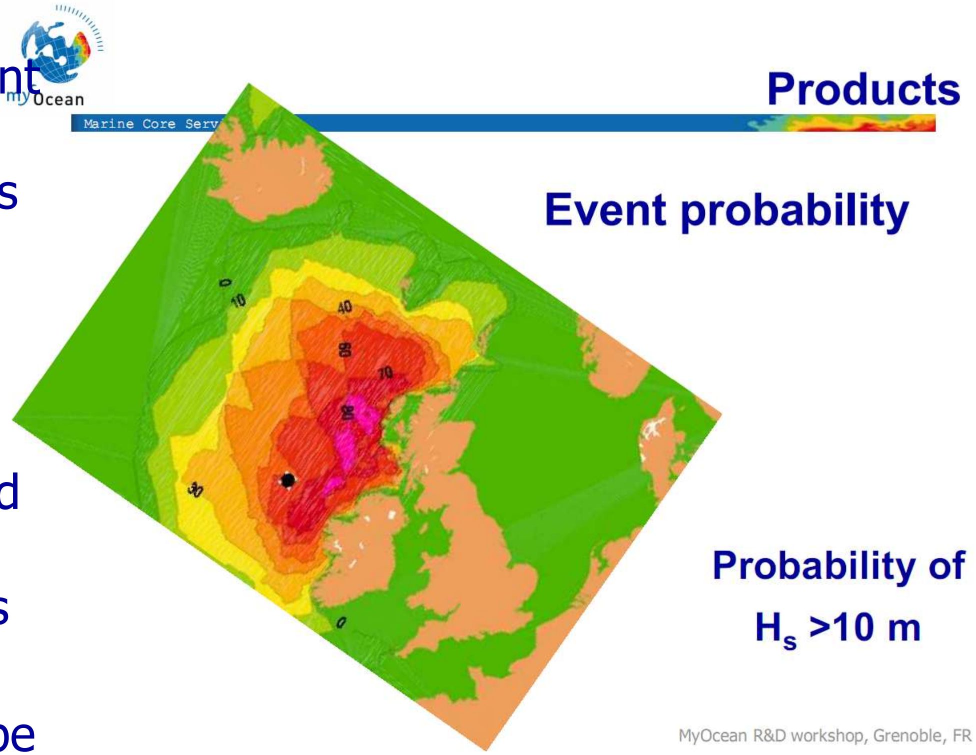
Stochastic response
to Gaussian wind errors
SYMPHONIE 3D BoB model
MERCATOR nesting - Summer 2004

-> SWOT can constrain coastal mesoscale errors + some HF errors on the shelf

(Le Hénaff & De Mey, ODYN, 2009)

Probabilistic model testing

- Array design results will be relevant only as long as the model is realistic enough, and as long as its error sources have been correctly identified
- When enough real observations become available, the validity of the stochastic model can be tested via the notion of reliability of its short- and medium-term forecasts (Toth et al., 2003)
- Its probability forecasts can also be validated with a similar approach (Melsom, MyOcean WP3)



(Melsom, Met.no, 2009)

- E.g. modified CRPS approach (CRPS: Continuous Ranked Probability Score: Hersbach, 2000)
  - Estimate the reliability of forecasts for specific events
  - Estimate the reliability of forecasts in a reduced state space, in a way parallel to the RMSpectrum approach

Conclusion

- Data assimilation in coastal models has a vital role to play
  - not only as a tool to provide short-term forecasts,
  - but also for the rigor it brings to the framework of model and data uncertainties, and to the design of the development of coastal observing systems.
- Tools reviewed
  - Short-term prediction (NIV)
  - Posterior ensemble variance
  - Internal consistency criterion (Jmin) to check data misfits
  - Array design for the detection of model errors (& later validation/assimilation)
  - Probabilistic model testing
- Because it is in a (relatively) early stage, coastal ocean forecasting provides an opportunity to develop arrays and models consistently to each other

Thank you.