Met Office

Assessing the impact of including a global ocean ensemble system in the Met Office coupled Numerical Weather Prediction system

Forecasting Research Technical Report 657

October 30, 2023

Daniel J. Lea[1], Matthew J. Martin[1], Martin Price[1], Jonah Roberts-Jones[1], Warren Tennant[1], and Chris Harris[1]

[1]Met Office, Exeter, UK

Met Office

Contents

Abstract

Improvements to the ocean and sea-ice components of the Met Office's coupled Numerical Weather Prediction (NWP) ensemble system have been implemented. These include the use of an ensemble of data assimilations (EDA) approach whereby the ocean and sea-ice components of each ensemble member include their own DA with perturbed observations. Stochastic perturbations were also added to the ocean model component of each coupled ensemble member forecast. These developments bring the ocean and sea-ice into line with the methods already used in the atmospheric component of the coupled ensemble. The availability of an ocean ensemble also allowed us to test the impact of using hybrid-3DEnVar data assimilation in the ocean component of the coupled ensemble together with an ensemble inflation scheme.

The impact of the ocean/sea-ice ensemble developments on the coupled ensemble was assessed by running some two-month trials. The ocean and sea-ice forecasts were shown to be more accurate in both a deterministic and ensemble sense compared to the existing coupled ensemble set-up whereby all the ocean/sea-ice ensemble members are initialised from the deterministic analysis on each cycle, with extra perturbations made to the sea surface temperature (SST) seen by the atmosphere. The sea-ice component was shown to have better spread with the developments included, but there were issues affecting the mean sea-ice concentration which had knock-on impacts on the coupled system due to the high sensitivity of the air-sea fluxes in areas of sea-ice. The atmospheric forecasts in the boundary layer were shown to need significantly larger perturbations than the uncertainty in the SSTs would suggest, which indicates that the perturbation methodology for the atmospheric boundary layer should be revisited so that it is not just the SST which generates the required spread. When we re–introduced some of the extra SST perturbations needed by the atmosphere on top of the ocean ensemble developments, we had a system with similar ensemble performance in the atmosphere compared to the existing system, a fairly reliable ocean ensemble, and significant improvements to the ocean performance through the use of hybrid-3DEnVar.

Keywords — data assimilation, ensembles, ocean, coupled, global

1 Introduction

The Met Office implemented a coupled model with coupled data assimilation (DA) in its operational global Numerical Weather Prediction (NWP) system in 2022. Prior to that the atmospheric forecasts were forced at the surface by external analyses of sea surface temperature (SST) and sea-ice from the Operational sea Surface Temperature and sea-Ice Analysis (OSTIA) product (Good et al. 2020). The initial coupled NWP implementation focussed largely on the deterministic aspects of the forecast system which were shown to be improved by the use of a coupled model (Vellinga et al. 2020, Guiavarc'h et al. 2019). The ensemble forecasts in the Met Office Global and Regional Ensemble Prediction System – Global (MOGREPS–G; Inverarity et al. (2023)) also now use the coupled model, but the ocean/sea-ice component of the coupled ensemble is not currently initialised well. The ocean/sea-ice component of each ensemble member is initialised from the deterministic ocean/sea-ice analysis on each cycle meaning that the spread in the ocean component, particularly in the early stages of the forecast, is very small. For that reason, additional perturbations were added to the SST fields seen by the atmospheric ensemble members in the same way that was done before the coupled model introduction, as described by Tennant & Beare (2014).

The current operational coupled NWP system performs well in the deterministic sense for both ocean and atmosphere, but the ocean components of the ensemble forecasts are not adequate for providing uncertainty information to ocean users, or for use in possible ensemble or hybrid ocean DA schemes. Also, the evolution of spread in the ocean, which forces the atmospheric ensemble, could be improved by implementing a proper ocean ensemble. An improved ocean ensemble could also be used to improve the initialisation of coupled seasonal predictions with the GloSea system which currently uses a lagged ensemble approach for initialisation, using the deterministic coupled NWP analyses without any explicit perturbations in the sub-surface ocean (MacLachlan et al. 2015). However, the inclusion of ocean ensemble developments in the coupled NWP system will add significant extra computational cost and extra complication which will also add to the cost of maintaining the system. The work described here therefore aims to assess the potential benefits of improving the ocean/sea-ice component of the coupled ensemble on both the ocean/sea-ice and atmospheric components.

Work to develop the ocean/sea-ice component of the ensemble system was previously carried out in an uncoupled framework by Lea et al. (2022). That work included the use of an ensemble of data assimilations (EDA) approach with perturbed observations and the use of stochastic ocean model perturbations.

It included work to assess the impact of using hybrid-3DEnVar data assimilation in the ocean, making use of the ensemble information to improve the DA, which necessitated the use of an ensemble inflation scheme. All these developments resulted in a system which was able to produce reliable short-range ocean forecasts (the ensemble spread was representative of the error in the ensemble mean), and the hybrid-3DEnVar scheme resulted in significant improvements to the accuracy of the ensemble mean.

Other centres around the world have implemented coupled data assimilation to initialise coupled NWP forecasts. The latest status of the ECMWF coupled data assimilation framework is described by de Rosnay et al. (2022) who also provide a description of the various flavours of coupled DA. Another major centre which has implemented coupled DA to initialise NWP forecasts is the US Navy whose system, including a coupled EDA system, is described by Barton et al. (2020). A weakly coupled DA system was also implemented in the Environment and Climate Change Canada (ECCC) global NWP system (Skachko et al. 2019).

The implementation in the Met Office system of coupled DA and coupled EDA, while similar in some respects, is somewhat different compared to those other systems. In section 2 we provide an overview of the present operational coupled NWP system at the Met Office together with a description of the changes to the ocean component of the coupled ensemble. Section 3 describes the experiments which are designed to assess the impact of those developments. An assessment of the results of the experiments is given in section 4 including the impact on the ocean, sea-ice and atmospheric components of the ensemble. Finally, section 5 gives a summary of the main results and discusses the way forward.

2 Ensemble system developments

2.1 Description of current coupled NWP system

2.1.1 Coupled model

The model used in the coupled NWP system is the GC4 configuration of the Met Office coupled modelling system. It consists of the Unified Model (UM) atmospheric model in the GA8 configuration, the JULES (Joint UK Land Environment Simulator) land surface model in the GL9 configuration, the NEMO ocean model in the GO6 configuration (Storkey et al. 2018) and the CICE sea-ice model in the GSI8.1 configuration (Ridley et al. 2018). Earlier versions of the atmosphere and land model components were described by Walters et al. (2019) with the the key changes since then being: prognostic based entrainment, which adds convective memory and improves precipitation rates and spatial structures; time-smoothed convective increments, which improves the convection-dynamics coupling and greatly reduces the dynamical effects of convective intermittency; a new riming parameterisation, which increases the amount of supercooled water and hence reduces southern ocean biases; and a package of surface changes, which improves the forecast of near-surface winds and removes the need for the aggregate tile in NWP. The atmosphere and land are run in one executable, as are the ocean and sea-ice, and the coupling fields are exchanged using the OASIS coupler every hour.

In the operational system the horizontal resolution of the atmosphere and land models is N1280 ($\sim$10 km) for the deterministic forecasts and N640 ($\sim$20 km) for the ensemble forecasts. However, these are very computationally expensive to run, so the experiments described later use lower resolution versions of N640 ($\sim$20 km) for the deterministic forecasts and N320 ($\sim$40 km) for the ensemble forecasts. The ocean and sea-ice model horizontal resolution is $1/4^\circ$ ($\sim$25 km) for both deterministic and ensemble components both operationally and in the experiments described later. The ocean model has 75 vertical levels which range in thickness from 1 m near the surface to 300 m at 6000 m depth.

2.1.2 Coupled data assimilation

The atmospheric component of the DA for the deterministic forecasting system is carried out using a hybrid–4DVar (four-dimensional variational DA) approach (Bowler et al. 2017). This uses a combination of the forecast-error covariances from the MOGREPS–G ensemble (described in section 2.1.3) with climatological error covariance information. It uses a linear Perturbation Forecast model to (in effect) evolve the forecast-error covariances through the assimilation window of 6 hours. The land surface DA uses an Extended Kalman Filter approach and is described by Gómez et al. (2020).

The ocean and sea-ice DA are both carried out using a 3DVar-FGAT (First Guess at Appropriate Time) scheme based on the NEMOVAR code (Waters et al. 2015, Mirouze et al. 2016). The DA for ocean and sea-ice are run separately. A hybrid-3DEnVar scheme (also using FGAT) has also been developed using NEMOVAR and tested by Lea et al. (2022) for the ocean DA (with the sea-ice still using 3DVar-FGAT), though this is not yet operational.

These separate atmosphere, land, ocean and sea-ice DA systems are combined in a coupled DA framework as described by Lea et al. (2015). The way in which the coupled DA cycles is shown in Fig. 1. All of the component DA systems are run on a 6-hour cycle where the inputs to the DA, including the innovations (observation-minus-model values in observation space) and non-linear model trajectory information, come from the coupled model forecast. Separate systems (as just described) are then used to produce the analysis increments for the ocean, sea-ice, atmosphere and land. Note that the sea-ice data assimilated are from a product which collects all data for each day and has a nominal time of 12Z, so only the 12Z cycle includes sea-ice DA. The atmospheric increments are added into the atmospheric component of the coupled model at the beginning of the time window with direct insertion (DI), while the ocean and sea-ice increments are added in using incremental analysis updates (IAU, Bloom et al. (1996)) over the first 3 hours of the time window. The analysis is valid in the middle of the time windows (00Z, 06Z, 12Z, 18Z) and forecasts are launched from these times.

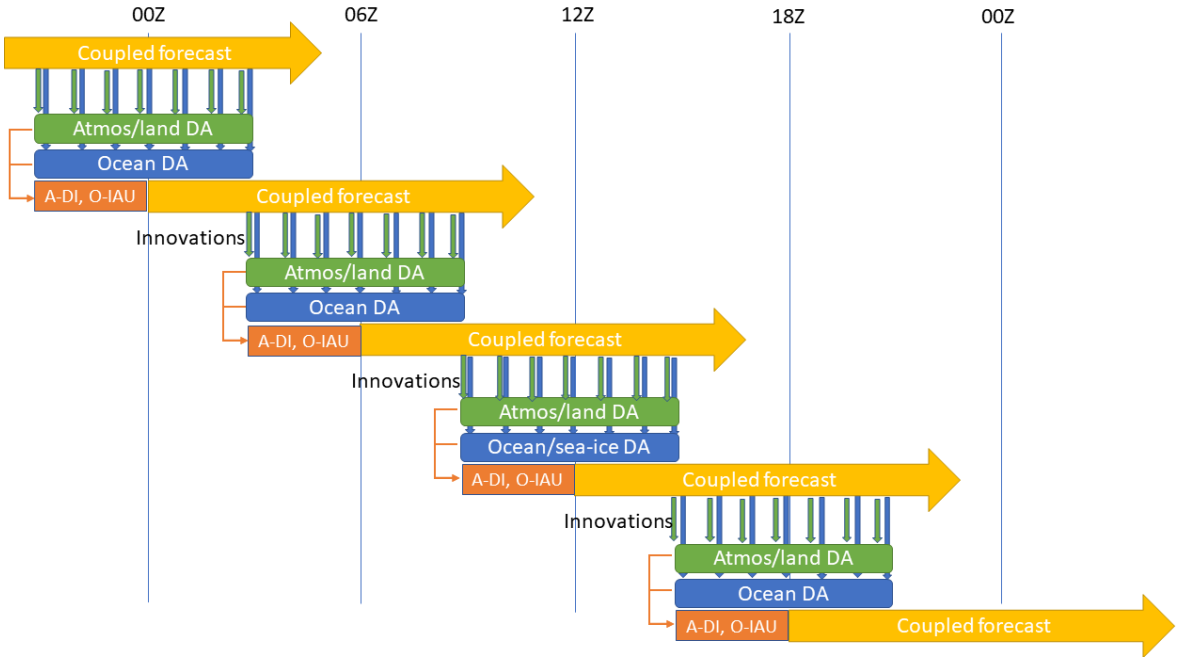


Figure 1: Schematic of the way in which the deterministic coupled data assimilation system is implemented.

In the operational implementation of this system, the main set of analysis/forecast cycles are run just before the end of the assimilation time window on each cycle in order to produce timely products. Reruns of the analyses are carried out later in the day to allow for late-arriving observations. For the

previous atmospheric-only system, only one of these re-runs was carried out. With the coupled system, multiple re-runs are carried out on some cycles in order to allow ocean observations, which typically arrive later than the atmospheric ones, to influence subsequent forecast cycles.

2.1.3 Coupled ensemble generation

The atmospheric component of the ensemble uses an ensemble of hybrid four-dimensional ensemble variational data assimilations (En-4DEnVar) as described by Inverarity et al. (2023). The Mean-Pert method of Lorenc et al. (2017) is used to reduce the cost of the EDA by generating an accurate analysis of the ensemble mean, with the analyses of the ensemble perturbations carried out using fewer minimisation iterations (and therefore lower computational cost). EDA schemes either need to perturb observations or adopt adjusted formulations of the analysis update equation. The atmospheric EDA takes the latter approach using the Relaxation to Prior Perturbations (RTPP) technique (Whitaker & Hamill 2012) (the ocean ensemble adopts the former approach, as described in section 2.2). Ensemble inflation is also used in the atmospheric EDA using the Relaxation to Prior Spread (RTPS) method to account for errors in the DA specifications. Stochastic atmospheric model perturbations are applied in all ensemble member forecasts using the Stochastic Kinetic Energy Backscatter (SKEB) scheme (Tennant et al. 2011) and Stochastic Perturbation of Tendencies (SPT) scheme (Sanchez et al. 2016). An additive inflation scheme (Piccolo et al. 2019) is also used both to provide additional perturbations to the ensemble forecast and to correct model biases.

As mentioned in the introduction, the ocean and sea-ice components of the ensemble are currently initialised from the deterministic analyses on each cycle. This means that there is zero spread in the ocean/sea-ice ensemble at the start of each forecast, and the spread will only grow during the forecast due to the atmospheric perturbations. The aim of this paper is to improve on this, and the developments to do so are described in the next sub-section.

The spread in SST is an important property of the coupled ensemble (see for example Hotta & Ota (2019)) and is crucial for generating spread in the lower troposphere over the oceans. The scheme used to generate these SST perturbations before the use of a coupled model was described by Tennant & Beare (2014). They used the day-to-day SST changes in the OSTIA product as a proxy for the uncertainty in the SST field. The power spectrum of these changes was then used to generate SST anomalies with the spatial covariance structures of the day-to-day SST variations, and with amplitude associated with each spectral component of those changes. Crucial to the work described in the current paper is that an additional calibration factor was used to increase the magnitude of these SST perturbations above the level of the derived uncertainties in OSTIA. This factor, which we call α here, multiplies the basic SST perturbations by two and was determined to give the best ensemble verification scores in the atmosphere (based on the Continuous Ranked Probability Score; CRPS). In the initial implementation of the coupled ensemble, this same SST perturbation method has been used to increase the magnitude of the SST spread in the ensemble as seen by the atmosphere, but this does not directly change the SSTs in the ocean model (only indirectly through changes to the surface fluxes).

2.2 Description of ocean ensemble changes to the coupled NWP system

The development of an improved ocean component of the coupled ensemble has been carried out in an ocean/sea-ice system forced by, but not coupled with, the Met Office atmospheric ensemble, and was reported by Lea et al. (2022). Here we recall the main developments from that work and describe how they have been applied in the coupled NWP ensemble framework.

The first change was to allow the transfer of ocean and sea-ice information from one cycle to the next for each ensemble member. Then a separate ocean/sea-ice data assimilation was implemented for each ensemble member where the observations used in the assimilation are perturbed in their values (by the expected measurement error for each observation) and locations (to mimic errors of

representation). These observation perturbations were shown by Lea et al. (2022) to generate good spread in the different observed variables.

Stochastic model perturbations were also included in the NEMO ocean model for each ensemble member based on schemes described by Storto & Andriopoulos (2021). These include the SPT scheme (Sanchez et al. 2016) which adds a perturbation to the total parametrised physics tendencies that is the product of the unperturbed tendencies and a random field. The stochastically perturbed parameters (SPP) scheme (Ollinaho et al. 2017) introduces stochastic perturbations to model parameters within parametrised processes. The SKEB scheme (Tennant et al. 2011) adds perturbations to the barotropic streamfunction proportionally to some sinks of energy on small scales in the model, which in a higher resolution model would otherwise have been backscattered to the scales we can resolve.

As well as the main changes to the ocean/sea-ice system to implement the above perturbation schemes, Lea et al. (2022) implemented additional changes to the data assimilation in each ensemble member, to use a hybrid-3DEnVar scheme. In that scheme the ensemble forecast-error covariance spatial structures are localised and used in conjunction with the existing 3DVar background error covariances. A multiplicative ensemble inflation scheme was also implemented alongside the hybrid-3DEnVar, based on the RTPS scheme to account for mis-specifications of the error covariances used in the data assimilation.

An assessment of the reliability of the ocean/sea-ice ensemble including the above changes was shown by Lea et al. (2022) to be good after the first 6 months of ensemble spin-up for all the observed variables. The use of hybrid-3DEnVar ocean data assimilation resulted in significant reductions in the error in the ensemble mean compared to the standard 3DVar scheme currently used operationally.

All of the above developments have been implemented in the ocean and sea-ice components of the coupled ensemble system and a summary of the coupled ensemble is given in table 1. The main difference between the coupled ensemble and the previous ocean/sea-ice ensemble system is that the coupled system runs on a 6-hour data assimilation cycle whereas the system of Lea et al. (2022) used a 24-hour cycle. We have not made any adjustments to the ocean/sea-ice ensemble perturbation strategies since they are independent of the cycle length. An assessment of the ocean spread at the shorter 6-hour forecast lead time will be shown in section 4. A schematic of a single member of the coupled ensemble is shown in Fig. 2. Including the ocean ensemble as we do means the ensemble members are now treated in a very similar way to the deterministic run as each now has its own ocean and atmosphere DA increments. One of the questions addressed in this work is whether the developments to the ocean component of the ensemble allow us to remove the previous method of adding SST perturbations as seen by the atmosphere discussed in section 2.1.3, represented by the red arrows in the figure.

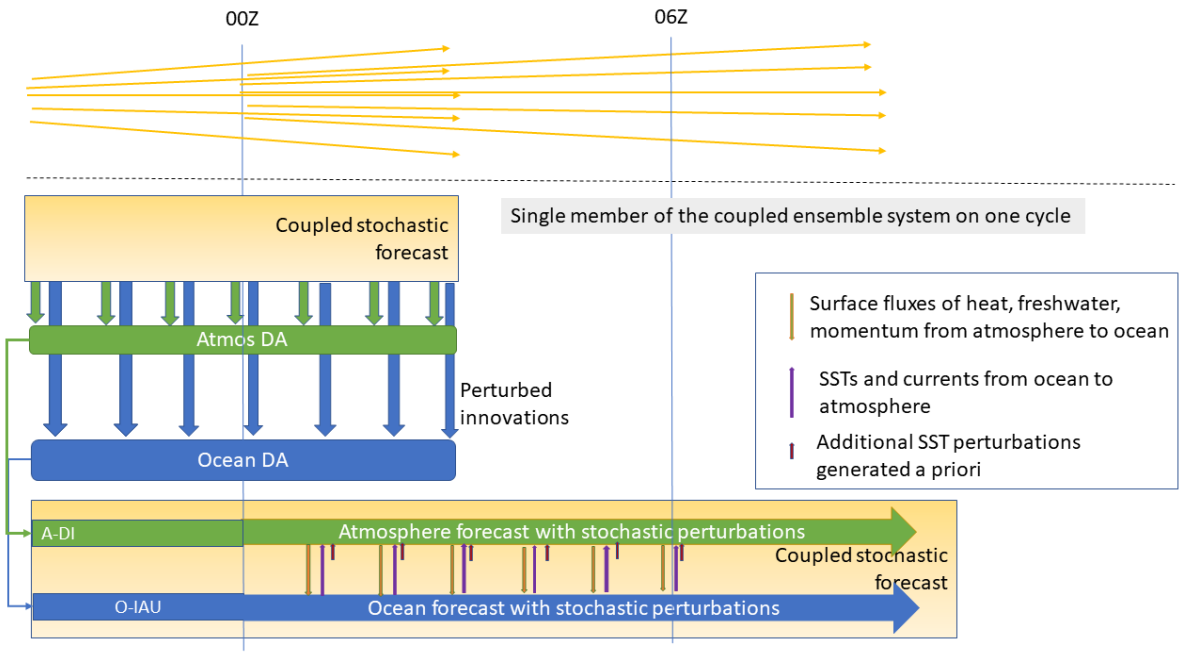


Figure 2: Schematic of the way in which the coupled ensemble system is implemented.

	Atmosphere	**Land**	**Ocean**	**Sea-ice**
Deterministic model configuration	N640 (∼20 km) UM	N640 (∼20 km) JULES	$1/4°$ (∼25 km) NEMO	$1/4°$ (∼25 km) CICE
Ensemble model configuration	N320 (∼40 km) UM	N320 (∼40 km) JULES	$1/4°$ (∼25 km) NEMO	$1/4°$ (∼25 km) CICE
Ensemble data assimilation approach	Ensemble of 4DEnVars with RTPP	Extended Kalman Filter	Ensemble of 3DVars or hybrid-3DEnVars with perturbed observation values and positions	Ensemble of 3DVars with perturbed observation values and positions
Stochastic model perturbations	SPPT, SKEB	Breeding method	SPP, SPPT, SKEB	–
Ensemble inflation	Additive inflation and RTPS	–	RTPS (when hybrid-3DEnVar activated)	–

Table 1: Summary of the coupled ensemble system and developments made to the ocean/sea-ice components (which are in red text). Note that the operational atmosphere/land resolution is N1280 (∼10 km) for the deterministic forecast and N640 (∼20 km) for the ensemble.

3 Experiment description

All experiments were run for the period 1st December 2019 to 31st January 2020. The system runs 44 ensemble members on each 6-hour cycle with forecasts out to 12 hours. Longer forecasts out to 8-days are carried out, but for only 18 members on each 6-hour cycle. Initial conditions for the ocean/sea-ice ensemble on 1st December were taken from an uncoupled ocean/sea-ice ensemble run which had been run for about 21 months. The first portion of the experiment is the same as described by Lea et al.

(2022) (experiment *ensda08_inf08* in that paper) and the experiment was continued on to 1st Dec 2019 to provide initial conditions for these experiments. That ensemble consisted of only 36 members, so we initialised members 37-44 of the new ensemble with the first 8 members of the uncoupled ensemble. The atmospheric initial conditions were taken from the operational MOGREPS–G ensemble.

All the experiments described in this paper are coupled ensemble experiments and only differ in the way the ocean and sea-ice components of the ensemble are designed and used. A control experiment (*Control*) was run which was set to be the same as the present operational system. None of the ocean ensemble developments are included in that experiment and the previous method for adding SST perturbations based on Tennant & Beare (2014) was used. The second experiment, *OEn3DVar*, included all the ocean ensemble developments described above with a 3DVar method used in the ocean DA, and the additional SST perturbations were removed. Since the SST spread in experiment *OEn3DVar* was much lower than the spread of the perturbations used in the current operational system, a tuning exercise was carried out to determine an α factor which could be applied to the additional SST perturbations such that the total SST spread from the ocean and SST perturbations would add up to a similar level globally as that of Tennant & Beare (2014). This resulted in an α factor of 1.8, and an experiment with both the ocean perturbations and the additional SST perturbations generated using this reduced α factor was carried out (*OEn3DVar_SSTperts*). The last experiment, *OHybrEn3DVar*, is one in which the ocean DA uses hybrid-3DEnVar rather than 3DVar in all ensemble members, with a hybrid weight of 0.5 (equal weight is given to the ensemble-based and modelled background error covariances), but no additional SST perturbations are included. The hybrid DA is described in more detail in section 4.3 and in Lea et al. (2022). A summary of these experiments is given in table 2.

Experiment name	**Description**	**SST perturbation factor*.**
Control	A coupled ensemble experiment where the ocean ensemble members are all initialised from the deterministic ocean analysis on every cycle.	2.0
OEn3DVar	Same as *Control* except the ocean ensemble is initialised each cycle with an ensemble of 3DVars with observation perturbations, and includes stochastic ocean model perturbations.	0.0
OEn3DVar_SSTperts	As for *OEn3DVar* except the additional SST perturbations are also included	1.8
OHybrEn3DVar	As for *OEn3DVar* but each ocean data assimilation uses a hybrid-En3DVar with a hybrid weight of 0.5.	0.0

Table 2: List of experiments. All experiments use a coupled model and differ only in the way the ocean component of the ensemble is configured and in the SST perturbations which the atmospheric component uses. *The SST perturbation factor is applied to a set of additional SST perturbations generated *a priori* which are applied in the atmospheric model on top of the SST generated by the ocean model.

4 Impact of improving the ocean component of the coupled ensemble

4.1 Impact on the ocean

4.1.1 Overall impacts on characteristics of the ensemble at 6-hour forecast lead times

In Fig. 3 we plot time-series of the innovations for the 6-hour deterministic forecasts of Sea Level Anomaly (SLA), SST, profile temperature and profile salinity comparing experiments *Control*, *OEn3DVar*, *OEn3DVar_SSTperts* and *OHybrEn3DVar*. In all the plots there is an initial adjustment taking only a few days in the statistics. This may be because the initial ocean restarts come from a 24 hour cycling uncoupled ocean system. For SLA there is a dramatic reduction in the SLA RMS innovations in *OHybrEn3DVar* compared to all the other experiments. This result is consistent with that seen in the ocean only results by Lea et al. (2022) when hybrid DA is used. There is little difference between the experiments with in situ SST. For the profile temperature and profile salinity, hybrid DA does result in some reduction in the RMS innovations.

The ensemble mean RMS innovations time-series results in Fig. 4 are O(10%) smaller than the deterministic for all experiments and SLA, profile temperature and salinity observation types. The exception is SST which shows no reduction in RMS innovations compared to the deterministic run.

The ensemble spread at 6-hour lead time in Fig. 5 is very similar for *OEn3DVar* and *OEn3DVar_SSTperts* except that the SST perturbation experiment does have a slightly increased spread in SST which suggests a weak feedback from perturbing the SSTs seen by the atmosphere into the ocean ensemble. The SST spread is lower for hybrid DA compared to those other 3DVar experiments. For SLA, even though we started the runs with a fully spun up ocean ensemble, the spread does increase significantly in the *OEn3DVar* and *OEn3DVar_SSTperts* experiments, taking about 1 month to stabilise. In contrast, the spread in the *OHybrEn3DVar* experiment remains fairly constant during the run. Perhaps the change in spread in the other experiments is reflecting the fact that they have larger forecast errors, as shown in Fig. 4 (note the ocean ensemble was generated by a previous ocean-only run which was doing hybrid DA). The profile temperature and salinity spread in the *OEn3DVar* and *OEn3DVar_SSTperts* experiments show an increase over the first month of the runs, similar to that of SLA but again the hybrid DA reduces the ensemble spread.

For the hybrid DA the spread is controlled partly by the RTPS (Return To Prior Spread) multiplicative inflation scheme. We used the same inflation factors here as in the ocean only experiments of Lea et al. (2022). The above results could indicate a need to retune the inflation factor for the coupled NWP context in order to improve the ensemble reliability. More discussion on the ensemble reliability can be found in section 4.1.5.

The spatial patterns of the RMS of the SLA ensemble mean innovations, Fig. 6, show that the SLA innovations are highest in regions with strong eddy activity like the Antarctic Circumpolar Current (ACC) and the Gulf Stream. Consistent with this, the ensemble spread is higher in these locations. Hybrid DA reduces the error and spread in these same locations. There is a small increase in error and spread in the tropical Atlantic and eastern tropical Indian ocean. More details on the impact of hybrid DA and the problems in this location can be found in section 4.3.

The spatial patterns of the RMS of the in situ SST innovations for the ensemble mean, Fig. 7, show that the SST innovations are large in regions with strong eddy activity like the ACC, the Gulf Stream, Kuroshio, and the Agulhas as was seen in the SLA results, but also in other regions such as the upwelling regions off the west coast of Africa, North and South America. The ensemble spread reflects the larger errors in some of these regions, such as the Gulf Stream and Kuroshio, but other areas of increased RMS do not show up in the ensemble spread. This indicates a deficiency in the ensemble generation for SST in the coupled NWP implementation which was not seen in the previous ocean-only

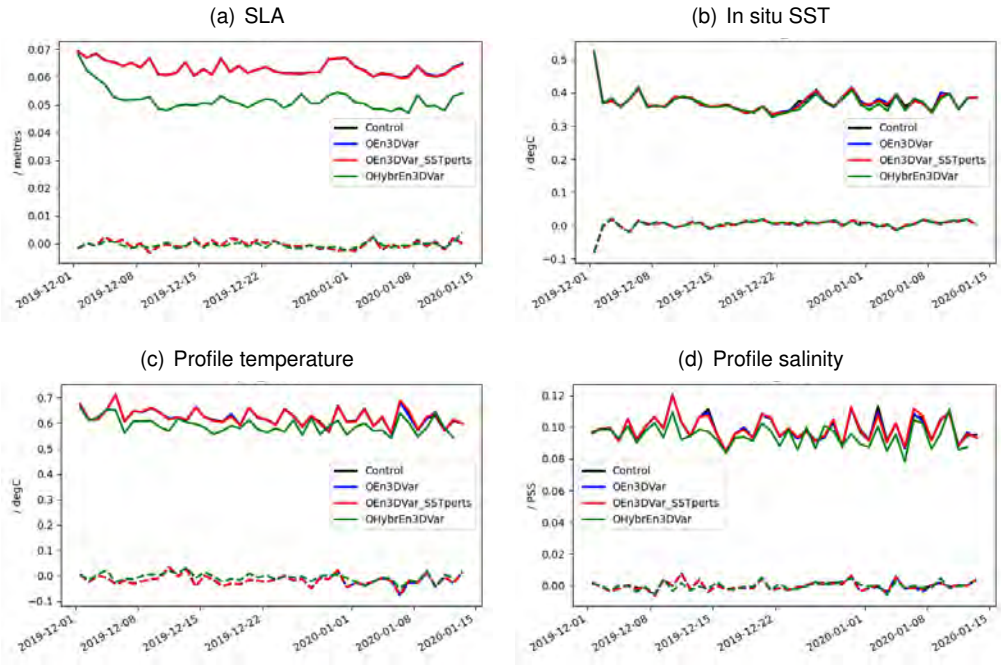


Figure 3: Time-series of the RMS (solid lines) and mean (dashed lines) in the deterministic 6-hour forecast errors compared to observations.

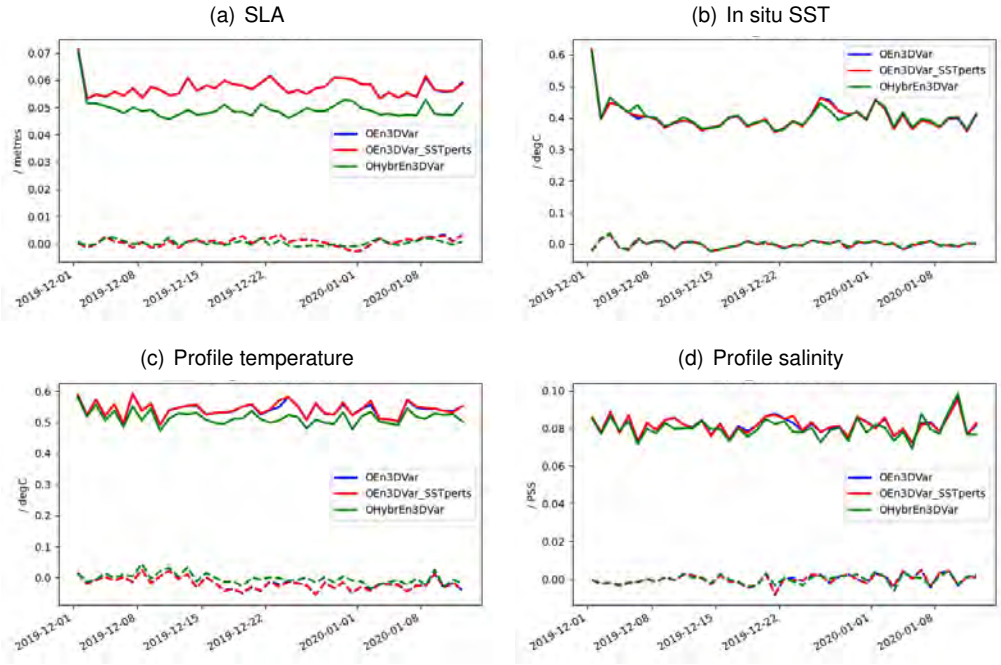


Figure 4: Time-series of the RMS (solid lines) and mean (dashed lines) in the ensemble mean 6-hour forecast errors compared to observations.

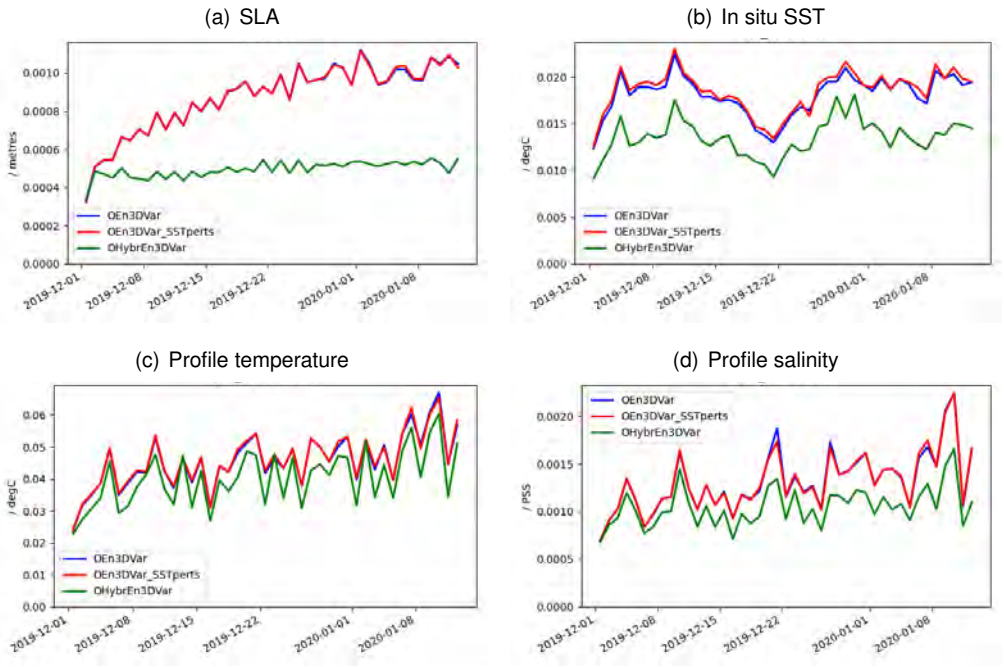


Figure 5: Time-series of the RMS (solid lines) and mean (dashed lines) in the 6-hour forecast ensemble spread.

work.

4.1.2 Ocean deterministic forecast verification

The impact of the ocean ensemble developments on ocean fields in the deterministic forecasts can be assessed using “Class 4” statistics. This is where the forecast is compared to observations at the observation locations. In Fig. 8 we compare the experiments. Encouragingly the improvement seen with hybrid DA is retained into the forecast for all variable types (SLA, SST, profile temperature and salinity). Indeed for SST, while the error still increases during the forecast, the improvement seen with hybrid DA also increases with forecast lead time (there is more improvement at the end of the 6-day forecast than in the analysis). This is perhaps due to the improvements in the other variables’ initial conditions affecting the SST later in the forecast.

The deterministic forecast performance differs little between the *Control* and ocean ensemble 3DVar experiments as might be expected since the ocean data assimilation on the deterministic is identical for these experiments. Small differences can arise due to impact of the ocean ensemble on the atmosphere ensemble which then affects the atmosphere assimilation and atmosphere model which then affects the ocean.

The other result of note is that there is an SST bias which develops in the forecast. This is particularly evident (not shown) in the tropical regions and the Southern Hemisphere (where the ocean mixed layer is shallower).

4.1.3 Impact on ensemble mean and spread in the ocean

The impact of the atmospheric SST perturbation changes on the mean and spread of the ensemble in the ocean is generally very small so the results shown below focus on the changes from the *Control* experiment to the *OEn3DVar* experiment.

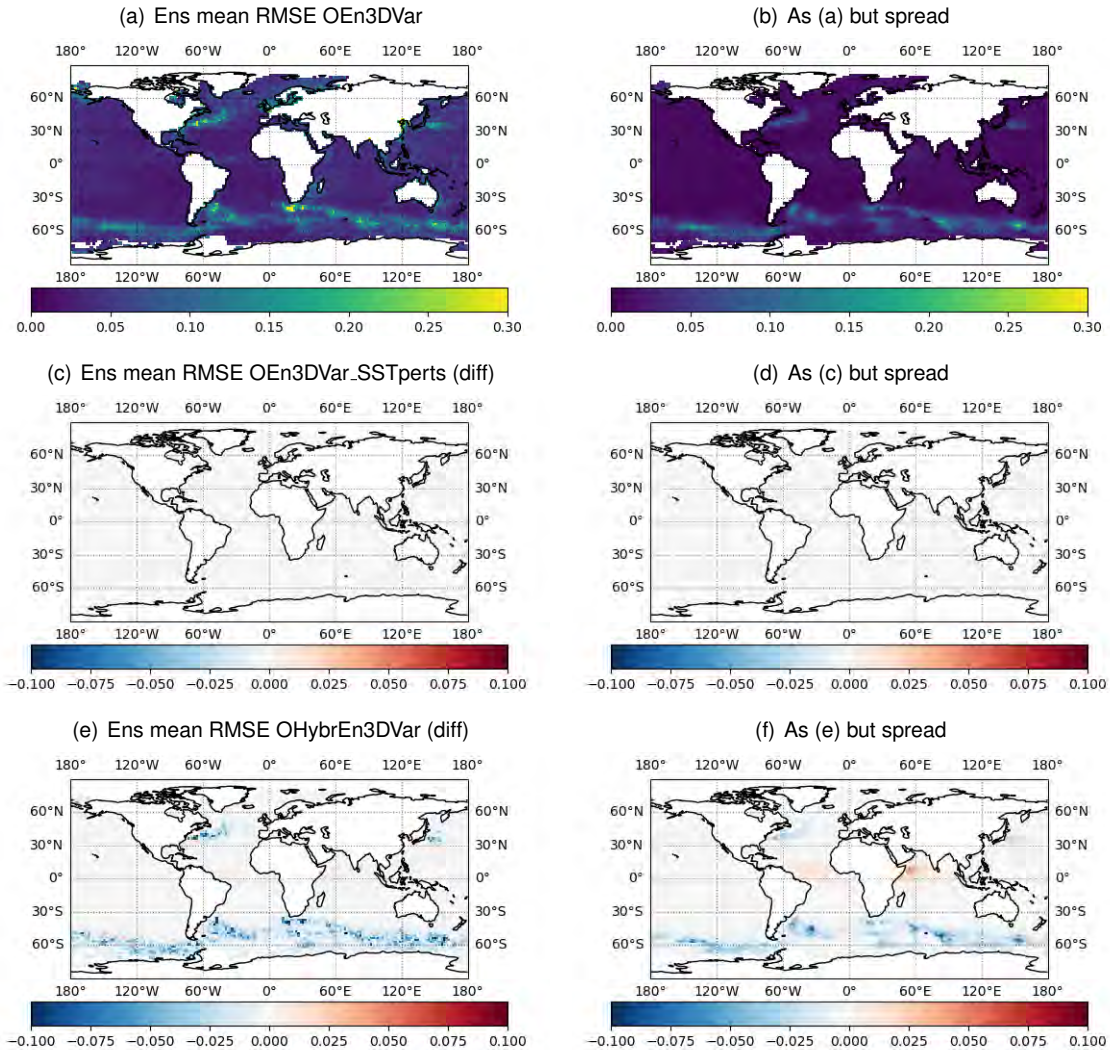


Figure 6: The SLA (m) 6-hour ensemble mean forecast error is shown in panel (a), and ensemble spread for *OEn3DVar* is shown in panel (b). Panels (c) and (d) show the difference in forecast error and spread to *OEn3DVar* for *OEn3DVar_SSTperts*, and (e) and (f) show the differences to *OEn3DVar* for *OHybrEn3DVar*. For (c)–(f) blue/red indicates a reduction/increase in error or spread compared to *OEn3DVar*.

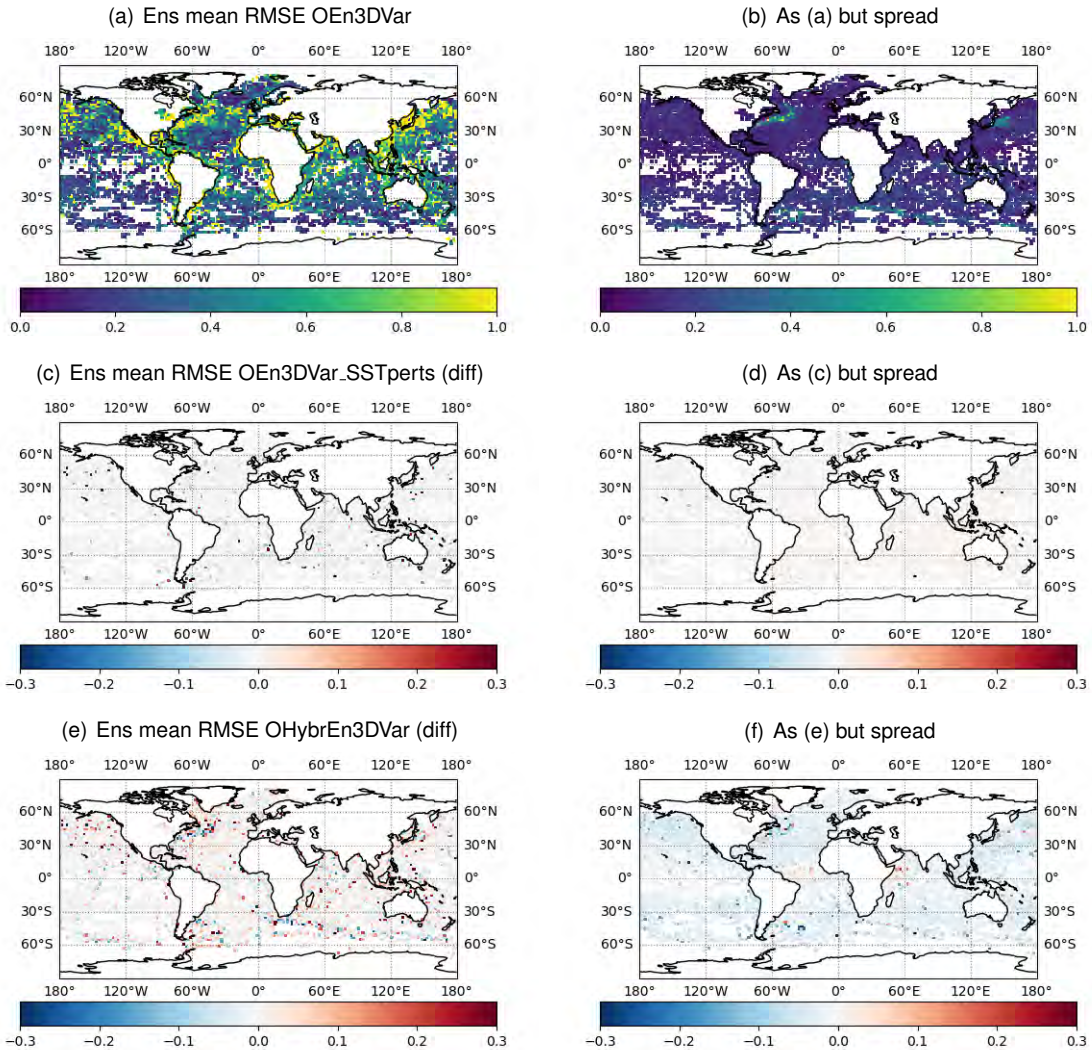


Figure 7: Spatial plots of SST (°C) 6-hour ensemble mean forecast error and ensemble spread. As Fig. 6 but for SST.

(a) SLA

(b) In situ SST

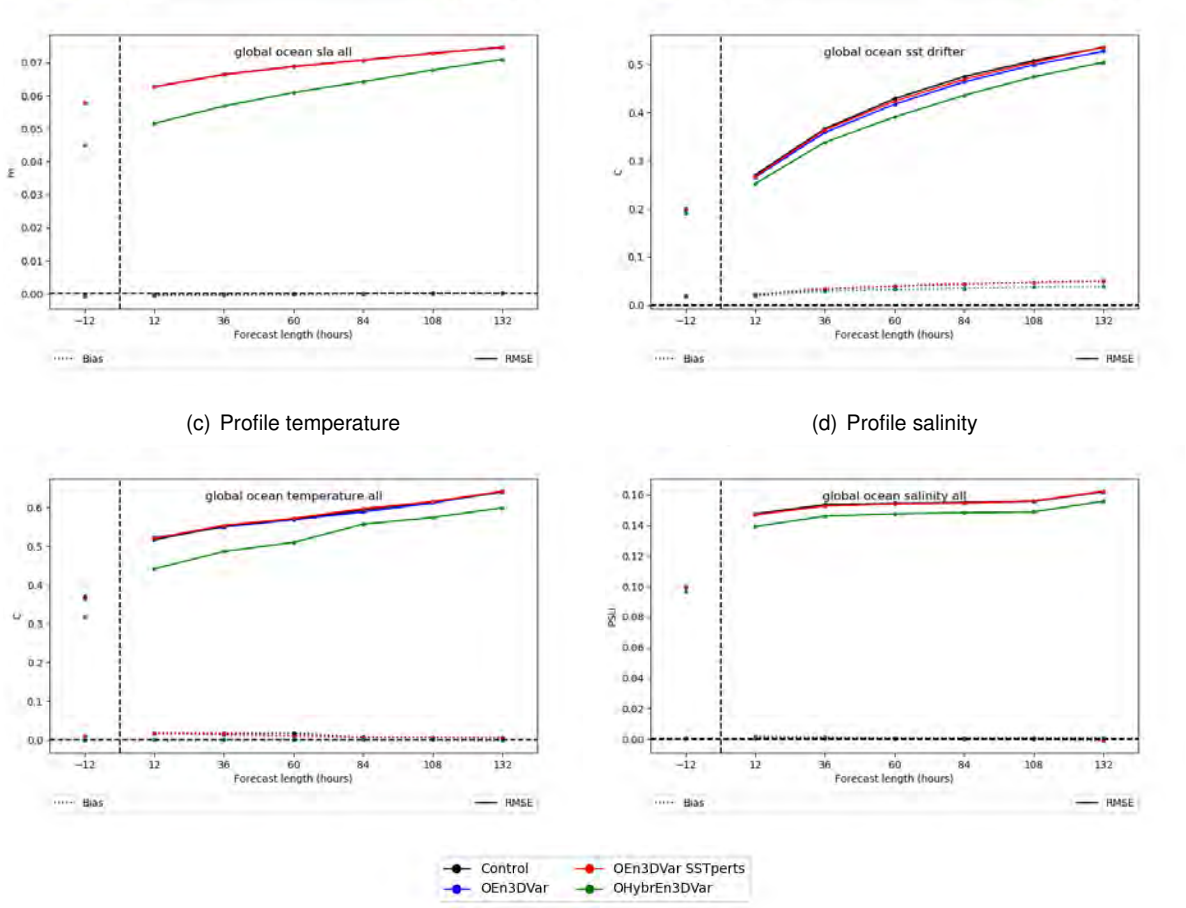


Figure 8: Deterministic ocean forecast RMSE error (solid lines) and bias (dotted lines) compared to various observation times as function of forecast lead time (also known as “Class 4 statistics”). Statistics calculated over 1 Dec 2019 – 12 Jan 2020.

Fig. 9 illustrates significant differences between the mean SST in the *OEn3DVar* and *Control* in particular at high latitudes where *OEn3DVar* is about 0.5°C cooler and there is a warming to a lesser degree at lower latitudes. This signal is sustained into the forecast.

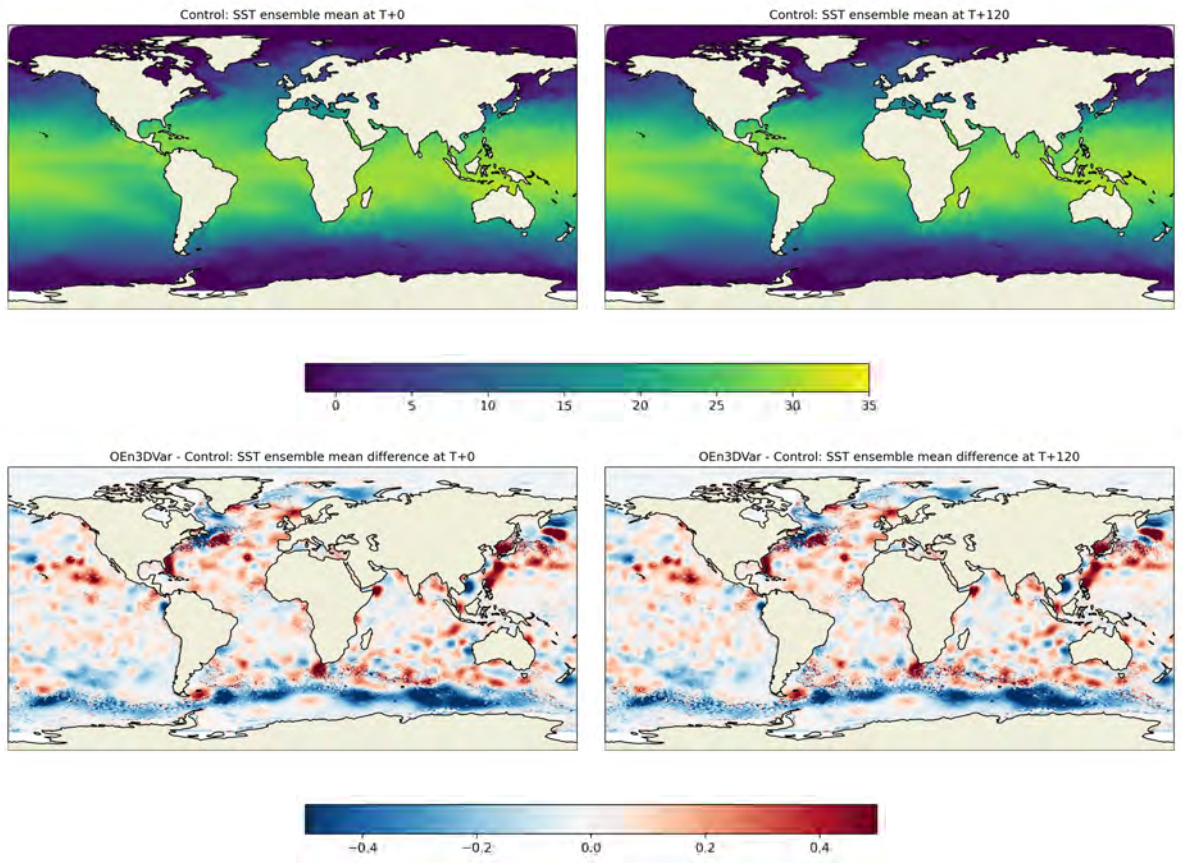


Figure 9: SST ensemble mean (°C) at T+0 (left) and T+120 (right) averaged over forecasts initialised between 1st – 15th January 2020. Top plots: *Control*, bottom plots: *OEn3DVar* minus *Control*.

The ensemble SST spread in the *Control* shown in Fig. 10 is very small since the ensemble is reinitialised from the deterministic analysis on every cycle. The spread does increase into the forecast somewhat due to the differences in atmosphere forcing from the atmosphere ensemble members, but only where the mixed layer is shallow. When we introduce the ocean ensemble in *OEn3DVar* the spread increases markedly, even at T+0 as the ensemble spread is now retained from cycle to cycle, with high spread in locations with strong currents and strong eddy activity.

For Sea Surface Salinity (SSS) the ensemble mean salinity is higher in *OEn3DVar* than *Control* in most of the Arctic where there is sea-ice (Fig. 11). This suggests a difference in the ensemble mean sea-ice concentration could be driving this salinity difference. At lower latitudes the differences in the ensemble mean are smaller scale resulting perhaps from chaotic changes in the circulation. The SSS spread again is very weak in the *Control* (Fig. 12). During the forecast the spread does increase in the control as a result of the forcing on the ocean from the spread in the atmosphere ensemble. The spread is much higher with the ocean ensemble included, even at T+0. There is particularly high spread in regions with strong currents, and in locations with strong river outflow like the Bay of Bengal, near the Amazon outflow and the Congo river outflow. There is also strong variability in parts of the Arctic perhaps associated with variations in the sea-ice concentration.

The ensemble mean mixed layer depth (MLD) is particularly deep in the North Atlantic and in some parts of the ACC (Fig.13). There is relatively little change in the mean MLD in the *OEn3DVar* experiment, though there is a small increase in mean MLD in the Arctic and south of the ACC. The MLD spread again

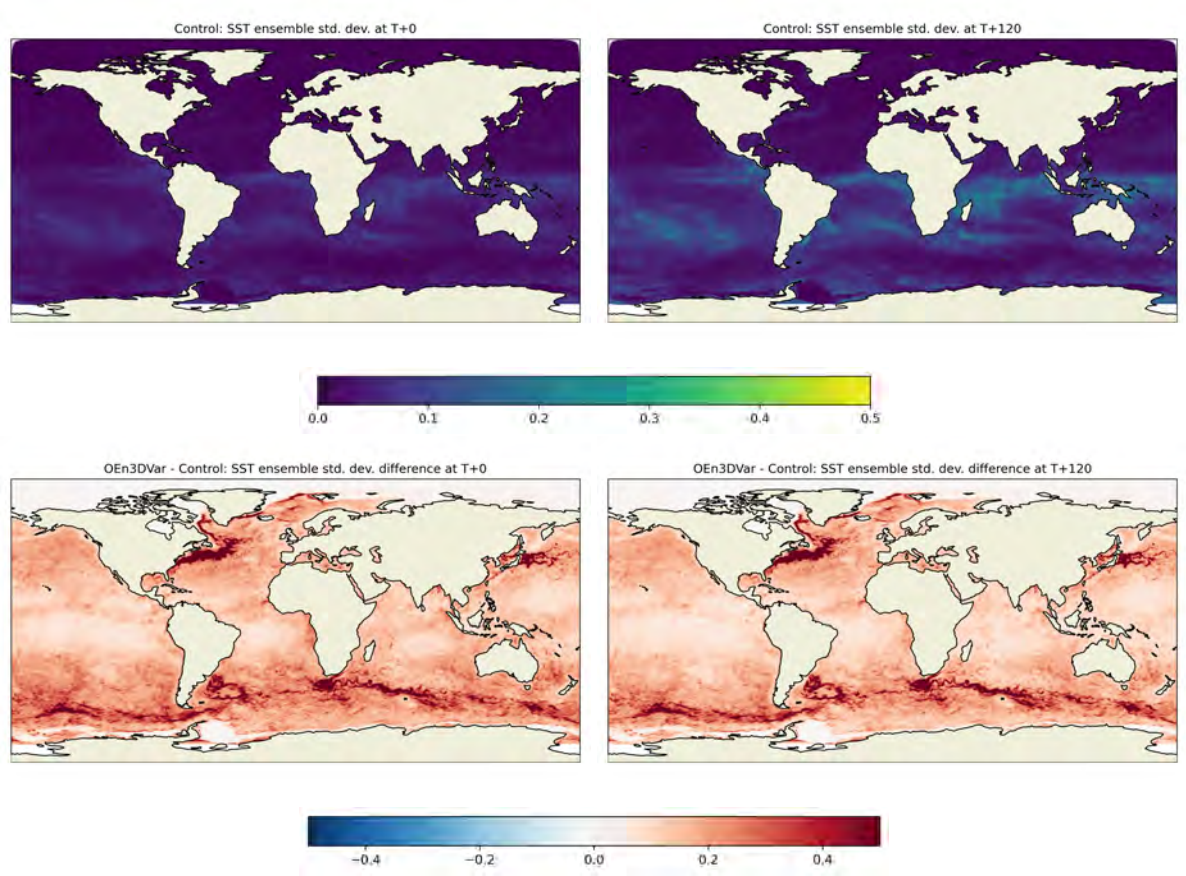


Figure 10: SST ensemble standard deviation (°C) at T+0 (left) and T+120 (right) averaged over forecasts initialised between 1st – 15th January 2020. Top plots: *Control*, bottom plots: *OEn3DVar* minus *Control*.

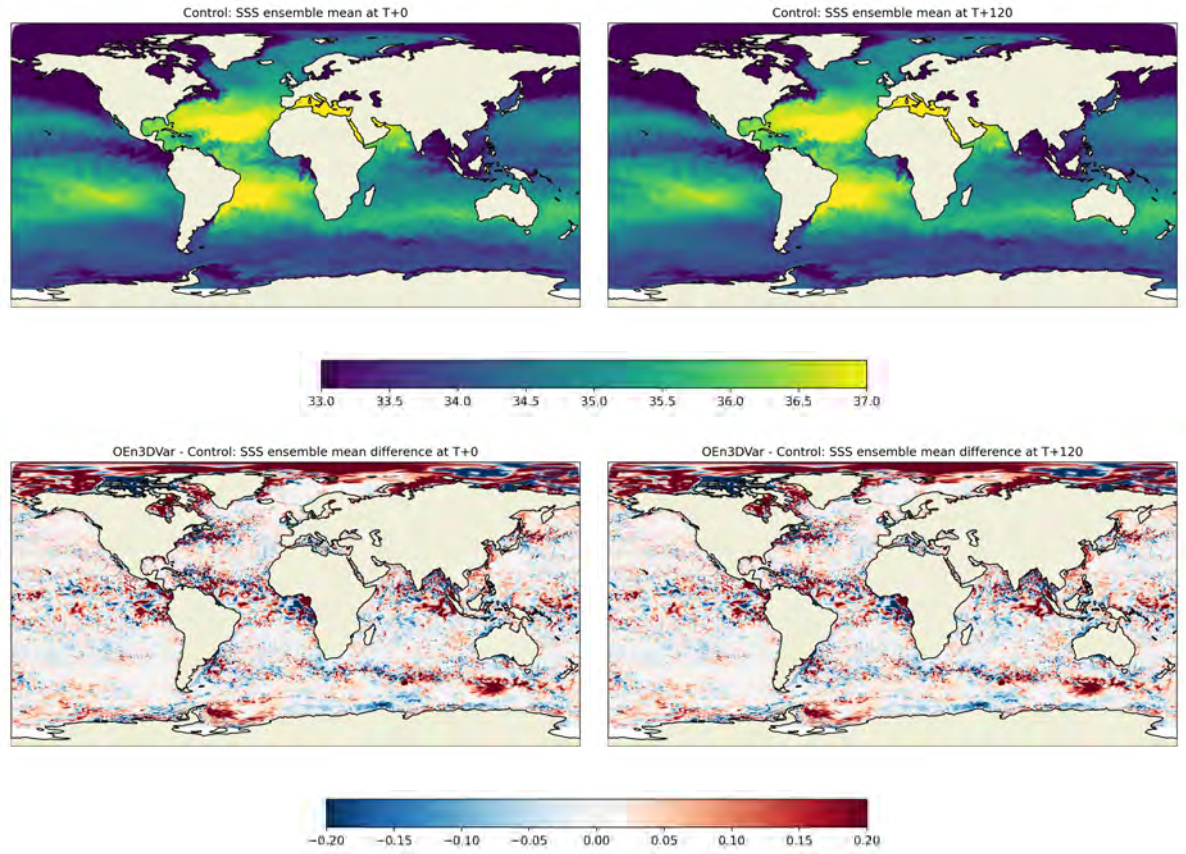


Figure 11: SSS ensemble mean (psu) at T+0 (left) and T+120 (right) averaged over forecasts initialised between 1st – 15th January 2020. Top plots: *Control*, bottom plots: *OEn3DVar* minus *Control*.

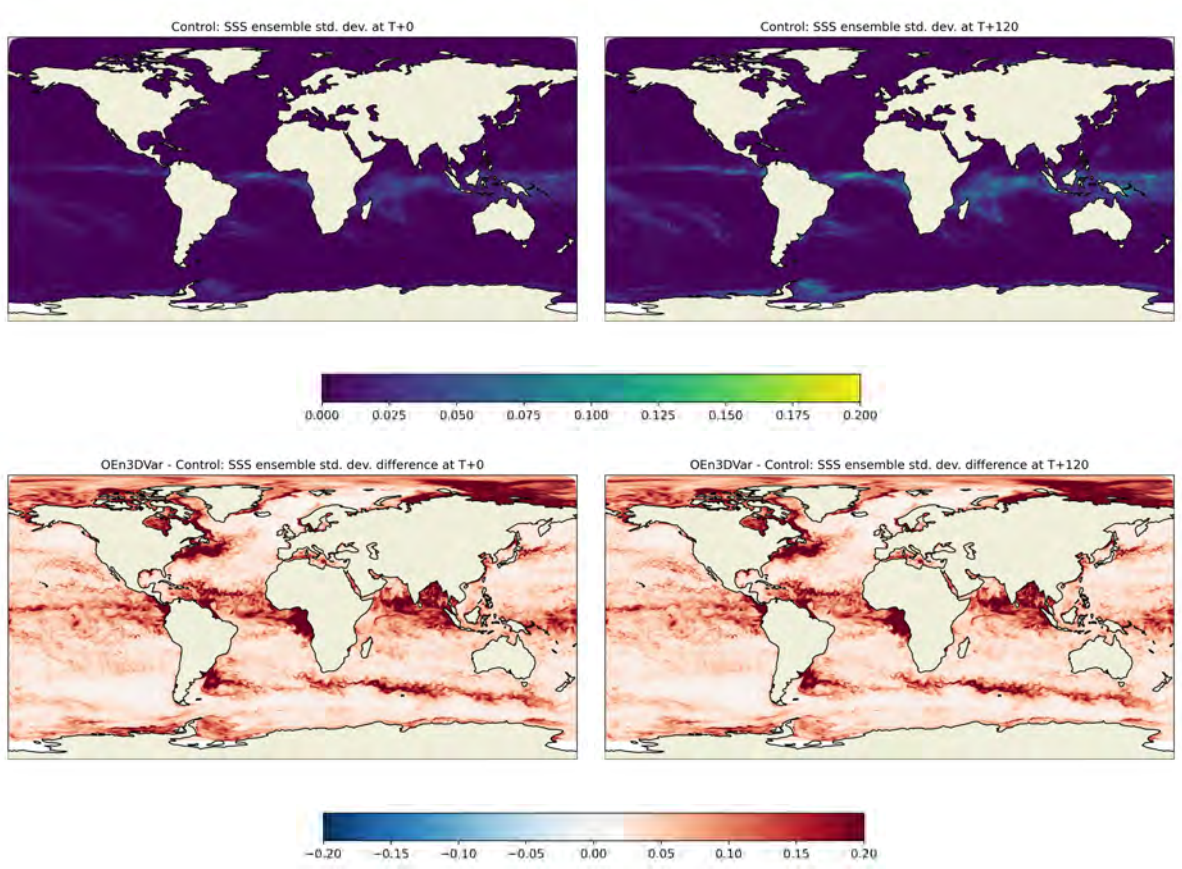


Figure 12: SSS ensemble standard deviation (psu) at T+0 (left) and T+120 (right) averaged over forecasts initialised between 1st – 15th January 2020. Top plots: *Control*, bottom plots: *OEn3DVar* minus *Control*.

is generally small in the *Control* (Fig. 14). As the forecast proceeds however it is high in some locations where the ocean is sensitive to the spread in the atmosphere forcing. The spread increases nearly everywhere with the ocean ensemble included, most notably in the ACC region and North Atlantic, and also the Gulf Stream and Kuroshio.

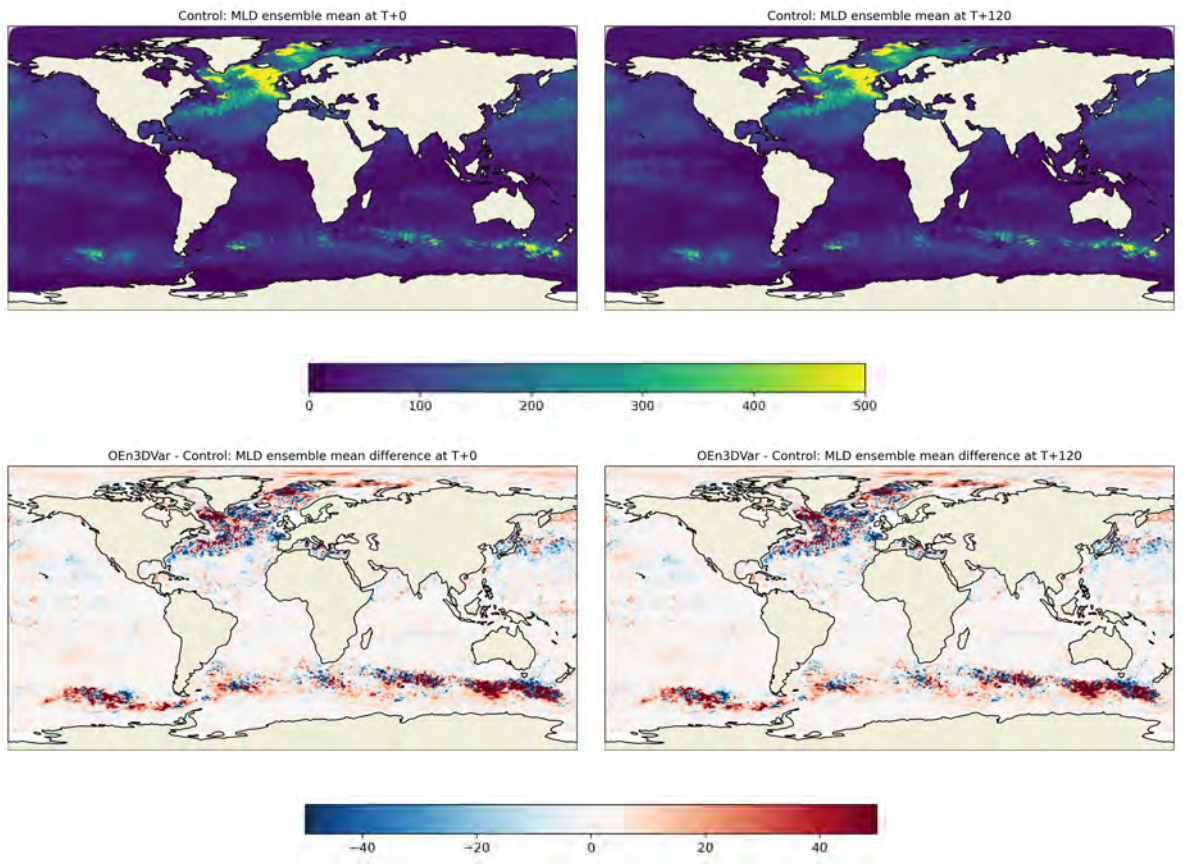


Figure 13: MLD ensemble mean (m) at T+0 (left) and T+120 (right) averaged over forecasts initialised between 1st – 15th January 2020. Top plots: *Control*, bottom plots: *OEn3DVar* minus *Control*.

The model Sea Surface Height (SSH) ensemble mean is changed at high latitudes comparing *Control* to *OEn3DVar* (Fig. 15). The SSH spread, again, is generally small in the *Control* (Fig. 16) and as the forecast proceeds it increases only in shelf locations where the spread in the atmosphere forcing induces variations in the barotropic SSH. The spread increases nearly everywhere with the ocean ensemble included. Particularly high spreads are seen in the ACC region and also the Gulf Stream and Kuroshio.

4.1.4 Impact on ensemble mean and spread in the sea-ice

For sea-ice concentration the ensemble mean is reduced in the Arctic at high latitudes in *OEn3DVar* compared to *Control* (Fig. 17) except near the ice edge where there is a small increase in ice concentration. A plausible reason for this pattern is the impact of the observation perturbations on the assimilation due to the fact that the ice concentration cannot exceed 100% or go below 0%. In the ice pack where the ice concentration is close to 100% the observation value perturbations are bound to reduce the average concentration. The opposite effect may hold around the ice edge where the ice concentration cannot go below 0% resulting in a positive mean change. To resolve this changes in the observations perturbation strategy for ice would be needed. It may be wise to reduce observation value perturbations where the values are close to hard limits.

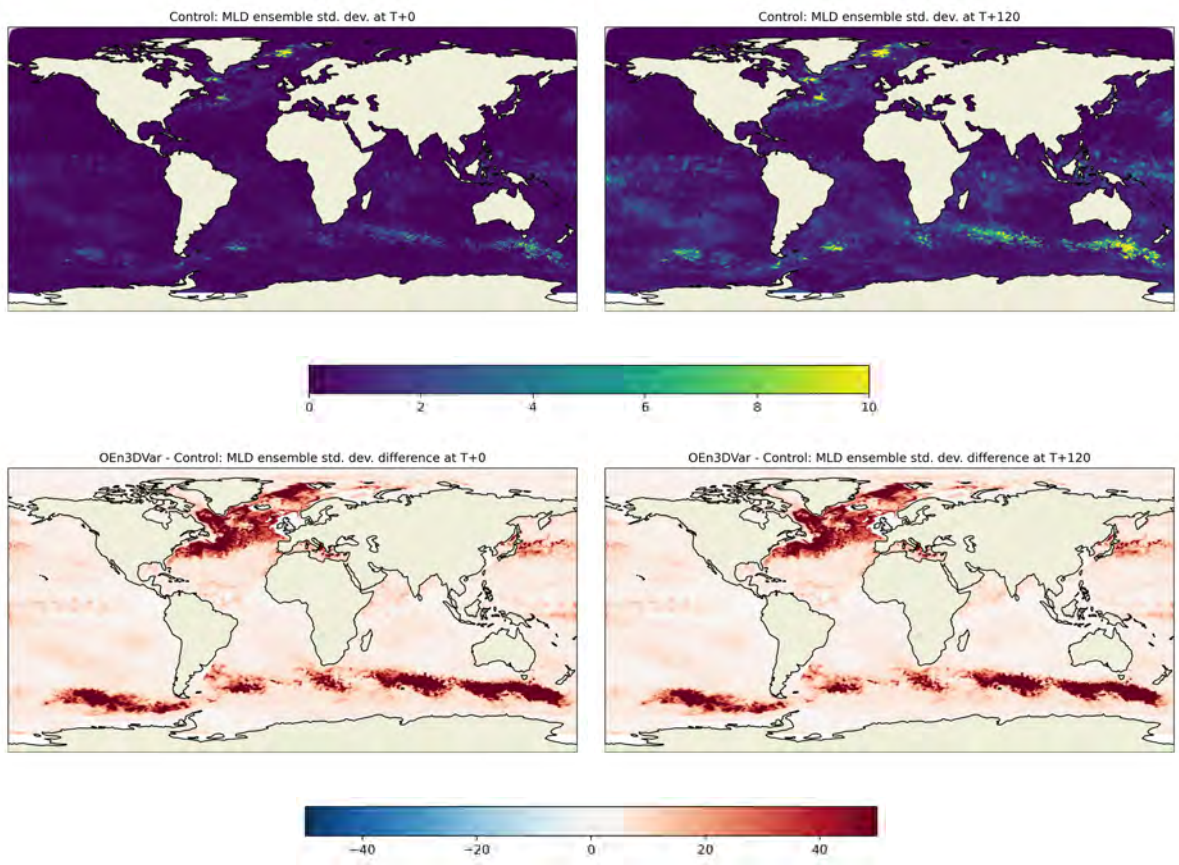


Figure 14: MLD ensemble standard deviation (m) at T+0 (left) and T+120 (right) averaged over forecasts initialised between 1st – 15th January 2020. Top plots: *Control*, bottom plots: *OEn3DVar* minus *Control*.

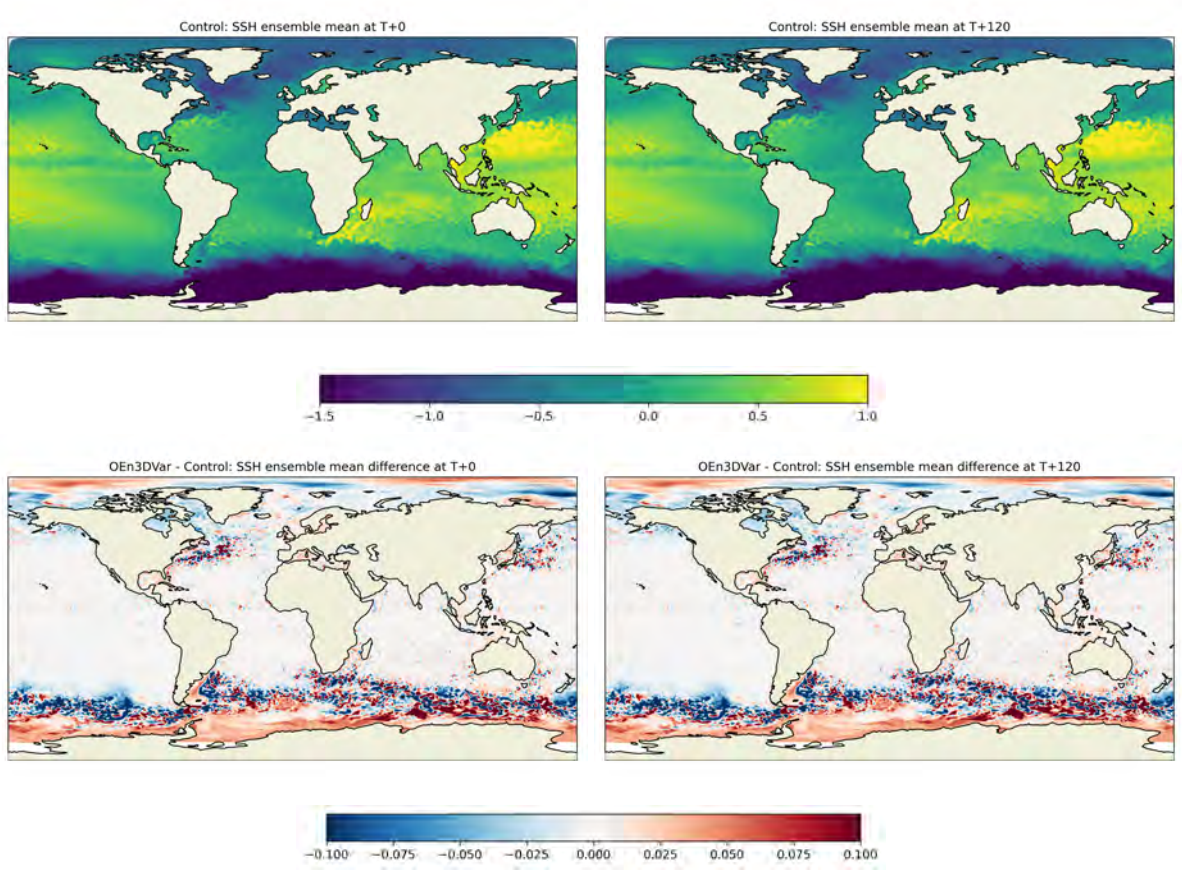


Figure 15: SSH ensemble mean (m) at T+0 (left) and T+120 (right) averaged over forecasts initialised between 1st – 15th January 2020. Top plots: *Control*, bottom plots: *OEn3DVar* minus *Control*.

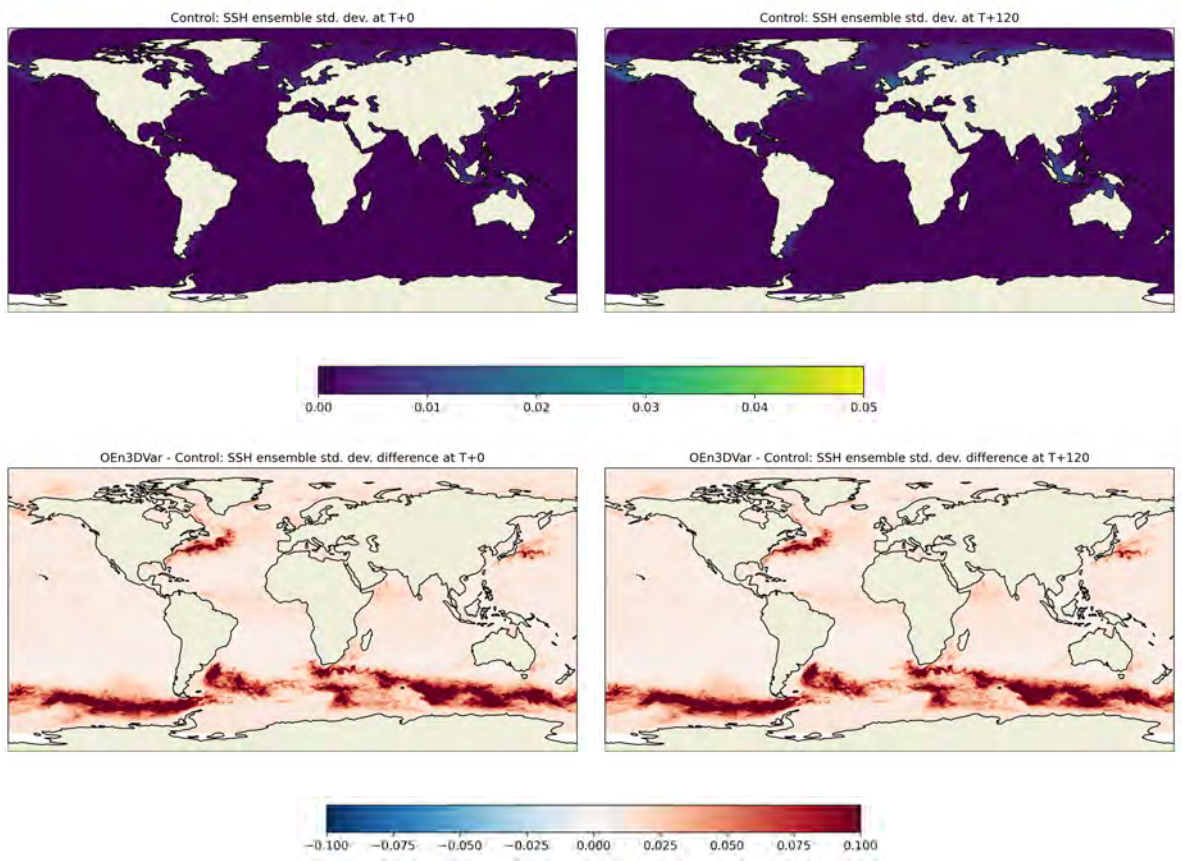


Figure 16: SSH ensemble standard deviation (m) at T+0 (left) and T+120 (right) averaged over forecasts initialised between 1st – 15th January 2020. Top plots: *Control*, bottom plots: *OEn3DVar* minus *Control*.

The sea-ice spread is generally small in the *Control* as the sea-ice ensemble, like the ocean ensemble, is reset each cycle in this experiment. The spread increases in *OEn3DVar* in the ice pack and particularly strongly near the ice edge (Fig. 18).

The ensemble mean sea-ice thickness is shown in Fig. 19. The thickness is generally reduced in *OEn3DVar* compared to *Control*. This is consistent with the ice concentration results above. The ensemble changes in *OEn3DVar* generate significant increases in ensemble spread in sea-ice thickness as shown in Fig. 20. This is particularly obvious to the east of Greenland and near the Canadian Archipelago. There are even larger increases in spread in thickness in the Antarctic.

The ensemble mean sea-ice temperatures are -20 to -30°C over the ice compared to water temperatures of about -2°C in the Arctic in Fig. 21. This means that the small reduction in the sea-ice concentration in the Arctic in *OEn3DVar* can increase the (grid cell average) ice surface temperature markedly (since more of the grid cell will be open water which is warmer than the ice). In this case the mean ice concentration is only reduced by 2-3%, but that leads to a change in ice surface temperature of 2-4°C (shown in Fig. 21). The sea-ice temperature ensemble standard deviation is shown in Fig. 22. This shows high spread in the Arctic but weak spread in the Antarctic where the ice temperature and ocean temperature are only minimally different in Southern Hemisphere summer.

In Fig. 23 we show time-series of sea-ice concentration innovation statistics comparing all the experiments in the deterministic runs where there are no significant differences. This is not too surprising since we do not include ice ensemble perturbations in the hybrid DA. For the ensemble mean statistics again the hybrid DA shows no difference to the equivalent 3DVar experiment. Including the additional SST perturbations to the atmosphere degrades the sea-ice concentration RMSE by around 20%. The SST perturbations also increase the ensemble spread of sea-ice concentration by around 20%. It may be that the SST perturbations are applied in the sea-ice covered regions of the ocean which through the coupling results in excessive sea-ice melting. If this is the case we will look to ramp down the SST perturbations over parts of the ocean which contain sea-ice.

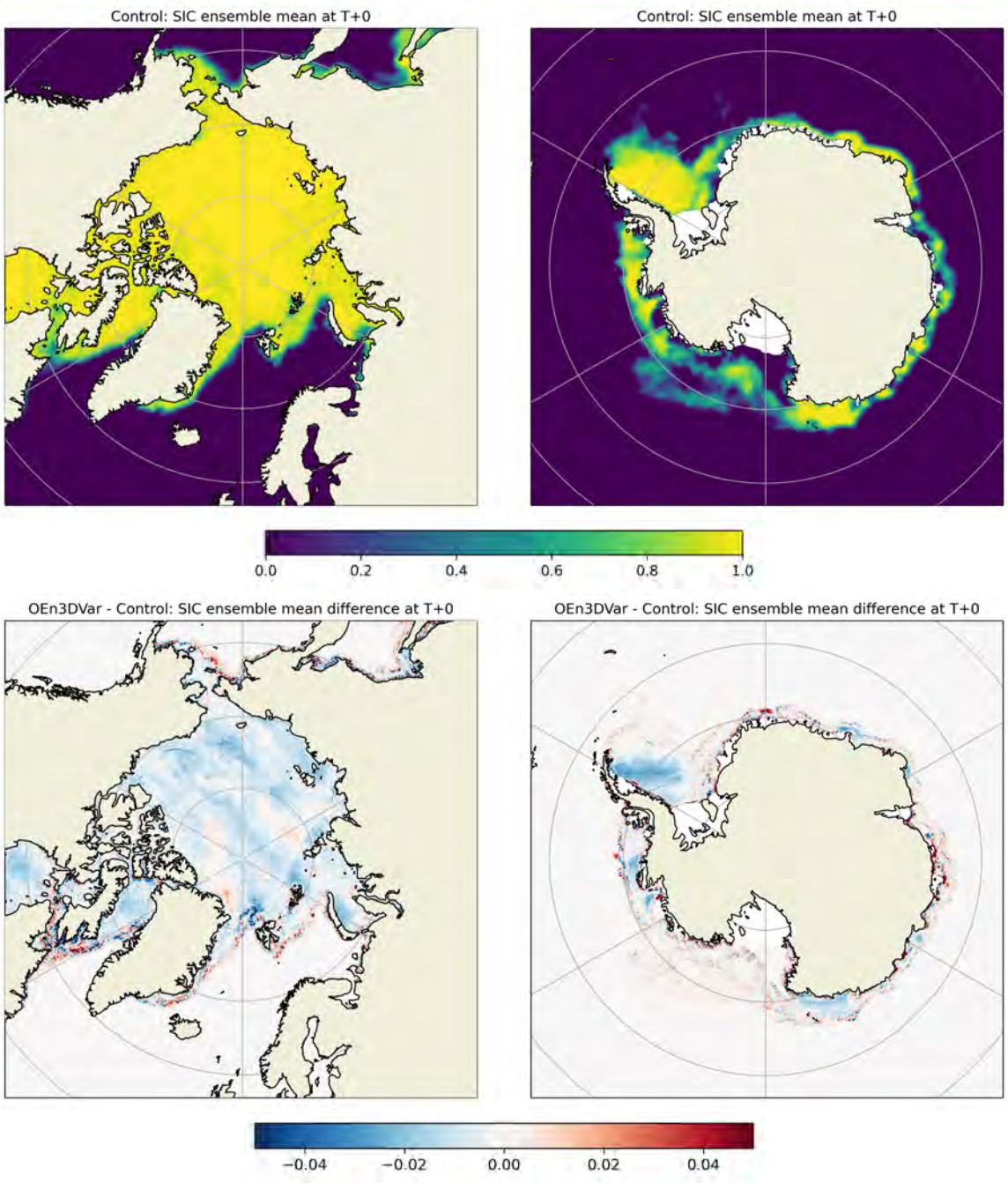


Figure 17: Sea-ice concentration ensemble mean (concentration) at T+0 for the Arctic (left) and Antarctic (right) averaged over forecasts initialised between 1st – 15th January 2020. Top plots: *Control*, bottom plots: *OEn3DVar* minus *Control*.

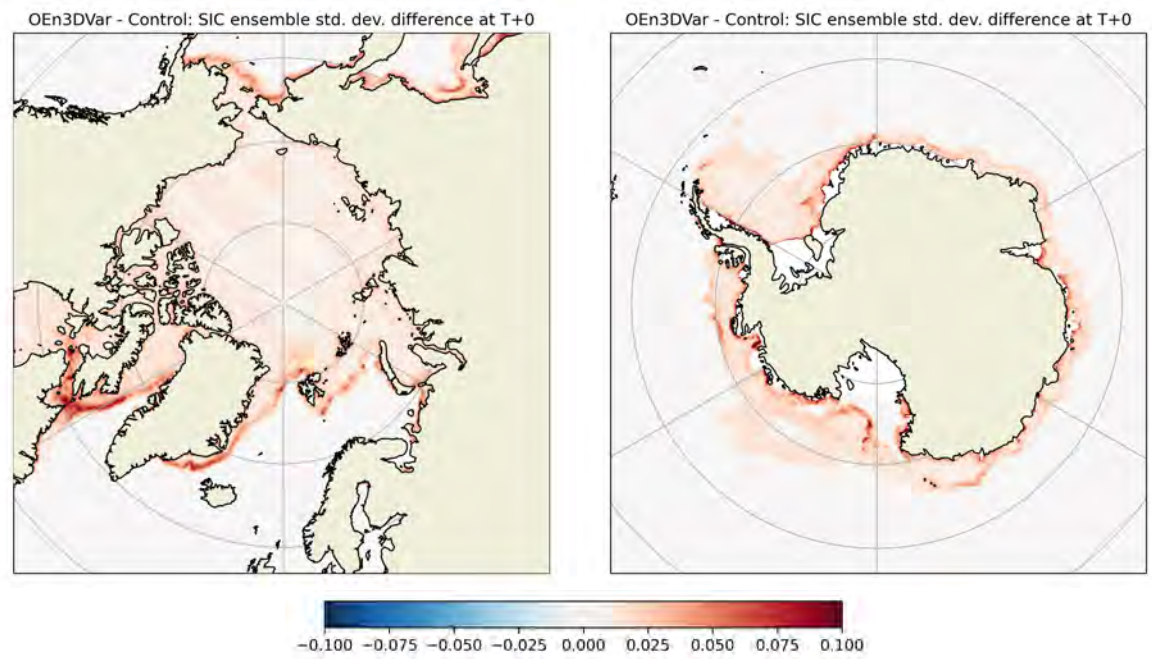


Figure 18: Change (*OEn3DVar* minus *Control*) in sea-ice concentration ensemble standard deviation (concentration) at T+0 for the Arctic (left) and Antarctic (right) averaged over forecasts initialised between 1st – 15th January 2020. The standard deviation of *Control* is very small.

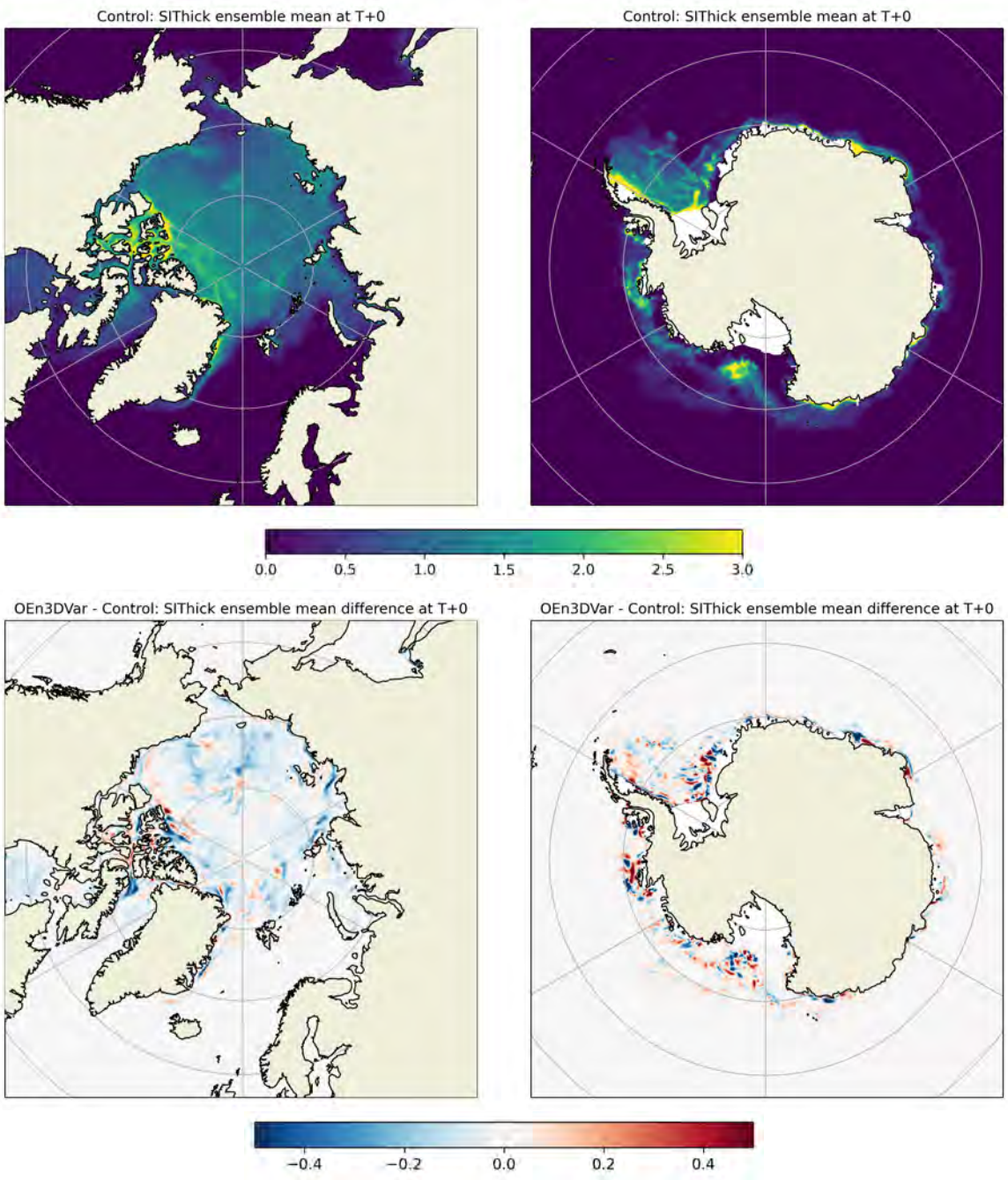


Figure 19: Sea-ice thickness ensemble mean (m) at T+0 for the Arctic (left) and Antarctic (right) averaged over forecasts initialised between 1st – 15th January 2020. Top plots: *Control*, bottom plots: *OEn3DVar* minus *Control*.

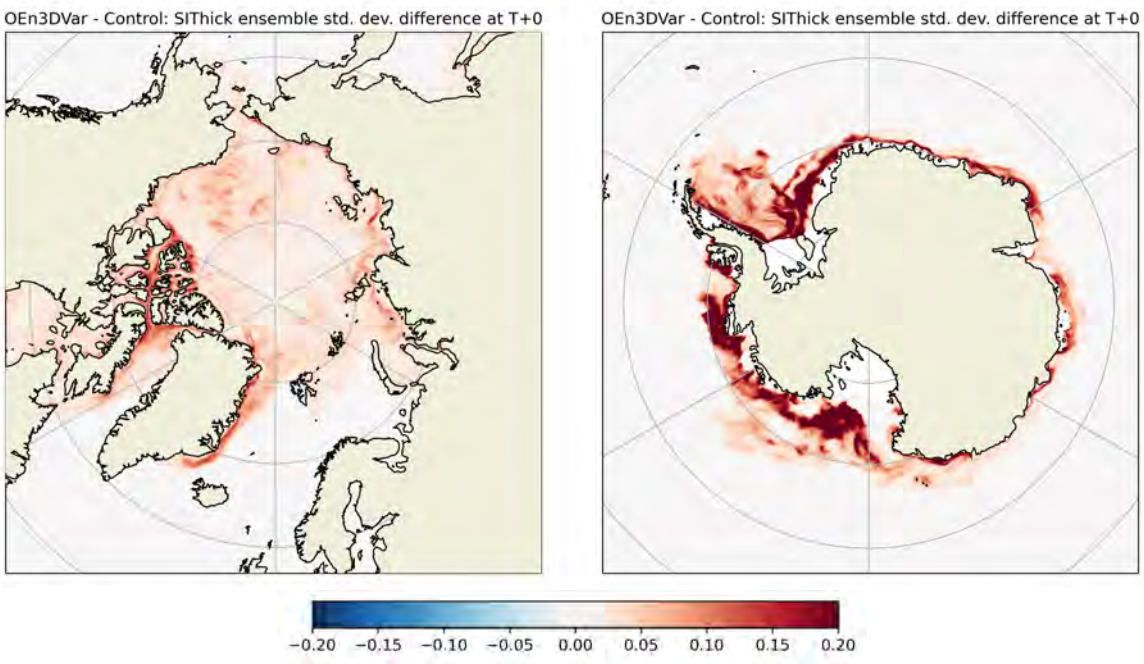


Figure 20: Change (*OEn3DVar* minus *Control*) in sea-ice thickness ensemble standard deviation (m) at T+0 for the Arctic (left) and Antarctic (right) averaged over forecasts initialised between 1st – 15th January 2020. The standard deviation of *Control* is very small.

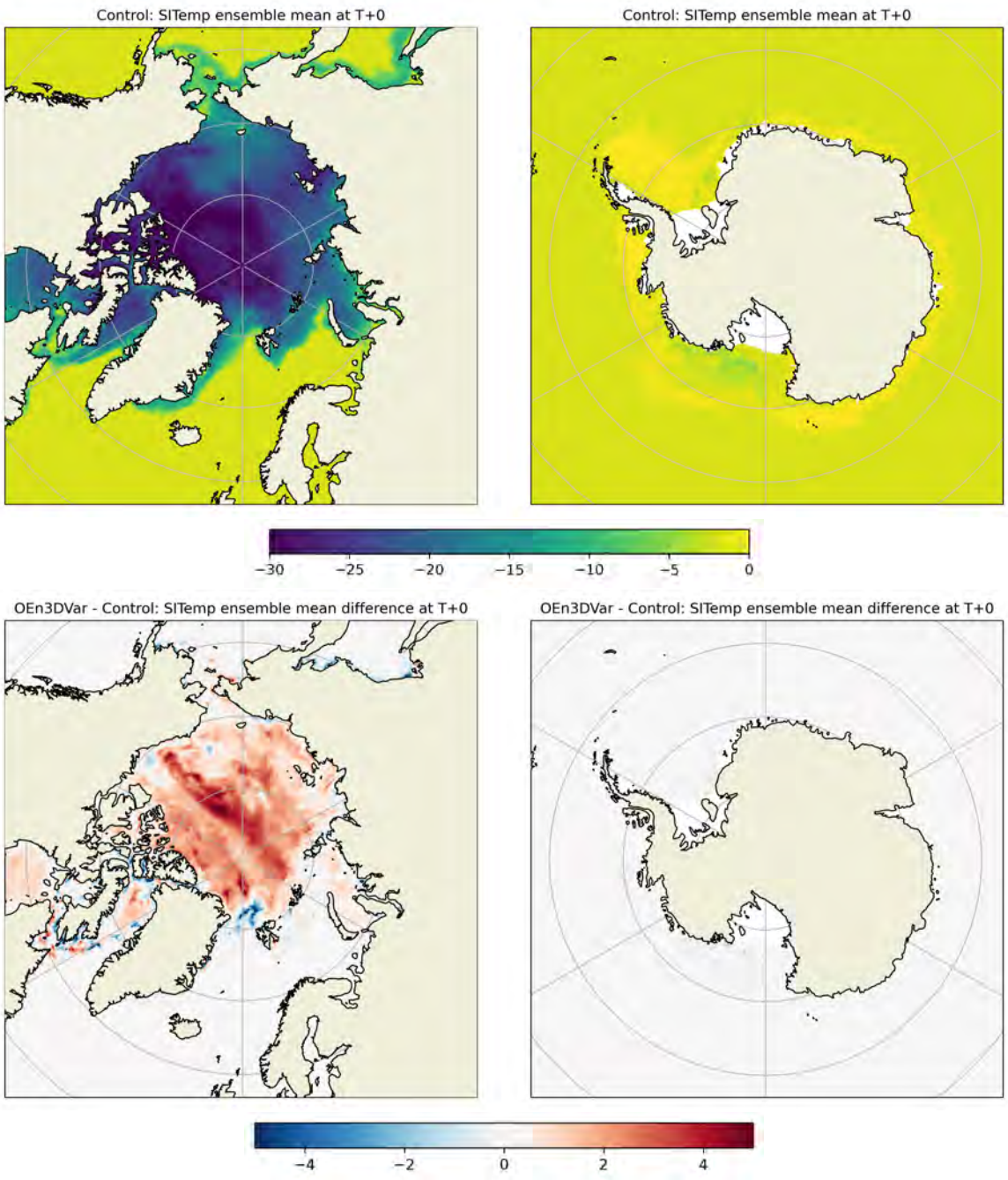


Figure 21: Sea-ice temperature ensemble mean (°C) at T+0 for the Arctic (left) and Antarctic (right) averaged over forecasts initialised between 1st – 15th January 2020. Top plots: *Control*, bottom plots: *OEn3DVar* minus *Control*.

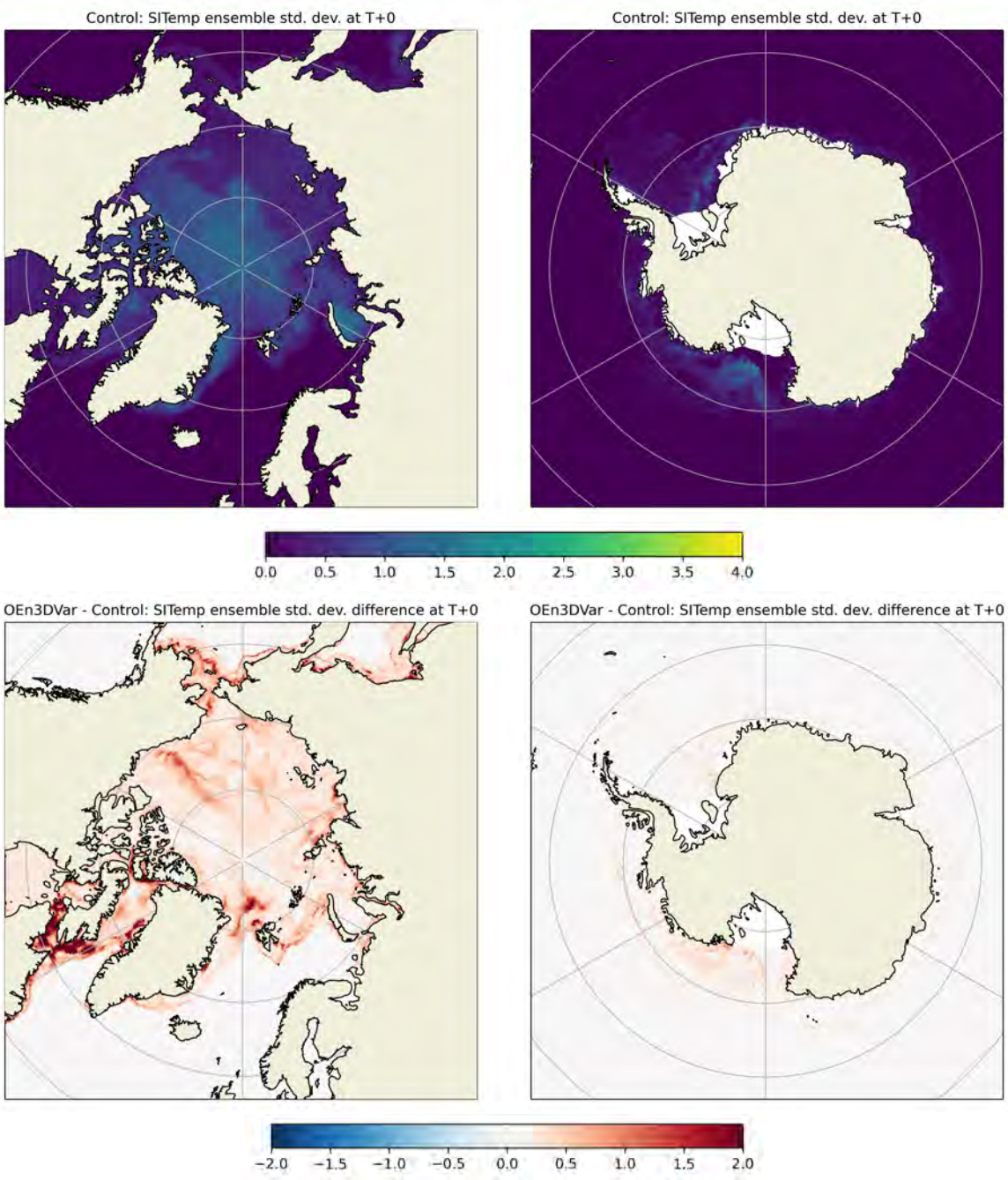
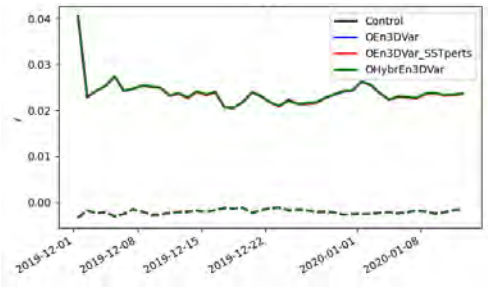
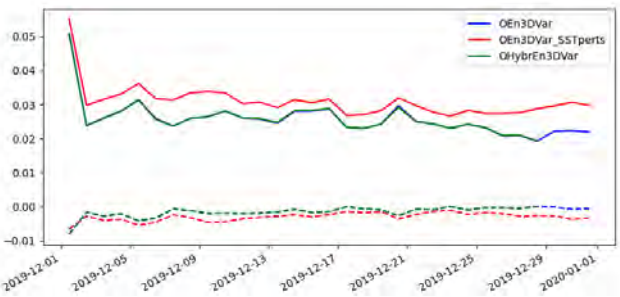


Figure 22: Sea-ice temperature ensemble standard deviation (°C) at T+0 for the Arctic (left) and Antarctic (right) averaged over forecasts initialised between 1st – 15th January 2020. Top plots: *Control*, bottom plots: *OEn3DVar* minus *Control*.

(a) Deterministic



(b) Ensemble mean



(c) Ensemble spread

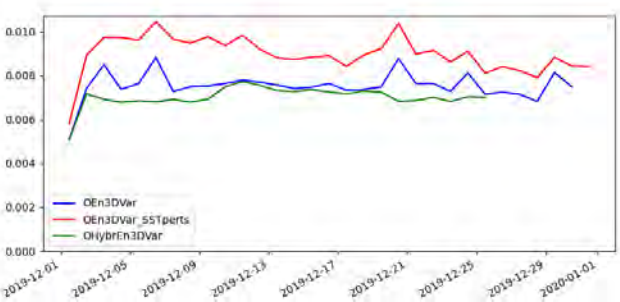


Figure 23: Time-series of the RMS (solid lines) and mean (dashed lines) in the ensemble mean 6-hour forecast errors of sea-ice concentration compared to observations.

4.1.5 Ensemble reliability in the ocean

The ensemble consistency test of Rodwell et al. (2016) is used to assess the reliability of the ensemble. For a completely reliable ensemble the ensemble spread should match the ensemble mean departure (a measure of short term forecast uncertainty) after accounting for observation errors and bias. A residual quantifies the ensemble reliability or the degree of mismatch.

First we focus on the reliability of the short-range (6-hour) forecast for the *OEn3DVar_SSTperts* experiment. The reliability statistics are shown for SLA in Fig. 24. A positive residual in the ACC, the Gulf Stream and North-West European shelf-seas indicates the ensemble spread is likely too low in those locations. Elsewhere the residual is small indicating a reliable ensemble for SLA. For SST, shown in Fig. 25, there is a positive residual in the Gulf Stream and Kuroshio in particular which is consistent with an ensemble spread which is too low. Over most of the global ocean, however, the residual is very small or slightly negative indicating a reasonable ensemble spread for SST.

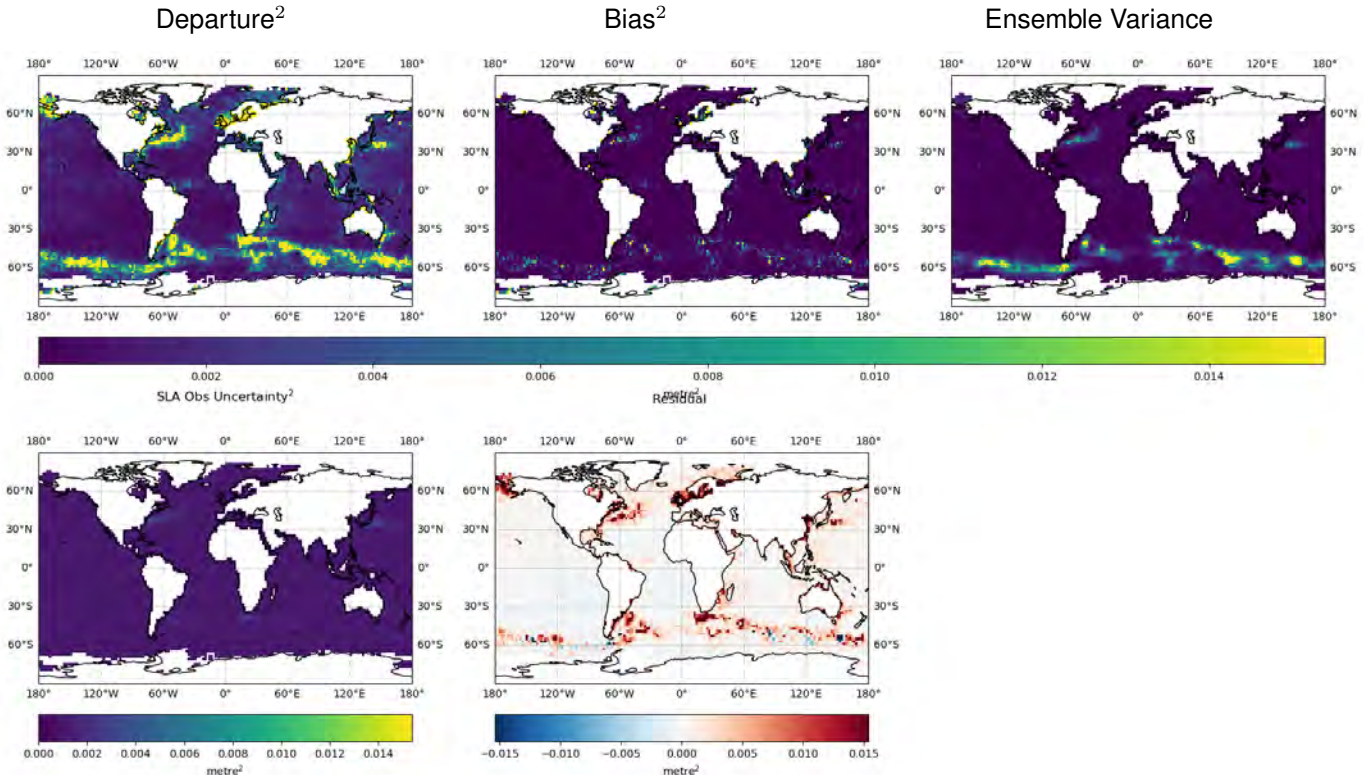


Figure 24: Reliability statistics of *OEn3DVar_SSTperts* for SLA (m^2) calculated over Dec 2019. Observation uncertainty as used in the assimilation (except for an extra near coastal amplification applied in the assimilation).

The impact of the hybrid-3DEnVar DA in experiment *OHybrEn3DVar* on the reliability is shown in Figs. 26 and 27 for SLA and SST respectively. The SLA has a lower ensemble spread and departure than *OEn3DVar_SSTperts* (Fig. 24), but a lower residual and a more reliable ensemble. The SST has a lower ensemble spread and similar departure to *OEn3DVar_SSTperts* (Fig. 25), but a higher residual and a slightly less reliable ensemble for SST with ocean hybrid DA. It may be that is possible to correct the problem here by appropriately tuning the ocean RTPS inflation scheme.

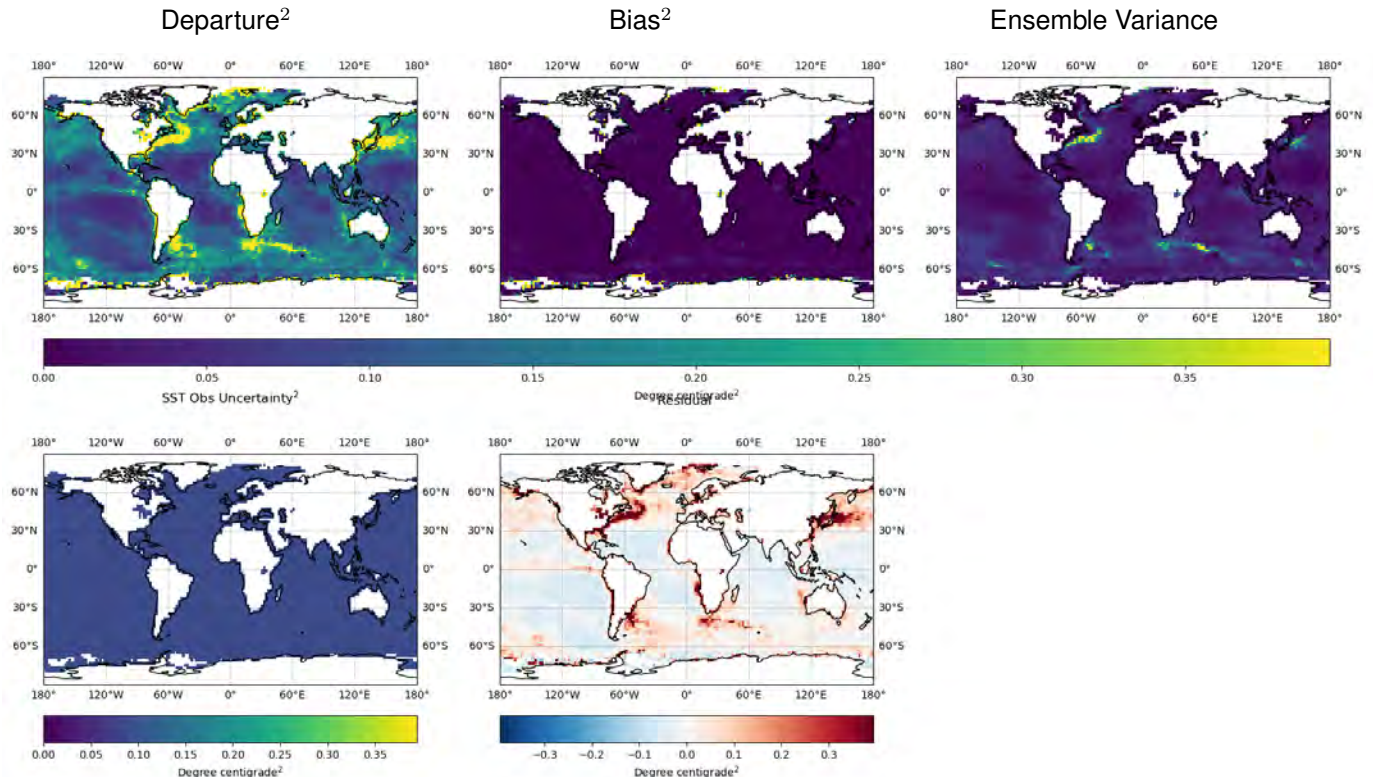


Figure 25: Reliability statistics of *OEn3DVar_SSTperts* for SST (°C ^{2}) calculated over Dec 2019. Note a uniform observation uncertainty of 0.3 °C is assumed.

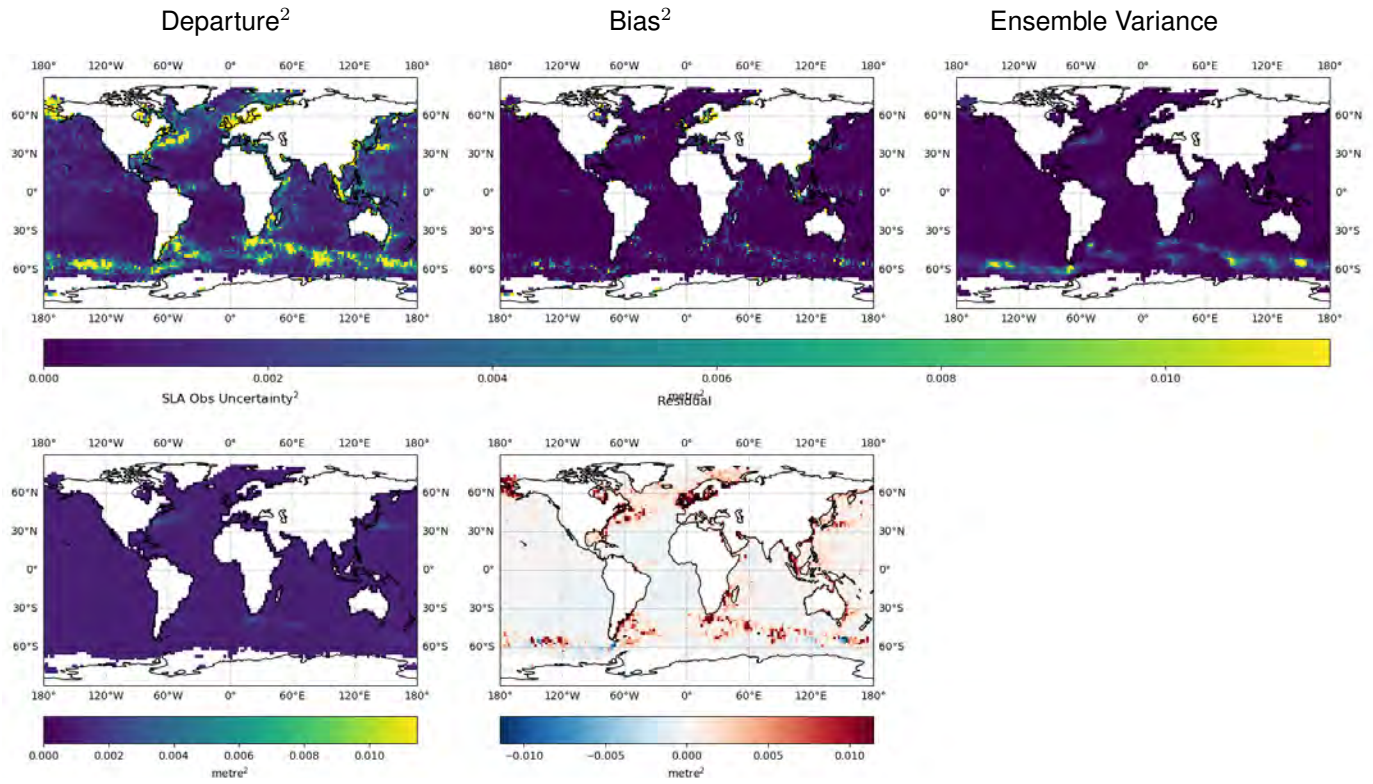


Figure 26: Reliability statistics of the hybrid DA experiment *OHybrEn3DVar* for SLA (m^2) calculated over 1-15 Dec 2019. Observation uncertainty as used in the assimilation (except no coastal amplification).

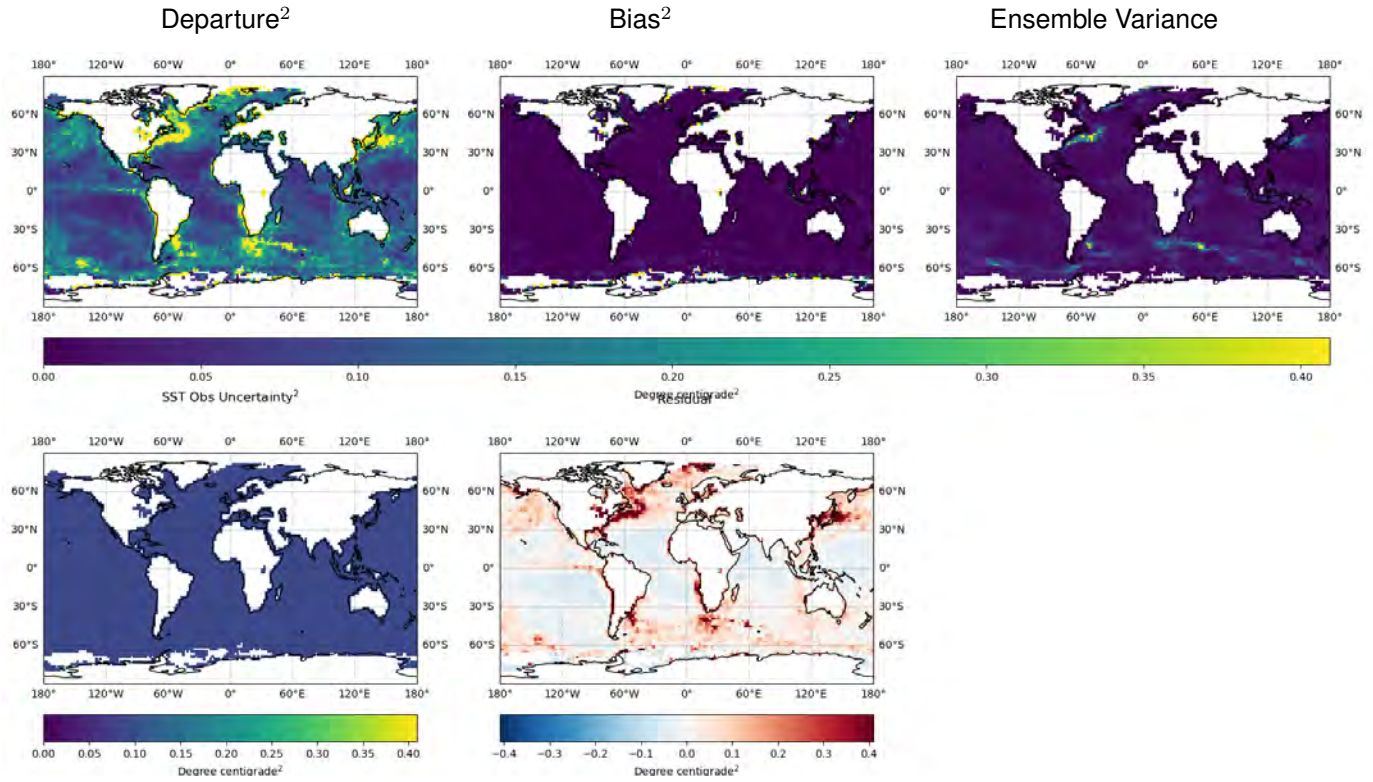


Figure 27: Reliability statistics of the hybrid DA experiment *OHybrEn3DVar* for SST (m^2) calculated over 1-15 Dec 2019. Observation uncertainty as used in the assimilation (except no coastal amplification).

4.2 Impact on the atmosphere

In this section we assess the impact of the ocean/sea-ice ensemble changes on the atmospheric component of the coupled system. Statistical verification of the results is presented first to understand where the changes are affecting the atmosphere and whether they are beneficial or detrimental. This verification is calculated over the two months of the experiments, 1st Dec 2019 to 31st Jan 2020. Deterministic and ensemble aspects of the system are verified using both observations and ECMWF analyses as references. Then we show the impact in terms of the changes in ensemble mean and ensemble spread for near-surface variables averaged over the period 1st - 14th Jan 2020.

4.2.1 Atmospheric deterministic forecast verification

The only effect of the ocean/sea-ice ensemble changes on the deterministic atmospheric forecast will come indirectly through the changes to the atmospheric ensemble fields which are used in the deterministic atmospheric DA through the hybrid forecast-error covariances. This is confirmed by Fig. 28 which shows a summary of the impacts for both the *OEn3DVar* and *OEn3DVar_SSTperts* experiments compared to the deterministic forecasts from the *Control* experiment. Very few of the scores have statistically significant results (indicated by the filled squares). The *OEn3DVar* experiment has marginally worse RMSE for some variables, most notably in the Northern Hemisphere when compared with ECMWF analyses, though overall there are only small changes. The *OEn3DVar_SSTperts* experiment has even smaller changes compared to the *Control*, the main exception being the tropical temperatures at 250 hPa compared to ECMWF analyses, though even there the changes are small.

4.2.2 Atmospheric ensemble forecast verification

We now focus on the impact on the ensemble atmospheric verification. Fig. 29 shows a summary of the changes to the CRPS for different atmospheric variables at different lead times. The CRPS gives a measure of the degree to which the probability density function of the observed forecast errors is consistent with the ensemble predicted forecast errors with higher CRPS values indicating a bigger mismatch between the ensemble and observations. There is a clear degradation to the CRPS, particularly for the 2 m temperatures in the Southern Hemisphere, Northern Hemisphere and tropics, when the ocean ensemble is used without any additional perturbations to the SSTs seen by the atmosphere (experiment *OEn3DVar*). This degradation is seen when the reference used in the verification is the observations and also when the reference is the ECMWF analyses. The degradations to the tropical 2 m temperatures are more obvious against the ECMWF analyses probably because this is a global gridded field including over the oceans, whereas the observation sampling is smaller over the oceans. Fig. 30 shows that this degradation in CRPS is likely due to a reduction in the near-surface temperature ensemble spread, with the RMSE of the ensemble mean unaffected. The reduction in temperature spread extends throughout the height of the atmosphere when looking at the 5-day forecast, also shown in Fig. 30.

When additional SST perturbations are included in experiment *OEn3DVar_SSTperts* the overall impact on the atmospheric CRPS is much smaller when compared with observations, though there is a small degradation to the 500 and 850 hPa geopotential height in the Northern Hemisphere as well as the 850 hPa geopotential height in the tropics. There are also some improvements compared to *Control* in the CRPS in the Southern Hemisphere for pressure at mean sea level (PMSL) and other variables. These patterns of changes follow through to the comparison with ECMWF analyses, with the exception being the larger degradation to the CRPS in tropical 2 m temperatures.

To learn more about the degradation of the tropical 2 m temperature in the *OEn3DVar_SSTperts* experiment, Fig. 31 shows how the CRPS evolves as a function of forecast lead time when comparing with observations and ECMWF analyses. The CRPS of the *Control* and *OEn3DVar_SSTperts* experiments start at the same value at the beginning of the forecast while the CRPS in experiment *OEn3DVar* is

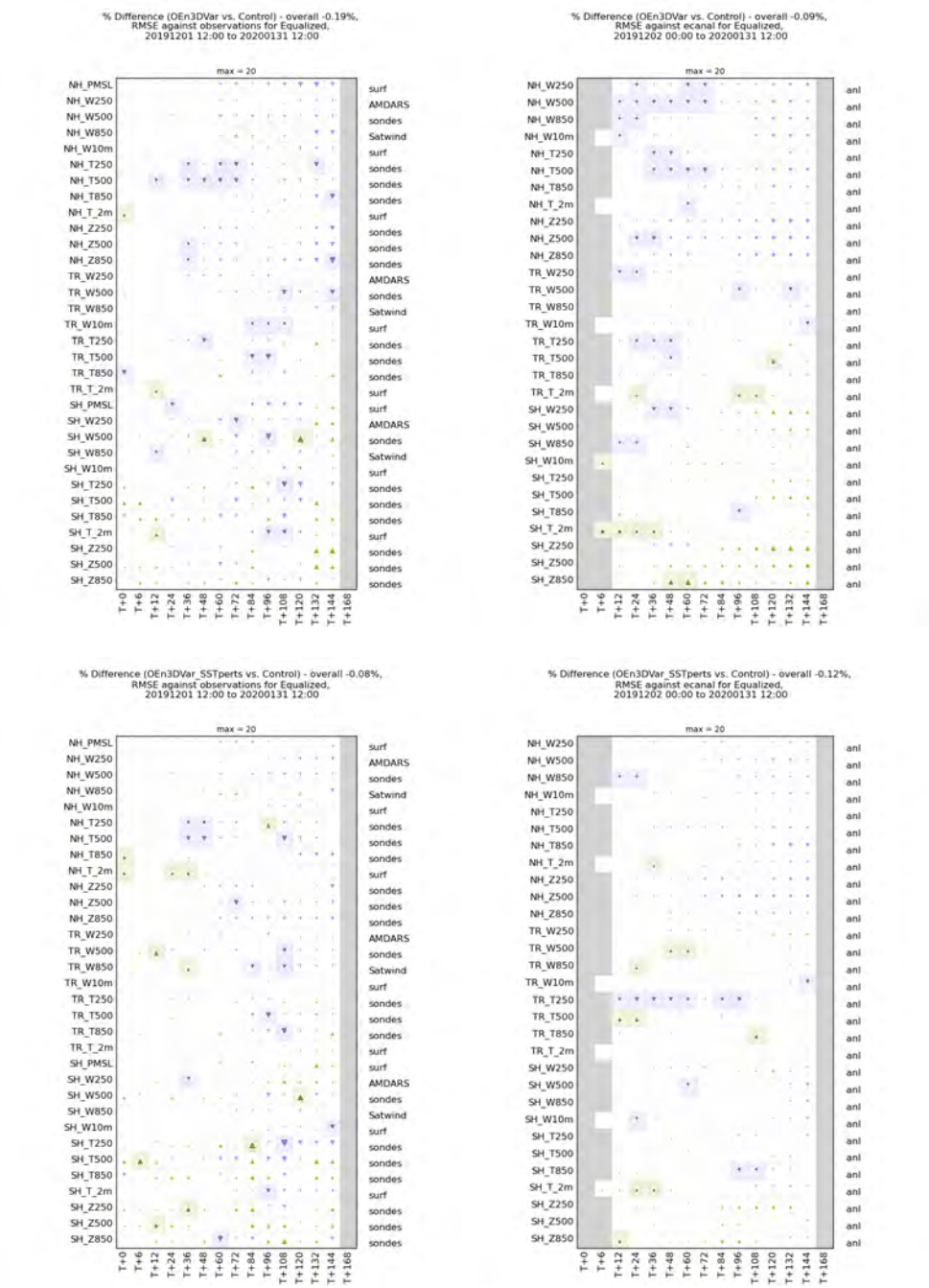


Figure 28: Deterministic scorecards showing the impact on RMSE for different variables, regions and forecast lead times. The left shows the impact on RMSE compared with observations while the right shows the impact compared with ECMWF analysis. The top plots show the change in scores of experiment *OEn3DVar* compared to *Control* while the bottom plots show the change in scores of experiment *OEn3DVar_SSTperts* compared to *Control*. An upward pointing green triangle indicates an improvement over the *Control* while a downward pointing purple triangle indicates a degradation. Boxes which are shaded have statistically significant changes.

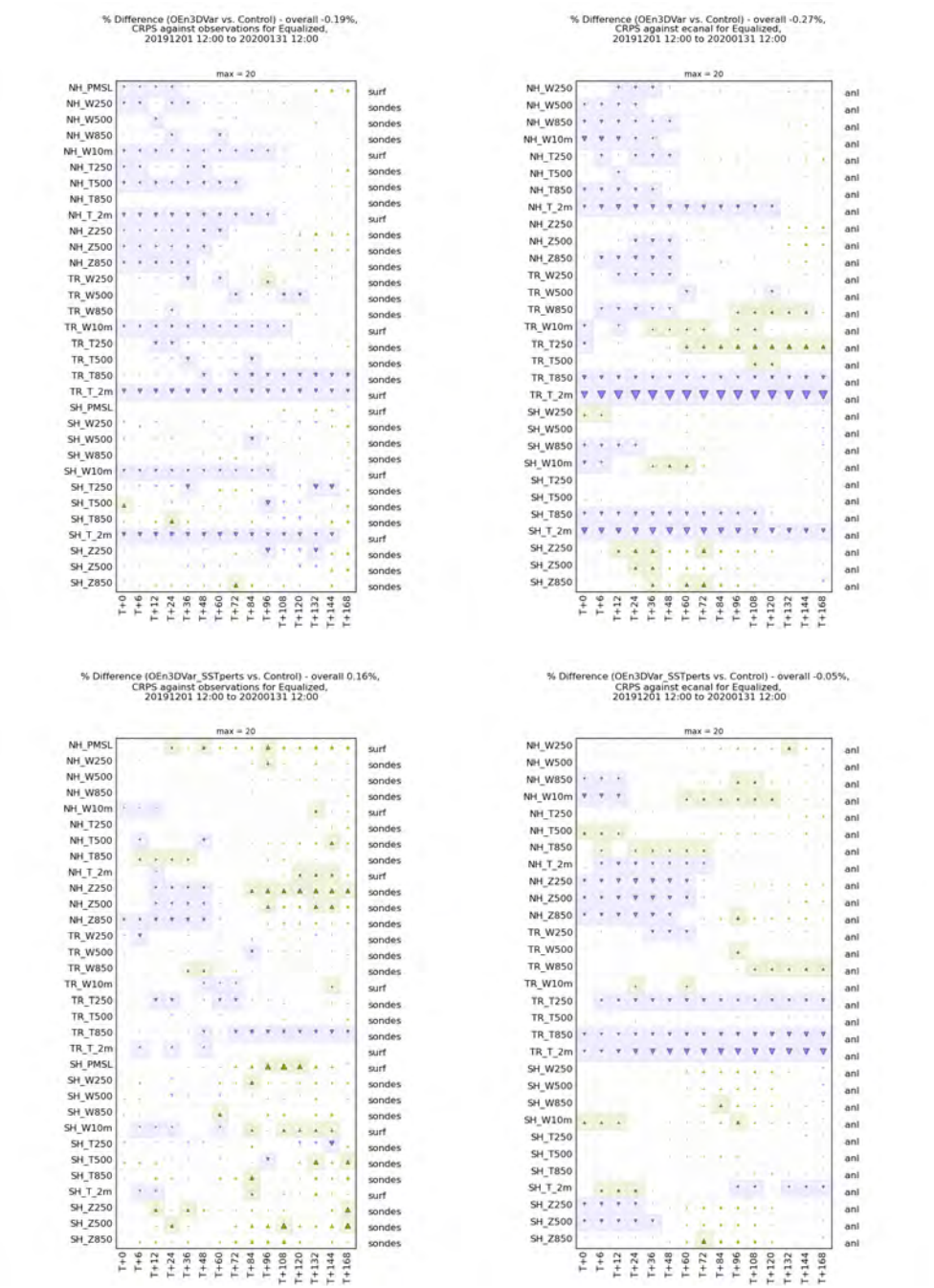


Figure 29: Ensemble scorecards showing the impact on CRPS for different variables, regions and lead times. The left shows the impact on CRPS compared with observations while the right shows the impact compared with ECMWF analysis. The top plots show the change in scores of experiment *OEn3DVar* compared to *Control* while the bottom plots show the change in scores of experiment *OEn3DVar_SSTperts* compared to *Control*. The symbols mean the same as in Fig. 28 but for CRPS rather than RMSE

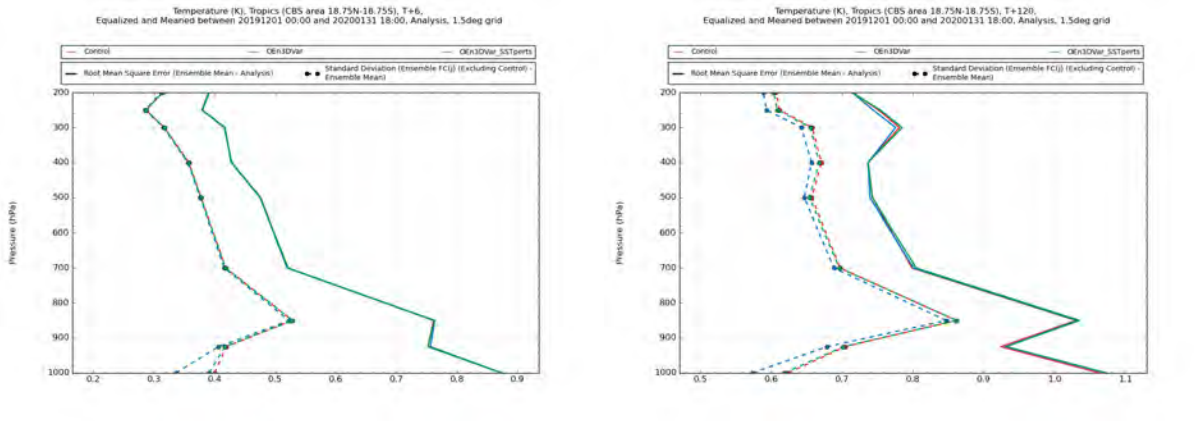


Figure 30: Ensemble spread (dotted) and RMSE of the ensemble mean (solid) for temperature profile compared with ECMWF analysis for the 6 hour forecast (left) and 5 day forecast (right) in the tropics.

larger at all lead times. The *OEn3DVar_SSTperts* CRPS increases more quickly than that of the *Control* experiment through the forecast. The RMSE and ensemble standard deviation are also shown on Fig. 31 which shows that the RMSE of the ensemble mean is similar in all of the experiments. The ensemble spread is smaller in the *OEn3DVar* experiment than the others (supporting what was shown in Fig. 30) but the spread in *OEn3DVar_SSTperts* is similar to that of the *Control* experiment, so the large change in CRPS is not easily explained. Fig. 32 shows that there is a small change in the mean error for the ensemble mean when compared to ECMWF analyses with the *OEn3DVar_SSTperts* experiment having a larger positive bias than the *Control* experiment, so perhaps this contributes to the increased CRPS.

The other aspect of the summary scorecards which degraded was in the Northern Hemisphere geopotential height. Fig. 33 provides more detail on this showing the ensemble spread and RMSE of the ensemble mean of the geopotential height in the Northern Hemisphere as a function of height at 24 hour lead-time. Against the observations, the *OEn3DVar_SSTperts* experiment has a larger RMSE compared to the *Control* experiment with a very similar spread. In the next section we look at the geographical differences in selected atmosphere fields to try understand more about this and other results from the assessment of the CRPS scorecards here.

4.2.3 Impact on atmospheric ensemble mean and spread

The time-averaged ensemble mean surface temperatures seen by the atmospheric model (including over the ocean, sea-ice and land) are different over large regions in the *OEn3DVar* and *OEn3DVar_SSTperts* experiments, as shown in Fig. 34. Both experiments have a very large positive difference compared to the *Control* experiment in the mean surface temperature in the Arctic, following as expected from the sea-ice results presented earlier. There is also a large negative difference in the Southern Ocean in both experiments. There is a dipolar structure to the changes in the North Atlantic Current south-east of Newfoundland.

The ensemble standard deviation for the surface temperature (as shown in Fig. 35) is significantly lower in the *OEn3DVar* experiment than the *Control*, except for over the Arctic where there is an increase in spread. At a lead-time of 5-days the spread in this experiment has increased somewhat, but is still much lower overall than the corresponding spread in the *Control*. At 5-days, there are some regions where the spread is larger than in the *Control* such as in the Gulf Stream, Kuroshio and ACC. However, the spread is still significantly lower, particularly in the northern extra-tropics and the upwelling regions (e.g. to the west of the Sahara Desert and to the west of California). The spread in the Arctic region at 5-day lead time is actually smaller than in the *Control*, in contrast to the change there at the start of the forecast. The surface temperature spread in the *OEn3DVar_SSTperts* experiment is much more similar to the

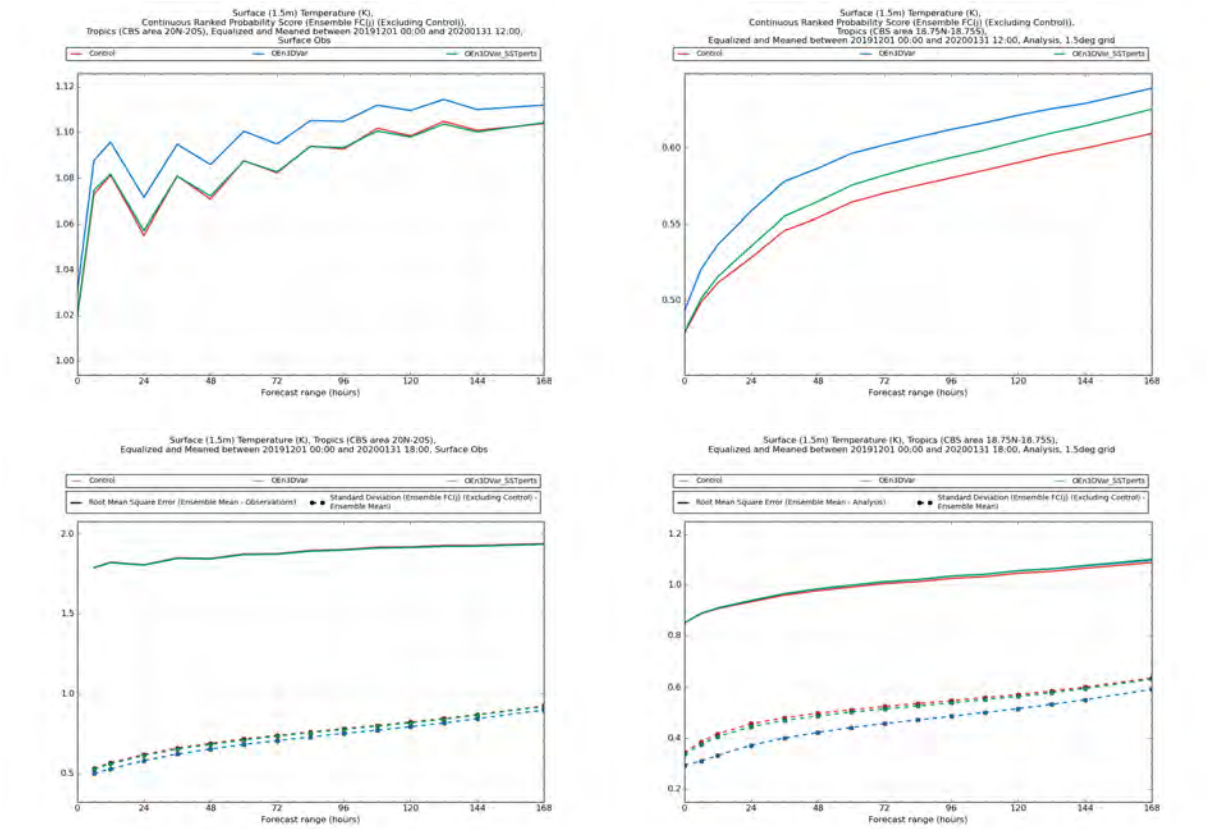


Figure 31: CRPS (top) and RMSE (bottom) in the tropical 2 m temperature (K) as a function of lead time calculated using observations (left) and ECMWF analyses (right).

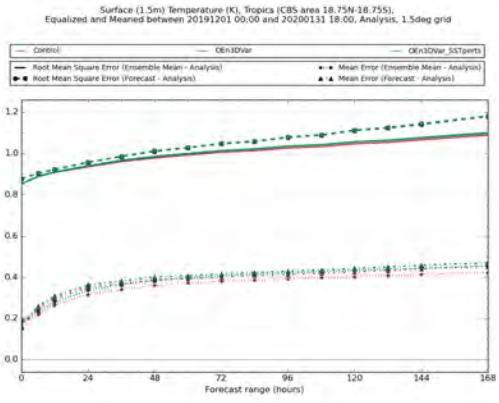


Figure 32: RMSE and mean error for the tropical 2 m temperature (K) ensemble mean and control member compared to ECMWF analyses as a function of lead time.

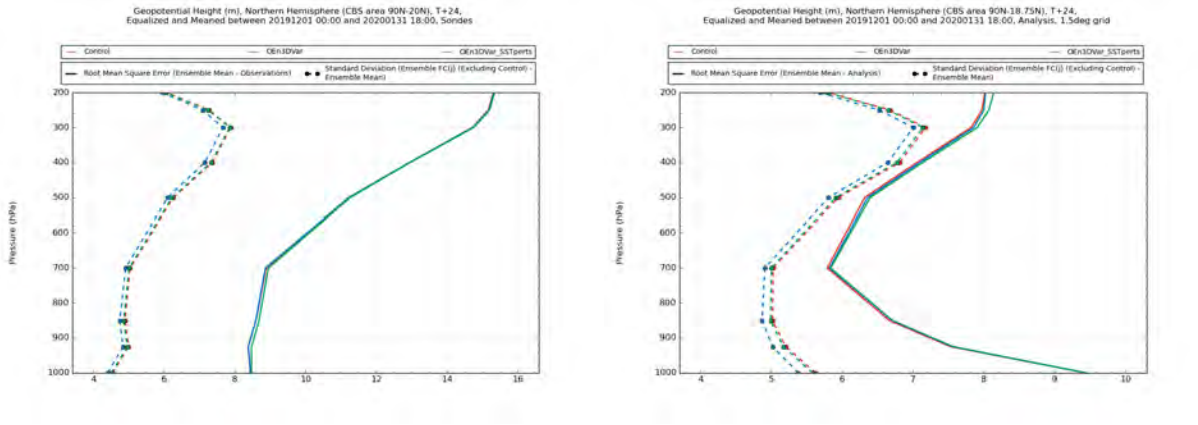


Figure 33: Ensemble spread (dashed) and RMSE of the ensemble mean (solid) of the geopotential height (m) compared to observations (left) and ECMWF analysis (right) in the Northern Hemisphere at 24 hour forecast lead-time.

Control, which is expected since we set the magnitude of the additional SST perturbations to achieve this. The pattern of differences in the Arctic region is similar to the pattern seen in the *OEn3DVar* experiment with an increase in spread at the start of the forecast and a reduction after 5-days in that region. At 5-days lead time the increase in spread in the Gulf Stream, Kuroshio and ACC regions is more obvious and is due to the more chaotic nature of those regions being excited by the ocean/sea-ice perturbations.

Figs. 36 and 37 are equivalent plots to the surface temperature plots above, but for the 1.5 m air temperature. The general results for the ensemble mean 1.5 m air temperature are similar to those seen with surface temperature. However, the magnitude of changes in the ensemble spread in 1.5 m air temperature is much lower than the change in spread seen in the surface temperature over most of the ocean regions, except for the Arctic. This low sensitivity of the spread in near-surface atmospheric temperature to the ensemble spread in the surface ocean could explain why such large perturbations to SST are needed to generate spread in the atmosphere. The spread in 1.5 m air temperature in the *OEn3DVar* experiment is actually quite similar to the *Control* experiment over many regions, with the upwelling regions and equatorial Pacific standing out as having reduced spread. It would be preferable to be able to dispense with the need for additional SST perturbations in the atmosphere if possible. Perhaps targeting increased spread in the ocean ensemble to these regions might be sufficient to improve the overall performance of the coupled ensemble rather than adding the extra SST perturbations based on OSTIA variability as was done in the *OEn3DVar_SSTperts* experiment.

The impact of the ocean/sea-ice changes on the ensemble mean atmospheric pressure at mean sea level is shown in Fig. 38. At T+0 lead time there is almost no change compared to the *Control* for both the *OEn3DVar* and *OEn3DVar_SSTperts* experiments since the PMSL is strongly constrained by the atmospheric DA. At T+120 lead time there are larger changes though still only up to a magnitude of about 80 Pa. Both experiments show a reduced PMSL over the Arctic. They also both show a dipolar structure in the PMSL changes with a positive change over the Greenland Sea (and Greenland itself in experiment *OEn3DVar*) and a negative change over Iceland. There are many changes in the Southern Ocean region though not anything very systematic, except a small reduction in PMSL over Antarctica in experiment *OEn3DVar_SSTperts*. The spread in PMSL is shown in Fig. 39 where there is almost no change in spread from the ocean/sea-ice changes at T+0 as might be expected. At T+120 there is a large reduction in PMSL spread in the *OEn3DVar* experiment in the Northern Hemisphere. The experiment *OEn3DVar_SSTperts* has less of a reduction in spread overall, though there is quite a large-scale reduction in the Greenland and Barents Seas. As with previous results the high sensitivity of the results seen at high latitudes motivates specific attention being paid to improving the sea-ice ensemble in future work.

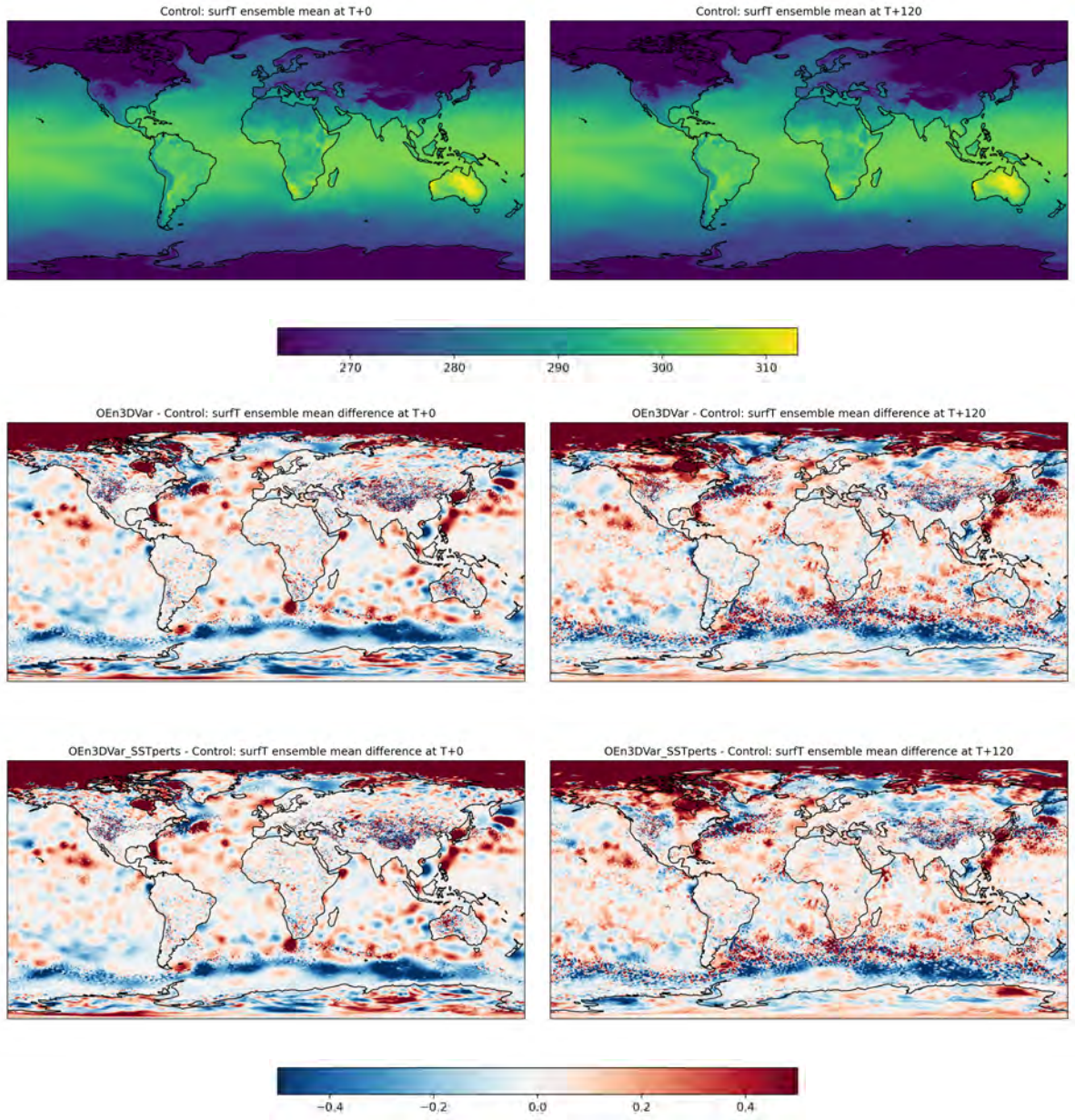


Figure 34: Surface temperature ensemble mean (K) at T+0 (left) and T+120 (right) averaged over forecasts initialised between 1st – 15th January 2020. Top plots: *Control*, middle plots: *OEn3DVar* minus *Control*, bottom plots: *OEn3DVar_SSTperts* minus *Control*.

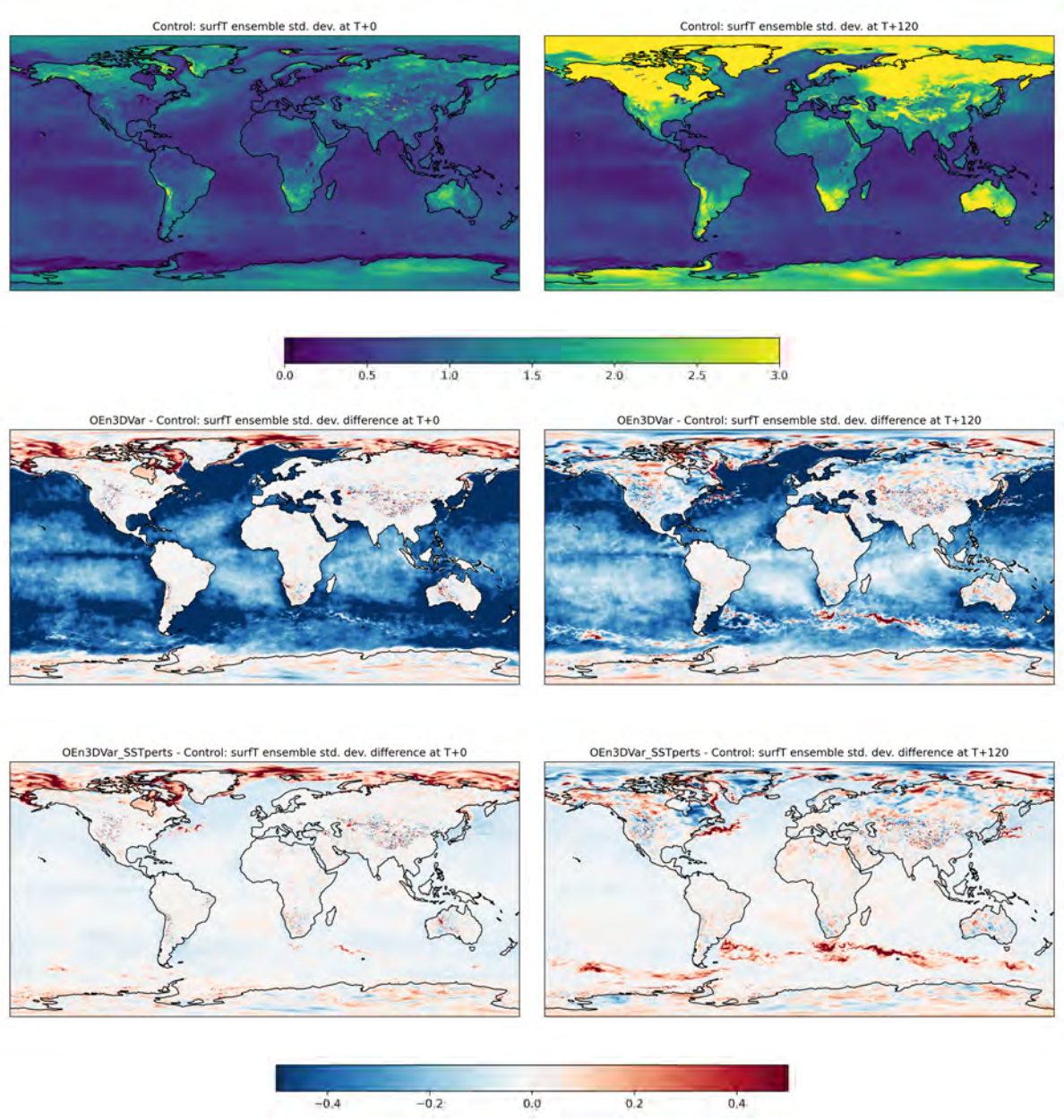


Figure 35: As for Fig. 34 but for surface temperature ensemble standard deviation (K).

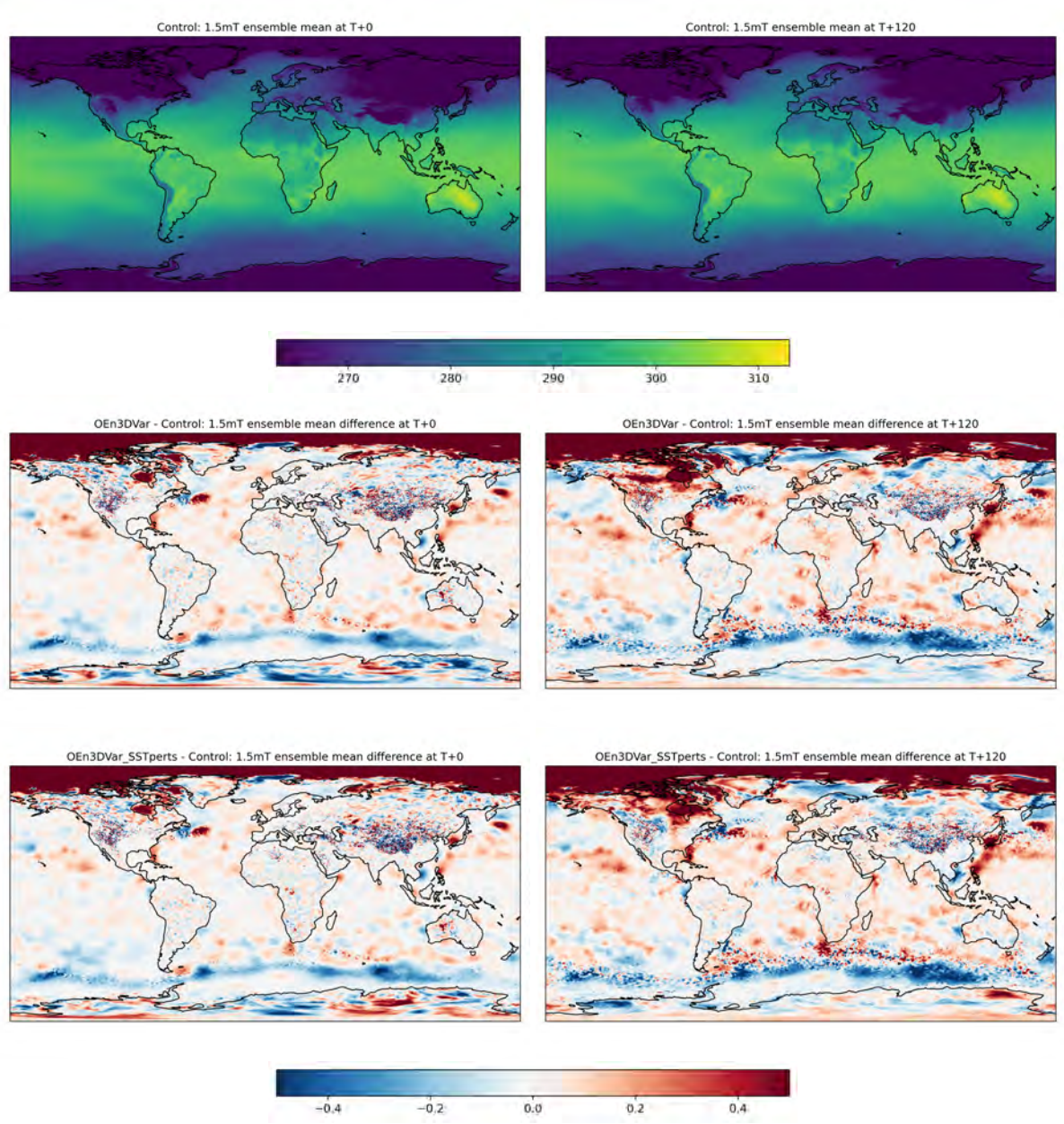


Figure 36: As for Fig. 34 but for 1.5 m air temperature ensemble mean (K).

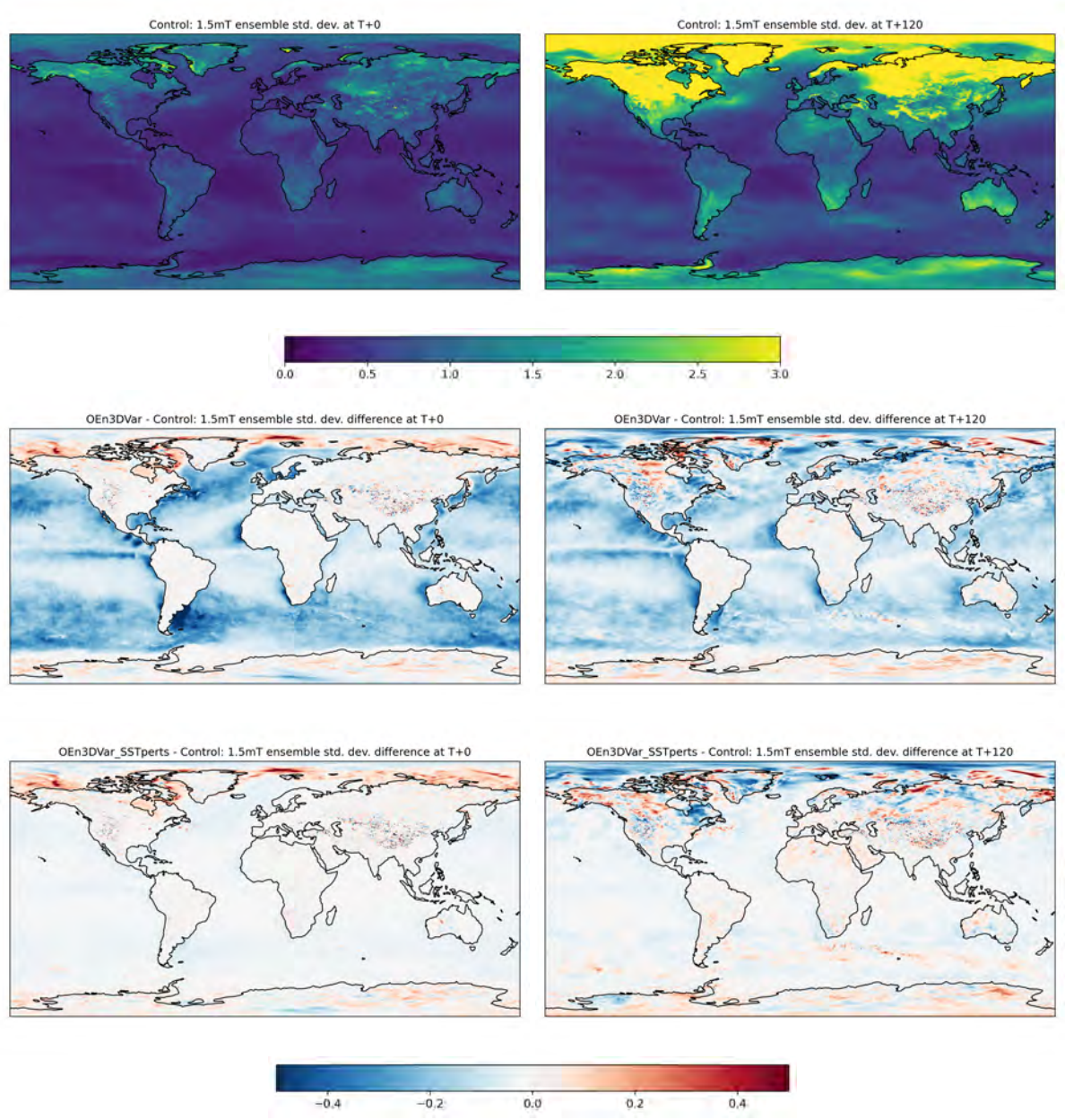


Figure 37: As for Fig. 34 but for 1.5 m air temperature ensemble standard deviation (K).

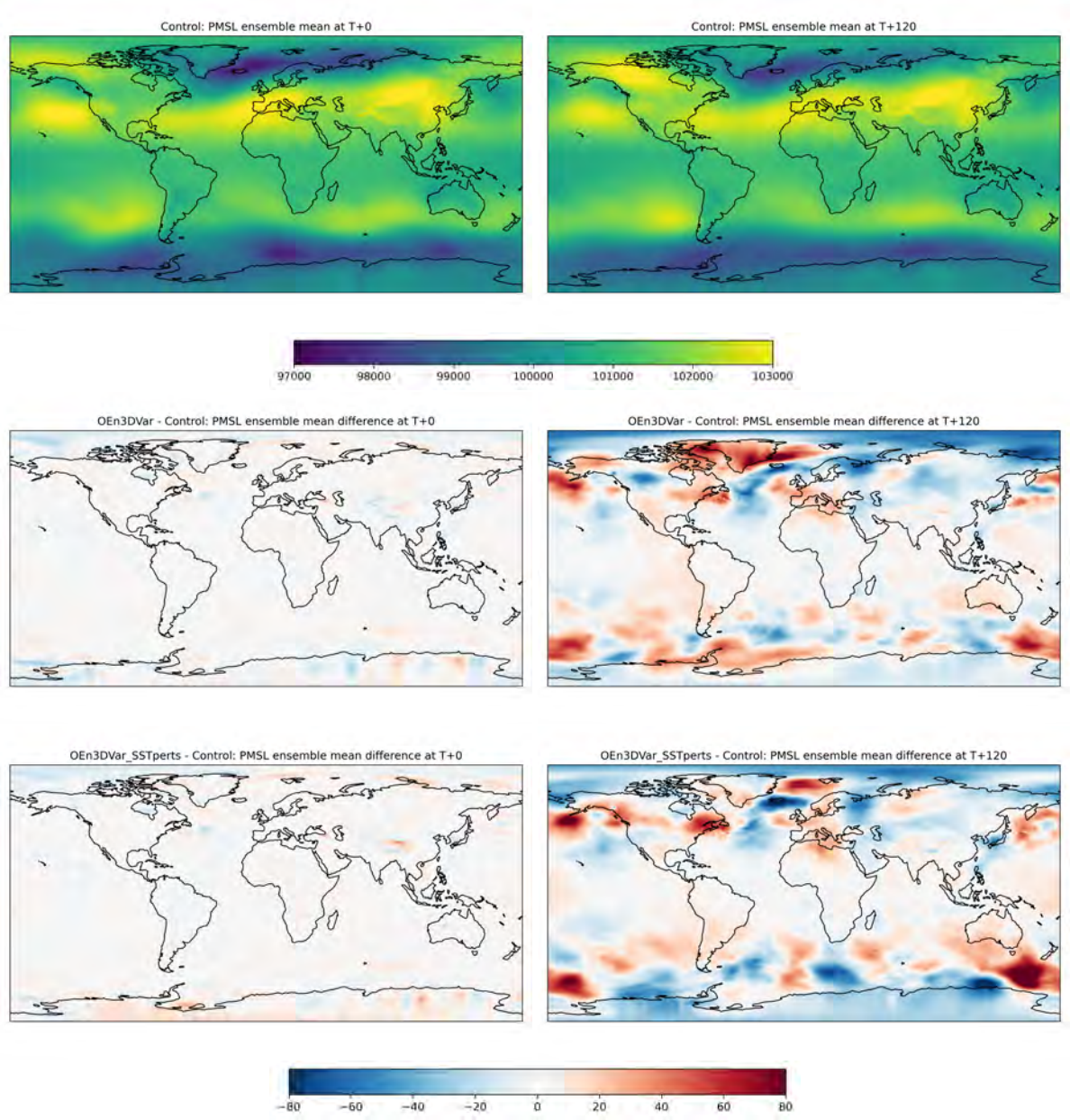


Figure 38: As for Fig. 34 but for pressure at mean sea level ensemble mean (Pa).

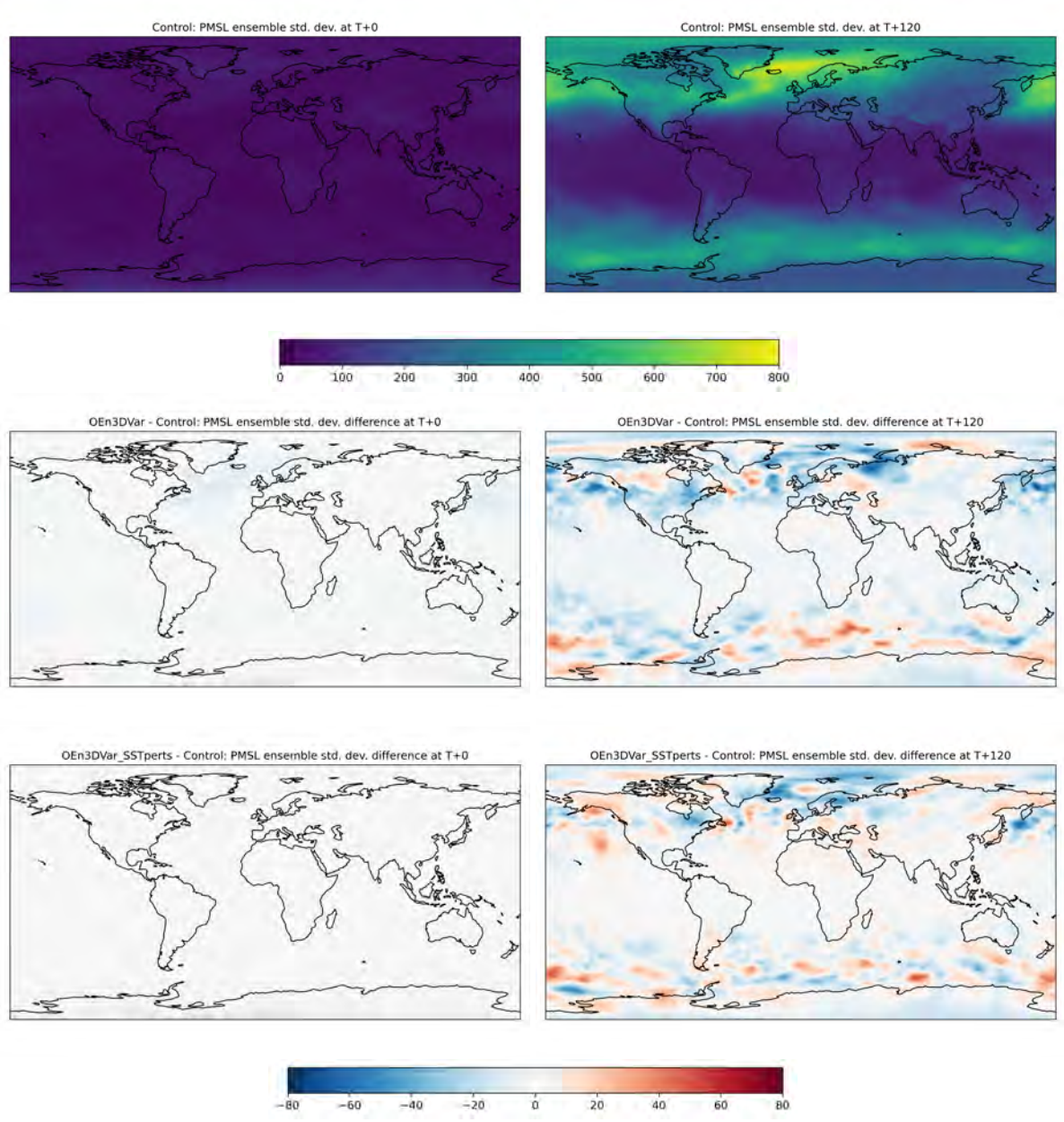


Figure 39: As for Fig. 34 but for pressure at mean sea level ensemble standard deviation (Pa).

One interesting result is that for most of the short range (6 hour) forecast statistics it appears that addition of SST perturbations to the SST that the atmosphere ensemble sees makes little difference to the ensemble spread of SSTs in the ocean. This most likely because the SST is heavily constrained by the assimilation due to the dense sampling of SST data. This interpretation is supported by looking at the difference in the ensemble spread of *OEn3DVar_SSTperts* compared to *OEn3DVar* for the short range and longer range forecasts (see Fig. 40). This shows a marked increase in SST spread when SST perturbations are applied on day 5 of the forecast, but little difference on day 0. The spread is particularly increased in the tropics and Southern Hemisphere (the summer hemisphere in January).

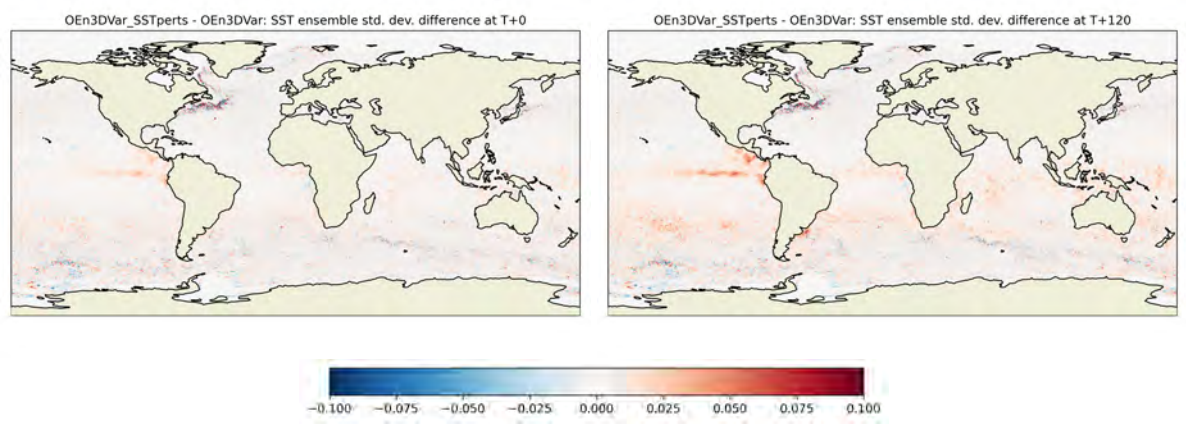


Figure 40: SST ensemble mean spread difference plots (°C) at T+0 (left) and T+120 (right) averaged over forecasts initialised between 1st – 15th January 2020 for *OEn3DVar_SSTperts* minus *OEn3DVar*.

4.3 Impact of ocean hybrid DA on the ensemble

One of the advantages of ensemble systems is that the ensemble gives uncertainty information about the forecasts. This information is also useful for data assimilation as it can be used to give an estimate of the background error covariance. A method to use this information in the data assimilation is hybrid ensemble variational assimilation (Hybrid EnVar).

A hybrid $\mathbf{B}$ matrix is a linear combination of the localised sample ensemble covariance matrix with the modelled covariance matrix:

$$\mathbf{B}_{\mathrm{h}} = (1 - \beta_{\mathrm{e}}^2)\mathbf{B}_{\mathrm{m}} + \beta_{\mathrm{e}}^2\mathbf{B}_{\mathrm{e}} \tag{1}$$

where β_{e}^2 is a weight which controls how much the modelled covariance $\mathbf{B}_{\mathrm{m}}$, used in standard 3DVar NEMOVAR, and the ensemble based covariance $\mathbf{B}_{\mathrm{e}}$, contribute to the overall covariance $\mathbf{B}_{\mathrm{h}}$. Note the overall weight for both components always adds up to 1. We call the case where $\beta_{\mathrm{e}} = 0$ standard 3DVar, the case where $\beta_{\mathrm{e}} = 1$ 3DEnVar and the case where $0 < \beta_{\mathrm{e}} < 1$ hybrid-3DEnVar, following Lorenc & Jardak (2018).

In section 4.1 we showed that the hybrid-3DEnVar experiment has significant improvements in ensemble reliability and the fit to observations when compared to the other experiments . We explore a little why this might be and discuss more about the impact of hybrid DA in this section.

We introduce a new experiment *OHybrEn3DVar08* where the ensemble component is given a weight of 0.8 compared to *OHybrEn3DVar* where the weight is 0.5. To aid distinguishing the hybrid experiments we denote the experiment named *OHybrEn3DVar* elsewhere in this report, as *OHybrEn3DVar05* in this section.

The SST increments for the first cycle on 1st Dec 2019 at 06Z are shown in Fig. 41 comparing the 3DVar ensemble results and the two hybrid experiments. The main thing to note is the increments are rather similar with a slight reduction in the increments away from the Gulf Stream front for the hybrid experiments along with smaller scale increments near the front.

South to north vertical sections across the Gulf Stream are plotted for the experiments in Fig. 42. A striking feature is the strong increments which extend to below 800 m along the Gulf Stream front which are not seen in the 3DVar case.

SST increments and temperature increments at 200 m depth in the Indian ocean on the first cycle are shown in Fig. 43. For both the hybrid DA experiments there is an area of strong small scale increments in the western Indian Ocean centred at around 5–10°N and 50–60°E. The increments are particularly strong in *OHybrEn3DVar08*.

After 15 days of the trials running we see some spurious small scale features in the surface temperature in the western Indian Ocean in *OHybrEn3DVar08*, as shown in Fig. 44. And a few days after this the model blew up. It is unclear why this problem was not seen previously in the ocean only experiments of Lea et al. (2022) but the change to 6 hourly cycling and the consequent reduction in data assimilated on each cycle is a possible reason we did not see this issue previously. The other option is that we are triggering a coupled instability, but given that we already see the problem starting on the first cycle using an ensemble initialised from previous ocean only runs this seems unlikely.

The idea of doing hybrid DA and not wholly ensemble based DA is that the modelled covariance stabilises the data assimilation results. If the ensemble is “noisier” with shorter time windows due to fewer observations being in the assimilation time-window there is an argument for increasing the weight given to the modelled part of the covariance and reducing the weight given to ensemble part of the covariance as we have done here.

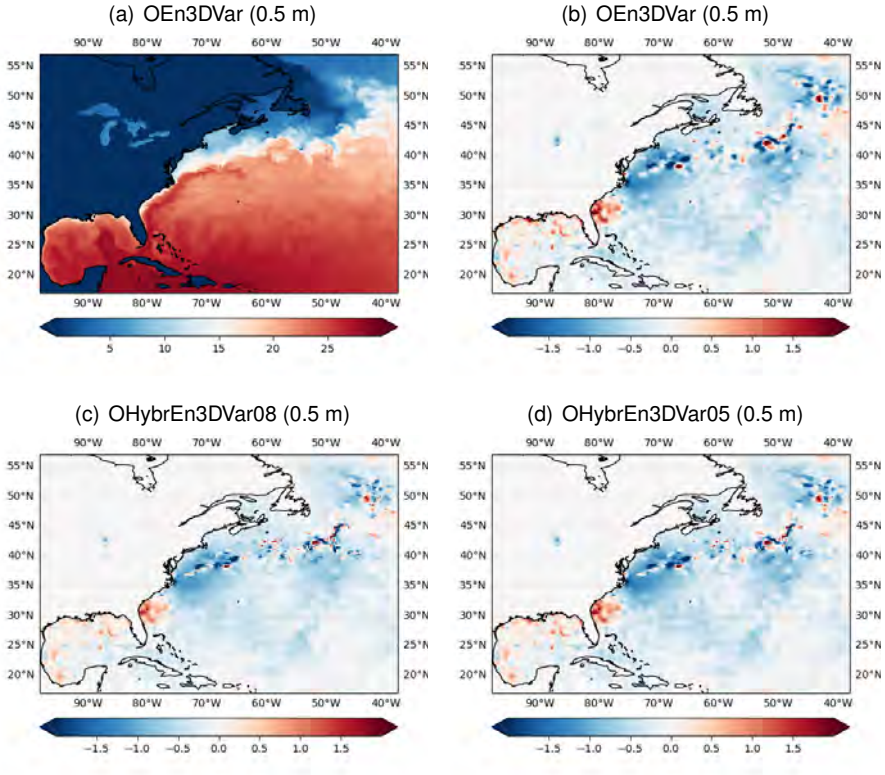


Figure 41: Background SST (in °C) and assimilation SST increments (°C), on the first cycle, comparing hybrid-3DEnVar experiments sand 3DVar for the unperturbed ensemble member.

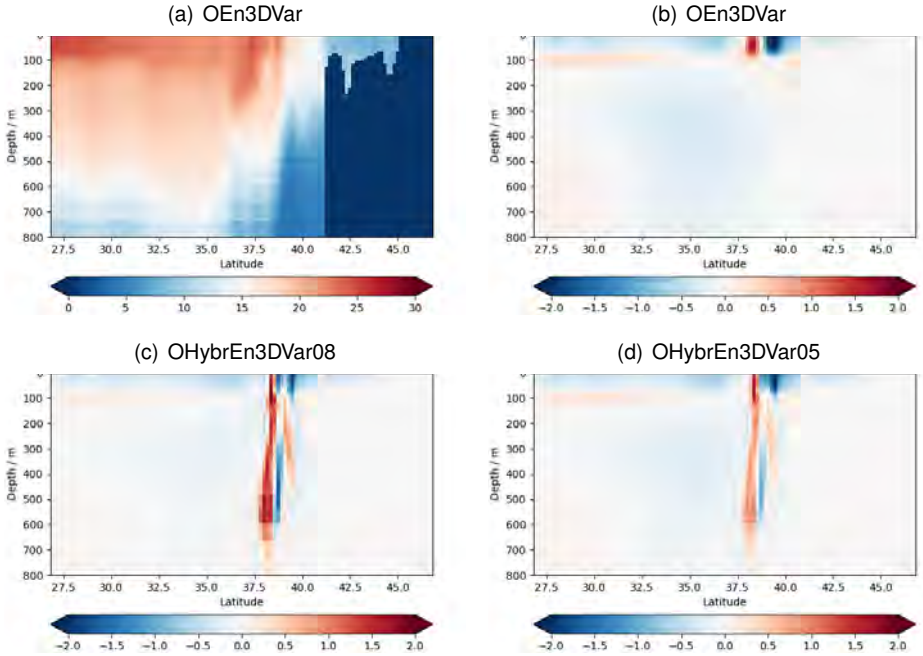


Figure 42: Vertical sections, on the first cycle, of background temperature from north to south across the Gulf Stream at 66°W (in °C) and vertical sections of ocean temperature increments (°C), at the same location and time, comparing hybrid-3DEnVar experiments and 3DVar for the unperturbed ensemble member.

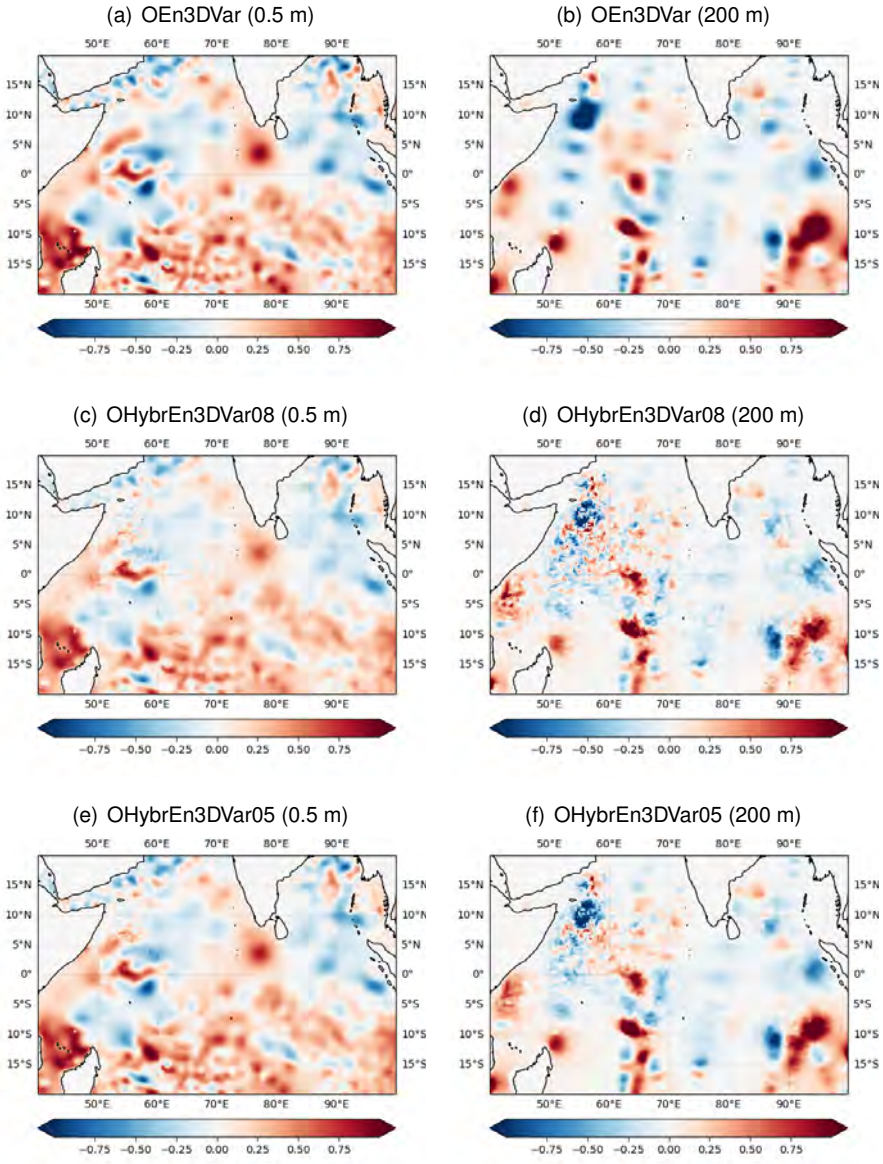


Figure 43: Temperature increments at the surface and at 200 m (in °C), on the first cycle, comparing hybrid EnVar experiments and 3DVar for the unperturbed ensemble member.

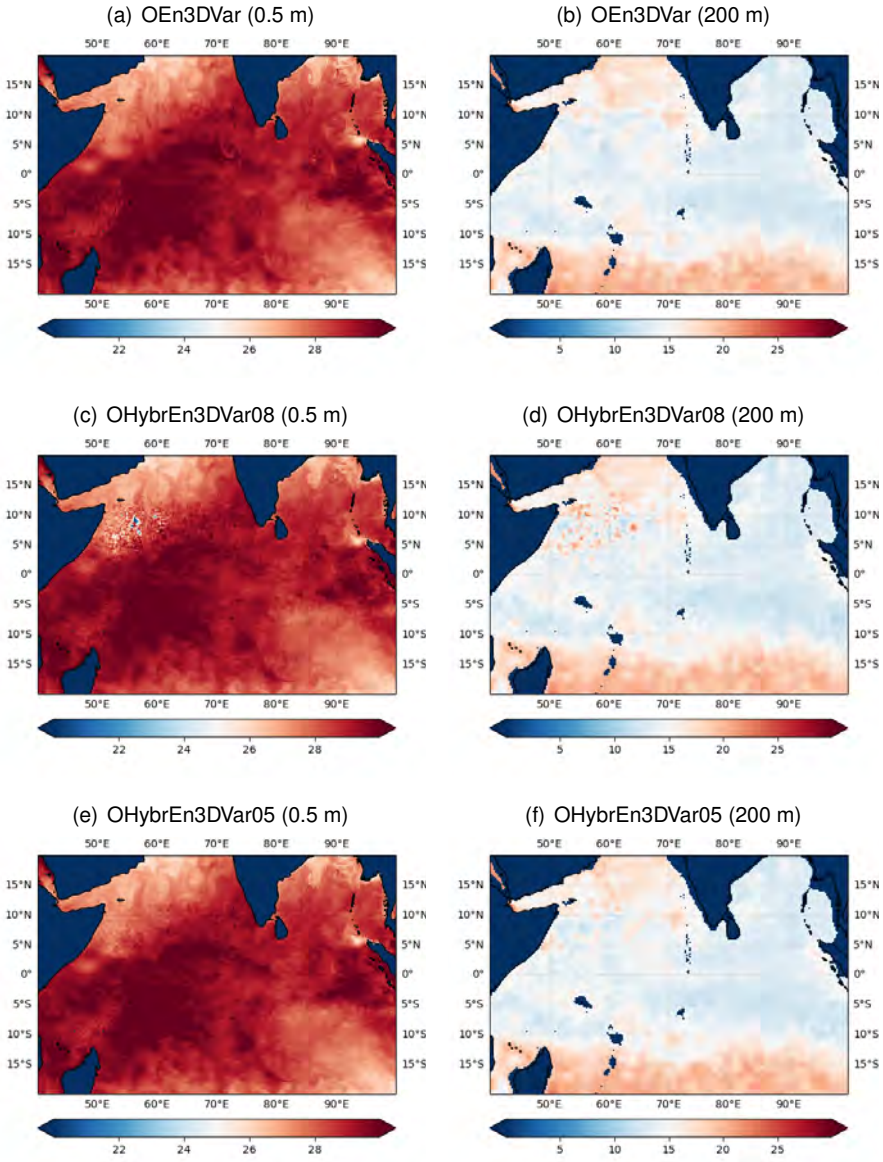


Figure 44: Ocean model temperature at the surface and at 200 m depth (in °C) comparing hybrid EnVar experiments and 3DVar for the unperturbed ensemble member.

5 Summary and discussion

5.1 Summary

We have included the ocean ensemble system developed previously (Lea et al. 2022) into the coupled NWP suite and run trials to assess the impact of ocean changes on the different components (ocean, sea-ice and atmosphere) of the coupled ensemble forecasts. The technical implementation has worked as expected and for the most part the upgraded system works well.

The ocean ensemble results are comparable to the previous uncoupled ocean ensemble results. In particular we have an ocean ensemble mean which gives smaller RMS errors against observations compared to the deterministic forecast for most fields. The hybrid-3DEnVar approach in the ocean delivers significantly improved SLA and profile statistics for the ocean both in the analysis and in the forecast. There are some areas which need further attention in the ocean ensemble generation though. The SST spread is somewhat too small compared to the uncertainty in some regions. The hybrid-3DEnVar scheme also generated some instabilities in the western Indian Ocean when too much weight was given to the ensemble part of the background error covariances, so some further work to understand this in more detail would be useful.

Sea-ice variables did not receive much attention in our previous ensemble development work so we have assessed some aspects here. The impact on the sea-ice itself was reduced RMS errors in the ensemble mean sea-ice concentration from the hybrid-3DEnVar in the ocean, probably due to improved ocean fields near the ice edge. However, the coupled ensemble was shown to be very sensitive to the way the sea-ice ensemble behaves. Small changes in the ensemble mean sea-ice concentration of only a few percent resulted in significant changes to the grid-box mean surface temperatures seen by the atmosphere (of a few degrees) in the Arctic winter due to the large surface sea-ice/open water temperature contrasts. This demonstrates the need to improve the sea-ice ensemble generation, particularly to develop improved methods for these non-Gaussian variables. Also, the addition of SST perturbations in the atmospheric part of the ensemble appears to degrade the sea-ice ensemble.

The atmosphere ensemble performance was shown to be sensitive to the ocean ensemble changes, with a large degradation in ensemble performance (CRPS scores) when the existing atmospheric SST perturbations were replaced by the ocean ensemble SSTs. While the ocean ensemble SSTs are slightly underspread compared to the uncertainty in the ensemble mean, the magnitude of the SST spread needed to obtain good performance in the near-surface atmospheric ensemble is significantly larger than the true SST uncertainty. The spread in 1.5 m air temperature is fairly insensitive to the SST spread (compare the middle panels of Fig. 35 and Fig. 37) which perhaps explains why such large SST perturbations are needed to generate spread in the atmosphere, which should ideally be introduced by improving the representation of uncertainty in other aspects of the atmospheric boundary layer. Still, when we re–introduce some of the extra SST perturbations needed by the atmosphere on top of the ocean ensemble developments, we end up with a system with similar ensemble performance in the atmosphere to the *Control*, a fairly reliable ocean ensemble, and improvements to the ocean performance possible through the use of hybrid-3DEnVar.

5.2 Directions for future work

We plan to explore and address particular aspects of the coupled NWP system which may have been degraded by the ocean ensemble. This will include work to increase the SST spread in the ocean so that it has improved reliability compared to the results shown here. Ideally we would also reduce or even eliminate the need for additional SST perturbations due to their adhoc nature. To understand how to do this will need more detailed understanding of the workings of the atmosphere boundary layer and requires a coordinated project with inputs from ocean, sea-ice and atmospheric boundary layer modelling and data assimilation scientists.

We also plan to work on developments to the ensemble aspects of the sea-ice component of the coupled ensemble. This includes modifying the observation perturbation code to deal more appropriately with non-Gaussianity as well as introducing sea-ice model perturbations. The move to the GC5 model in upcoming operational updates means that the sea-ice model is changing from CICE (as used here) to the SI3 model which is part of the NEMO software. This will make it easier to use the existing ocean model perturbation code for the sea-ice part and make use of sea-ice ensemble perturbation methods developed by other NEMO developers.

There will need to be work on the new JEDI-based Observations Processing Application (JOPA), in order to implement ocean and sea-ice observation perturbations in future versions of the coupled NWP suite. A longer term goal is to explore running the ocean data assimilation in JEDI which will provide a convenient framework for potentially implementing more strongly coupled data assimilation.

Due to the prioritisation of other developments to the coupled NWP system there will be a delay of several years before the ocean ensemble developments explored here will be implemented operationally. We will need to evolve the ocean ensemble system to keep up with these developments. The plan for coupled NWP is for all the coupled ensemble forecasts to run the ocean at 1/12° resolution (compared to the current 1/4° resolution). In order to be ready for this we will explore, with the ocean only ensemble suite, running our ocean ensemble at 1/12° resolution. To make this feasible, we will improve the computational performance of NEMOVAR, and hybrid-3DEnVar in particular, by running the ensemble localisation on a coarser grid.

References

Barton, N., Metzger, E. J., Reynolds, C. A., Ruston, B., Rowley, C., Smedstad, O. M., Ridout, J. A., Wallcraft, A., Frolov, S., Hogan, P., Janiga, M. A., Shriver, J., McLay, J., Thoppil, P., Huang, A., Crawford, W., Whitcomb, T., Bishop, C., Zamudio, L. & Phelps, M. (2020), 'The Navy's Earth System Prediction Capability: a new global coupled atmosphere-ocean-sea ice prediction system designed for daily to subseasonal forecasting', *Earth and Space Science* p. e2020EA001199.
URL: *https://doi.org/10.1029/2020EA001199*

Bloom, S. C., Takacs, L. L., da Silva, A. M. & Ledvina, D. (1996), 'Data assimilation using incremental analysis updates.', *Monthly Weather Review* **124**, 1256–1271.
URL: *https://doi.org/10.1175/1520-0493(1996)124¡1256:DAUIAU¿2.0.CO;2*

Bowler, N. E., Clayton, A. M., Jardak, M., Jermey, P. M., Lorenc, A. C., Wlasak, M. A., Barker, D. M., Inverarity, G. W. & Swinbank, R. (2017), 'The effect of improved ensemble covariances on hybrid variational data assimilation', *Quarterly Journal of the Royal Meteorological Society* **143**(703), 785–797.
URL: *https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2964*

de Rosnay, P., Browne, P., de Boisséson, E., Fairbairn, D., Hirahara, Y., Ochi, K., Schepers, D., Weston, P., Zuo, H., Alonso-Balmaseda, M., Balsamo, G., Bonavita, M., Borman, N., Brown, A., Chrust, M., Dahoui, M., Chiara, G., English, S., Geer, A., Healy, S., Hersbach, H., Laloyaux, P., Magnusson, L., Massart, S., McNally, A., Pappenberger, F. & Rabier, F. (2022), 'Coupled data assimilation at ecmwf: current status, challenges and future developments', *Quarterly Journal of the Royal Meteorological Society* **148**(747), 2672–2702.
URL: *https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.4330*

Gómez, B., Charlton-Pérez, C. L., Lewis, H. & Candy, B. (2020), 'The met office operational soil moisture analysis system', *Remote Sensing* **12**(22).
URL: *https://www.mdpi.com/2072-4292/12/22/3691*

Good, S., Fiedler, E., Mao, C., Martin, M. J., Maycock, A., Reid, R., Roberts-Jones, J., Searle, T., Waters, J., While, J. & Worsfold, M. (2020), 'The current configuration of the ostia system for operational production of foundation sea surface temperature and ice concentration analyses', *Remote Sensing* **12**(4).
URL: *https://www.mdpi.com/2072-4292/12/4/720*

Guiavarc'h, C., Roberts-Jones, J., Harris, C., Lea, D. J., Ryan, A., & Ascione, I. (2019), 'Assessment of ocean analysis and forecast from an atmosphere-ocean coupled data assimilation operational system', *Ocean Sci.* **15**, 1307–1326.
URL: *https://doi.org/10.5194/os-15-1307-2019*

Hotta, D. & Ota, Y. (2019), 'Statistical generation of SST perturbations with spatio-temporally coherent growing patterns', *Quarterly Journal of the Royal Meteorological Society* **145**(721), 1660–1673.
URL: *https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3518*

Inverarity, G. W., Tennant, W. J., Anton, L., Bowler, N. E., Clayton, A. M., Jardak, M., Lorenc, A. C., Rawlins, F., Thompson, S. A., Thurlow, M. S., Walters, D. N. & Wlasak, M. A. (2023), 'Met Office MOGREPS-G initialisation using an ensemble of hybrid four-dimensional ensemble variational (En-4DEnVar) data assimilations', *Quarterly Journal of the Royal Meteorological Society* **149**(753), 1138–1164.
URL: *https://doi.org/10.1002/qj.4431*

Lea, D. J., Mirouze, I., Martin, M. J., King, R. R., Hines, A., Walters, D. & Thurlow, M. (2015), 'Assessing a New Coupled Data Assimilation System Based on the Met Office Coupled Atmosphere-Land-Ocean-Sea Ice Model', *Monthly Weather Review* **143**(11), 4678–4694.
URL: *https://doi.org/10.1175/MWR-D-15-0174.1*

Lea, D. J., While, J., Martin, M. J., Weaver, A., Storto, A. & Chrust, M. (2022), 'A new global ocean ensemble system at the met office: Assessing the impact of hybrid data assimilation and inflation settings', *Quarterly Journal of the Royal Meteorological Society* **148**(745), 1996–2030.
URL: *https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.4292*

Lorenc, A. C. & Jardak, M. (2018), 'A comparison of hybrid variational data assimilation methods for global NWP', *Quarterly Journal of the Royal Meteorological Society* **144**(717), 2748–2760.

Lorenc, A. C., Jardak, M., Payne, T., Bowler, N. E. & Wlasak, M. A. (2017), 'Computing an ensemble of variational data assimilations using its mean and perturbations', *Quarterly Journal of the Royal Meteorological Society* **143**(703), 798–805.
URL: *https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2965*

MacLachlan, C., Arribas, A., Peterson, K. A., Maidens, A., Fereday, D., Scaife, A. A., Gordon, M., Vellinga, M., Williams, A., Comer, R. E., Camp, J., Xavier, P. & Madec, G. (2015), 'Global seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system', *Quarterly Journal of the Royal Meteorological Society* **141**(689), 1072–1084.
URL: *https://doi.org/10.1002/qj.2396*

Mirouze, I., Blockley, E. W., Lea, D. J., Martin, M. J. & Bell, M. J. (2016), 'A multiple length scale correlation operator for ocean data assimilation', *Tellus A: Dynamic Meteorology and Oceanography* **68**.
URL: *https://doi.org/10.3402/tellusa.v68.29744*

Ollinaho, P., Lock, S.-J., Leutbecher, M., Bechtold, P., Beljaars, A., Bozzo, A., Forbes, R. M., Haiden, T., Hogan, R. J. & Sandu, I. (2017), 'Towards process-level representation of model uncertainties: stochastically perturbed parametrizations in the ECMWF ensemble', *Quarterly Journal of the Royal Meteorological Society* **143**(702), 408–422.
URL: *https://doi.org/10.1002/qj.2931*

Piccolo, C., Cullen, M. J. P., Tennant, W. J. & Semple, A. T. (2019), 'Comparison of different representations of model error in ensemble forecasts', *Quarterly Journal of the Royal Meteorological Society* **145**(718), 15–27.
URL: *https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3348*

Ridley, J. K., Blockley, E. W., Keen, A. B., Rae, J. G. L., West, A. E. & Schroeder, D. (2018), 'The sea ice model component of HadGEM3-GC3.1', *Geoscience Model Development* **11**, 713–723.
URL: *https://doi.org/10.5194/gmd-11-713-2018*

Rodwell, M. J., Lang, S. T. K., Ingleby, B. N., Bormann, N., Hólm, E., Rabier, F., Richardson, D. S. & Yamaguchi, M. (2016), 'Reliability in ensemble data assimilation', *Quarterly Journal of the Royal Meteorological Society* **142**, 443–454.
URL: *https://doi.org/10.1002/qj.2663*

Sanchez, C., Williams, K. D. & Collins, M. (2016), 'Improved stochastic physics schemes for global weather and climate models', *Quarterly Journal of the Royal Meteorological Society* **142**(694), 147–159.
URL: *https://doi.org/10.1002/qj.2640*

Skachko, S., Buehner, M., Laroche, S., Lapalme, E., Smith, G., Roy, F., Surcel-Colan, D., Bélanger, J.-. & Garan, L. (2019), 'Weakly coupled atmosphere–ocean data assimilation in the canadian global prediction system (v1)', *Geosci. Model Dev.* **12**, 5097–5112.
URL: *https://doi.org/10.5194/gmd-12-5097-2019*

Storkey, D., Blaker, A. T., Mathiot, P., Megann, A., Aksenov, Y., Blockley, E. W., Calvert, D., Graham, T., Hewitt, H. T., Hyder, P., Kuhlbrodt, T., Rae, J. G. L. & Sinha, B. (2018), 'UK Global Ocean GO6 and GO7: a traceable hierarchy of model resolutions', *Geoscientific Model Development* **11**, 3187–3213.
URL: *https://doi.org/10.5194/gmd-11-3187-2018*

Storto, A. & Andriopoulos, P. (2021), 'A new stochastic ocean physics package and its application to hybrid-covariance data assimilation', *Quarterly Journal of the Royal Meteorological Society* .
URL: *https://doi.org/10.1002/qj.3990*

Tennant, W. & Beare, S. (2014), 'New schemes to perturb sea-surface temperature and soil moisture content in MOGREPS', *Quarterly Journal of the Royal Meteorological Society* **140**(681), 1150–1160.
URL: *https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2202*

Tennant, W. J., Shutts, G. J., Arribas, A. & Thompson, S. A. (2011), 'Using a stochastic kinetic energy backscatter scheme to improve MOGREPS probabilistic forecast skill', *Monthly Weather Review* **139**(4), 1190–1206.
URL: *https://doi.org/10.1175/2010MWR3430.1*

Vellinga, M., Copsey, D., Graham, T., Milton, S. & Johns, T. (2020), 'Evaluating benefits of two-way ocean–atmosphere coupling for global nwp forecasts', *Weather and Forecasting* **35**(5), 2127 – 2144.
URL: *https://journals.ametsoc.org/view/journals/wefo/35/5/wafD200035.xml*

Walters, D., Baran, A. J., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J., Furtado, K., Hill, P., Lock, A., Manners, J., Morcrette, C., Mulcahy, J., Sanchez, C., Smith, C., Stratton, R., Tennant, W., Tomassini, L., Van Weverberg, K., Vosper, S., Willett, M., Browse, J., Bushell, A., Carslaw, K., Dalvi, M., Essery, R., Gedney, N., Hardiman, S., Johnson, B., Johnson, C., Jones, A., Jones, C., Mann, G., Milton, S., Rumbold, H., Sellar, A., Ujiie, M., Whitall, M., Williams, K. & Zerroukat, M. (2019), 'The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations', *Geoscientific Model Development* **12**(5), 1909–1963.
URL: *https://gmd.copernicus.org/articles/12/1909/2019/*

Waters, J., Lea, D., Martin, M., Mirouze, I., Weaver, A. & While, J. (2015), 'Implementing a variational data assimilation system in an operational 1/4 degree global ocean model', *Quarterly Journal of the Royal Meteorological Society* **141**, 333–349.
URL: *https://doi.org/10.1002/qj.2388*

Whitaker, J. S. & Hamill, T. M. (2012), 'Evaluating methods to account for system errors in ensemble data assimilation', *Monthly Weather Review* **140**(9), 3078–3089.
URL: *https://doi.org/10.1175/MWR-D-11-00276.1*

Met Office
FitzRoy Road
Exeter
Devon
EX1 3PB
United Kingdom

Produced with the Met Office LaTeX template v3.0.0