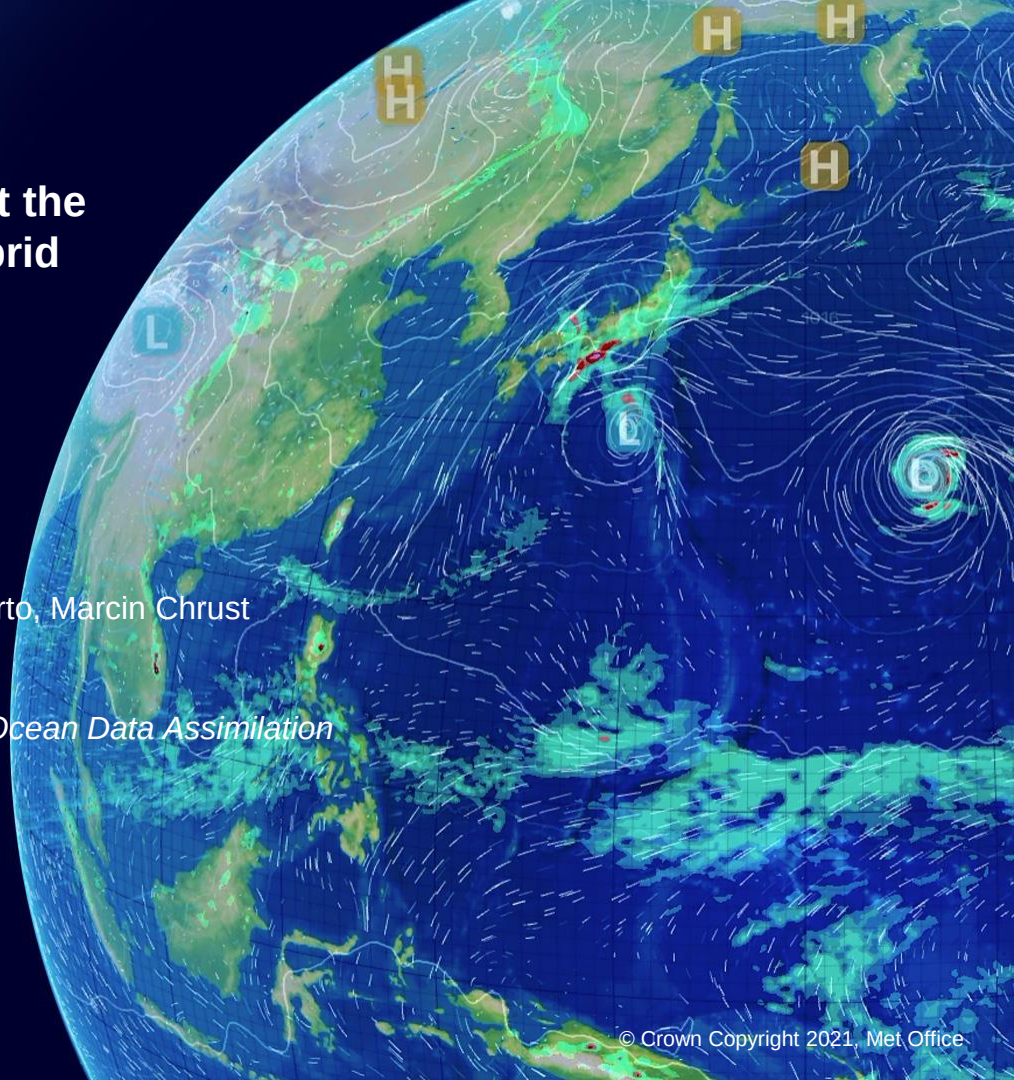
A new global ocean ensemble system at the Met Office: Assessing the impact of hybrid data assimilation and inflation settings

Dan Lea,

James While, Matt Martin, Anthony Weaver, Andrea Storto, Marcin Chrust

Joint ECWMF/OceanPredict workshop on Advances in Ocean Data Assimilation

17 May 2021



Overview

Two main aspects of the work done:

1. Developed a global ocean/sea-ice ensemble system which could be used in a coupled ensemble forecasting system.
2. Use of the ensemble information in the ocean data assimilation.

System details:

- Ensemble of data assimilations (EDA) approach
- Perturbed observation locations and values
- Each member forced by a different atmospheric ensemble (MOGREPS) member
- Stochastic model perturbations in NEMO (SPPT, SPP, SKEB)
- Multiplicative inflation (relaxation to prior spread - RTPS) to deal with mis-specifications in the DA. ($\alpha = 0.8$)
- Hybrid ensemble/3DVar data assimilation in all members (β_e^2 gives the weight given the ensemble in the DA)

Ongoing work this

- **Paper in preparation**
- Put our ocean ensemble system into the coupled NWP system

Relaxation To Prior Spread Scheme (RTPS) ensemble inflation scheme.

- Multiplicative inflation scheme – it increases the analysis variance, but does not change the subspace spanned by the ensemble.

Relaxes the analysis spread towards the forecast spread controlled by a parameter α (α = 0: Spread unchanged α = 1: Inflated spread = forecast spread)

- To keep the model balanced inflation is applied to the 'unbalanced'* (i.e. uncorrelated) increments used in the data assimilation, rather than the real variables. (* temperature, unbalanced salinity, and unbalanced SSH.)

Stochastic model perturbations

Code for NEMO (v3.6) by Andrea Storto with these types of stochastic perturbations

1. **Stochastically Perturbed Parameterization Tendencies** (***SPPT***), which adds perturbations collinear to the unresolved process tendencies.

2. **Stochastically Perturbed Parameters** (***SPP***), which adds perturbation to supposedly uncertain parameters within parameterized processes

3. **Stochastic Kinetic Energy Backscatter** (***SKEB***), which add perturbations to the barotropic streamfunction proportionally to some sources of small-scale dissipated energy, backscattered to the model resolved scales:

Perturbation fields for each type of scheme are correlated in time and space, according to user-specified de-correlation time- and length-scales.

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.

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

where the ϵ_i are the difference between the state in the i^{th} 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 can affect the dynamical balances, and that observations can only have limited spatial influence (which isn't good for sparse observing systems).

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

Localisation 2x Rossby radius 300 km cap

- 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.

$$\mathbf{B}_h = \beta_m^2 \mathbf{B}_m + \beta_e^2 \mathbf{B}_e$$

Dan Lea

List of experiments

3DVar uses parametrised/model background error
3DEnVar uses ensemble information instead
Hybrid 3DEnVar uses a weighted mixture of the two

Experiments

Experiment Name	Assimilation type	Ensemble hybrid weight (β_e^2)	Inflation factor (α)
ensda00_inf00	3DVar	0	0
ensda02_inf08	hybrid 3DEnVar	0.2	0.8
ensda05_inf08	hybrid 3DEnVar	0.5	0.8
ensda08_inf08	hybrid 3DEnVar	0.8	0.8
ensda10_inf08	pure 3DEnVar	1.0	0.8
ensda08_inf00	hybrid 3DEnVar	0.8	0

Reliability

Reliability: the degree to which forecast probabilities agree with outcome frequencies.

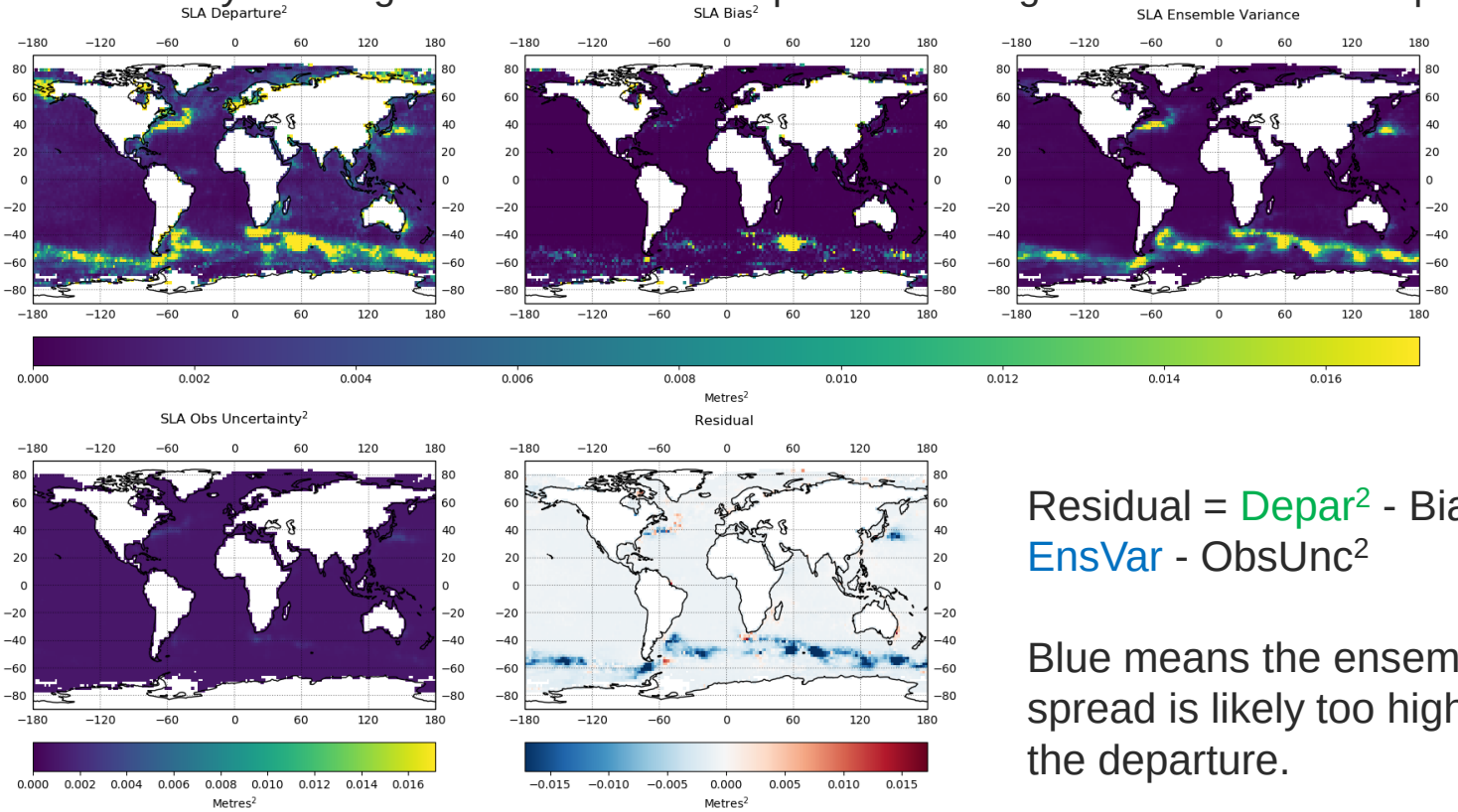
Rodwell (2016) ensemble consistency test.

$Depar^2 = Bias^2 + EnsVar + ObsUnc^2 + Residual$

The residual should be zero if the ensemble is reliable. That means the Ensemble Spread is consistent with the Departure (Ensemble mean model values minus observations) and bias (generally quite small in our examples).

SLA reliability (3DVar: ensda00_inf00)

Reliability: the degree to which forecast probabilities agree with outcome frequencies.

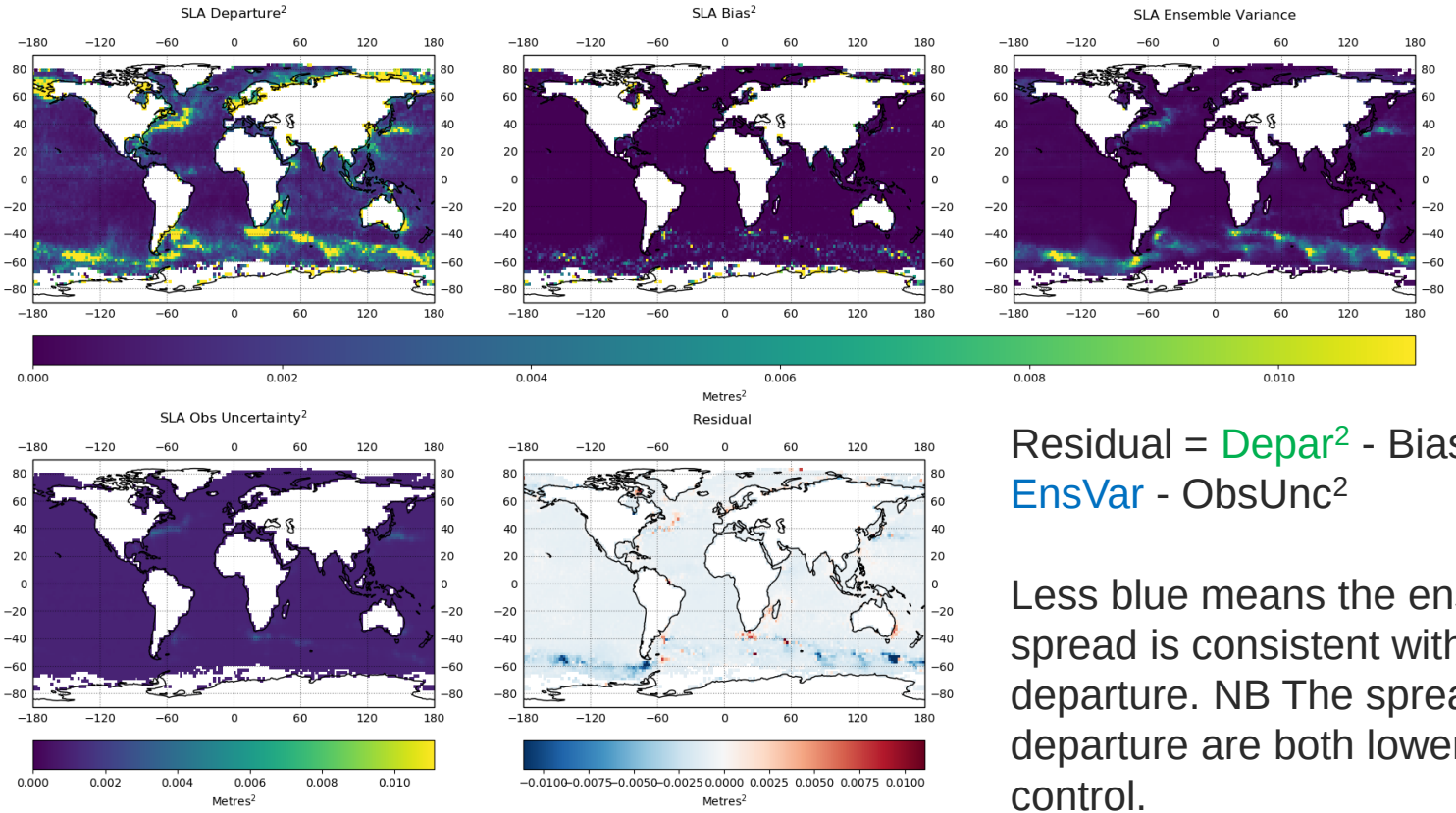


Residual = Depar² - Bias² - EnsVar - ObsUnc²

Blue means the ensemble spread is likely too high given the departure.

Dan Lea

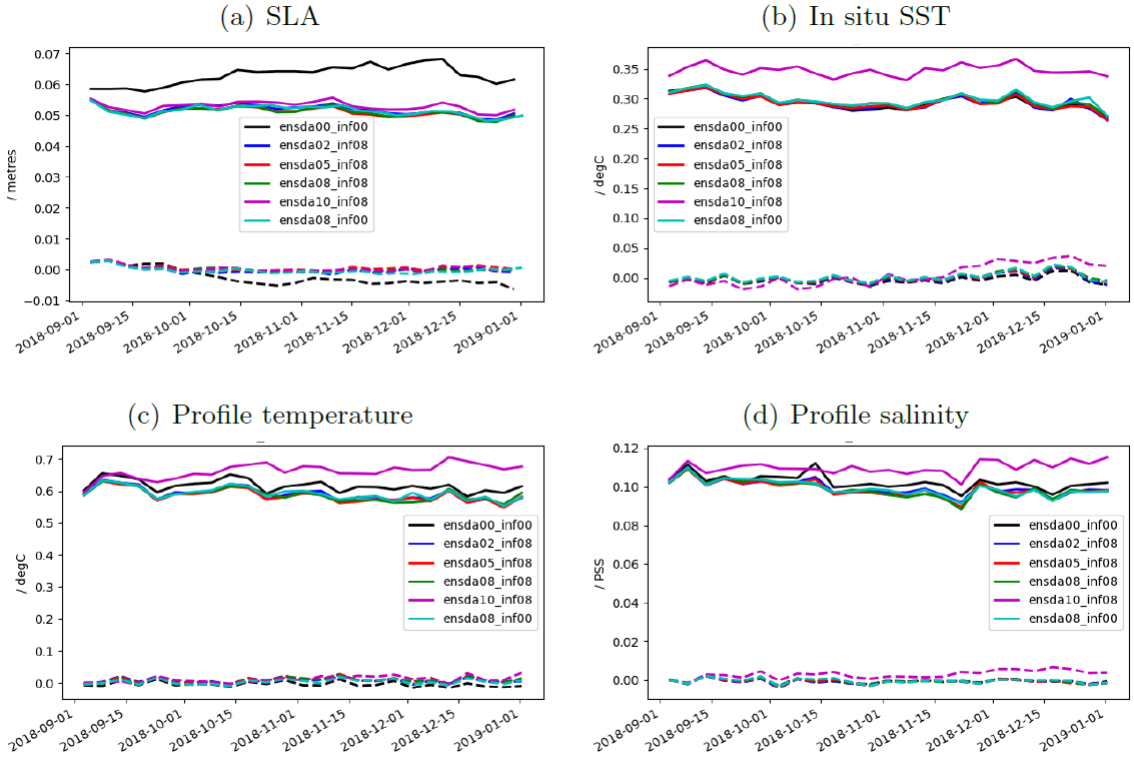
SLA reliability (hybrid 3DEnVar: ensda08_inf08)



Residual = Depar² - Bias² - EnsVar - ObsUnc²

Less blue means the ensemble spread is consistent with the departure. NB The spread and departure are both lower than the control.

Ensemble mean vs obs (rmse/bias)

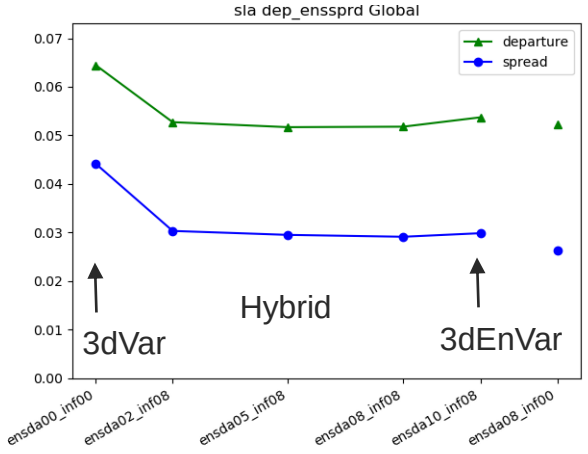


4 month experiments

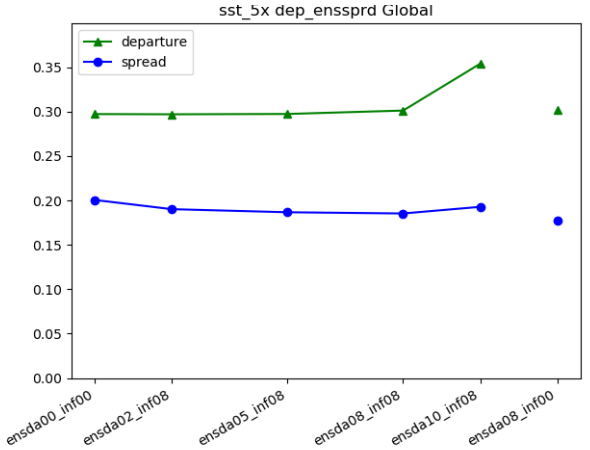
Hybrid best for all variables

Ensemble departure/spread global statistics

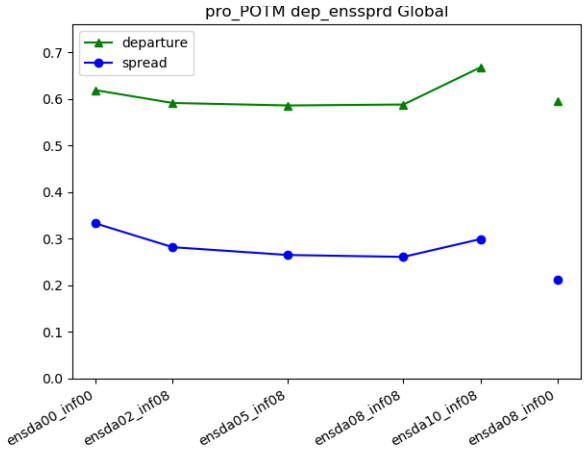
SLA /m



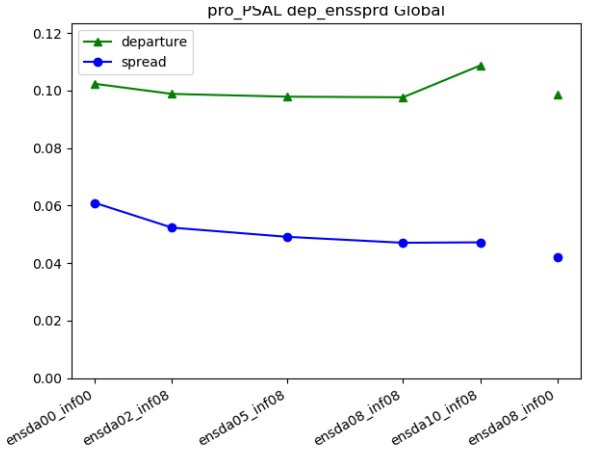
SST °C



Temp. profiles °C

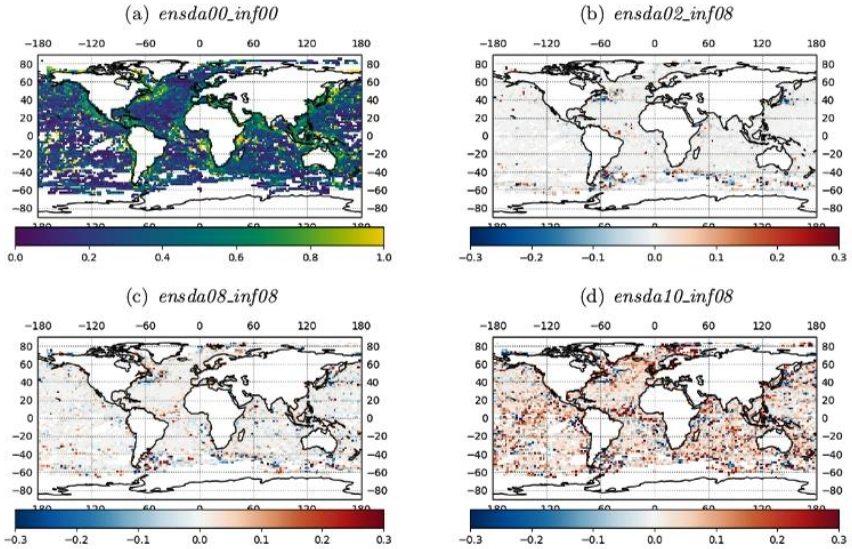


Salinity profiles PSS

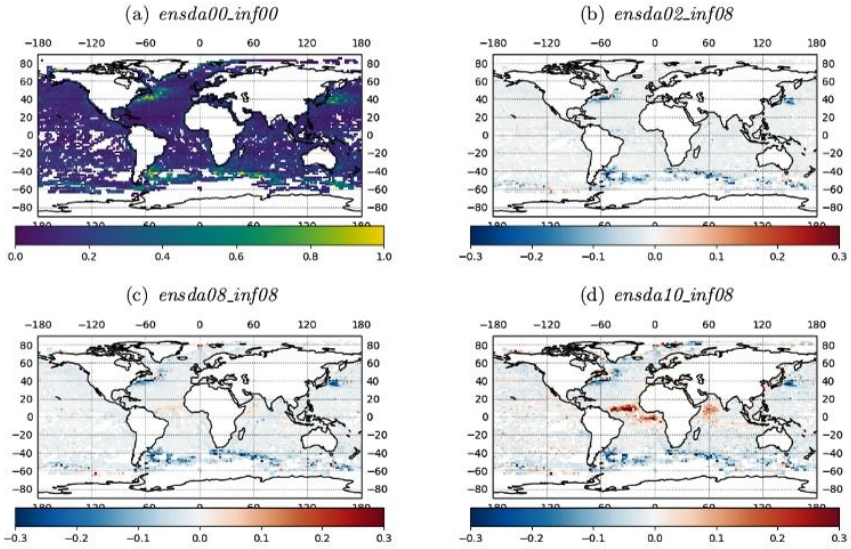


SST departure and spread for various expts

Ensemble mean vs obs (RMS)



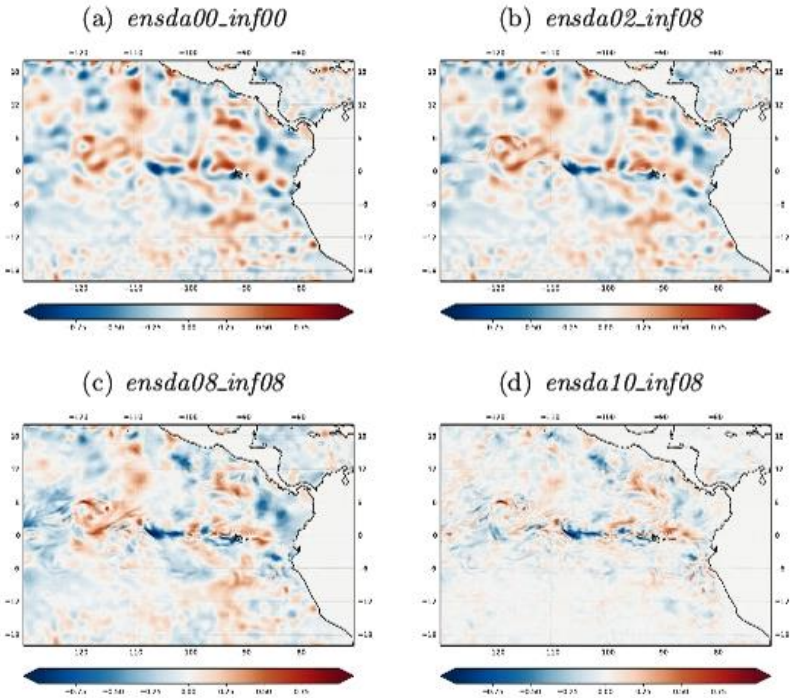
Ensemble spread



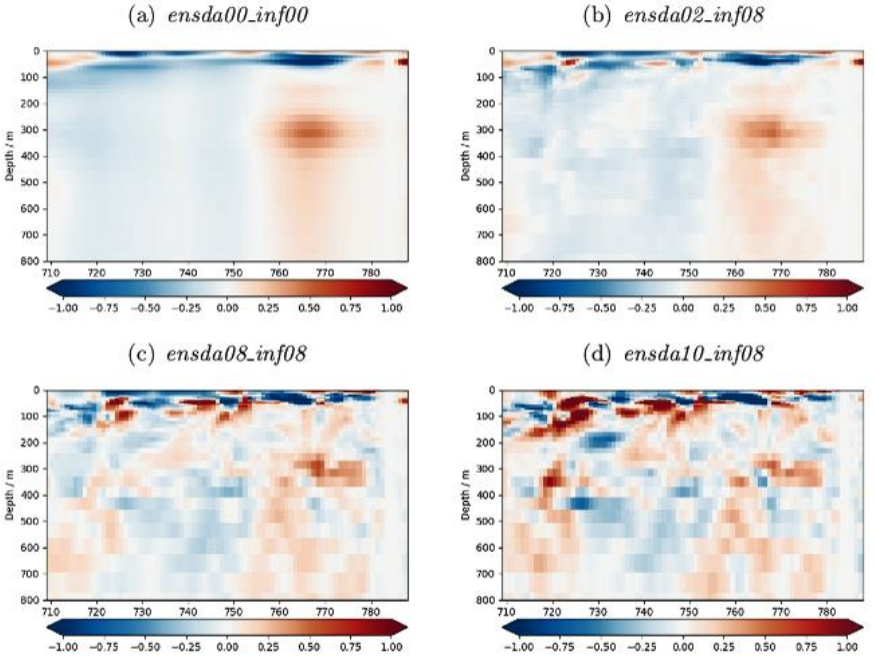
Panels show control and the expt minus control. Blue areas indicate the expt value smaller then the control expt.

Dan Lea

SST Increments (1st day of run)

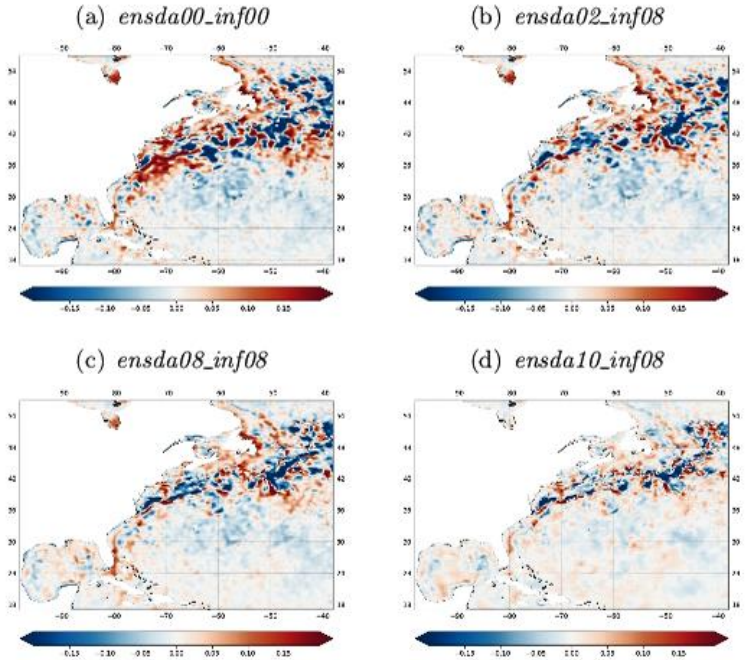


T incs Equatorial section 180W -120W

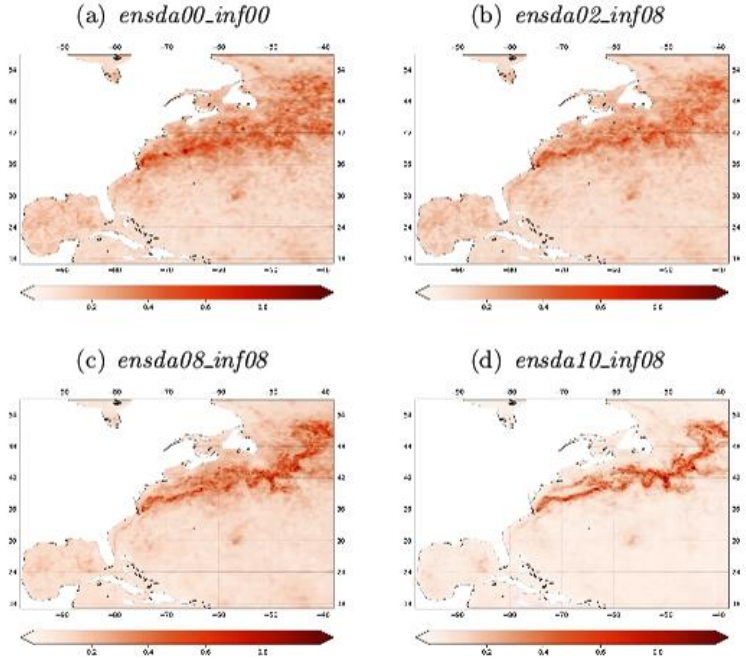


Considering reducing the localisation length scale (up to 300 km) in the tropics because of the small length scale features in the ensemble increments

Mean SST increments



Std dev SST increments



The pure 3DEnVar does not seem to do enough to correct the SST bias south of the Gulf Stream

Summary and ongoing work

- We have developed a new ocean ensemble system.
- Demonstrated the ensemble system works well.
- Showed the benefits of including ensemble information in the data assimilation. (but maybe not too much - 0.5-0.8 hybrid ensemble weight seems optimal)
- Drawback of hybrid DA is additional expense at least 2x cost of non-hybrid DA)

- The coupled NWP system has an atmosphere ensemble. Only a single ocean analysis used at present – this year we will include our ocean ensemble system.
- We might need to find a way to make hybrid DA cheaper. Although an ensemble of 3DVars will still be better than what is done now.

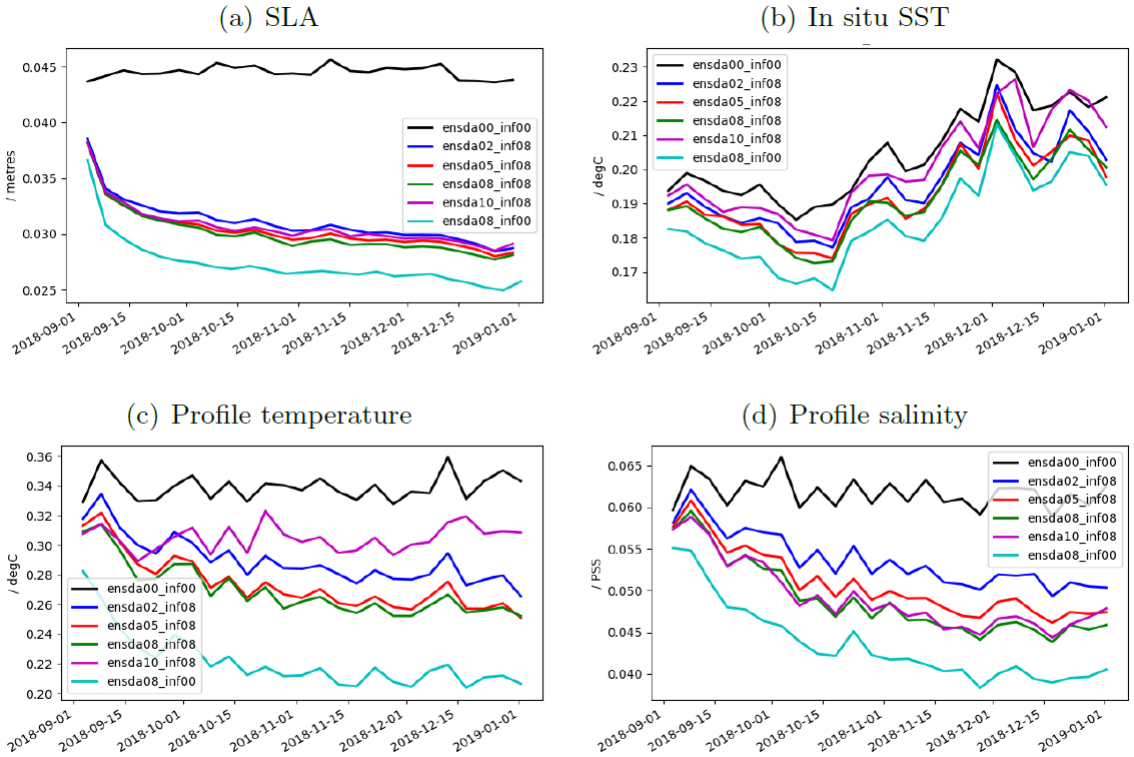
Fin

-

Spare slides

-

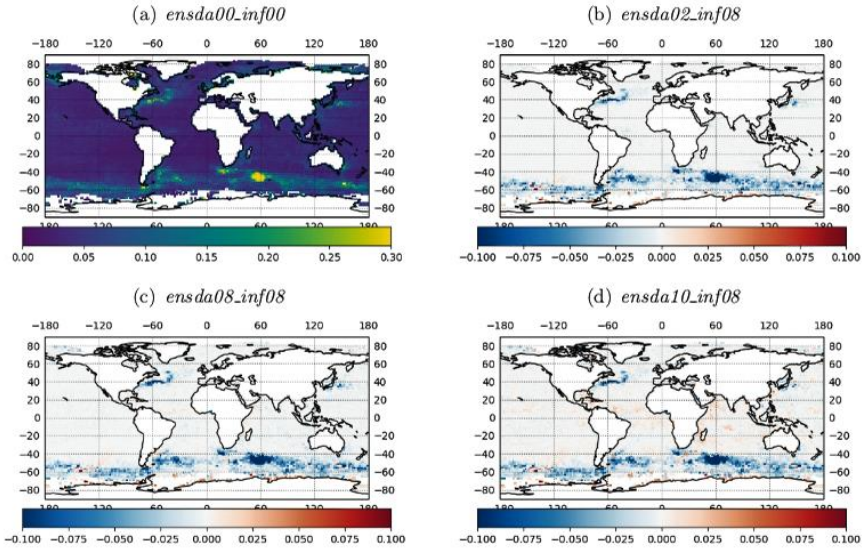
Ensemble spread



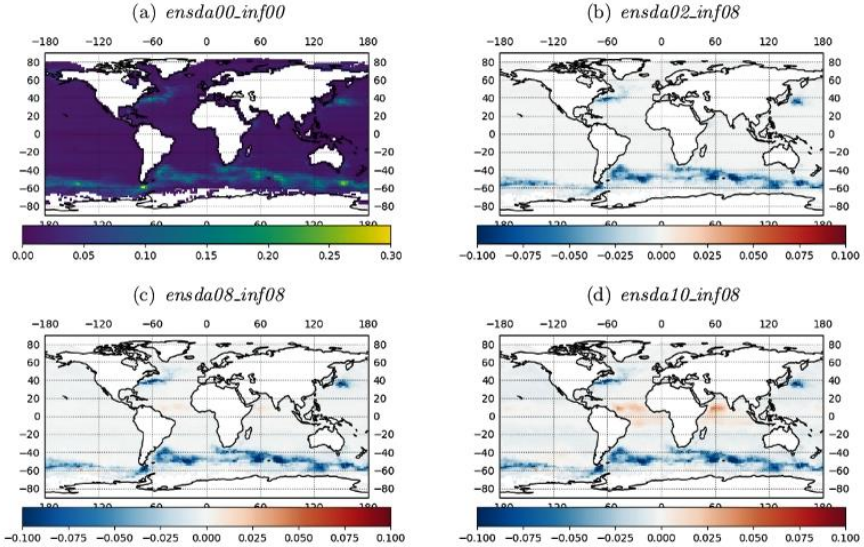
Hybrid DA reduces ensemble spread. Inflation reverses this to some extent.

SLA

Ensemble mean vs obs (rmse)



Ensemble spread



Panels show control and the expt minus control. Blue means the expt value smaller then the control expt.

Met Office Increments (1st day of run)

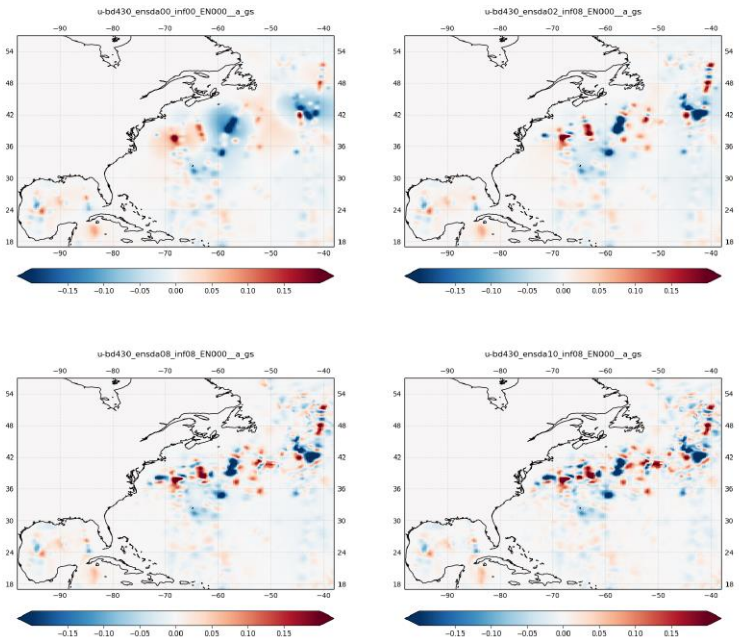


Figure 21: SSH increments on the first day of the experiments /m.

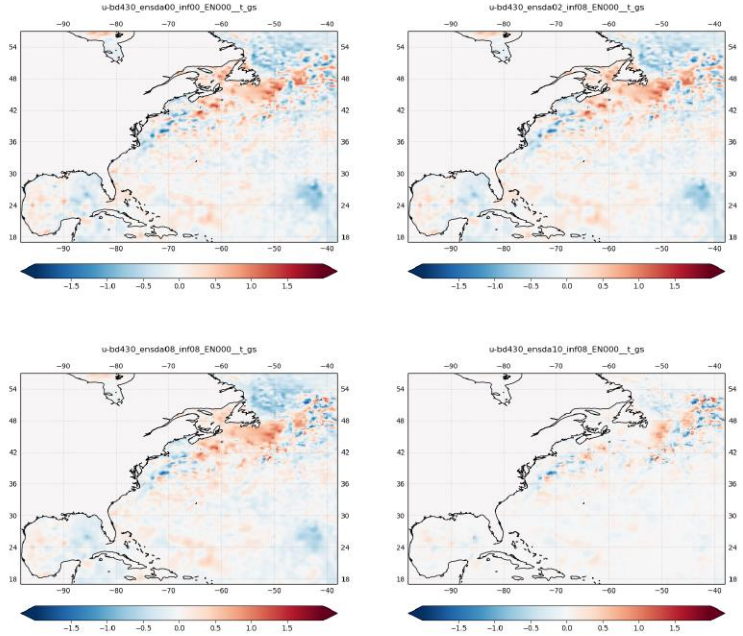
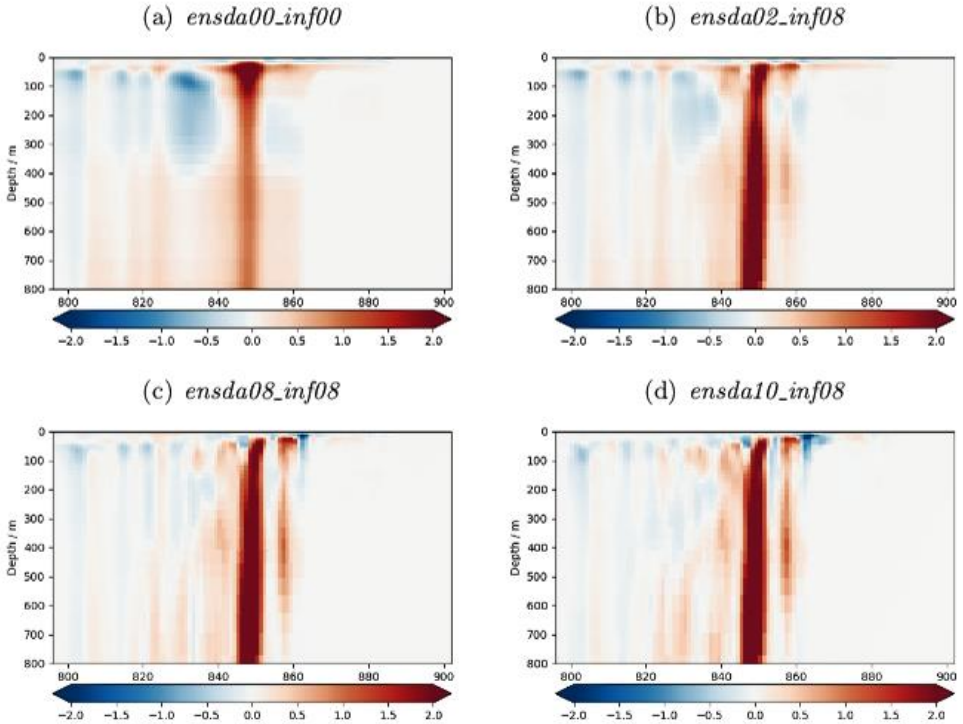


Figure 22: Surface temperature increments on the first day of the experiments /°C .

NB On the first day all the assimilation inputs are the same for each experiment differences are thus wholly due to the assimilation settings

T Increments (1st day of run) South-North section 17N -57N at 68W



Note sloping in the vertical structure of increments with (hybrid) 3DEnVar