Met Office

Initial experiences in using ensemble information in data assimilation with the Met Office ocean forecasting (FOAM) system

Daniel Lea[1], Matthew Martin[1], James While[1], Jennifer Waters[1], Robert King[1], Anthony Weaver[2]

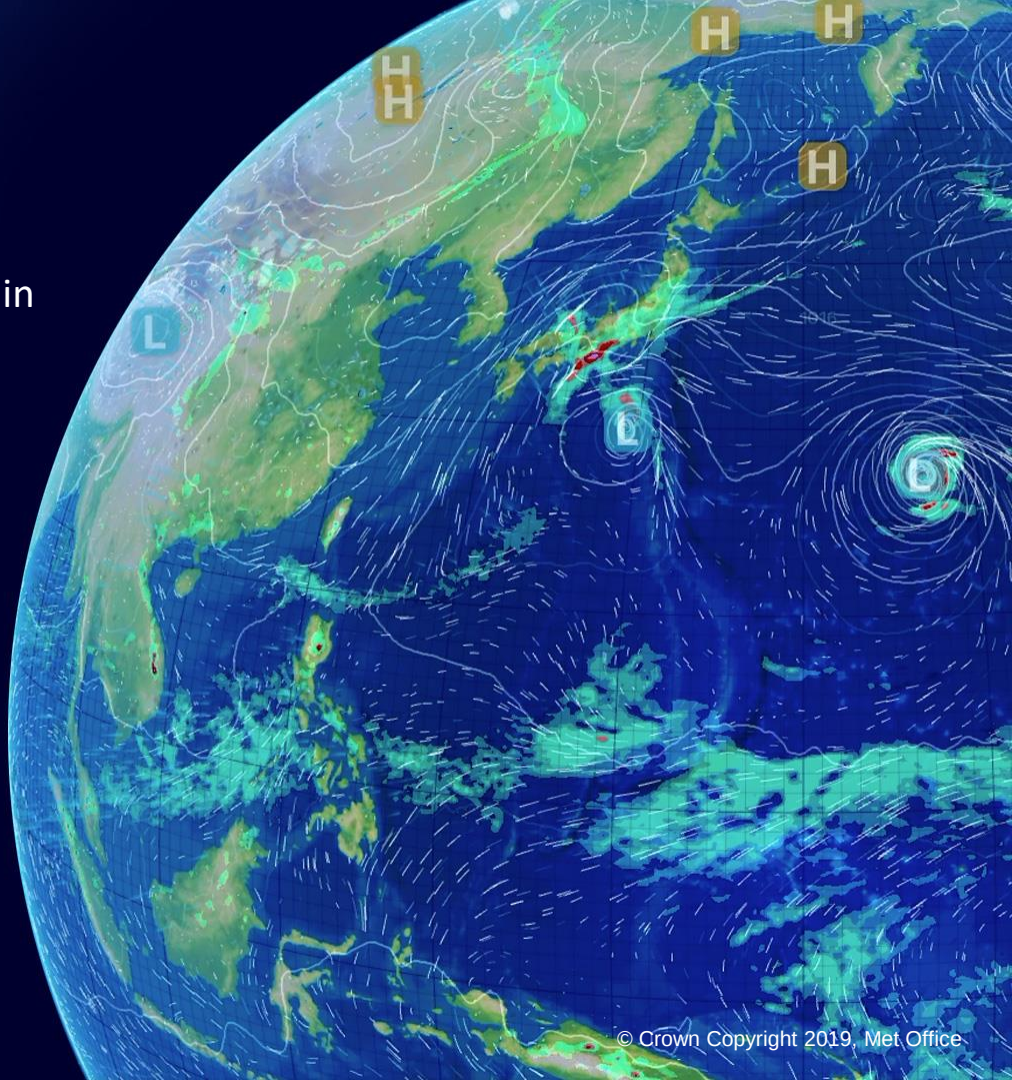
[1]Met Office, Exeter, UK

[2]CERFACS, Toulouse,

OceanPredict19

Halifax, Nova Scotia, Canada

7 May 2019



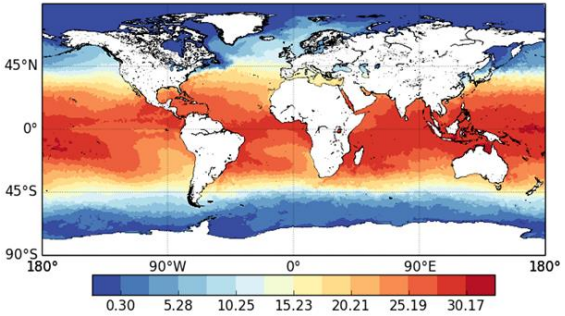
FOAM ensemble member

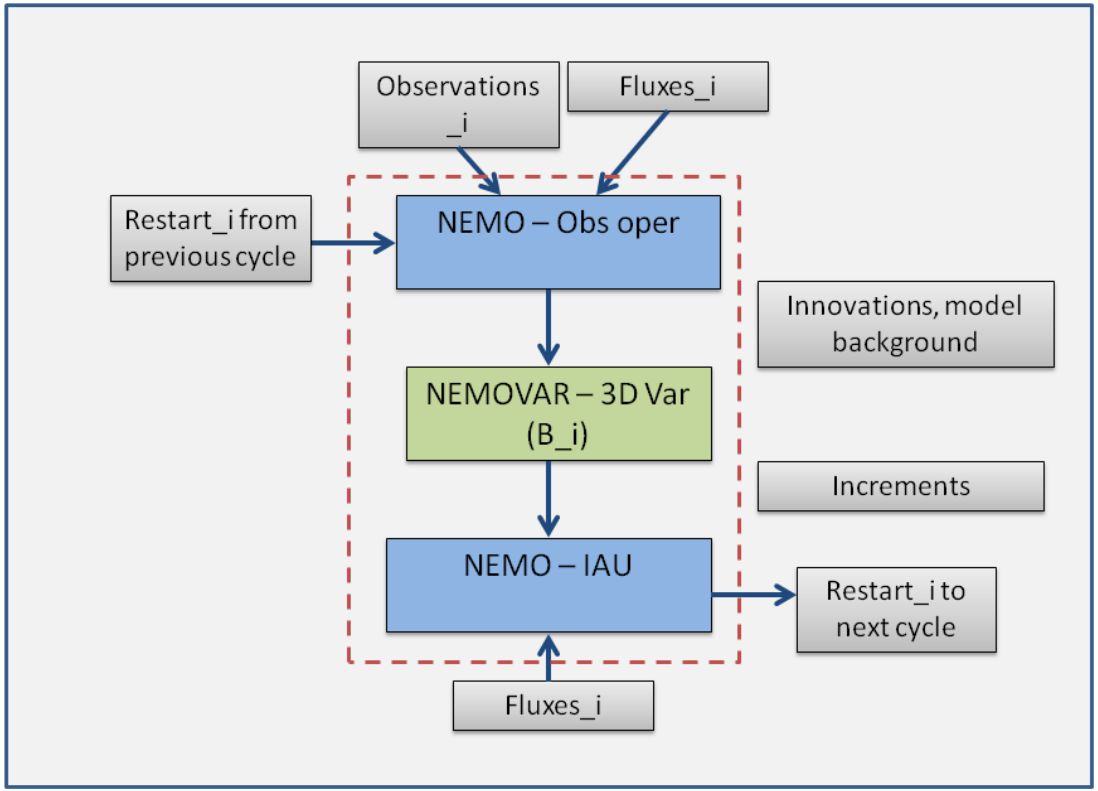
One of 37 forced by NWP MOGREPS fluxes (see James' talk for more details)

NEMO vn3.6 ¼° ORCA025

NEMOVAR v5

Run in 3DVar FGAT mode





Standard 3DVar data assimilation

- The standard incremental 3DVar cost function penalises:

  (1) the difference between the analysis and the background (forecast), and

  (2) the difference between the analysis and the observations.

$$J(\delta\mathbf{x}) = \overbrace{1/2(\delta\mathbf{x}^{\mathrm{T}}\mathbf{B}^{-1}\delta\mathbf{x})}^{(1)} + \overbrace{1/2((\mathbf{d}-\mathbf{H}\delta\mathbf{x})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{d}-\mathbf{H}\delta\mathbf{x}))}^{(2)}$$

- The vector δ**x** is the difference between the analysis and the model forecast, also known as the increments.
- The matrix **B** is the background error covariance matrix which is the size of the model state squared, i.e. the covariance between each variable at each grid point and all other variables at all other grid points. It is too large to estimate explicitly all the elements of **B** and we couldn't store them even if we could estimate them.
- **d** is the difference between the observations and the model background interpolated to the observation locations, also known as the innovations.
- **R** is the observation error covariance matrix.

Standard 3DVar data assimilation

Modelled background error covariance

- In the NEMOVAR system the **B** matrix is decomposed in the following way:

$$\mathbf{B}_{\mathrm{m}} = \mathbf{K}_{\mathrm{b}} \mathbf{D}_{\mathrm{m}}^{1/2} \mathbf{C}_{\mathrm{m}} \mathbf{D}_{\mathrm{m}}^{1/2} \mathbf{K}_{\mathrm{b}}^{T}$$

- The off-diagonal elements of **B** (determined by the correlations **C**) are modelled by assuming a functional form: a combination of 2 Gaussian functions each with their own length-scales where one is the first baroclinic Rossby radius and the second is 400km. An *implicit diffusion operator* is used to efficiently model these spatial correlations.
- We specify the diagonal elements (variances, **D**) as spatially and seasonally varying estimates based on previous reanalyses at the surface. The sub-surface error variances of temperature are parameterised based on the vertical temperature gradients (which change from day-to-day and depending on location).
- Physically-based multi-variate relationships (**K)** are specified to transfer information between variables in **B** so that the resulting increments (δx) are dynamically balanced. These are based on geostrophic and hydrostatic balances.

Hybrid ensemble/variational data assimilation

Ensemble based background error covariance

- If we have an ensemble system, the ensemble provides extra information about the forecast error covariances on each day.

$$\mathbf{X} = \frac{1}{\sqrt{N_\mathrm{e} - 1}}(\epsilon_1 \ldots \epsilon_{N_e}) \qquad \mathbf{B_e} = \mathbf{X}\mathbf{X}^\mathrm{T}$$

where the ϵ_i are the difference between the state in the ith member and the ensemble mean.

- However, the ensemble is usually only of limited size [O(10-100)] - leads to sampling errors if the ensemble were to be used directly to specify **B** in the data assimilation.
- There are ways to reduce sampling errors by localising the spatial influence of the ensemble near each observation. However, this means that dynamical balances aren't maintained, and that observations can only have limited spatial influence (which isn't good for sparse observing systems).

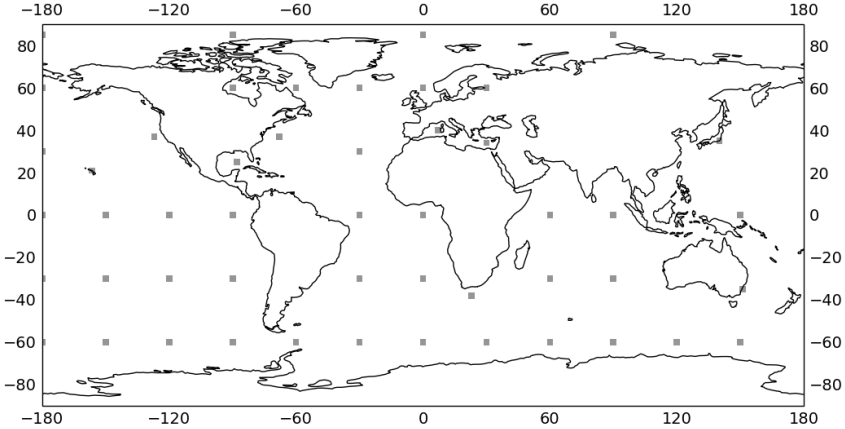
$$\mathbf{B}_\mathrm{e} = \mathbf{C}_L \circ \mathbf{X}\mathbf{X}^\mathrm{T}$$

- A way to allow data assimilation schemes to gain the benefits of the robustness of the existing *modelled* $\mathbf{B_m}$, and the benefits of the errors-of-the-day from the *ensemble-based* $\mathbf{B_e}$ is to linearly combine them in the variational cost function and use the existing infrastructure in NEMOVAR to minimise the new cost function.

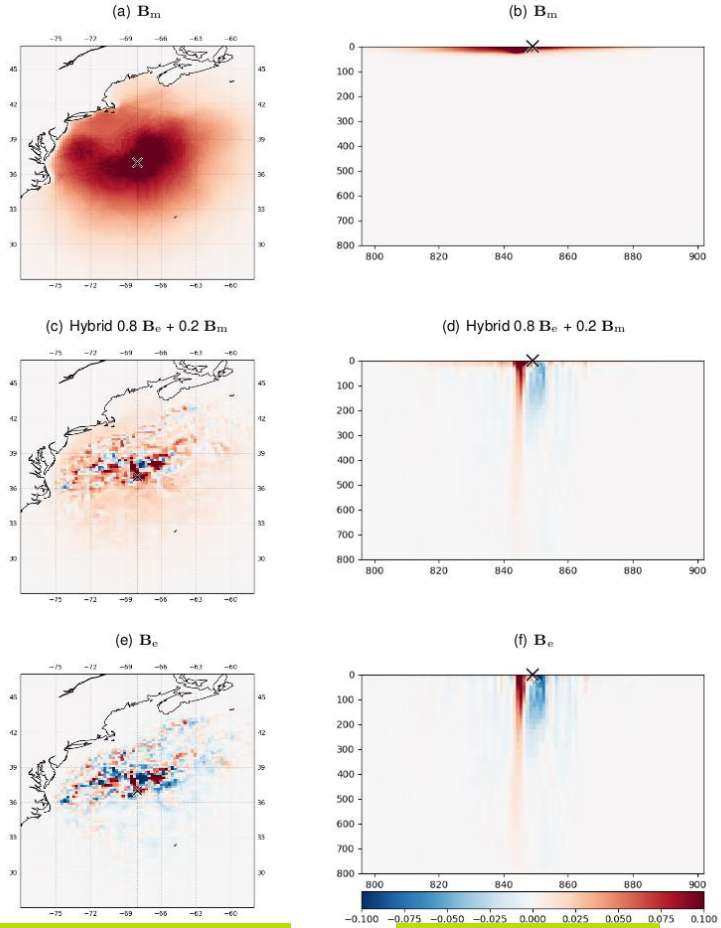
$$\mathbf{B}_\mathrm{h} = \beta_\mathrm{m}^2 \mathbf{B}_\mathrm{m} + \beta_\mathrm{e}^2 \mathbf{B}_\mathrm{e}$$

Testing hybrid DA with idealised observations

- We have started investigating the use of the ensemble in the DA using **hybrid ensemble/3DVar** using the ensemble with localisation.
- The ensemble used here is based on the output from the 37 member ensemble described in James' talk
- Idealised observations have been set up to illustrate how the hybrid DA works.



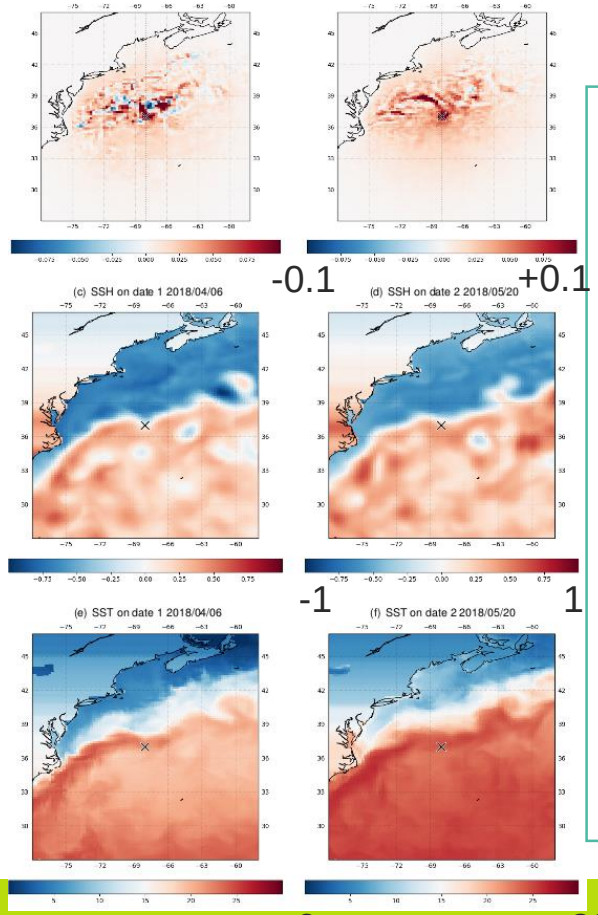
Increments /°C from idealised SST observations



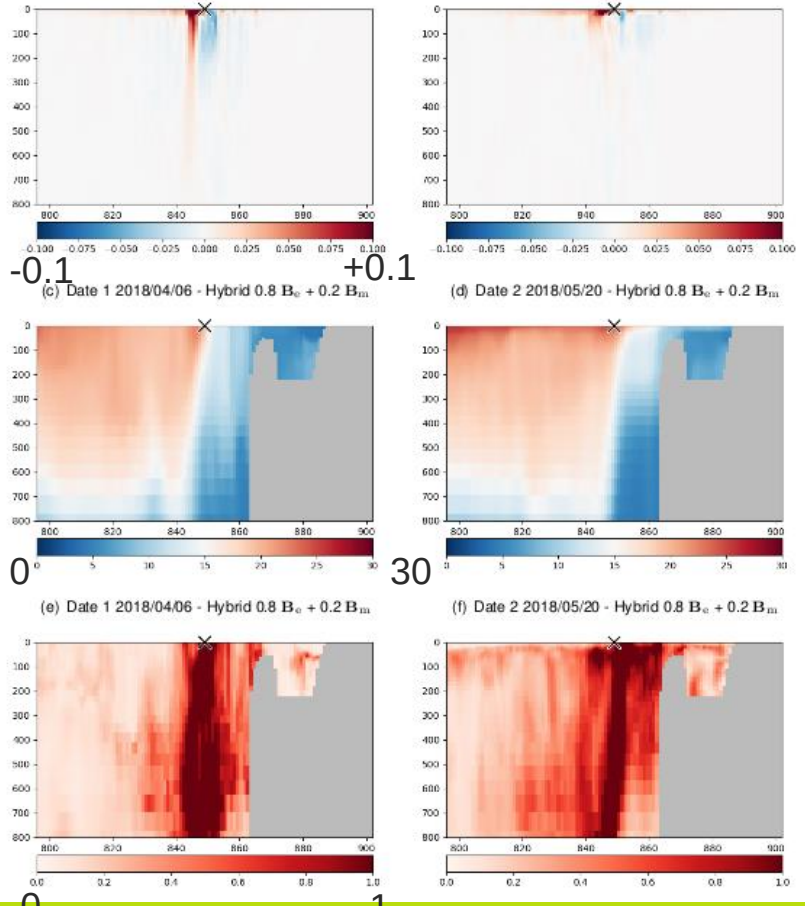
- One SST observation in the Gulf Stream (innovation 0.5°C)
- Standard NEMOVAR projects the information vertically based on the background mixed layer depth.
- Large localisation scales here of 4° just to illustrate the method. These need estimating/tuning.
- The ensemble covariance allows the observation to adjust the location of the front.
- The hybrid retains some larger scale correction based on the standard NEMOVAR system, but still makes changes based on the ensemble.

Dan Lea

Increments /°C, ensemble mean SSH (m)/SST, standard deviation from idealised SST observations



- One SST observation in the same location in the Gulf Stream.
- Two different dates (6th April and 20th May 2018).
- On the second date, the location of the Gulf Stream front has changed, so the ensemble error covariances produce quite a different increment.



Dan Lea

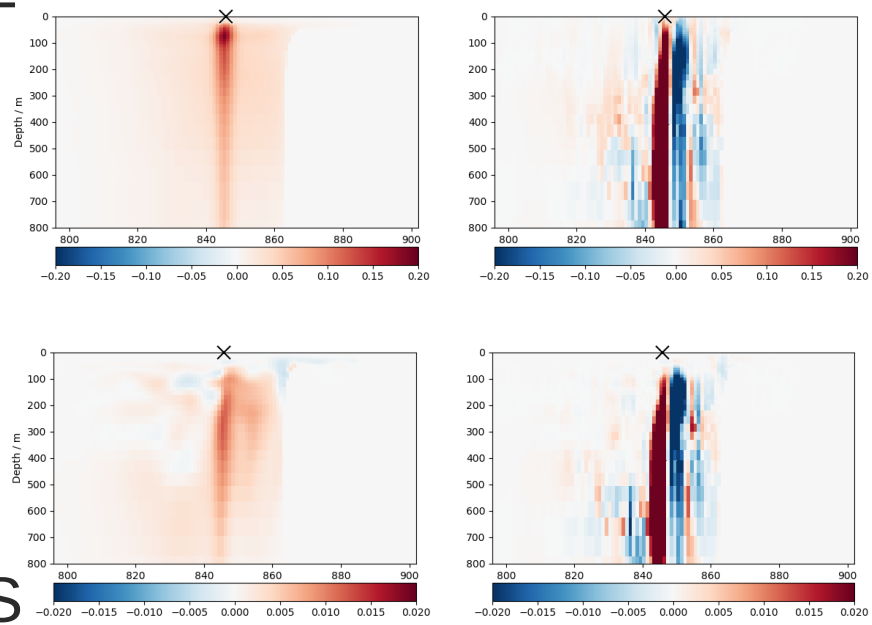
Temperature and Salinity Increments from idealised SLA observations

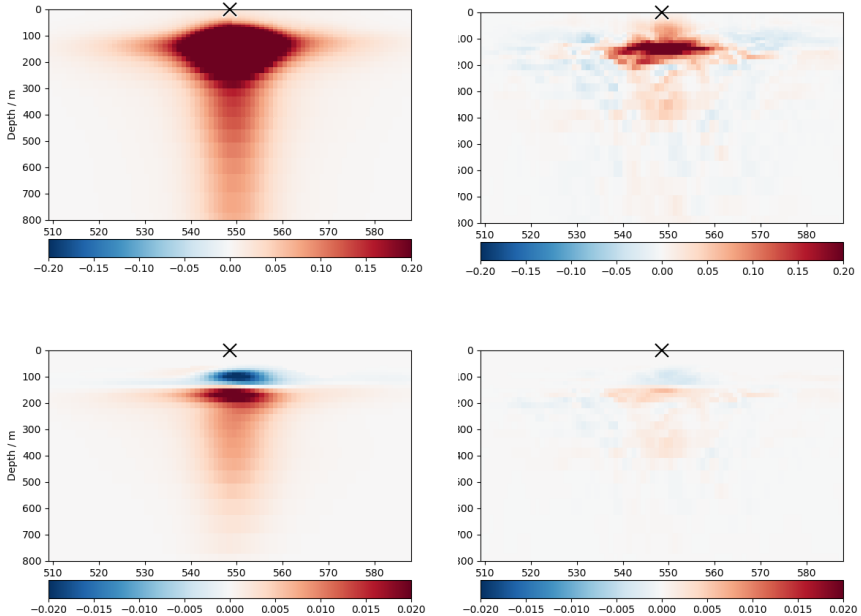
Gulf Stream

Tropical Pacific

T

S





$\mathbf{B}_m$ $\mathbf{B}_e$

SLA Innovation
8cm

$\mathbf{B}_m$ $\mathbf{B}_e$

Dan Lea

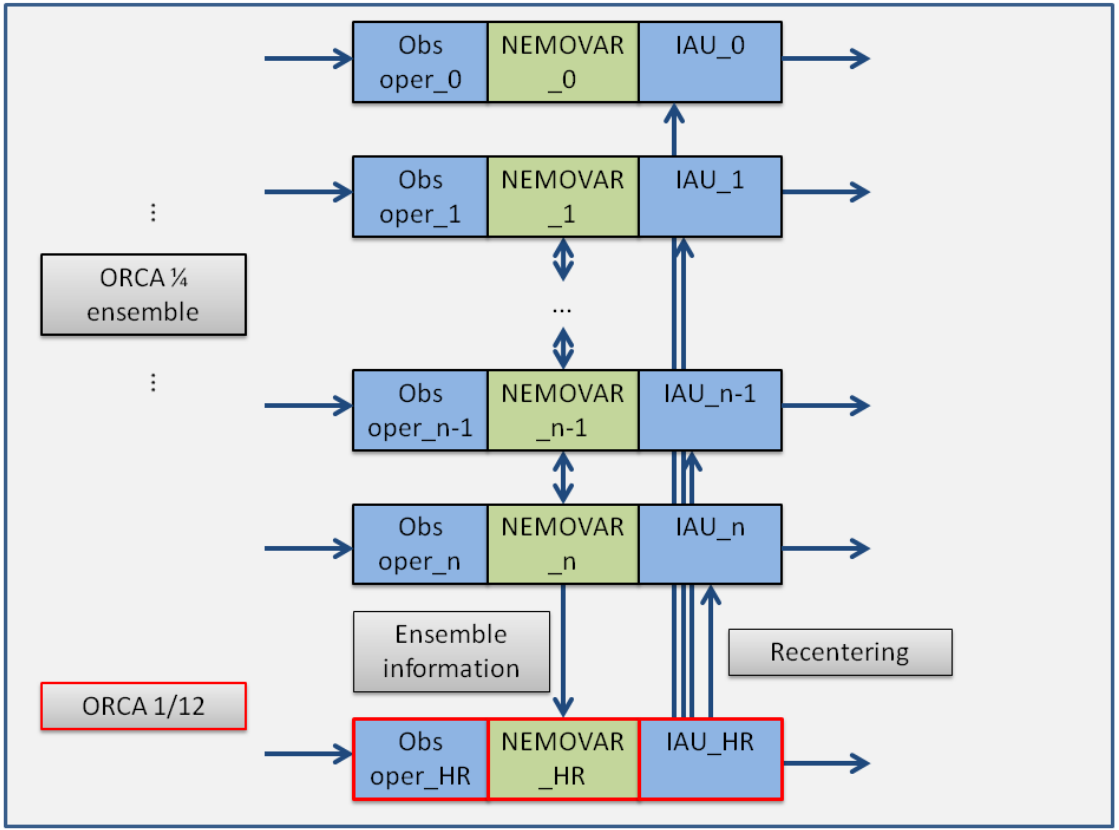
Summary

- Physically plausible increments produced using hybrid ensemble variational assimilation.
- Preliminary results.
- Tuning of the localisation and hybrid weights needed.
- An alternative approach available is to use the ensemble information to adjust the modelled covariances rather than localising the sample covariances.
- The results (relatively small ensemble based increments) suggest that the ensemble spread is a bit small particularly at the surface.
- The spread is also expected to be small in the deep ocean but this has not yet been tested in these idealised experiments.

Future work

- Improve the ensemble spread – test model perturbations, ensemble inflation (see James' talk).
- Further tests and tuning of the hybrid DA system. E.g. test the ocean model response to hybrid increments.
- Test the hybrid DA with real observations in a cycling ensemble system.
- Investigate methods for tuning the hybrid weights (Menetrier and Auligne, MWR 2015).
- Normalized Interpolated Convolution on an Adaptive Subgrid (NICAS) in the BUMP (Background error covariance on Unstructured Mesh Package; Menetrier) or multigrid diffusion based (Vidard) ensemble localisation.
- Develop hybrid DA ORCA12 (global 1/12°) with ORCA025, also coupled.
- Related work on hybrid DA in the North West European Shelf.

The dream



Coupled ocean atmosphere

Thanks

Spare slides

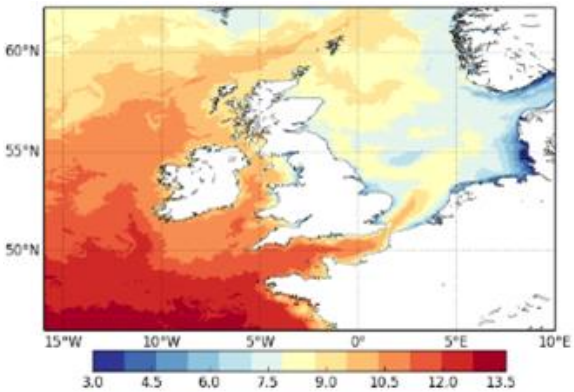
Implementation in the NWS system

- The use of a hybrid DA scheme in the shelf-seas should allow us to make much better use of information from gliders. (As both the global and NWS systems use NEMOVAR developments in the DA can be used in both systems).

- It is critical to have a “good” ensemble forecasting system in place. The various different sources of uncertainty in the system need to be represented.

- NWS ensemble system which uses:
  - An ensemble of 10 different surface forcings from ERA5
  - Perturbed observations
- Will test perturbation schemes to the model, e.g. rivers, boundary conditions, mixing scheme parameters, …..

- The ensemble could potentially be used to decide where to position gliders to have the most impact in the data assimilation.



Hybrid ensemble/variational data assimilation

- If we have an ensemble system, the ensemble provides extra information about the forecast error covariances on each day.

$$\mathbf{X} = \frac{1}{\sqrt{N_\mathrm{e} - 1}}(\epsilon_1 \ldots \epsilon_{N_e}) \qquad \mathbf{B_e} = \mathbf{X}\mathbf{X}^\mathrm{T}$$

- However, the ensemble is usually only of limited size [O(10-100)] and much much smaller than the dimension of the state [O(10^8) in ORCA025] .
- This leads to sampling errors if the ensemble were to be used directly to specify **B** in the data assimilation.
- There are ways to reduce sampling errors by localising the spatial influence of the ensemble near each observation. However, this means that dynamical balances aren't maintained, and that observations can only have limited spatial influence (which isn't good for sparse observing systems).

$$\mathbf{B}_\mathrm{e} = \mathbf{C}_L \circ \mathbf{X}\mathbf{X}^\mathrm{T}$$

- A way to allow data assimilation schemes to gain the benefits of the robustness of the existing *modelled* $\mathbf{B_m}$, and the benefits of the errors-of-the-day from the *ensemble-based* $\mathbf{B_e}$ is to linearly combine them in the variational cost function and use the existing infrastructure in NEMOVAR to minimise the new cost function.

$$\mathbf{B}_\mathrm{h} = \beta_\mathrm{m}^2 \mathbf{B}_\mathrm{m} + \beta_\mathrm{e}^2 \mathbf{B}_\mathrm{e} \qquad \mathbf{B}_\mathrm{e} = \mathbf{K}\mathbf{D}_\mathrm{e}^{1/2}(\mathbf{C}_L \circ \tilde{\mathbf{X}}\tilde{\mathbf{X}}^\mathrm{T})\mathbf{D}_\mathrm{e}^{1/2}\mathbf{K}^\mathrm{T}$$

Balance relationships

Change of variables in the assimilation to balanced and unbalanced variables.

$$\delta T = \delta T$$

$$\delta S = \Gamma_{ST}\delta T + \delta S_u$$

$$\boxed{\delta \eta = \Gamma_{\eta\rho}\delta\rho + \delta\eta_u}$$

$$\delta u = \Gamma_{p\rho}\delta p + \delta u_u$$

$$\delta v = \Gamma_{p\rho}\delta p + \delta v_u$$

Water mass preservation
Troccoli and Haines (1999)

hydrostatic balance

geostrophic balance

Where

$$\delta\rho = \Gamma_{\rho T}\delta T + \Gamma_{\rho S}\delta S$$

$$\delta p = \Gamma_{p\rho}\delta\rho + \Gamma_{\rho\eta}\delta\eta$$

Weaver et al (2005)

This means that an SLA (eg) observation will produce increments in SSH, temperature and salinity.