

# **Short-range ocean forecast error characteristics in high resolution assimilative systems**

March 25, 2021

Technical Report No: 645

Davi M. Carneiro<sup>1</sup>, Robert King<sup>1</sup>, Matthew Martin<sup>1</sup> and  
Ana Aguiar<sup>1</sup>

<sup>1</sup>Met Office, Exeter, UK

## Executive summary

This work assesses short-range forecast error characteristics of high-resolution assimilative ocean forecasting systems based on NEMO. The resolution of the models used ranges from  $1/4^\circ$  to  $1/12^\circ$  globally and down to 1.5 km in the North-West European Shelf-Seas region. The regional model includes processes not represented in the global models such as tides and a better representation of the complex bathymetry in the region. A common set of high-resolution satellite data (as well as in situ observations) are assimilated into each of these configurations which are all used in CMEMS forecasting systems. A common two-year period is used for all the integrations with a common set of surface atmospheric forcing so that the forecast error characteristics can be estimated in a consistent way. Additionally, the 1.5 km model is forced by two different sets of atmospheric fields to understand their impact on the magnitude and scales of the forecast errors.

Two methods are used to estimate the forecast error characteristics for sea surface temperature (SST). The first is based on the observation-forecast differences (called the innovations) and the second is based on differences in two forecasts of different lengths, both valid at the same time. Both sets of information about the forecast errors are used to extract seasonal information about the forecast error covariances and the results are summarised by estimating length-scales and variances associated with two components of the SST errors. The results show a clear seasonality in the SST forecast errors in both global configurations, with increased errors in the summer hemisphere. Although both global configurations have similar seasonal variations, the system at  $1/12^\circ$  resolution gives larger SST small-scale errors than the one at  $1/4^\circ$  resolution. This increase of the small-scale errors in the  $1/12^\circ$  global configuration is mostly associated with the increase of the error covariances within 25 km of separation distance, consistently seen in the boundary currents. Further improvements should be made to the assimilation in the  $1/12^\circ$  global system to properly constrain fine-scale structures, which are not present at a resolution of  $1/4^\circ$ . The impact of the surface forcing on the representation of the forecast errors is unclear with different results given by the error covariance methods, and further investigation is necessary to understand these differences. However, in contrast to the  $1/12^\circ$  global system, the benefits of increasing the model resolution to 1.5 km in the North-West European shelf is clearly seen, with an overall reduction of the small-scale SST forecast errors and their associated length-scales in the shelf seas and deep part of the domain. Therefore, the positive impact of 1.5 km systems on properly reproducing fine-scale SST features reinforces the large potential for future higher-resolution global models that are currently being developed by operational centres. In addition to the impact of the resolution on the forecast errors, the error covariance estimates produced here will also be used to update the model background error parameters in the data assimilation system, which is expected to reduce the forecast errors of these model configurations in the future.

## Contents

1. Introduction .....	4
2. Model and data assimilation system description.....	6
2.1 Global models.....	8
2.2 Shelf-seas model .....	8
2.3 Data assimilation .....	10
2.4 Description of assimilated observations .....	11
3. Methods for estimating error covariances.....	12
3.1 Innovations-based method.....	14
3.2 Forecasts of different lengths.....	16
4. Results.....	17
4.1 Global error covariances .....	17
4.1.1 Comparison of HL estimates between VIIRS and drifter innovations.....	17
4.1.2 Small-scale forecast error variances .....	19
4.1.3 Large-scale forecast error variances .....	22
4.1.4 Forecast error length-scales .....	24
4.1.5 Forecast error covariances at short separation distances .....	28
4.2 North-west European shelf-seas error covariances.....	31
4.2.1 Small-scale forecast error variances .....	31
4.2.2 Large-scale forecast error variances .....	33
4.2.3 Small-scale forecast error length-scales .....	33
4.2.4 Large-scale forecast error length-scales .....	34
4.2.5 Forecast error covariances at short separation distances .....	34
5. Discussion and conclusions .....	39
References .....	43

## 1. Introduction

Short-range ocean forecasts are produced operationally at many centres around the world with the forecasts used for various applications including ship routing, oil spill modelling, search and rescue, and Naval operations (Davidson et al., 2019). The ocean analyses from such systems are also used to initialise coupled ocean/atmosphere forecasts from short-range out to seasonal and decadal timescales (Guiavarc'h et al., 2019; Penny et al., 2019). A good estimate of the current state of the ocean is required to produce these forecasts, and this is done using data assimilation methods which combine the available in situ and satellite observations with the ocean model (see Moore et al., 2019 for a recent review).

For short-range ocean and coupled prediction in particular, many applications need high resolution information about the ocean. For global scales, eddy resolving models are now run at many operational centres with Mercator Ocean and the Met Office now running  $1/12^\circ$  global systems operationally for instance (Lellouche et al., 2018; Aguiar et al., 2021) and CMCC running a  $1/16^\circ$  global system (Iovino et al., 2016). Higher resolution global systems are being run elsewhere: for instance, the US Navy currently runs a  $1/25^\circ$  global version of HYCOM to produce operational forecasts (Barton et al., 2020); Mercator Ocean is currently developing a  $1/32^\circ$  global system for future implementation.

Regional forecasting systems are often run at even higher resolution than the above global systems (De Mey-Frémaux et al., 2019). This extra resolution is often needed to represent processes which are important for regional and coastal applications which attempt to represent the ocean state in regions including sharp bathymetric features, complex coastlines and in shallower waters where tides are important. The Met Office for instance currently runs a 1.5 km resolution version of NEMO to produce operational forecasts for the North West European Shelf Seas (NWS) as part of the Copernicus Marine Environment Monitoring Service (CMEMS; Tonani et al., 2019).

The aim of this work is to improve our understanding of the error characteristics of high-resolution global and shelf-seas forecasting systems. This includes understanding not only the magnitude (variance) of the forecast errors, but also their spatial characteristics (their covariances). Such an understanding is useful for several reasons. Users of the forecast data often want to know how accurate the data are, so the forecast error variance information is useful for them. When designing observing systems for assimilation and forecasting applications, information about the spatial scales of the errors can help to determine the optimal sampling strategy in different regions. Data assimilation systems also need information about the forecast error covariances to make best use of the observations. Information about the forecast errors can also help to identify areas where the forecasting systems can be improved through model and data assimilation development and can be used to understand whether increases in model resolution lead to better forecasts.

Estimation of forecast errors is a large topic which has been addressed by many authors both for verification/validation purposes and for data assimilation applications. The fundamental issue with this subject is that there is no spatially and temporally complete estimate of the “true” ocean with which to compare the forecasts. Observations are the only source of direct

information about the ocean, but they are spatially and temporally incomplete and suffer from various types of error. Proxies for the true error can be obtained by comparing model forecasts of differing lengths, valid at the same time. This has the advantage of providing information everywhere at the resolution of the model grid but can suffer from issues when the observations being assimilated are sparse compared to the scales of the errors and/or when there is model bias. Ensembles can also be used to estimate uncertainties in the forecast, but the results depend on the ensemble generation strategies adequately representing the various sources of uncertainties in the forecast, and there being a large enough ensemble to avoid sampling issues, which is very expensive to do when dealing with high resolution models. A comprehensive review of these various methods for estimating forecast error covariances has been written by Bannister et al. (2008) with a focus on atmospheric data assimilation. Another approach to tune the error covariances in a data assimilation system is the one originally proposed by Desroziers et al. (2005) which uses differences between observations, analysis and forecast solutions at the observation locations. This was shown by Mattern et al. (2018) to improve the quality of a regional ocean data assimilation system, but their approach scales an existing forecast error field for each variable, assuming that its underlying spatial structure is correct. We do not apply the Desroziers method here since we aim to provide a good estimate of the underlying spatial structure of the forecast error, and compare those structures between different assimilative systems. Further work could include its application to further tune the error covariances estimated in this work.

Roberts-Jones et al. (2016) used observation-forecast differences to estimate the error covariances of sea surface temperature (SST) in the Operational sea Surface Temperature and sea Ice Analysis (OSTIA) system. The work described here follows on from that work by applying similar methods to understand the forecast errors in high resolution global and regional forecasting systems, as well as using the forecast-difference approach mentioned above. The observation operator code in NEMO (which interpolates model information to the observation locations at the nearest model time-step) is used to generate the information for the observation-based method. Python code has been developed to perform these calculations and has been made freely available via a Github repository.

Outputs from three different configurations of the NEMO model (Madec, 2016), all employed for operational forecasting in the Met Office, are used: i) a global  $1/4^\circ$  resolution system, ii) a  $1/12^\circ$  resolution global system and (iii) two versions of a 1.5 km resolution system for the NWS region, with different sets of atmospheric forcing. A common experiment framework is used to produce analysis and forecast data over two years for these configurations: the same version of NEMO (version 3.6; Madec, 2016) is used for all systems, the observations input to the data assimilation are the same, and the data assimilation systems use a common 3DVar-FGAT (First Guess at Appropriate Time) scheme based on the NEMOVAR code (Mogensen et al., 2012). High resolution observational datasets need to be assimilated for the forecast error covariance estimation methods to work well, and we assimilate swath SST data from many satellites into these systems along-side in situ SST data, observations from satellite altimeters, in situ profiles of temperature and salinity, and satellite sea-ice concentration data (in the global systems).

More detail on the model and data assimilation systems as well as the observations

assimilated is given in section 2. The methods used to estimate the forecast error characteristics are described in section 3. The estimated error covariances are presented in section 4, including: a comparison of the results from the different methods; an assessment of the impact of the model resolution on the error covariances and their spatial scales, both globally and for the NWS region; and an assessment of the impact of the atmospheric forcing. A discussion of the main results and the conclusions are presented in section 5.

## **2. Model and data assimilation system description**

The experiments described in this report use different configurations of the same underlying ocean analysis and forecasting system: the Met Office Forecasting Ocean Assimilation Model (FOAM) system. For the global experiments, we use the 1/12° ORCA12 and 1/4° ORCA025 configurations of FOAM, while for the higher resolution regional experiments, we use the 1.5 km resolution AMM15. All these systems are run operationally to produce analyses and forecasts, and we use the most recent version of the systems. All the FOAM configurations use the NEMO ocean model and the NEMOVAR assimilation scheme with a 24-hour assimilation window. Each assimilates broadly the same observations, but with configuration-specific observation averaging and quality control (as detailed below). On each day a two-day forecast was run to allow the calculation of short-term forecast errors.

To illustrate the differences in characteristics of the three different configurations used here, Fig. 1 shows the daily mean surface currents in the region of the AMM15 configuration. Although the main dynamical features are in similar locations, there is a clear change in character between the ORCA025 and ORCA12 configurations, with ORCA12 having sharper features with stronger currents. Comparing ORCA12 with AMM15, the higher resolution system can represent much smaller scales in the deeper part of the region and exhibits different characteristics to the flow in the shallower waters.

All simulations were initialised in June 2016 from a spun-up state from previous reanalysis runs with an older version of the system and different sets of observations, and they were run for a period of 6 months prior to being used in the error covariance calculations from 1<sup>st</sup> January 2017. The global simulations were then run to the end of 2018, while the AMM15 simulations were run to June 2018. After discarding the 6-month spin-up period, this provided a 2-year ORCA12, a 2-year ORCA025, and 18-month AMM15 simulations. A summary of the experiments used in the forecast error calculations is given in Table 1.

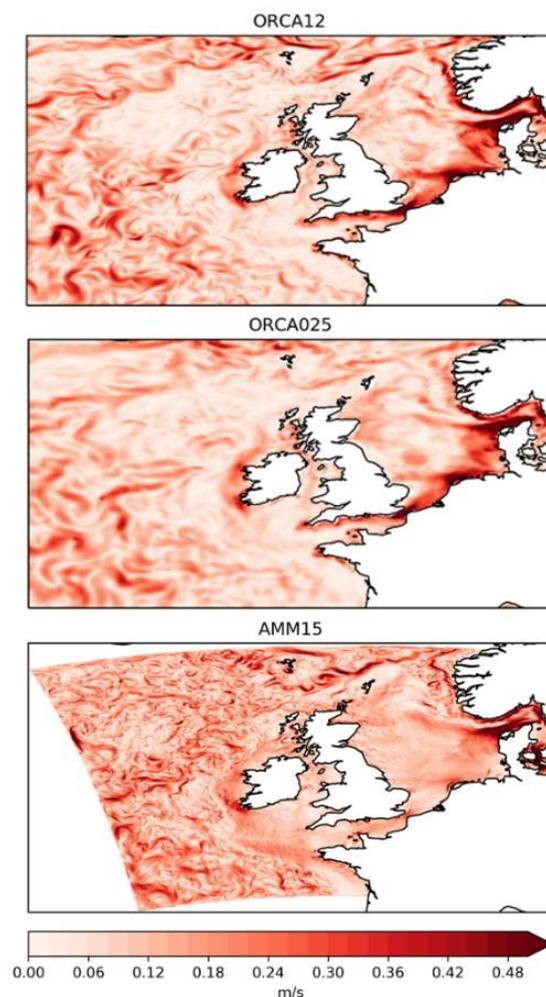


Figure 1: Daily mean surface current speed in the AMM15 domain from the three FOAM configurations for 1<sup>st</sup> January 2018.

Configuration	Model	Grid resolution of model/assimilation	Surface forcing and resolution	Period
ORCA025	NEMO/CICE	1/4° (1/4°)	MetUM (17 km -> 10 km)	1/1/2017-31/12/2018
ORCA12	NEMO/CICE	1/12° (1/4°)	MetUM (17 km -> 10 km)	1/1/2017-31/12/2018
AMM15-MO	NEMO/WWIII	1.5 km (1.5 km)	MetUM (17 km -> 10 km)	1/1/2017-30/6/2018
AMM15-ECMWF	NEMO/WWIII	1.5 km (1.5 km)	ECMWF IFS (14 km)	1/1/2017-30/6/2018

Table 1. Summary of the experiments used as input to the error covariance calculations.

## 2.1 Global models

The FOAM-ORCA12 and FOAM-ORCA025 configurations use the NEMO ocean model version 3.6 (Madec, 2016) coupled to the Los Alamos sea ice model (CICE; Hunke et al., 2010) in the GO6 configuration described by Storkey et al. (2018) and the GSI8 configuration described by Ridley et al. (2018). Both configurations have a non-linear-free surface and 75 vertical levels, with 1 m vertical resolution at the surface and decreasing resolution with increasing depth. The three-fold increase in horizontal resolution from ORCA025 to ORCA12 encompasses the transition from eddy-permitting to eddy-resolving scales at most latitudes (see Fig. 2). In both configurations, the ocean is forced at the surface using hourly momentum and 3-hourly freshwater and heat fluxes from the operational Met Office Unified Model (MetUM) global Numerical Weather Prediction (NWP) system. The resolution of the atmospheric forcing is about 17 km for the period up to mid-July 2017, and then increases to about 10 km after that. Both these configurations are currently run operationally at the Met Office. The ORCA025 configuration is used as part of a coupled forecasting system (Guiavarc'h et al., 2019) which provides ocean forecasts to CMEMS, while the ORCA12 configuration is producing operational forecasts for the UK Royal Navy (Aguiar et al., 2021). A similar ORCA12 configuration is used to produce global ocean forecasts for CMEMS by Mercator Ocean.

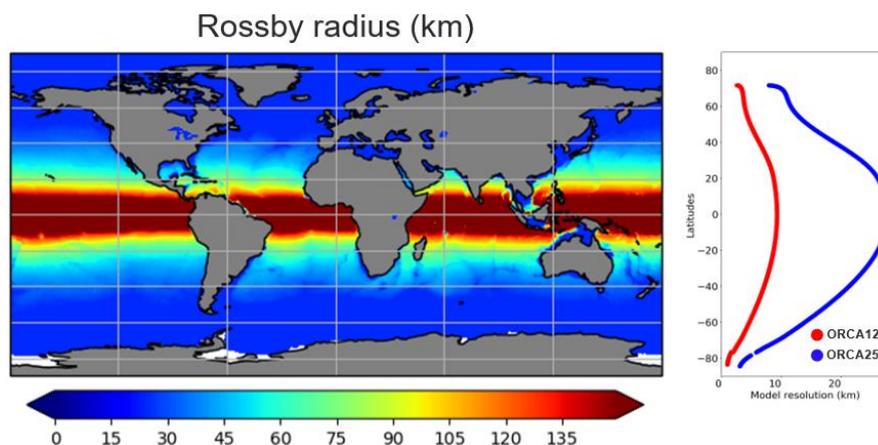


Figure 2: Rossby radius (km) (left) and model grid resolution (km) as a function of latitude (right) for ORCA025 and ORCA12.

## 2.2 Shelf-seas model

The FOAM-AMM15 configuration employs the same system as used in the operational Met Office North-West European Shelf ocean forecasting system (Tonani et al., 2019) which delivers forecast products to CMEMS. This system couples the NEMO ocean model, also at version 3.6, with the WAVEWATCH III spectral wave model (Saulter et al. 2017; Lewis et al. 2019). The domain (shown in Fig. 2) extends from approximately 45°N, 20°W to 63°N, 11°E on a rotated pole grid, covering the north-west European Shelf and part of the north-east Atlantic Ocean. This domain has a complex bathymetry, containing regions of shallow water,

a continental shelf, and regions of deep ocean (i.e. depths greater than 5000 m). The resulting dynamics produce mesoscale eddies along with seasonally stratified and well-mixed regions with marked boundaries. The large changes in bathymetry mean that there is a wide range of scales associated with the first baroclinic Rossby radius (also shown in Fig. 3). In the shallow waters of the shelf the Rossby radius is smaller than 5 km, while in the deeper ocean it is closer to 25 km.

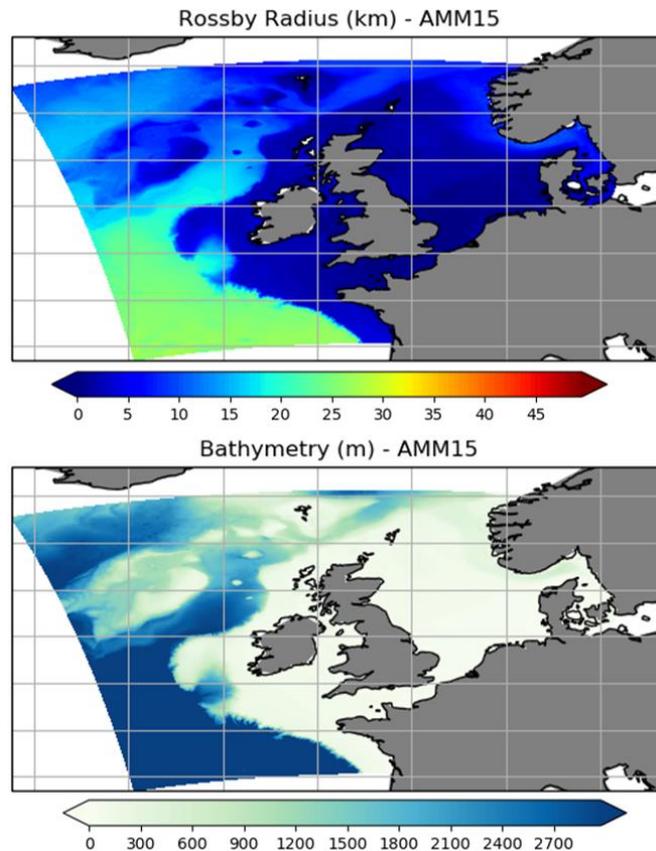


Figure 3: Rossby radius (km) (top) and Bathymetry (m) (bottom) for AMM15.

The ocean model component uses the eddy-resolving 1.5 km Atlantic Margin Model configuration (AMM15, Graham et al., 2018) of NEMO version 3.6 (Madec 2016). The model employs a non-linear free surface, 51 vertical terrain-following levels and includes 11 tidal constituents, specified on the open boundary via the Flather radiation condition (Flather et al., 1981). The AMM15 is nested within the Met Office 1/12° North Atlantic model (NATL12; Storkey et al. 2010; Blockley et al. 2014) which provides the lateral boundary conditions, except for the Baltic boundary which is taken from the CMEMS Baltic Sea model (Berg et al., 2012). The ocean and wave components are coupled hourly using the OASIS3-MCT coupler (Craig et al., 2017).

The operational version of this system is forced at the surface using ~14 km resolution, 3-hourly wind and atmospheric pressure fields, radiative and E-P (evaporation-minus-precipitation) fluxes from the ECMWF operational Integrated Forecasting System (IFS). However, as detailed above, the FOAM-ORCA12 system uses surface forcing from the operational Met Office global atmospheric model (~10 km horizontal resolution, hourly wind

and pressure, and 3-hourly radiative and E-P fluxes). Therefore, to investigate the impact of the different forcing, we have run two AMM15 simulations: one forced by the ECMWF IFS fluxes, and the other forced by the Met Office UM fluxes. In each case, the wave and ocean models are forced by the same wind fields, and the wave model uses the surface currents computed by the ocean model.

### 2.3 Data assimilation

The NEMOVAR data assimilation scheme is used in all configurations of the operational Met Office ocean analysis and forecasting systems. This is a multi-variate, multi-lengthscale assimilation scheme developed collaboratively by the Met Office, Centre Europeen de Recherche et de Formation Avancee en Calcul Scientifique (CERFACS), the European Centre for Medium Range Weather Forecasts (ECMWF), and Institut National de Recherche en Informatique et en Automatique (INRIA) (Mogensen et al. 2009; Mogensen et al. 2012; Waters et al. 2015). The implementation in the global configurations is based on that used in the ORCA025 system as described by Waters et al. (2015) and Mirouze et al. (2016). In the AMM15 system the implementation is as described for a lower resolution regional AMM7 system by King et al. (2018), with modifications for application to AMM15 as described by Tonani et al. (2019).

An assimilation window of 24 hours is used to assimilate observations including satellite and in situ SST, satellite altimeter observations of the sea level anomaly (SLA), and in situ temperature and salinity profiles from a variety of platforms. The systems include a variational bias correction scheme for SST and SLA observations (While and Martin, 2019; Lea et al., 2008) and the assimilation increments are added incrementally over 24 hours through an incremental analysis update (Bloom et al. 1996). In the ORCA025 and AMM15 configurations, the assimilation is carried out on the same grid as the ocean model. However, in the current implementation of ORCA12, the assimilation takes place on the ORCA025 grid and increments are interpolated to ORCA12 ahead of the incremental analysis update.

NEMOVAR uses linearised balance relationships to account for correlations between ocean variables as defined by Weaver et al. (2005). The temperature-salinity balance operator is a linearised version of the water-mass property conservation scheme of Troccoli & Haines (1999). The sea surface height (SSH) balance operator is a linearised version of the Cooper & Haines (1996) scheme in which SSH changes are associated with the lifting and lowering of the entire water column. The velocity balance operators are based on the geostrophic relation, density balance is the linearised equation of state, and the pressure balance is the vertically integrated hydrostatic equation.

The background error covariance is approximated using an estimate of the model variable variances and with background-error correlations modelled using a diffusion equation (Mirouze & Weaver, 2010). Various options exist in NEMOVAR for how to apply the diffusion equation to model the spatial correlations. It can be solved explicitly or implicitly, and it can be used to represent isotropic or anisotropic correlation structures (Weaver and Mirouze, 2013). Various functions belonging to the class of Matérn functions (which contains the Gaussian

function) can be represented efficiently using the implicit solution to the diffusion equation. In the FOAM system we use the implicit solution to the diffusion equation to model a combination of two isotropic Gaussian functions. This uses a combination of a short and a long length-scale to account for errors in both internal ocean processes and larger-scale errors from the atmospheric forcing (Mirouze et al., 2016). In each configuration, the short horizontal correlation length-scales are prescribed through the first baroclinic Rossby radius. In each case, the minimum length-scale is capped to avoid numerical issues (25 km for ORCA025, and 4.5 km for AMM15). Furthermore, in the global configurations the maximum short length-scale is capped at 150 km near the equator. The long background error correlation length-scale is set to 400 km in the global configurations and to 100 km in the AMM15.

The current observation and background error variances in NEMOVAR are specified using a combination of statistical estimates and parameterisations. The surface temperature, salinity, and unbalanced SSH error variances are specified based on spatially and seasonally varying estimates calculated from earlier model runs with ORCA025 and AMM7. The sub-surface temperature/salinity background errors are parameterised as follows. The surface value is set by the statistical estimate mentioned above and a minimum value at each model level is defined by a tapered function of the surface value. To introduce some flow dependence, the vertical temperature/salinity gradient is then calculated from the background field at all grid-points. Below the base of the mixed layer, the background error standard deviation is defined as the larger of the tapered surface value or a parameterisation of the temperature/salinity gradient (e.g.  $\delta z^*(dT/dz)$ , where  $\delta z=20$  m, as described by Waters et al. 2015). This results in the largest errors occurring at the thermocline/halocline, allowing the observations to have greater weight to shift the depth of the thermocline/halocline.

In this work we are producing new estimates of the forecast error covariances based on the methods described in section 3. This will lead to new spatially and seasonally varying estimates of the error covariances which can then be used to improve the various specifications in the data assimilation described above. The work is therefore expected to feedback into improvements in the data assimilation which should lead to reduced forecast errors in these configurations in the future.

## 2.4 Description of assimilated observations

Each of the three configurations assimilates broadly the same set of observations: in situ and satellite SST observations, satellite altimeter observations of the sea level anomaly, and in situ temperature and salinity profiles from several platforms. The global configurations also assimilate satellite observations of the sea-ice concentration (the regional AMM15 has no sea-ice component).

The in-situ SST observations assimilated include observations from ships, surface drifters, and moorings. We also assimilate L2p/L3c satellite swath observations of SST from NOAA/AVHRR, MetOp/AVHRR, MSG/SEVIRI, Sentinel-3/SLSTR, and Suomi-NPP/VIIRS. All these data were distributed in Near-Real Time (NRT) by the Group for High Resolution SST (GHRSSST). In the global configurations, these data were used with a spatial median averaging

over a 13 km radius for each satellite type, while in the higher resolution AMM15 no observation averaging was done.

A background check was performed for the SST observations where each observation was compared to the model background at the nearest time-step and interpolated to the observation location. Using the specified observations and background errors, any observation with a probability of gross error exceeding 50% was rejected. Biases in the satellite observations are accounted for using an on-line bias correction scheme described by While and Martin (2017) which uses the in situ drifters and a sub-set of the VIIRS data (night-time data near the centre of the swath) as anchor observations.

Along-track satellite altimeter SLA data from CMEMS are assimilated with an on-line bias correction scheme to account for errors in the mean dynamic topography (MDT; see Lea et al., 2008) and residual errors due to atmospheric effects. While the global configurations assimilate the standard SLA data which removes the effects of tides and the ocean barotropic response to the wind forcing, the AMM15 uses supplied corrections to ensure the assimilated SLA includes the effects of wind and pressure as these physical processes are included in the regional model. Although the tidal contribution of the observed SLA is also provided, this is not used as the tides are removed from the AMM15 model counterpart before calculating the SLA innovation.

Quality-controlled in situ profile observations of temperature and salinity were taken from the EN4 dataset (Good et al., 2013). This dataset includes temperature and salinity profiles from Argo floats, moored buoy arrays, XBTs, CTDs, gliders, and marine mammals.

Sea-ice concentration observations from Special Sensor Microwave Imager/Sounder (SSM/I/S) satellite instruments, processed by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) Ocean and Sea-Ice Satellite Application Facility (OSI-SAF) are assimilated in the global ORCA025 and ORCA12 configurations.

### **3. Methods for estimating error covariances**

We use two methods for estimating the forecast error covariances. The first method is based on observation-minus-forecast data, where the model data are interpolated to the observation location at the nearest model time-step during a one-day forecast of NEMO. It was originally proposed by Hollingsworth and Lonnberg (1986) so we refer to it as the HL method here. More recently, Roberts-Jones et al. (2016) used that method to estimate error covariances for the OSTIA system and we follow their approach. The HL method relies on a dense sampling by the observations which is a challenge for the ocean since it is relatively sparsely observed compared to the atmosphere. The aim of this work is to investigate forecast errors for high resolution systems, so we chose to focus on sea surface temperature (SST) which is the variable with the densest observing network. One drawback of the HL method is that it assumes that any spatial covariances in the observation-model differences are due to forecast errors, and that the observation errors are uncorrelated spatially, so we have to be careful about which SST satellite data types we use. The advantage of the HL method is that it

provides direct estimates of the forecast errors without needing to make assumptions about the behaviour of the model forecast errors.

To complement the HL method with another method that can provide information about the forecast errors at higher spatial resolution, we use the NMC method, originally proposed by Parish and Derber (1992). This method uses forecasts of different lengths valid at the same time as a proxy for the forecast error. This has the advantage that we get information about the forecast errors at the resolution of the model grid, but it does make assumptions about the behaviour of the model in order for this to be a good proxy for the real error. A more detailed description about the HL and NMC methods and how we apply them are given in the next two sub-sections.

As discussed in the introduction, other methods are available for estimating forecast error covariances, including ensemble-based methods. We discount this type of method for this work since it is too computationally expensive to run a reasonable size ensemble of the high-resolution assimilative forecasting systems we are interested in. For example, an ensemble system for ORCA025 is currently being developed at the Met Office, but applying the same strategy is still too computationally expensive for the high-resolution model configurations, such as ORCA12 and AMM15. Error covariances for ORCA12 could still be produced based on the ORCA025 ensemble, however they would not be fully optimal as the ORCA025 ensemble is not expected to represent the same fine-scale structures generated by the ORCA12 system.

The HL method allows us to produce estimates of the forecast errors on a pre-defined (determined by having adequate observation sampling) regular grid, while the NMC estimates are made on the model grid. We calculate the covariance information on each day and accumulate the information over seasons (with 2-years' worth of data as described in the previous section). Both methods allow us to produce "local" statistics of the covariances of the forecast error either with observation locations near a particular point (for the HL method) or model grid-points near a particular point (for the NMC method). We cannot afford computationally to produce "global" error covariance estimates for every point with all other points around the globe. We restrict the distances over which the error covariances are calculated to be within 500-1000 km depending on the configuration and assume that the error covariances at longer distances are small. Even with this simplification, the error covariance calculations produce very large volumes of data from which we need to distil information about the magnitude and spatial characteristics of the errors.

To summarise the information about the error covariances from both methods, we assume a functional form of the error covariances, and produce estimates of the function's parameters by fitting it to the covariance data at each grid point. The function we use is a combination of two Gaussian functions, each with their own amplitude and their own length-scale:

$$f(r) = V_1 e^{-r^2/(2L_1^2)} + V_2 e^{-r^2/(2L_2^2)}. \quad (1)$$

Here  $V_k$  and  $L_k$  are the forecast error variance and the length-scale for component  $k$ , and  $r$  is the separation distance between two points. At zero separation distance, the forecast error variance is the sum of the variance associated with each component. It has been shown in

previous studies (e.g. Martin et al., 2007) that the errors in ocean forecasts often have two components to their errors and that using a combination produces a much better fit to the covariance data than a single function. Physically we expect there to be errors on scales associated with internal ocean model dynamics which may occur at the mesoscale, and errors associated with atmospheric forcing which may have scales similar to those of atmospheric phenomena (e.g. synoptic systems). These two length-scales are deduced iteratively from the fit with Eq. 1, and the initial guesses for the small and large length-scales are defined as 40 and 400 km, respectively. There is a question as to whether more than two length-scales should be used for very high-resolution systems which may have a third error component associated with sub-mesoscale processes.

A summary of the methods used is:

1. On each day, generate information about the forecast errors using either the HL or NMC methods described in the next sub-sections. This information is accumulated on a predefined grid (regular grid for the HL method and the model grid for the NMC method). For the HL method a pre-defined set of separation bins is used to accumulate the covariance statistics. For the NMC method the model grid-points surrounding the current grid point are used with the number of grid-points described in the next sub-sections.
2. Accumulate the daily forecast error covariance information over each month, then combine the monthly statistical information into seasons and into files for the entire experiments.
3. Fit the function defined in equation (1) to the covariance statistics produced at each grid point.

Python code has been developed to do the above calculations and has been made publicly available from a Github repository ([https://github.com/MetOffice/ocean\\_error\\_covs](https://github.com/MetOffice/ocean_error_covs)). This currently includes only the code to perform the HL calculations, but will be updated soon to include the NMC code.

### 3.1 Innovations-based method

As described above, the HL method uses information at observation locations to determine the error covariances. The differences between the observations,  $\mathbf{y}$ , and the model forecast,  $\mathbf{x}^f$ , interpolated to the observation locations using the observation operator,  $h$ , are known as the innovations and are denoted  $\mathbf{d}$ . They are comprised of the observation errors,  $\boldsymbol{\varepsilon}^o$ , and the model forecast errors,  $\boldsymbol{\varepsilon}^f$ :

$$\mathbf{d} = \mathbf{y} - h(\mathbf{x}^f) = (\mathbf{y} - h(\mathbf{x}^t)) - (h(\mathbf{x}^f) - h(\mathbf{x}^t)) = \boldsymbol{\varepsilon}^o - \boldsymbol{\varepsilon}^f,$$

where  $\mathbf{x}^t$  is the true ocean state represented on the model grid. Here we use systems with a one-day assimilation cycle and therefore use the model values during a one-day forecast to form the innovations. Pairs of innovations at locations  $i$  and  $j$  are used to estimate the error covariances. Assuming that the observation errors and forecast errors are not cross-

correlated, and that spatial correlations in the observation errors are much smaller-scale than in the forecast errors, we can write the following:

$$\begin{aligned} \langle d_i d_j \rangle &= \langle (\varepsilon_i^o - \varepsilon_i^f)(\varepsilon_j^o - \varepsilon_j^f) \rangle = \langle \varepsilon_i^o \varepsilon_j^o - \varepsilon_i^o \varepsilon_j^f - \varepsilon_i^f \varepsilon_j^o + \varepsilon_i^f \varepsilon_j^f \rangle \\ &= \begin{cases} \langle (\varepsilon_i^o)^2 \rangle + \langle (\varepsilon_i^f)^2 \rangle & \text{where } i = j \\ \langle \varepsilon_i^f \varepsilon_j^f \rangle & \text{where } i \neq j \end{cases} \end{aligned}$$

where the brackets  $\langle . \rangle$  denotes the expectation operator. So, under the above assumptions, the variance of the innovations is comprised of the sum of the observation error variance and the forecast error variance. The covariances of the innovations, for separation distances between locations  $i$  and  $j$  greater than zero, provides information about the forecast error covariances. In the HL method, parameters of a pre-defined function (here defined in equation (1)) are estimated by fitting the function to the forecast error covariance data with separation distances greater than zero. The forecast errors can then be extrapolated to the zero-separation point using that function to estimate the forecast error variance. The observation error variances can also be determined by subtracting the forecast error variance from the total variance of the innovations, though we focus here only on the forecast errors.

There are two aspects to the spatial resolution of the resulting error covariance estimates. One is the grid on which the estimates are accumulated. Here we use a  $1^\circ$  grid which was determined to allow enough observations to avoid issues with sampling noise, yet still allow us to estimate useful spatial variations in the covariance parameters. The second aspect is the resolution on which to accumulate the spatial covariance statistics, i.e. the separation distances. Again, we tested various options for these, and ended up using 18 bins with the distances for the global estimates of 10, 15, 20, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900 and 1000 km. For the estimates in the NWS region, 24 bins were used with the following distances of 10, 15, 20, 30, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 350, 400, 450, 500, 550, 600, 700, 800 and 900 km. It would be possible to estimate anisotropic aspects of the error covariances by looking in different directions when accumulating the error covariance statistics, but we do not do that here. Isotropic error covariances may be questioned in specific regions of the ocean (e.g. in the tropics or along land boundaries), but chi-squared tests were calculated to evaluate the goodness of the function fitting for each point of the grid. Although chi-squared values are slightly larger along the land boundaries, their values are very close to zero almost everywhere (not shown), indicating that the fitting of isotropic error covariances is well suited for the purposes of this report.

In terms of the observation types which we used for these estimates, there are several SST satellite instruments which we can choose from. We decided to use the VIIRS data for these calculations since they are very numerous, and have been shown to be unbiased (they are used as reference data in the variational bias correction scheme carried out as part of the data assimilation process). The HL method assumes that correlations in the observation errors are much smaller-scale than in the forecast errors. For high-resolution satellite data such as VIIRS, this condition is more likely to be achieved by thinning the observations (i.e. using a subset of available observations) or by calculating super-observations (i.e. averaging subsets

of available observations). As mentioned in section 2.4, the latter is applied to VIIRS data in the global configurations, but in higher resolution AMM15 no observation averaging is carried out. It is worth noting that only assimilated VIIRS observations are considered for the HL error covariance calculation. In order to understand whether the assumption of small spatial correlations in the observation errors is valid, HL error covariances estimated from VIIRS data are compared to those from in situ drifters (see section 4.1.1).

### 3.2 Forecasts of different lengths

The second method is based on forecasts of different lengths, both valid at the same time. If we denote the forecast of length  $n$  hours from an analysis at time  $m$  as  $\mathbf{x}_{m,n}^f$ , the difference of two forecasts at time  $T0 + 48$  each initiated from analyses at times  $T0$  and  $T0 + 24$  is

$$\mathbf{x}_{T0,48}^f - \mathbf{x}_{T0+24,24}^f = (\mathbf{x}_{T0+48}^t + \boldsymbol{\varepsilon}_{T0,48}^f) - (\mathbf{x}_{T0+48}^t + \boldsymbol{\varepsilon}_{T0+24,24}^f) = \boldsymbol{\varepsilon}_{T0,48}^f - \boldsymbol{\varepsilon}_{T0+24,24}^f,$$

where  $\mathbf{x}_{T0+48}^t$  is the true state at time  $T0 + 48$  and  $\boldsymbol{\varepsilon}_{m,n}^f$  is the forecast error growth over  $n$  hours from an analysis at time  $m$ . We aim to estimate forecast errors for a 24-hour forecast so choose forecast lengths 24 hours apart.

If we assume that, on average, the error growth over a two day forecast is some factor,  $\gamma$ , multiplied by the error growth over a one day forecast then  $\boldsymbol{\varepsilon}_{T0,48}^f \approx \gamma \boldsymbol{\varepsilon}_{T0+24,24}^f$ . If we also assume that  $\boldsymbol{\varepsilon}_{T0+24,24}^f$  has the same covariance as the 24-hour forecast errors  $\boldsymbol{\varepsilon}^f$ , then:

$$\begin{aligned} < (\mathbf{x}_{T0,48}^f - \mathbf{x}_{T0+24,24}^f)(\mathbf{x}_{T0,48}^f - \mathbf{x}_{T0+24,24}^f)^T > = < (\boldsymbol{\varepsilon}_{T0,48}^f - \boldsymbol{\varepsilon}_{T0+24,24}^f)(\boldsymbol{\varepsilon}_{T0,48}^f - \boldsymbol{\varepsilon}_{T0+24,24}^f)^T > \\ & \approx < (\gamma - 1)^2 \boldsymbol{\varepsilon}_{T0+24,24}^f \boldsymbol{\varepsilon}_{T0+24,24}^{fT} > \approx (\gamma - 1)^2 < \boldsymbol{\varepsilon}^f \boldsymbol{\varepsilon}^{fT} > = (\gamma - 1)^2 \mathbf{B}. \end{aligned}$$

We can therefore use the 24-hour forecast differences to estimate climatological statistics for the background error covariances. If there is approximately linear error growth in the forecast over these timescales, then  $\gamma$  will be close to 2 and the resulting statistics will be representative of  $\mathbf{B}$ . If there is a large growth in error over the first day and the growth over the second day is comparatively small then  $\gamma$  will be smaller (but always larger than one), and the magnitude of the resulting error covariance estimates will be smaller than the true  $\mathbf{B}$ . This method assumes that the biases in the 24-hour and 48-hour forecasts are similar. If the biases grow significantly over that period, then the method will also underestimate the overall error. The method also underestimates the magnitude of forecast errors in sparsely observed regions since the analysis at time  $T0+24$  will not be very different to the 24-hour forecast valid at that time, and the two sets of forecasts will not differ significantly. Due to these various assumptions, it is important for the NMC method to be combined with another one such as the HL method to ensure the magnitude of the error covariance estimates can be assessed, as we do here.

In the implementation of the NMC method here, we only do the spatial covariance estimation in the directions of the model grid. So at each location we do the calculations in the positive and negative  $i,j$  directions, and then average the resulting covariances before doing the

function fitting to produce isotropic estimates. It would be possible to also obtain the spatial structures of the covariances in these different directions separately to estimate some anisotropy (even if still aligned with the grid directions) but this is not done here. For ORCA025/ORCA12/AMM15 we chose to use 25/100/500 grid-points in each direction giving a maximum separation distance of about 700/900/750 km.

## 4. Results

### 4.1 Global error covariances

In this section we compare the forecast error covariances estimated by the HL and NMC methods for ORCA025 and ORCA12 systems. The seasonal aspects of the forecast errors and their respective length-scales are also evaluated based on error covariances estimated over 2017-2018. The seasons are defined here as follows: December, January and February (DJF); March, April and May (MAM); June, July and August (JJA); and September, October and November (SON). The HL results are generated using a common grid and a common specification of the separation distances for both model configurations.

#### 4.1.1 Comparison of HL estimates between VIIRS and drifter innovations

In this section the HL estimates from VIIRS high-resolution satellite data are compared with those estimated from in situ drifters. The similarities and discrepancies between these two innovation datasets are evaluated and possible reasons for the differences are also discussed, including the HL assumption about small-scale spatial correlations in the observation errors. This assumption must be carefully checked for satellite data since correlations between them may exist due to correlations along the scanning geometry and satellite swath, and due to correlated errors in the atmospheric correction of the satellite retrieval. ORCA025 and ORCA12 total background errors estimated from VIIRS innovations are compared in Fig. 4 to those estimated from drifter buoy innovations. The representation of the main background error features is consistent between the two innovation datasets. Both clearly show areas of increased SST background errors within boundary current regions and the Antarctic Circumpolar Current (ACC) for ORCA025 and ORCA12. Similar magnitudes of the background errors can also be found in the subtropical gyres and tropical regions, which is reinforced by the similar magnitudes of their zonal averages between 40°S and 40°N in Fig. 4e.

Nevertheless, differences in the SST background errors estimated from VIIRS and drifter innovations can be spotted in coastal upwelling regions, such as along the western coast of Africa, South and North Americas. Fig. 4e also shows that zonally averaged SST background errors estimated by the two innovation datasets diverge towards high latitudes, particularly

poleward of 40°S and 40°N. These are approximately the latitudes where the number of VIIRS and drifter observations begin to drop quite significantly (Fig. 5). This drop in the number of observations matches quite consistently the increasing poleward differences between the SST background errors estimated from VIIRS and drifters in Fig. 4e. As mentioned previously, the robustness of HL error covariance estimates relies on the number of observations available. If there are less VIIRS and drifter observations at high latitudes, one would expect greater uncertainties around the estimation of HL error covariances, and therefore more significant differences between the two innovation datasets are likely to appear there when compared to lower latitudes.

Another relevant point to mention is that we only consider night-time VIIRS observations. As a result, the number of VIIRS observations decreases at high latitudes in the summer hemisphere where the daytime is extremely long. This results in HL error covariance estimates that may be skewed towards winter results and therefore can be a factor contributing to explain the discrepancies seen at high latitudes in Fig. 4e. Although this is perhaps not as relevant as the points mentioned above, VIIRS and drifter measurements are done in different ways. While drifters directly measure temperatures at depths of around 30 cm, VIIRS SSTs are originally skin measurements. They need to be adjusted to represent the sub-skin through a thorough calibration with drifters, which is a step performed by the data producers.

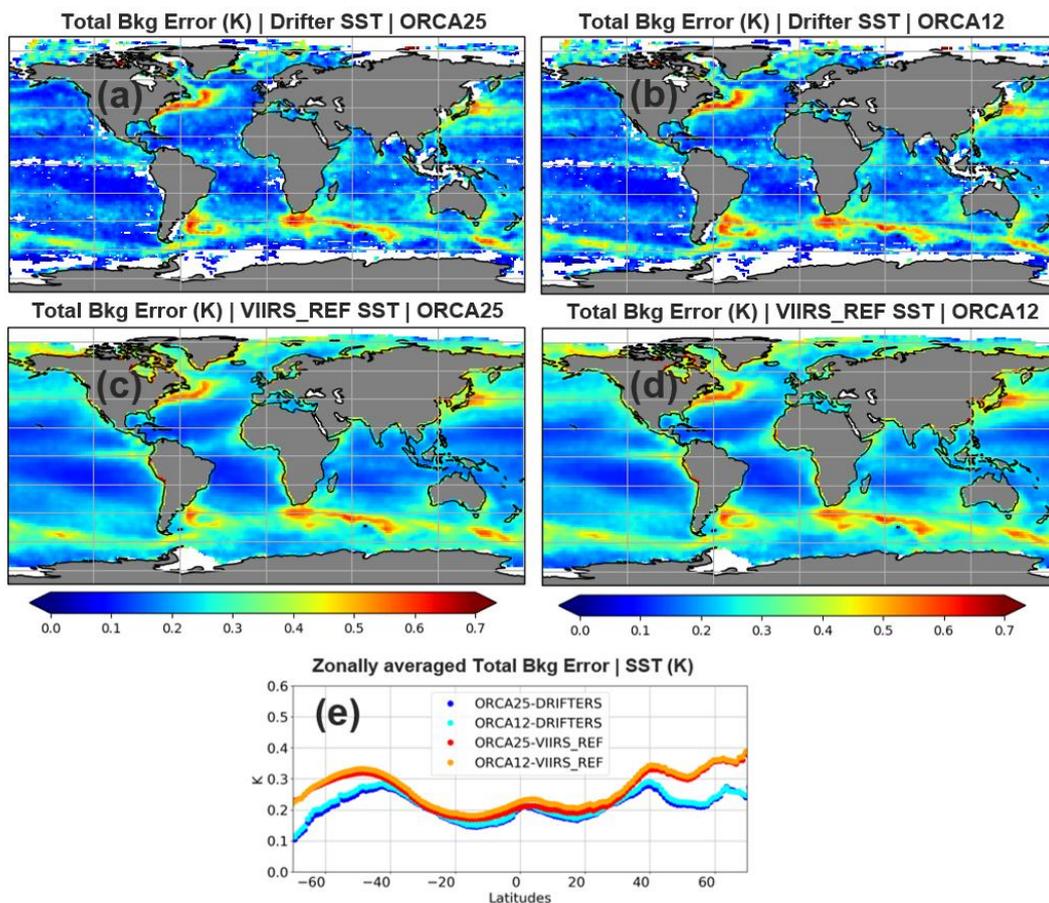


Figure 4: (a-d) 2017-2018 SST background errors (K) estimated by the HL method for drifters and VIIRS observations, considering both ORCA025 and ORCA12 systems, respectively. The zonal averages of the SST background errors for each global configuration and the respective observation types are shown in (e).

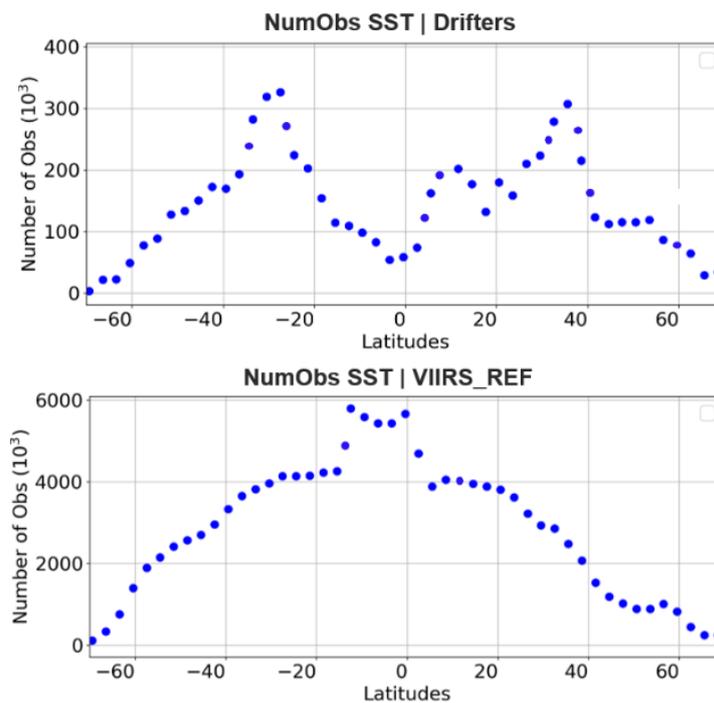


Figure 5: 2017-2018 number of SST observations ( $10^3$ ) integrated for each latitude for drifters (top) and VIIRS (bottom).

#### 4.1.2 Small-scale forecast error variances

After validating the HL SST background errors generated by VIIRS satellite data, the seasonal aspects of the HL as well as the NMC background errors and their differences in both ORCA025 and ORCA12 are evaluated. In order to understand the different characteristics of the forecast errors, we decompose them into a small-scale component associated with the mesoscale ocean variability and a large-scale component representing the background errors that can occur over the scale of atmospheric synoptic systems (see section 3 for more details). The magnitude of the former is shown for the summer and winter seasons in ORCA025 and ORCA12, estimated by HL (Fig. 6) and NMC (Fig. 7) methods. Regardless of the seasons, areas of increased small-scale background errors can be discerned in ocean regions with large SST gradients, such as the Brazil-Malvinas Confluence (BMC), Gulf Stream and Kuroshio current regions. This increase is also captured by the zonal averages of the small-scale errors in both NMC and HL methods (Fig. 8), which peak at mid-latitudes where these boundary currents are located.

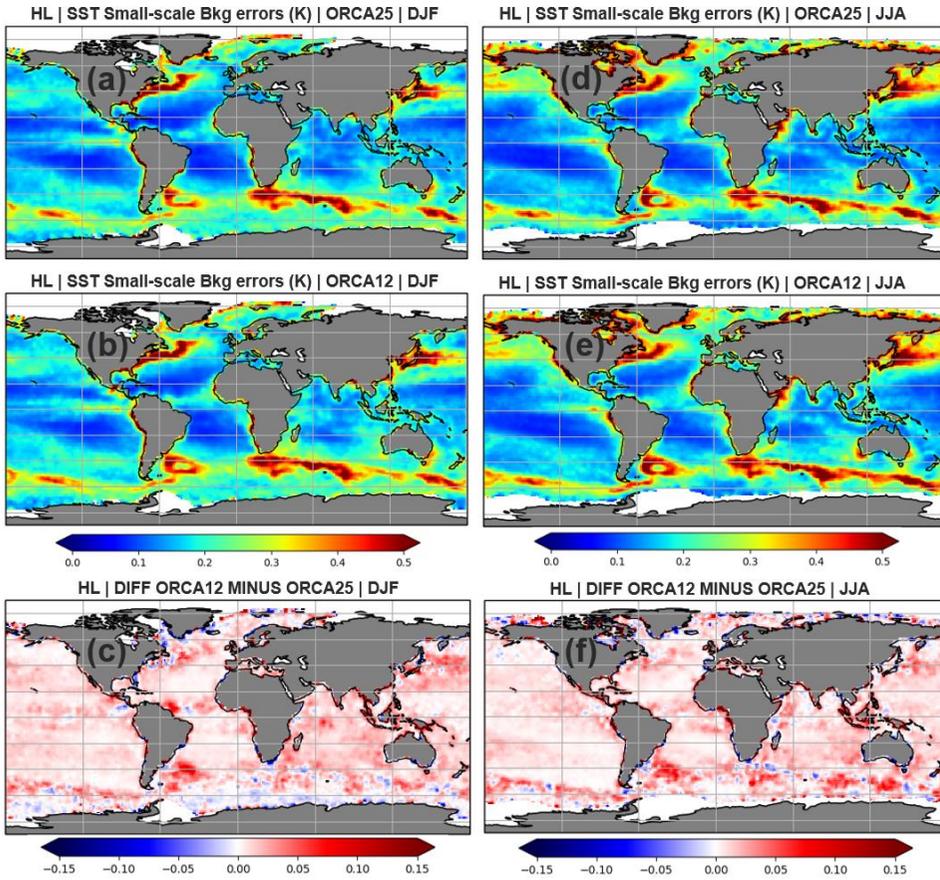


Figure 6: ORCA025 and ORCA12 small-scale SST background errors (K) estimated by the HL method using VIIRS observations for DJF (a-b) and JJA (d-e) seasons, respectively. The differences between ORCA12 and ORCA025 for DJF and JJA are shown in (c) and (f), respectively.

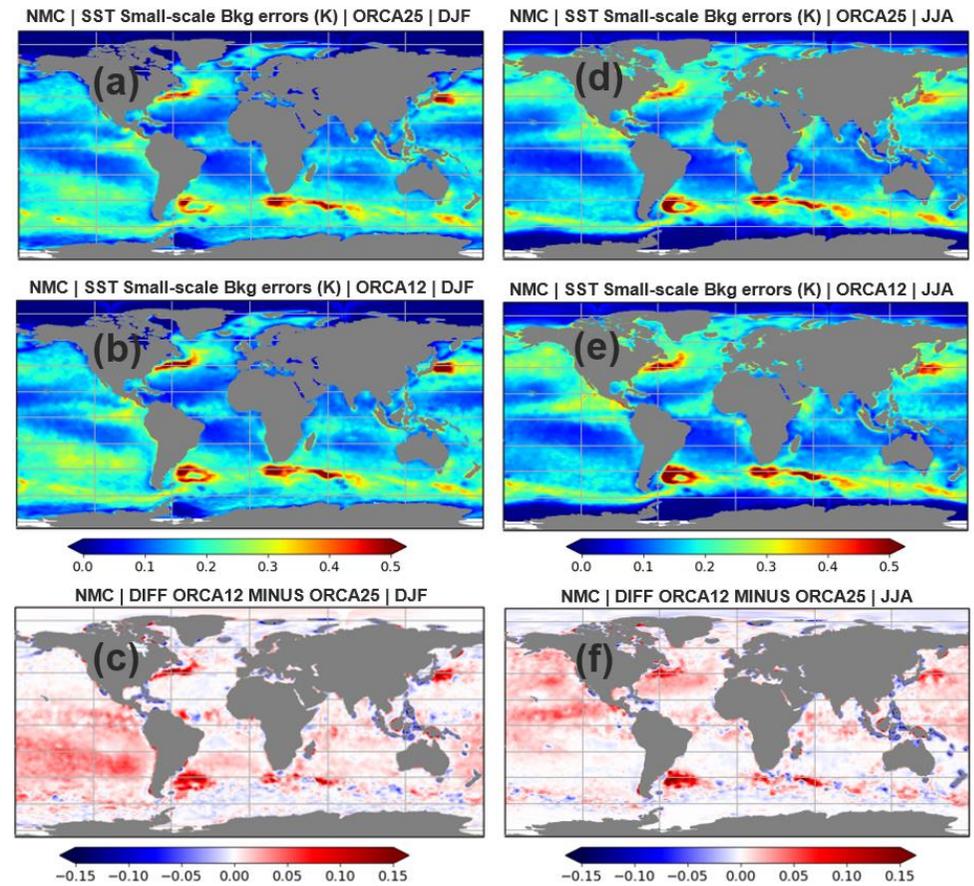


Figure 7: ORCA025 and ORCA12 small-scale SST background errors (K) estimated by the NMC method for DJF (a-b) and JJA (d-e) seasons, respectively. The differences between ORCA12 and ORCA025 for DJF and JJA are shown in (c) and (f), respectively.

Although boundary currents are persistent features throughout all seasons, Figs. 6, 7 and 8 show a clear seasonal pattern of increased small-scale background errors in the summer hemisphere, which is consistently seen in both error covariance methods and model configurations. Roberts-Jones et al. (2016) found similar patterns of seasonal variability in SST background errors estimated from the HL method. We agree with Roberts-Jones et al. (2016) that a possible explanation for this is that the mixed layer depth (MLD) in both hemispheres is shallower in summer months compared to winter. A shallower MLD in summer leads to more sensitivity in the mixed layer due to wind and surface heat fluxes, which in turn increase the SST variability. More localised seasonal patterns in the SST small-scale background errors are also clearly picked up by the different methods and model configurations in Figs. 6 and 7. For example, the Somali current in the Indian Ocean shows enhanced SST small-scale background errors in JJA, which is consistent with the seasonality caused by the Indian monsoon.

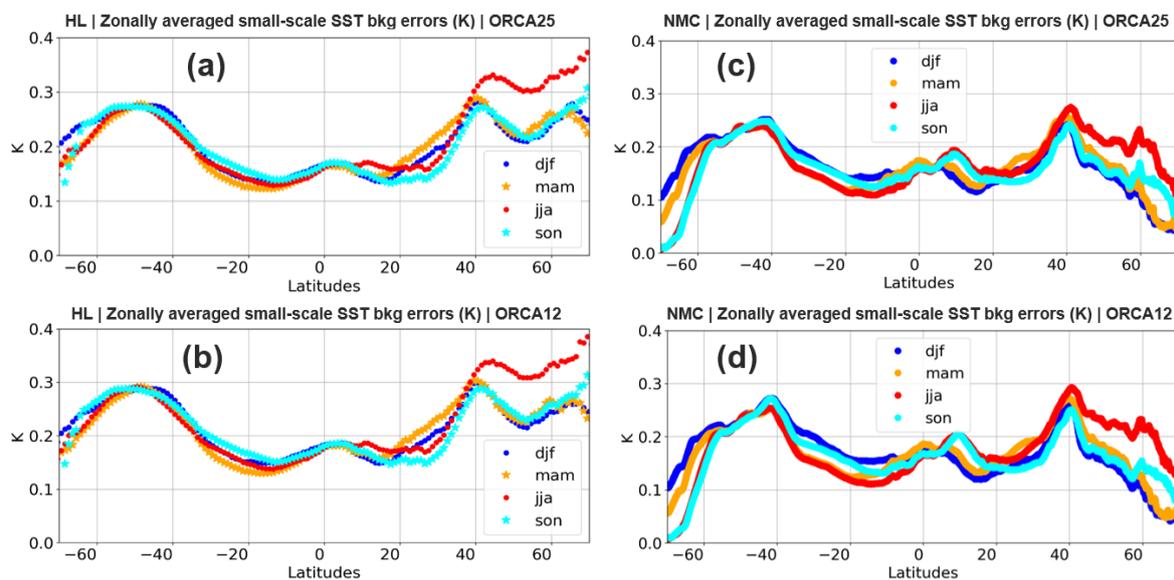


Figure 8: Zonal averages of HL (a-b) and NMC (c-d) small-scale SST background errors (K) for ORCA025 and ORCA12, respectively. The error covariances in the HL method are estimated using VIIRS observations. The zonal averages are calculated for DJF, MAM, JJA and SON seasons.

Even though similar seasonal patterns are reproduced across distinct model configurations, some clear differences can be seen between ORCA12 and ORCA025 in both HL and NMC methods for the small-scale background errors. In Figs. 6 and 7 ORCA12 produces larger SST small-scale errors than ORCA025 almost everywhere, and this is clearly significant within the boundary current regions for both DJF and JJA seasons. High-resolution models commonly produce forecasts with seemingly realistic small-scale patterns that can be slightly misplaced. Traditional point-matching verification measures would penalise such misplacements twice: for not having the pattern where it should be, and for having a pattern where there should not be one. For instance, when using neighbourhood methods for verification of the forecast errors, Aguiar et al. (2021) found that ORCA12 performed better than ORCA025 if double penalty effects are removed. The double penalty effects in ORCA12 SSTs might be explained

by the fact that ORCA12 resolves smaller scales, and it has larger SST gradients at those scales than ORCA025. While SST observations are very high resolution and can constrain those scales, the SLA observations are not as high resolution as the SSTs. Therefore, one would expect that SLA assimilation is not able to constrain the same SST small-scale features, and this might lead to an inability of the SST assimilation to properly constrain ORCA12 small-scale features in the forecast results. This issue is exacerbated by the fact that the assimilation in ORCA12 is performed on the ORCA025 grid, as described in Section 2.3. More results about the small-scale component of the SST forecast errors in ORCA12 are discussed in Section 4.1.5.

It is also interesting to note that differences in the small-scale background errors between ORCA12 and ORCA025 follow a clear seasonal pattern in the NMC estimates and are larger in summer compared to winter, particularly in the Pacific Ocean (Figs. 7c, f). The same pattern is not obvious in the HL estimates (Figs. 6c, f). A possible explanation for this could be related to diurnal cycle effects since the forecast difference for the NMC method is valid at 00 UTC, which corresponds to the middle of the day in the Pacific Ocean. Errors at the peak of the diurnal cycle are larger than other times of the cycle, and this becomes evident in summer where the diurnal cycle is larger. In the HL technique the innovations only consider night-time VIIRS observations and therefore the HL method only estimates the error covariances at night. This might explain some of the discrepancies in the seasonal differences between ORCA12 and ORCA025 given by HL and NMC methods in the Pacific Ocean.

#### 4.1.3 Large-scale forecast error variances

Figs. 9 and 10 show the large-scale component of ORCA025 and ORCA12 SST background errors estimated from the HL and NMC methods, respectively. As in the small-scale component of the SST errors, areas of increased large-scale errors are also found in the high SST gradient regions, although their magnitude are smaller. Some of the mesoscale variability in these regions may be accounted for in the large-scale error component, however as these same regions encompass the storm tracks the increased errors could be a true large-scale synoptic signal. The seasonal patterns in the large-scale component of the errors also closely follow those from the small-scale component, with increased magnitudes in the summer hemisphere. This seasonality is also reinforced by the zonal averages of the large-scale SST background errors in Fig. 11. Regions with enhanced large-scale background errors are also seen near India and Asia in JJA compared to DJF, which is consistent with the seasonal cycle caused by the monsoons in these places. It is also worth noting that the spatial distribution of the SST large-scale background errors is noisier compared to its small-scale counterpart. This is probably due to the greater sensitivity of the function fitting to the large-scale component – the signal is smaller at longer separation distances resulting in a smaller signal-to-noise ratio there. Another explanation may be that the correlation structure is more anisotropic for the large length-scales and therefore more difficult to fit with an isotropic function.

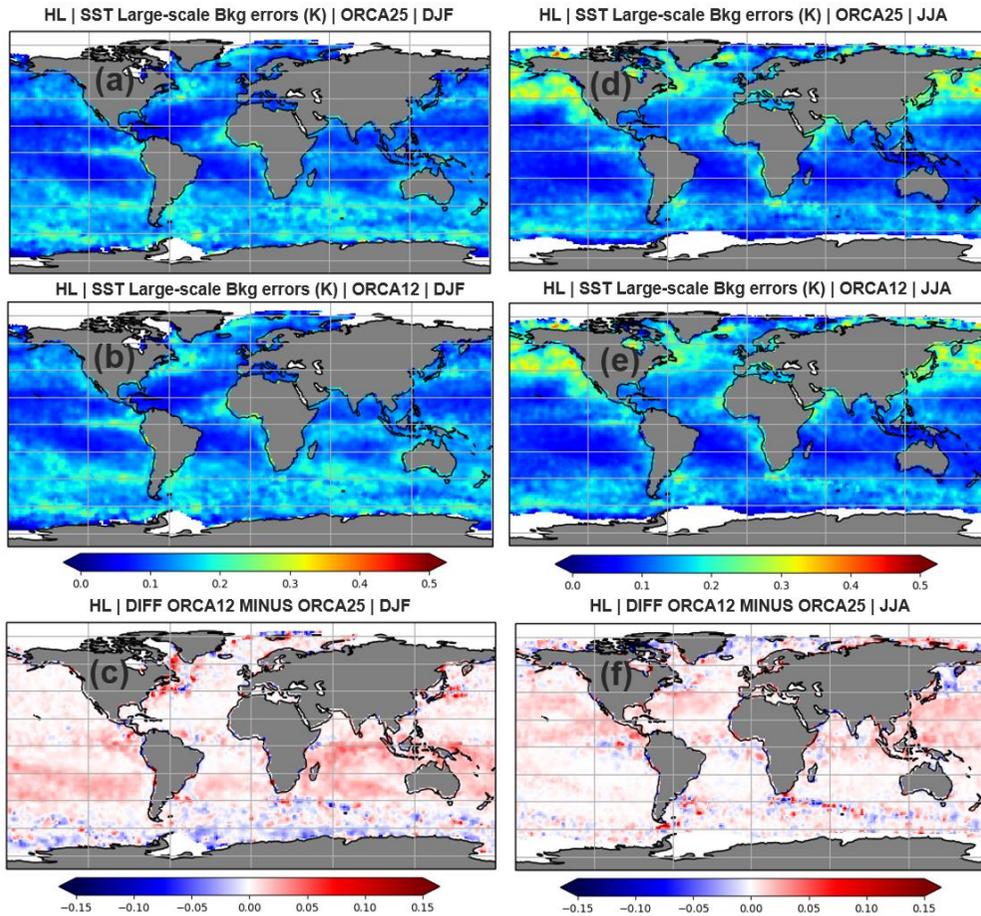


Figure 9: ORCA025 and ORCA12 large-scale SST background errors (K) estimated by the HL method using VIIRS observations for DJF (a-b) and JJA (d-e) seasons, respectively. The differences between ORCA12 and ORCA025 for DJF and JJA are shown in (c) and (f), respectively.

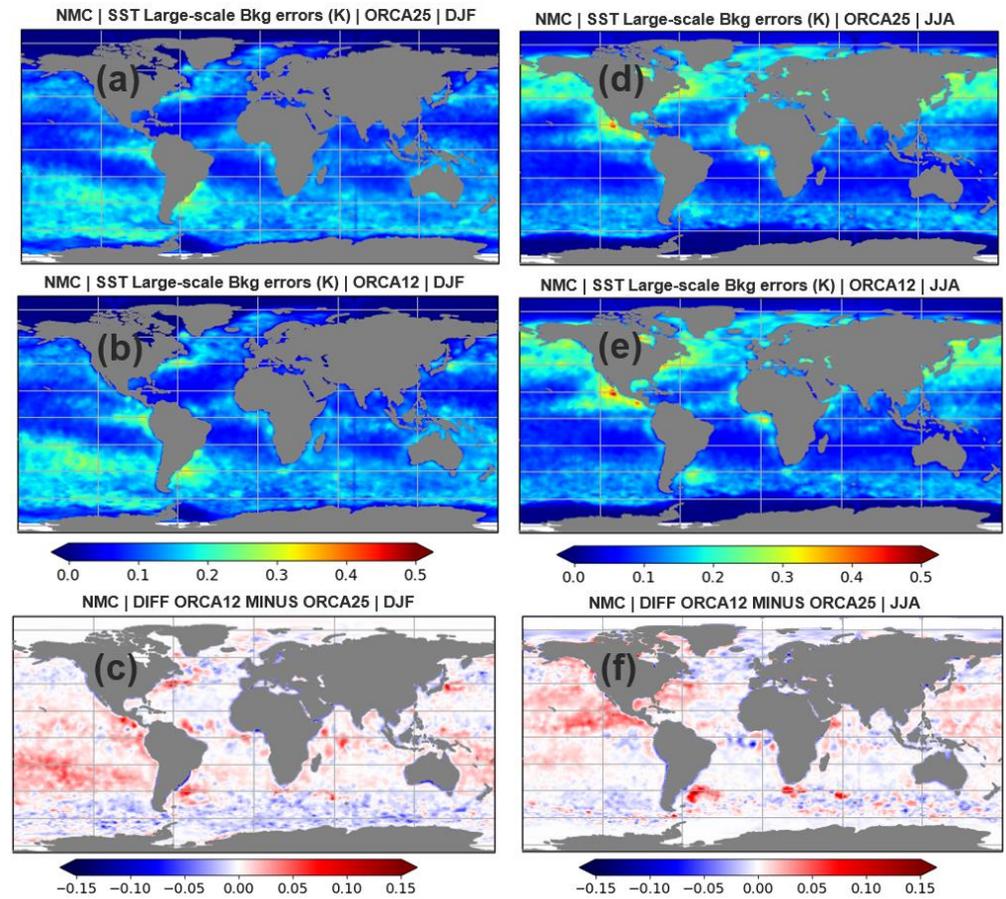


Figure 10: ORCA025 and ORCA12 large-scale SST background errors (K) estimated by the NMC method for DJF (a-b) and JJA (d-e) seasons, respectively. The differences between ORCA12 and ORCA025 for DJF and JJA are shown in (c) and (f), respectively.

Analogous to its small-scale component, ORCA12 also shows larger SST large-scale background errors than ORCA025 at mid-latitudes in the summer hemisphere with both NMC and HL methods. This is perhaps surprising as both systems are forced by the same atmospheric forcing, as well as having the same vertical levels and mixing schemes. However, clear improvements in the large-scale component of SST errors are seen for ORCA12 in the southern and northern high latitudes, particularly in the NMC estimates. This is consistent with results found by previous works intercomparing models and ocean reanalyses, which reported improvements in the large-scale circulation at high latitudes when increasing the model resolution (e.g., Jackson et al., 2019).

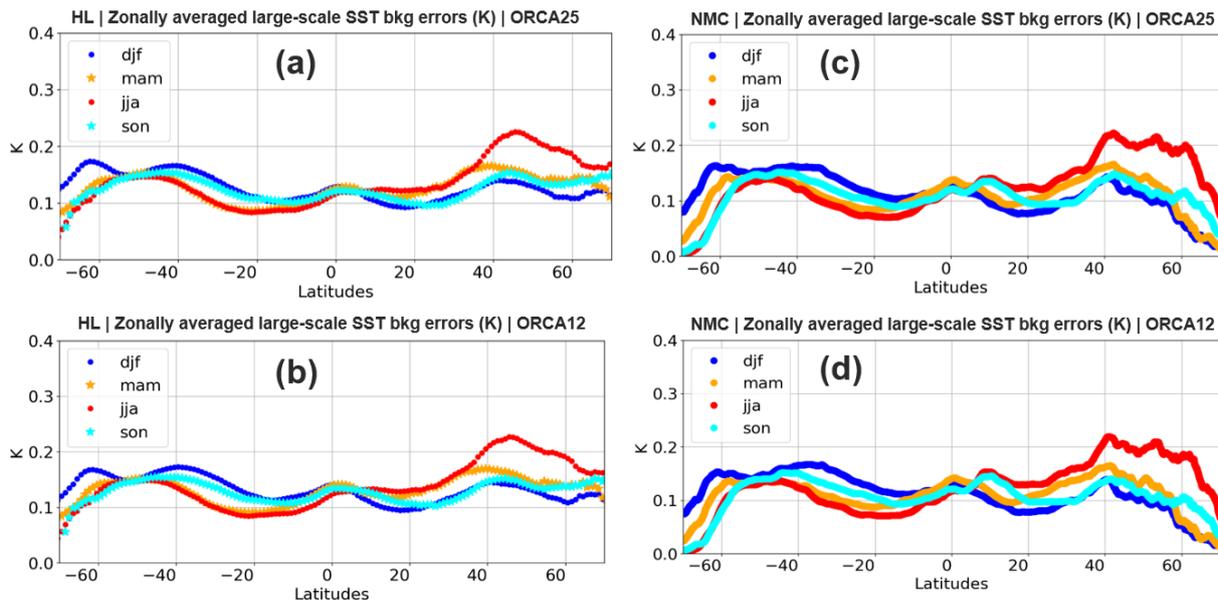


Figure 11: Zonal averages of HL (a-b) and NMC (c-d) large-scale SST background errors (K) for ORCA025 and ORCA12, respectively. The error covariances in the HL method are estimated using the VIIRS observations. The zonal averages are calculated for DJF, MAM, JJA and SON seasons.

#### 4.1.4 Forecast error length-scales

In order to assess the correlation length-scales of the SST background error components, the small and large length-scales are computed for ORCA025 and ORCA12 by the HL (Fig. 12) and NMC (Fig. 13) methods. Both Fig. 12 and Fig. 13 estimate the correlation length-scales based on the entire 2017-2018 period. Although the SST small-scale background errors show relatively similar magnitudes across both methods, their associated length-scales are much larger in NMC than in HL, particularly within the tropical and subtropical regions. This is reinforced by the zonal averages of the small and large length-scales given by each method for ORCA025 and ORCA12 in Fig. 14. Regardless of the model resolution, the small length-scales rise to approximately 100 km near the equator in NMC, whereas for HL the rise of the SST length-scales is less pronounced, reaching up to approximately 60 km in the tropics. This can be explained by the way in which the error covariances are estimated in both methods. While HL focus on the innovations, the NMC method takes the difference between 48 h and 24 h forecasts. Therefore, one would expect the assimilation information to spread out further, leading to larger scales in NMC, particularly in the tropics where the equatorial waves can

quickly propagate the assimilation increments. In contrast, the large length-scales look more similar between the HL and NMC methods. Both methods estimate large length-scales which are within 300-400 km for ORCA12 and ORCA025, although the NMC length-scales do get larger variations as a function of latitude, particularly at high latitudes (Fig. 14).

As well as seen in the short length-scales, ORCA12 also has longer length-scales than ORCA025 estimated by the NMC method, except for the regions near the poles (Fig. 13f). It is a common issue reported by previous works that the NMC method overestimates covariances, particularly at large scales (e.g., Bannister et al., 2008; Berre et al., 2006). One of the causes for this overestimation is related to the fact the NMC method uses longer (48 and 24 h) forecast differences. The reasons for using forecasts staggered by 24 h is to eliminate errors in modelling the diurnal cycle, but according to Bannister et al. (2008) the need for these particular forecast lengths is not necessary and other forecasts lengths can be tested, such as 36 h and 12 h forecasts or even 30 h and 6 h forecasts as in Berre et al. (2000). The reason why we have chosen the minimum forecast length to be 24 h is to avoid contamination of the covariance statistics by assimilation shocks that might happen when the forecast is initialised.

In terms of seasonal variations in the small and large length-scale components, Fig. 15 shows the zonal averages of ORCA12 and ORCA025 length-scales for each season. Regardless of the method used, the large length-scales do not show a clear seasonal pattern in ORCA025 and ORCA12. However, it is worth noting that HL small length-scales in the autumn hemisphere are smaller than other seasons around 60° of latitude, reaching a magnitude of approximately 200 km. This value is similar to the large length-scales found in the NMC method for the same latitudinal band. Unlike its counterpart, the small length-scales are always larger in summer compared to winter, having a seasonal pattern with similar characteristics of the small-scale background errors, which are also greater in the summer hemisphere. Given our previous explanation that the shallowness of the MLD in summer might make the SSTs more sensitive to the wind and surface heat fluxes, one would expect the variability in atmospheric fluxes not only to increase the SST error covariances but also the length-scales associated with them.

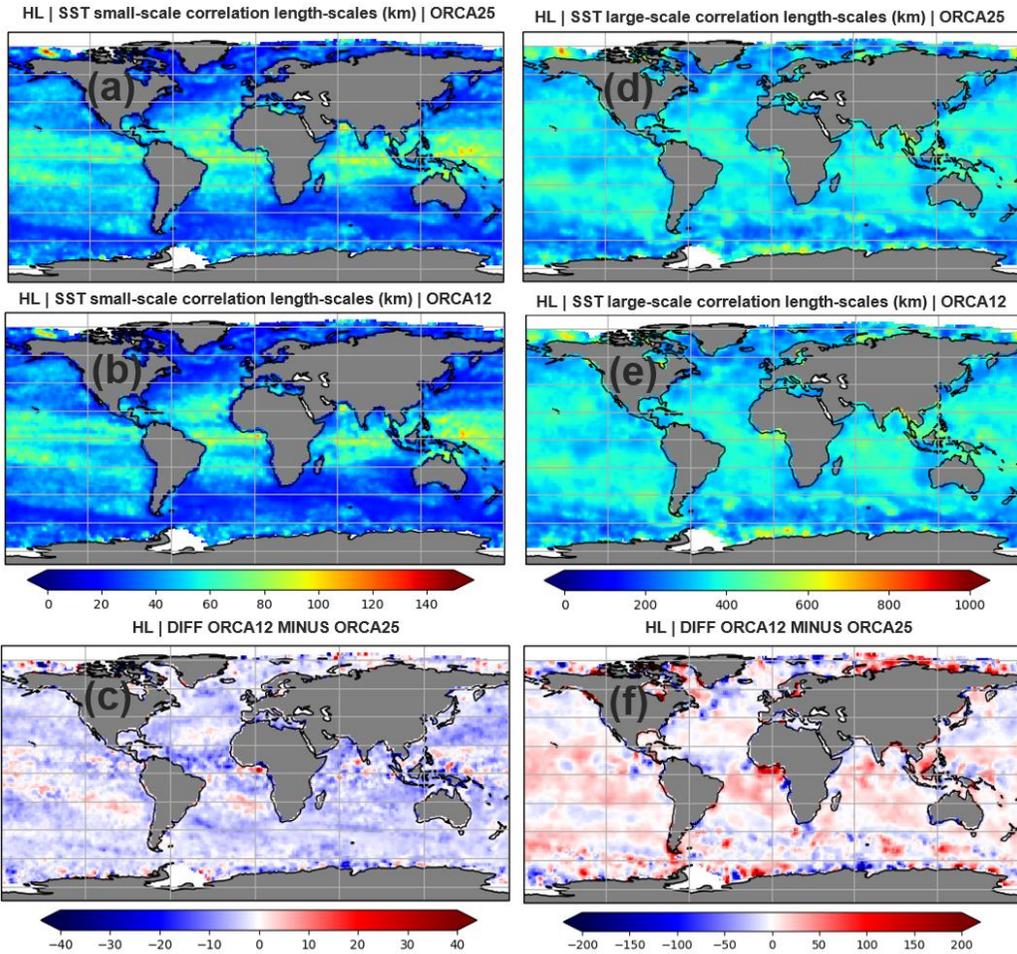


Figure 12: 2017-2018 ORCA025 (a, d) and ORCA12 (b, e) small and large-scale SST correlation length-scales (km) estimated by the HL method using VIIRS observations, respectively. The differences between ORCA12 and ORCA025 for the small and large SST length-scales are shown in (c) and (f), respectively.

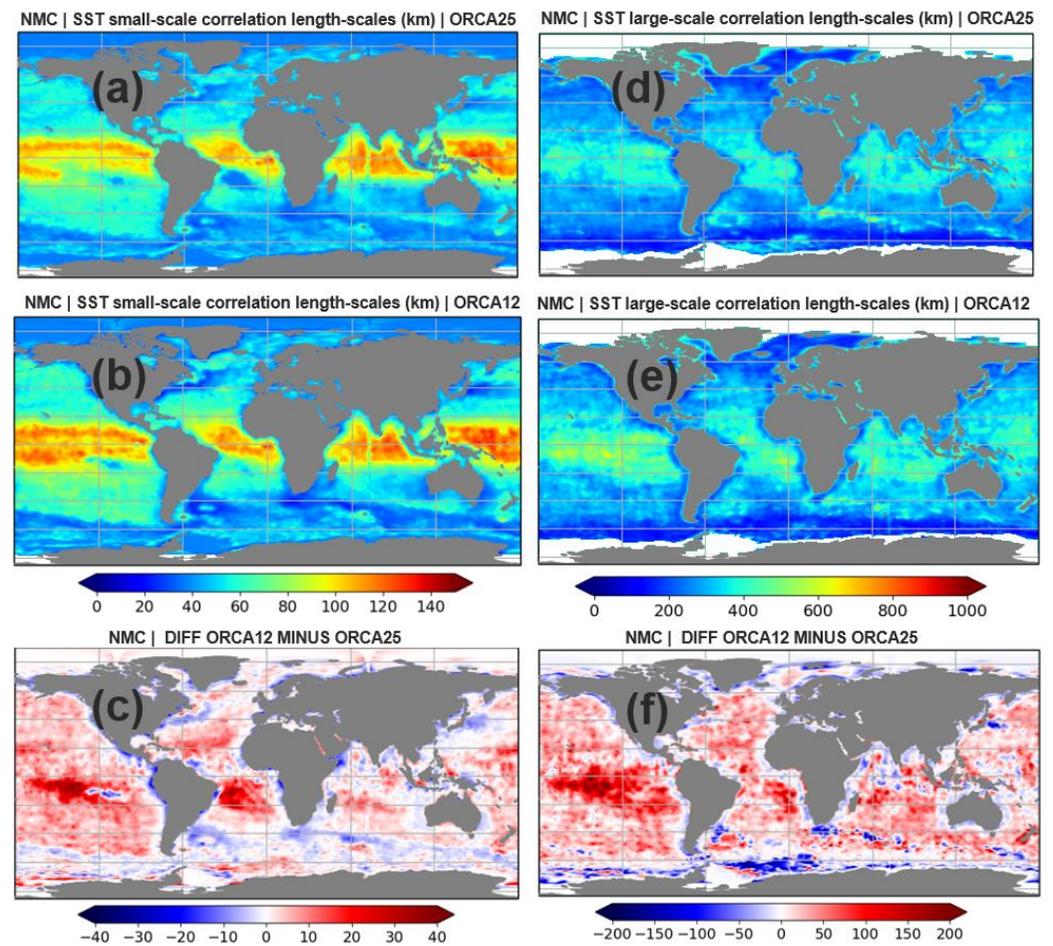


Figure 13: 2017-2018 ORCA025 (a, d) and ORCA12 (b, e) small and large-scale SST correlation length-scales (km) estimated by the NMC method, respectively. The differences between ORCA12 and ORCA025 for the small and large SST length-scales are shown in (c) and (f), respectively.

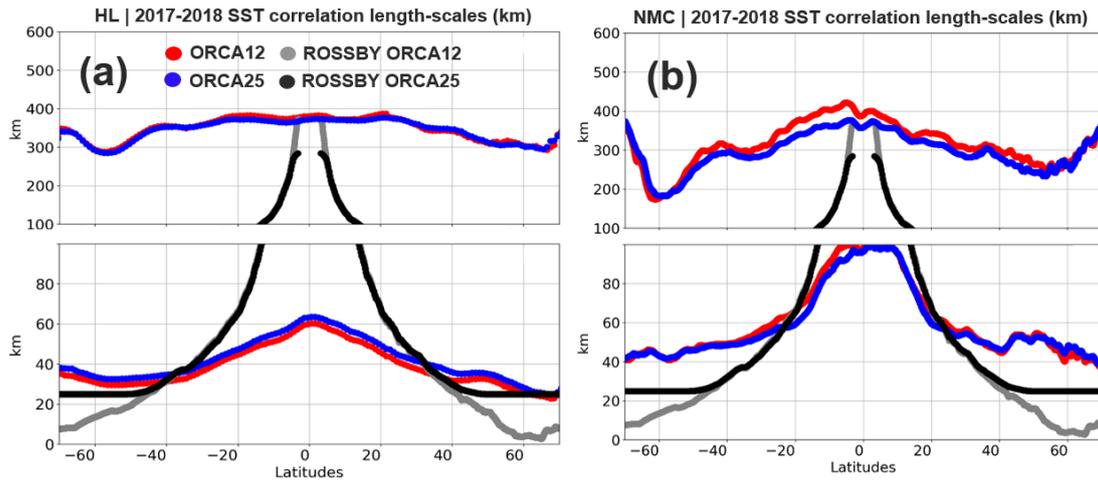


Figure 14: 2017-2018 HL (a) and NMC (b) zonal averages of small and large-scale SST correlation length-scales (km) for ORCA025 and ORCA12. The error covariances in the HL method are estimated using VIIRS observations.

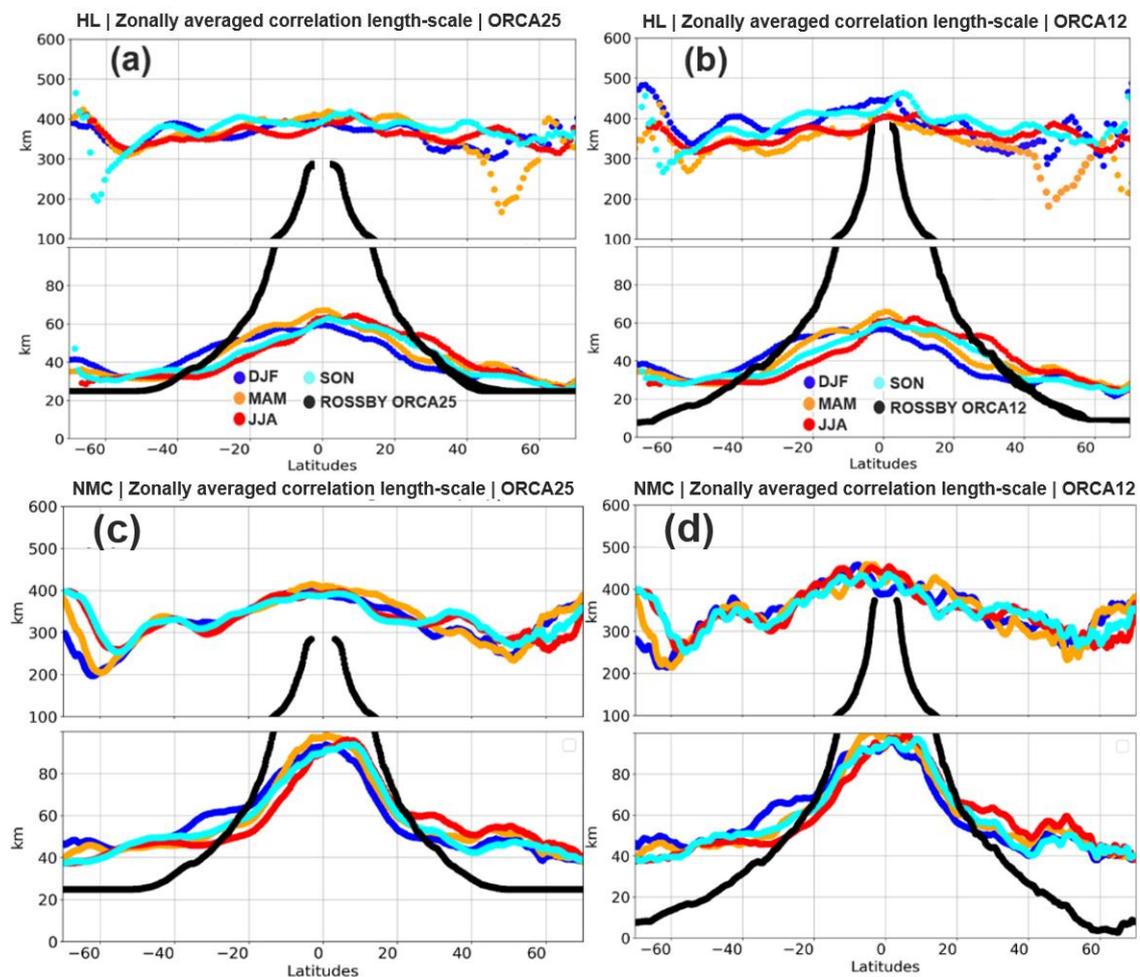


Figure 15: Zonal averages of HL (a-b) and NMC (c-d) correlation length-scales (km) for ORCA025 and ORCA12, respectively. The error covariances in the HL method are estimated using VIIRS observations. The zonal averages are calculated for DJF, MAM, JJA and SON seasons.

#### 4.1.5 Forecast error covariances at short separation distances

High-resolution systems are expected to better resolve regions with high mesoscale variability, such as the boundary currents. Both HL and NMC methods show that ORCA12 can represent smaller SST scales in the boundary currents, however they are not properly constrained by assimilation since the forecast errors associated with those scales increase in ORCA12 relative to ORCA025. This is reinforced by the HL (Fig. 16) and NMC (Fig. 17) function fitting of the SST error covariances for points located at the Gulf Stream and BMC. In both methods and points ORCA12 shows a sharp decay in the errors close to the zero-separation distance, however their magnitudes are larger compared to ORCA025.

In order to illustrate the global patterns of what we mentioned above, Figs. 18 and 19 show the spatial distribution of the errors for the first few bins of separation distance. In the case of the HL method (Fig. 18), the bins of separation distance are the same between the model configurations and we have chosen to represent the error covariances from the separation distances of 10, 20 and 100 km. It is clear how quickly the differences in the errors between the model configurations decrease within 100 km of separation distance, with most of the differences being concentrated within 20 km. The errors in ORCA12 are usually larger than in ORCA025 within separation distances of 20 km, particularly in the western boundary currents. This is likely to be the reason why the SST small-scale background errors are larger in ORCA12, as consistently seen in Figs. 6 and 7.

Since the NMC method is dependent on the model resolution, in Fig. 19 we have chosen equivalent bins of separation distance between model configurations to compare their error covariances. For ORCA12, we chose the first, third and twelfth bins of separation distance, which represent roughly 8, 25 and 100 km at the equator. For simplicity, the separation distances of ORCA12 and ORCA025 are defined here considering equatorial distances, since the model resolution varies with latitude (see Fig. 2). Since ORCA025 cannot represent distances smaller than 25 km at the equator, the selected bins for this system were the first (25 km) and fourth ones (100 km). As in the HL estimates, the model errors in ORCA12 are generally larger than in ORCA025 within 25 km of separation distance, especially in the Gulf Stream, BMC and Kuroshio current. Looking at ORCA12 individually, it is seen that the model errors almost double in the boundary current regions going from 25 to 8 km of separation distance. This reinforces the ORCA12 assimilation issues in constraining the forecast errors on such small scales.

In contrast to the HL method, the NMC results show a persistent signature of larger errors in ORCA12 in the Pacific Ocean compared to ORCA025 (Fig. 19). This is still visible within 100 km of separation distance. As discussed previously, the NMC forecast differences are valid at 00 UTC, which corresponds to the middle of the day in the Pacific Ocean and therefore to the peak of the diurnal cycle there. So, it is expected that SST model errors will be larger in the Pacific, but somehow this seems to affect more ORCA12 than ORCA025 in the NMC results.

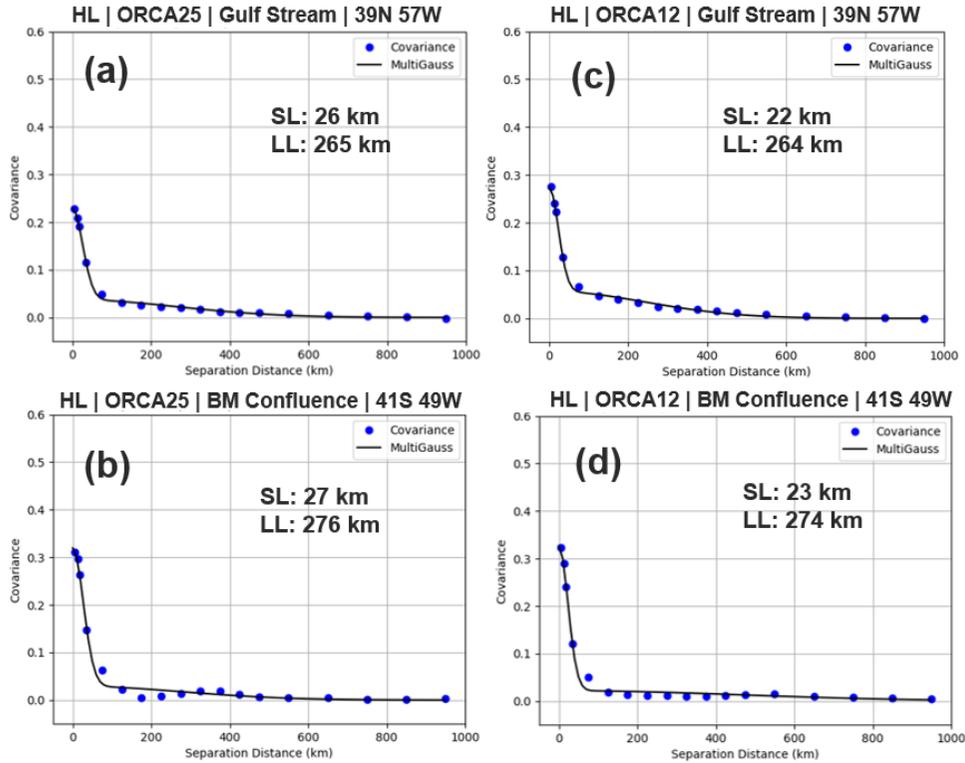


Figure 16: ORCA025 (a-b) and ORCA12 (c-d) HL SST function fitting for two points located at the Gulf Stream and the Brazil-Malvinas confluence, respectively. The blue dots represent the SST error covariances ( $K^2$ ) estimated from VIIRS observations as a function of the separation distance, whereas the black line corresponds to the Multi-Gaussian function fit.

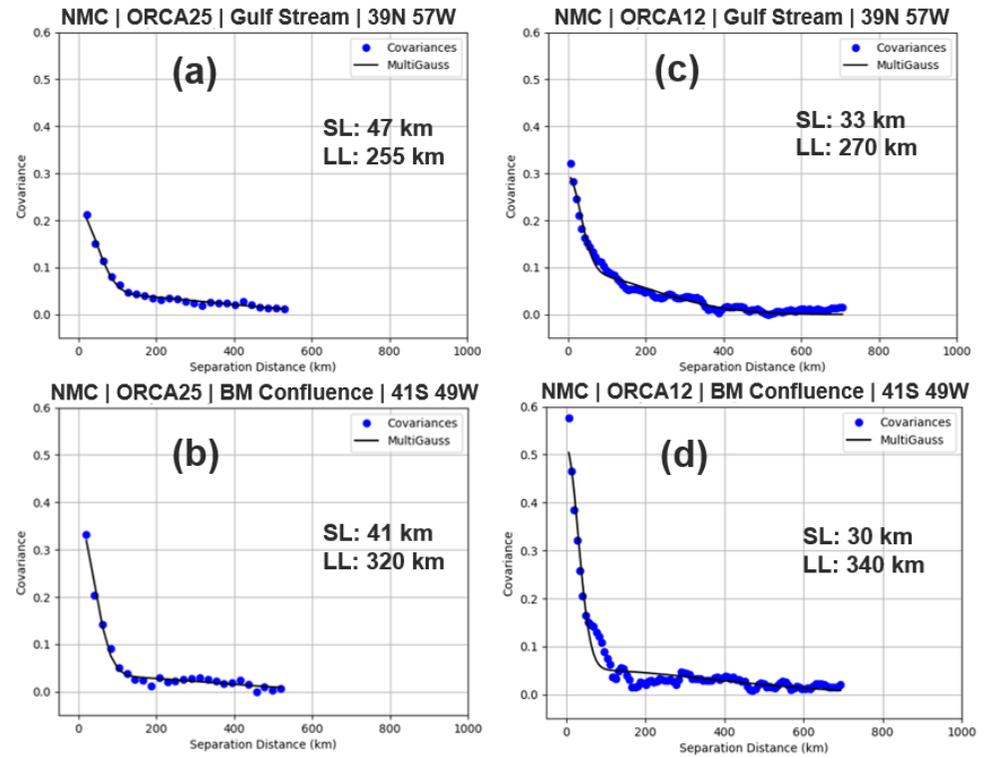


Figure 17: ORCA025 (a-b) and ORCA12 (c-d) NMC SST function fitting for two points located at the Gulf Stream and the Brazil-Malvinas confluence, respectively. The blue dots represent the SST error covariances ( $K^2$ ) estimated as a function of the separation distance, whereas the black line corresponds to the Multi-Gaussian function fit.

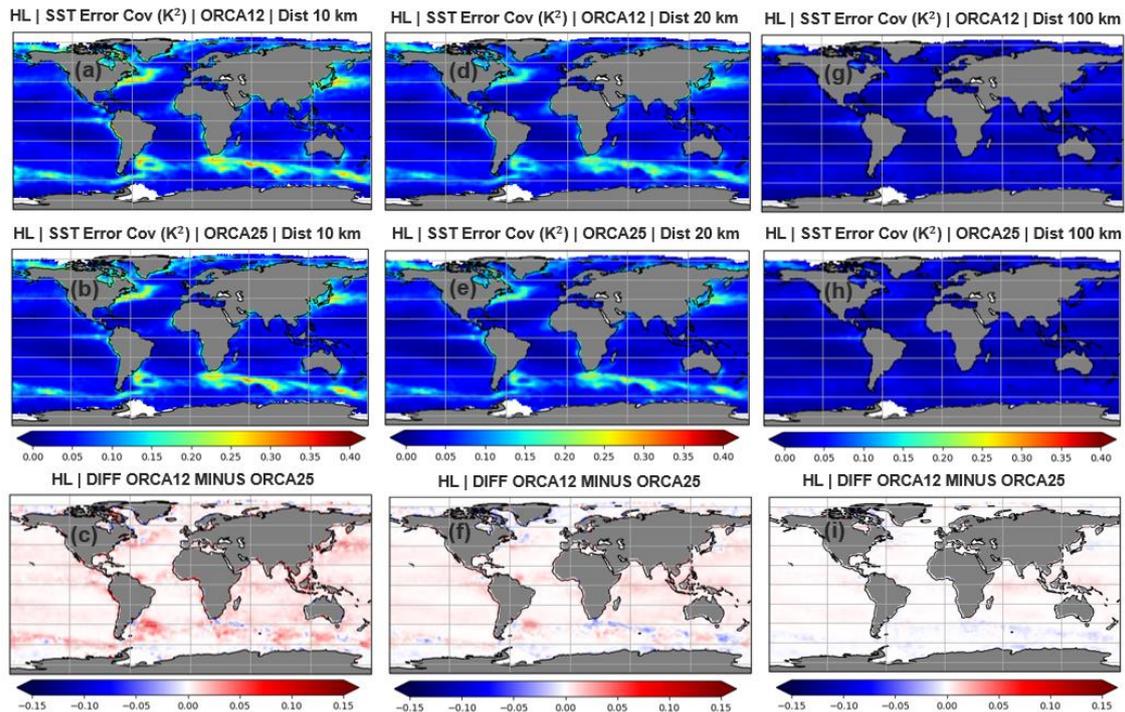


Figure 18: ORCA025 and ORCA12 HL SST error covariances ( $K^2$ ) for the separation distances of 10 km (a-b), 20 km (d-e) and 100 km (g-h), respectively. The SST error covariance difference between ORCA12 and ORCA025 for the separation distances of 10, 20 and 100 km are shown in (c), (f) and (i), respectively.

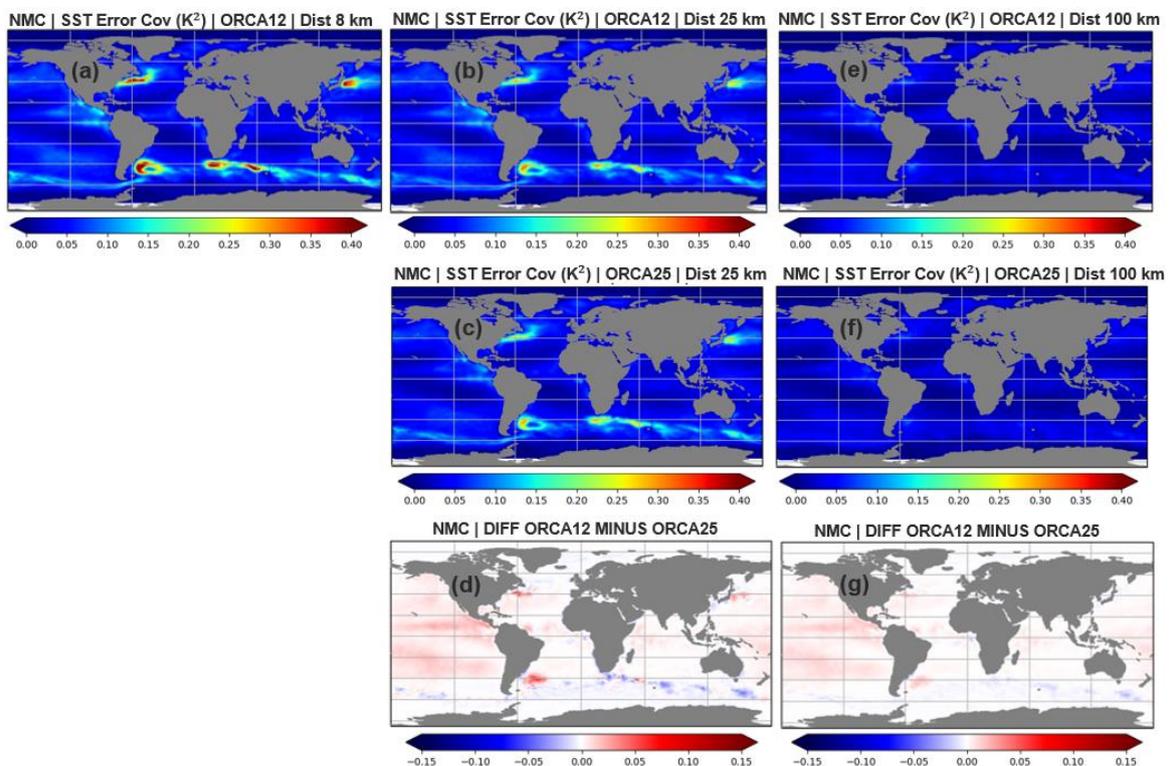


Figure 19: ORCA025 and ORCA12 NMC SST error covariances ( $K^2$ ) for the separation distances of approximately 25 km (b-c) and 100 km (e-f), respectively. The SST error covariance difference between ORCA12 and ORCA025 for the separation distances of 25 and 100 km are shown in (d) and (g), respectively. In (a) the separation distance of 8 km is shown for ORCA12 only due to its higher resolution. It is important to note that the NMC separation distances are defined at the equator for both model configurations.

## 4.2 North-west European shelf-seas error covariances

In this section we compare the forecast error covariances estimated in the NWS region for the two high-resolution AMM15 systems with those from the global ORCA025 and ORCA12 systems. We do this for the error covariances estimated over 18 months from Jan 2017 to June 2018 (and use the same period for all experiments). The HL results are generated using a common grid and common specification of the separation distances for all configurations. In the figures below the edge of the shelf-break region is highlighted by a black dashed line which marks the 300 m isobath.

### 4.2.1 Small-scale forecast error variances

The magnitude of the small-scale background errors (see Figs. 20 and 21) shows a similar pattern in all the systems with larger errors near the coast, smaller errors in the middle of the North Sea and smaller errors in the off-shelf regions. In terms of the impact of the surface forcing on the small-scale errors, there is a slightly larger error when using the Met Office forcing compared with the system which uses the ECMWF forcing, with a more marked increase in errors near the coasts when using the Met Office forcing. One explanation for this could be the way in which the Met Office atmospheric model defines the proportion of land and sea within a grid point. This is not very sophisticated in the atmosphere-only system and might result in larger errors near the coast. This is improved in the coupled NWP system currently being developed for operational implementation (Guiavarc'h et al., 2019) which has a more sophisticated coastal tiling scheme and should result in improved forcing fields around the coasts.

As expected, the ORCA12 and ORCA025 configurations both have larger small-scale errors than the AMM15 configuration over most of the domain. The main exceptions to this are very close to the coasts of Ireland, Wales, Scotland and north-east England as well as in the German Bight, Skagerrak and Kattegat regions at the eastern edge of the domain. The change in behaviour near the coast could be related to the inclusion of tides in the AMM15 configurations which leads to higher variability in those regions and is perhaps not so well constrained by the assimilation. Errors in the Baltic boundary condition could also be contributing to the increase in errors in the eastern part of the domain. Overall, the increase in resolution and the inclusion of additional process in the AMM15 model are contributing to lower small-scale errors.

The results of the NMC method (Fig. 21) generally support those of the HL method and provide additional fine-scale information, although the overall magnitude of the errors is lower in the NMC results. The NMC method allows us to see an increase in the errors in the AMM15 system in the Norwegian Trench compared with ORCA12 and ORCA025. It also highlights an increase in error in the AMM15 system in the region of the Flamborough Head tidal front in the central North Sea. This increase in error is likely due to the large gradients associated with the front, which is present in the AMM15 system, but not captured at all by ORCA12 and ORCA025 which do not include the tidal processes needed to represent it.

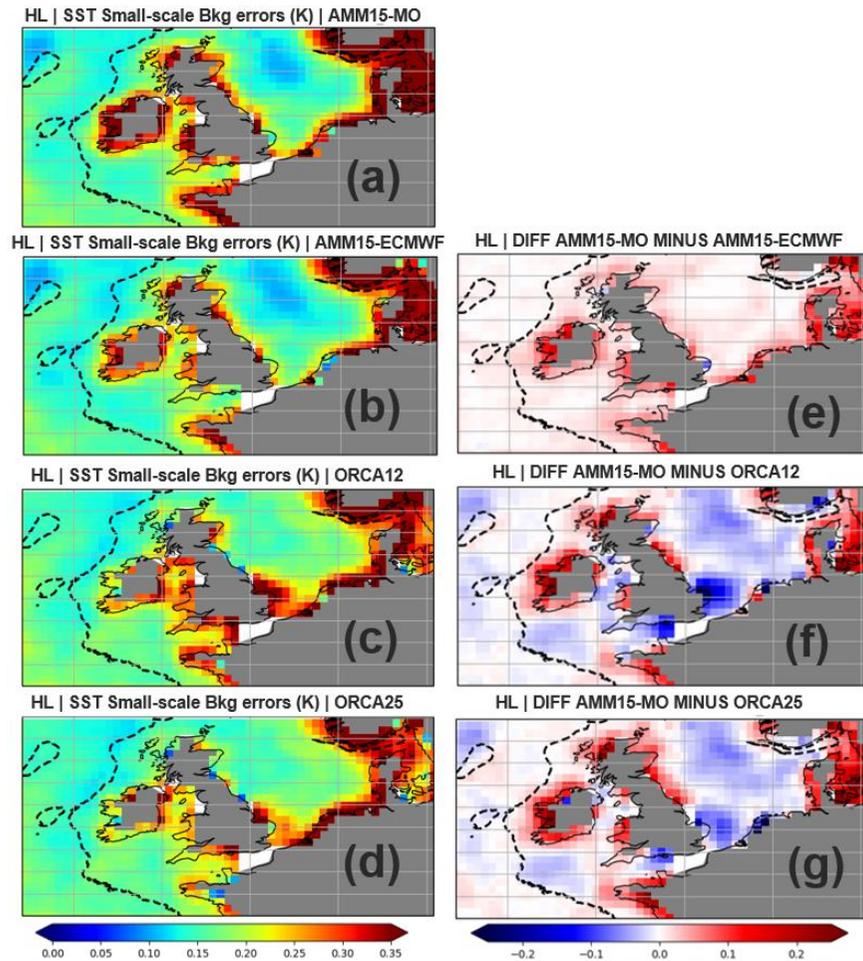


Figure 20: (a-d) 201701-201805 small-scale SST background errors (K) estimated by the HL method using VIIRS observations for AMM15-MO, AMM15-ECMWF, ORCA12 and ORCA025, respectively. The differences between AMM15-MO and the other model configurations are shown in (e-g). The black dashed contour represents the 300 m isobath.

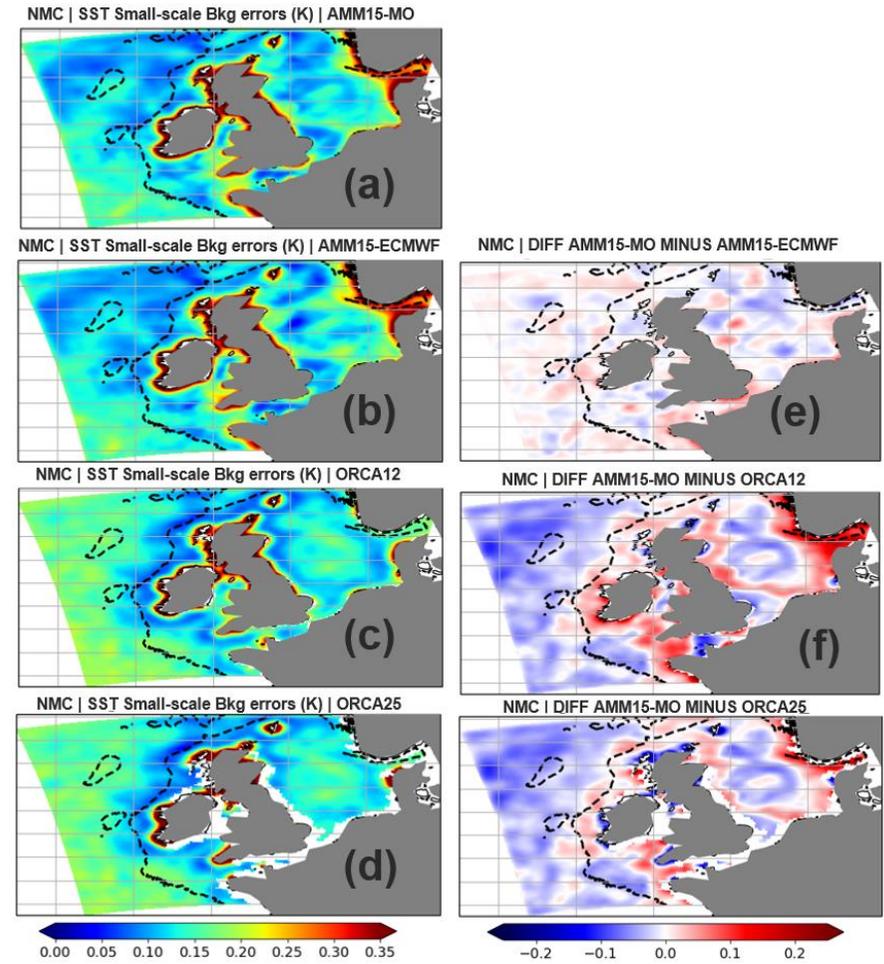


Figure 21: (a-d) 201701-201805 small-scale SST background errors (K) estimated by the NMC method for AMM15-MO, AMM15-ECMWF, ORCA12 and ORCA025, respectively. The differences between AMM15-MO and the other model configurations are shown in (e-g). The black dashed contour represents the 300 m isobath.

### 4.2.2 Large-scale forecast error variances

The magnitude of the large-scale errors seen by the HL method (Fig. 22) is generally slightly larger over most of the on-shelf region for the AMM15 experiment using the Met Office fluxes compared with the one forced using ECMWF fluxes. The only exception to this is the German Bight and Kattegat/Skagerrak regions where the Met Office forcing produces smaller errors at the large scale. The NMC method gives quite different results for the magnitude of the large-scale errors (Fig. 23). It shows the Met Office forcing to have lower errors than the ECMWF forcing over most of the domain. The cause of this discrepancy is not obvious and merits further investigation.

In most of the on-shelf part of the domain the AMM15 configurations have larger magnitude errors for the large-scale component than the ORCA12 and ORCA025 configurations. This could be due to the inclusion of tides in the AMM15 system which could increase the overall level of error at the large-scale. Another explanation could be the different specification of the error covariances used in the data assimilation systems for AMM15 and the global systems: the long length-scale specified in AMM15 is 100 km which is shorter than the one specified in the global systems (400 km) and shorter than the length-scales estimated in this work (see below). In contrast, the off-shelf region shows smaller magnitude large-scale errors for the AMM15 than the global configurations.

### 4.2.3 Small-scale forecast error length-scales

In terms of the small length-scales (Figs. 24 and 25), these are much smaller for both AMM15 configurations than for ORCA12 and ORCA025. As might be expected of a system with much higher horizontal resolution, the AMM15 is able to represent smaller-scale features and its dominant errors are on those smaller-scales. The NMC method (Fig. 25) shows that the scales are shorter in AMM15 particularly around the coasts, with larger scales along the shelf-break and in the central North Sea. The scales in the central North Sea are spatially quite variable.

The variations in the length-scales are consistent in both the HL and NMC results, although the NMC method produces larger length-scales, particularly along the shelf-break, the central North Sea and near the Rockall Plateau, while it produces smaller scales along the Flamborough Head front and the Norwegian Trench. The ORCA12 and ORCA025 scales are much less spatially variable than the AMM15 ones.

The source of the atmospheric forcing has relatively little systematic impact on the short length-scales of the AMM15 configuration, although the Met Office forcing produces slightly increased length-scales overall.

#### 4.2.4 Large-scale forecast error length-scales

For the HL method, the larger length-scales are fairly constant over the NWS region except for increases around some coasts and decreases in the central and eastern North Sea and the Norwegian Trench. The Met Office forcing gives a longer length-scale than the ECMWF forcing which is quite surprising. The AMM15 configuration with the Met Office forcing has much longer length-scales than the global configurations ORCA12 and ORCA025 over most of the domain, except for the Norwegian Trench.

The NMC method gives quite different results compared to those of the HL method. The NMC method gives much less variation in the length-scales for the different configurations, except for the Norwegian Trench where there is a reduction in scales in AMM15 but not in ORCA12 or ORCA025. The NMC large-scale background errors are very small near the coast in Fig. 23 (except for the Norwegian Trench). Therefore, the function fitting results for the large length-scales are expected to be very close to their initial guess (i.e. close to 400 km) in coastal regions, as clearly seen in Fig. 27.

#### 4.2.5 Forecast error covariances at short separation distances

Here we further investigate the error covariances for each configuration estimated by the HL and NMC methods for two example locations. One is in the deeper, off-shelf part of the AMM15 domain (55.8°N, 13.8°W) while the other is in the shallow waters of the North Sea (57.8°N, 2.8°E). The error covariances are shown as a function of separation distance from these points in Figs. 28 and 29. The HL results in Fig. 28 show the off-shelf point to have a much clearer distinction between the short and long length-scale components than the on-shelf point, for all configurations. This results in a shorter length-scale component being clearly identified by the HL method at the off-shelf location. The shorter scales present in the AMM15 configuration are clear at this location with a sharp decay in the errors close to the zero-separation point compared with the global configurations.

At the on-shelf location, the HL method does not show such a clear distinction between two scales. There is some indication in the AMM15 results of a levelling-off of the errors at around 150 km separation with a sharper decay in the errors beyond that separation, and this feature is not obvious in either of the global configurations.

In contrast to the HL results, the NMC results shown in Fig. 29 give evidence of a very short length-scale to the errors in the on-shelf location for the AMM15 configurations, which is not present in the global configurations. The Rossby radius of deformation is very small on the shelf (between 2 and 4 km in the North Sea; Baden et al., 2009) which we wouldn't expect the global configurations to capture, and the HL method does not seem able to provide information about errors on this scale. This feature might merit the inclusion of a third component to the error covariance function given in equation (1) for the AMM15 system so that we could properly represent the error covariances at very small scales using the NMC method.

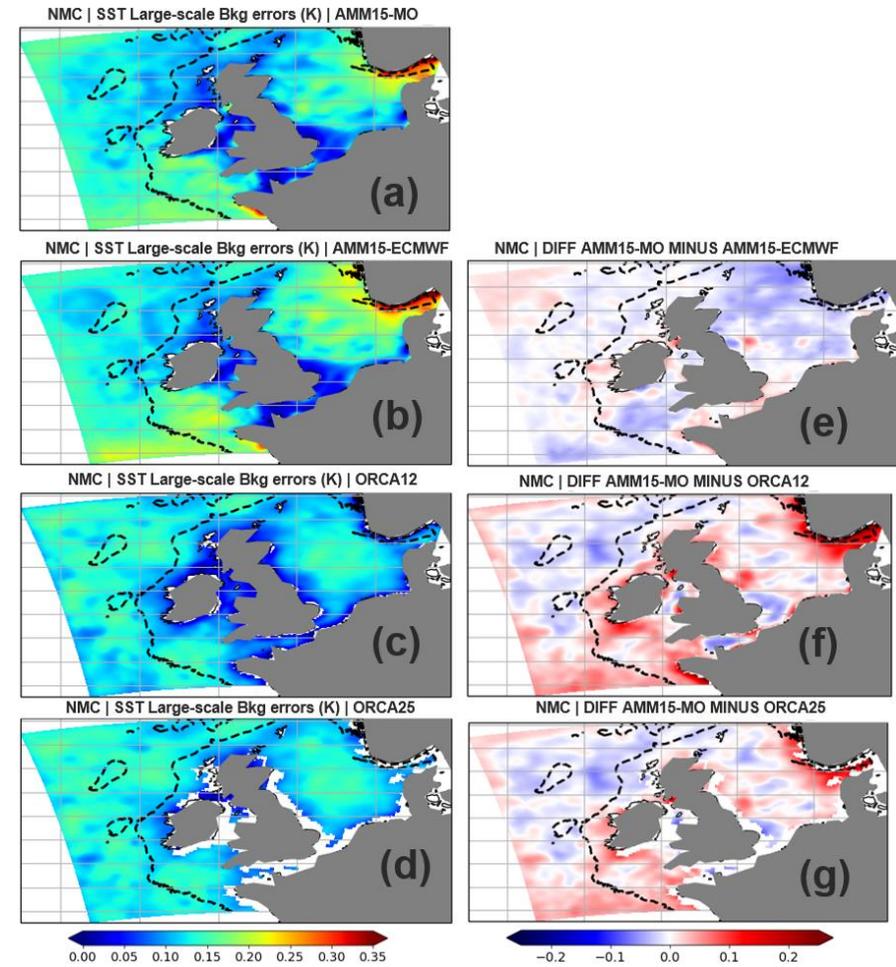
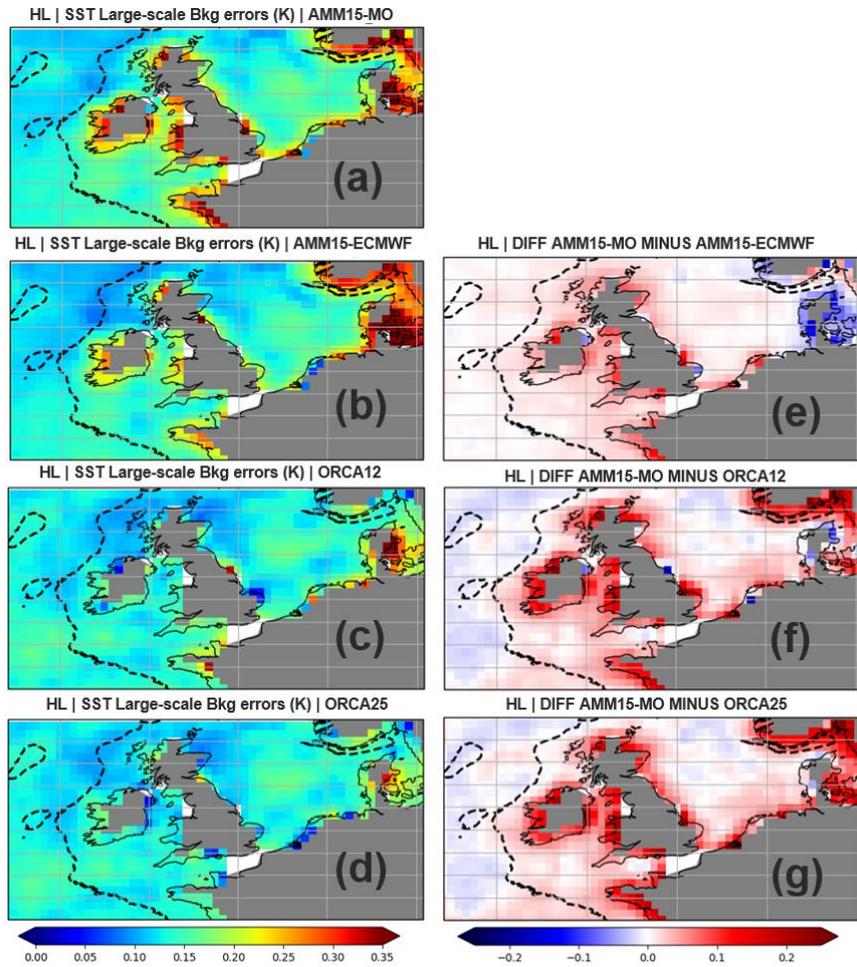


Figure 22: (a-d) 201701-201805 large-scale SST background errors (K) estimated by the HL method using VIIRS observations for AMM15-MO, AMM15-ECMWF, ORCA12 and ORCA025, respectively. The differences between AMM15-MO and the other model configurations are shown in (e-g).

Figure 23: (a-d) 201701-201805 large-scale SST background errors (K) estimated by the NMC method for AMM15-MO, AMM15-ECMWF, ORCA12 and ORCA025, respectively. The differences between AMM15-MO and the other model configurations are shown in (e-g).

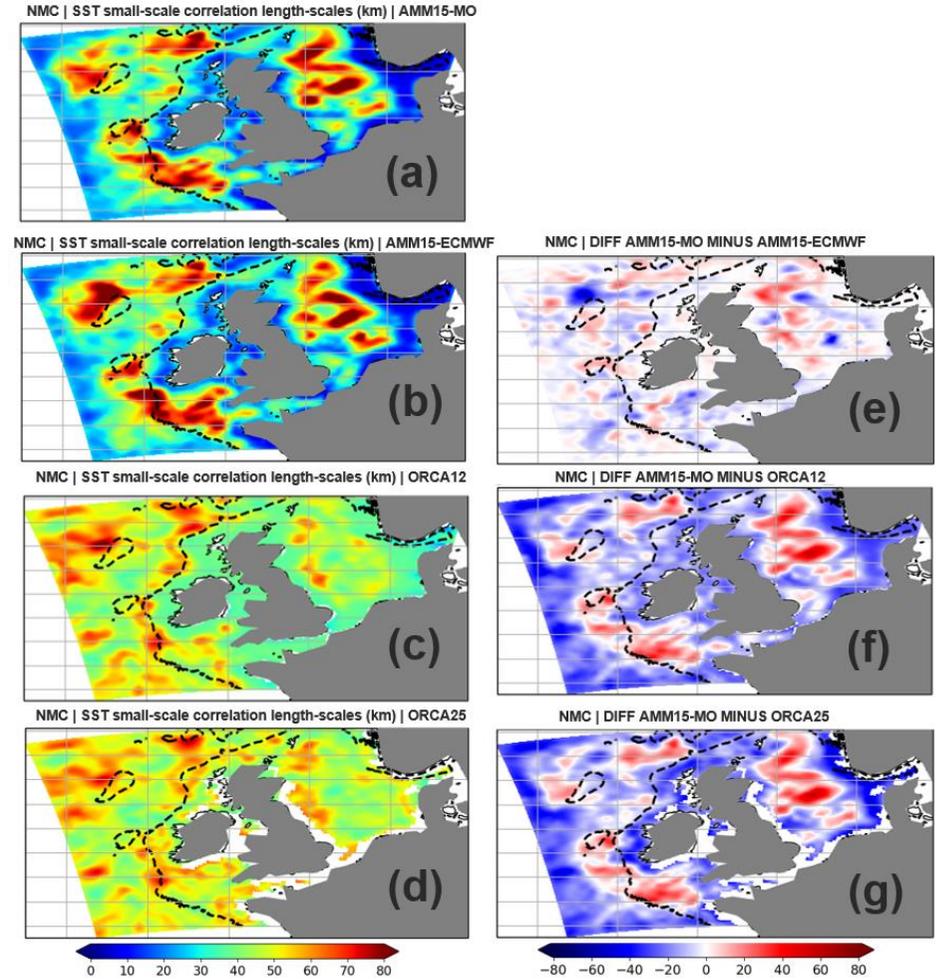
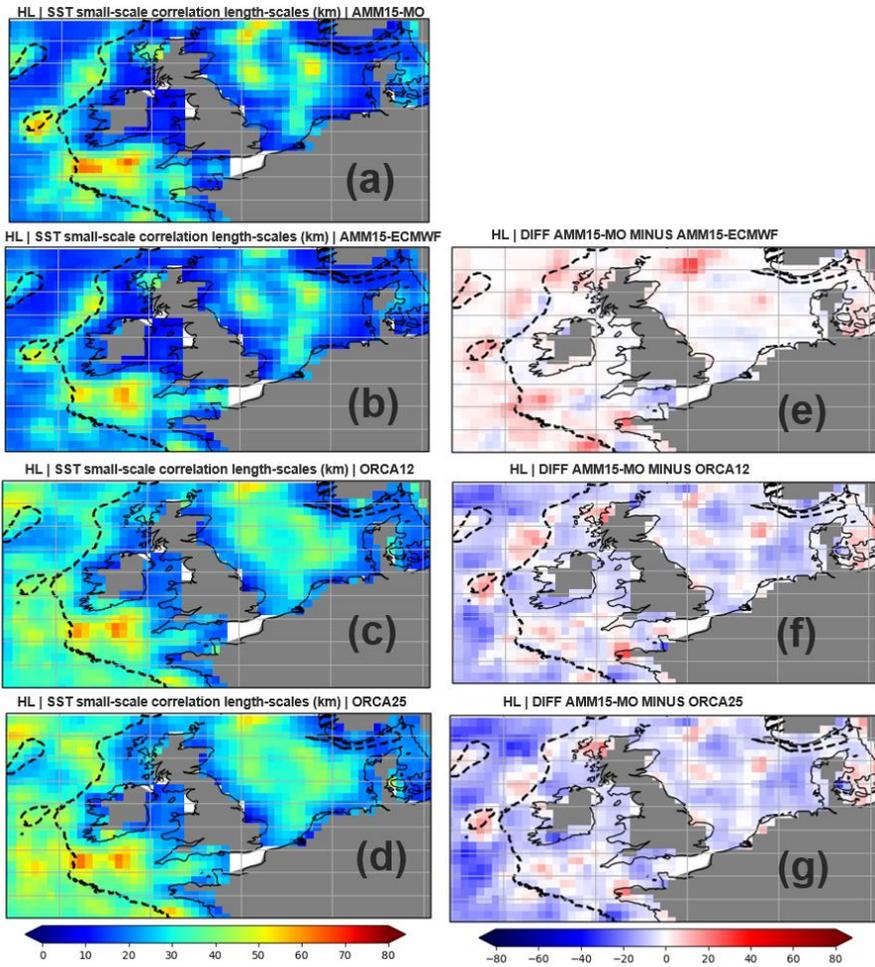


Figure 24: 201701-201805 small-scale SST correlation length-scales (km) estimated by the HL method using VIIRS observations for (a) AMM15-MO, (b) AMM15-ECMWF, (c) ORCA12 and (d) ORCA25. The differences between AMM15-MO and the other model configurations are shown in (e-g).

Figure 25: 201701-201805 small-scale SST correlation length-scales (km) estimated by the NMC method for (a) AMM15-MO, (b) AMM15-ECMWF, (c) ORCA12 and (d) ORCA25. The differences between AMM15-MO and the other model configurations are shown in (e-g).

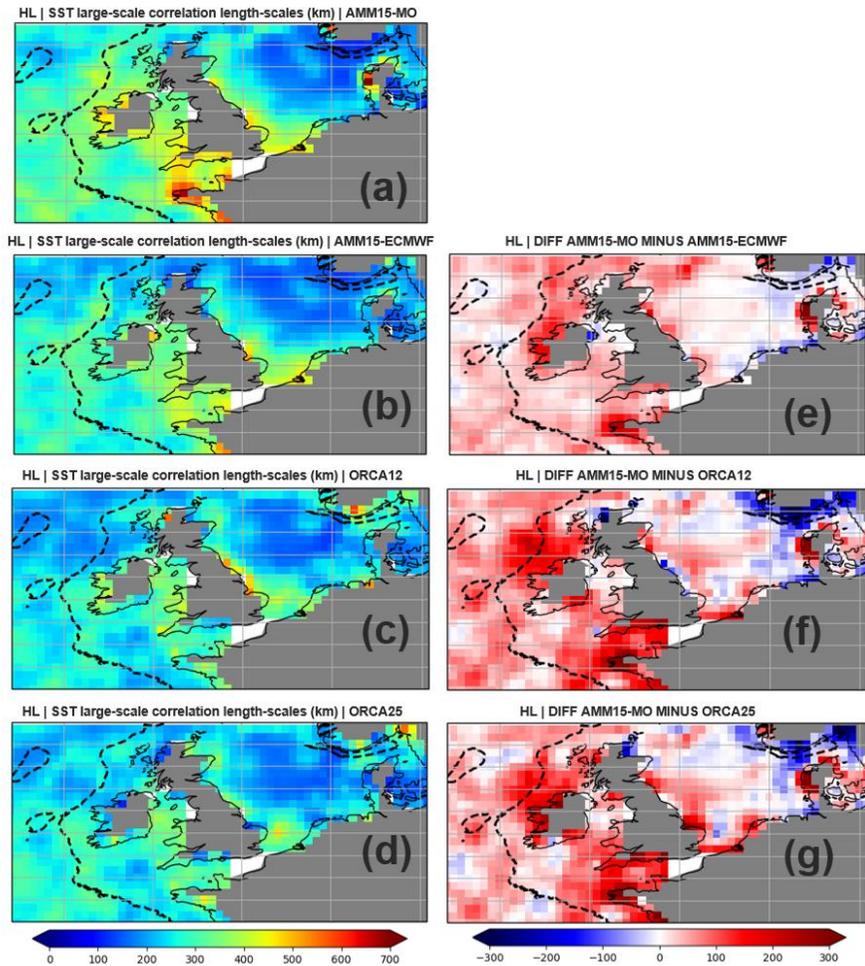


Figure 26: 201701-201805 large-scale SST correlation length-scales (km) estimated by the HL method using VIIRS observations for (a) AMM15-MO, (b) AMM15-ECMWF, (c) ORCA12 and (d) ORCA25. The differences between AMM15-MO and the other model configurations are shown in (e-g).

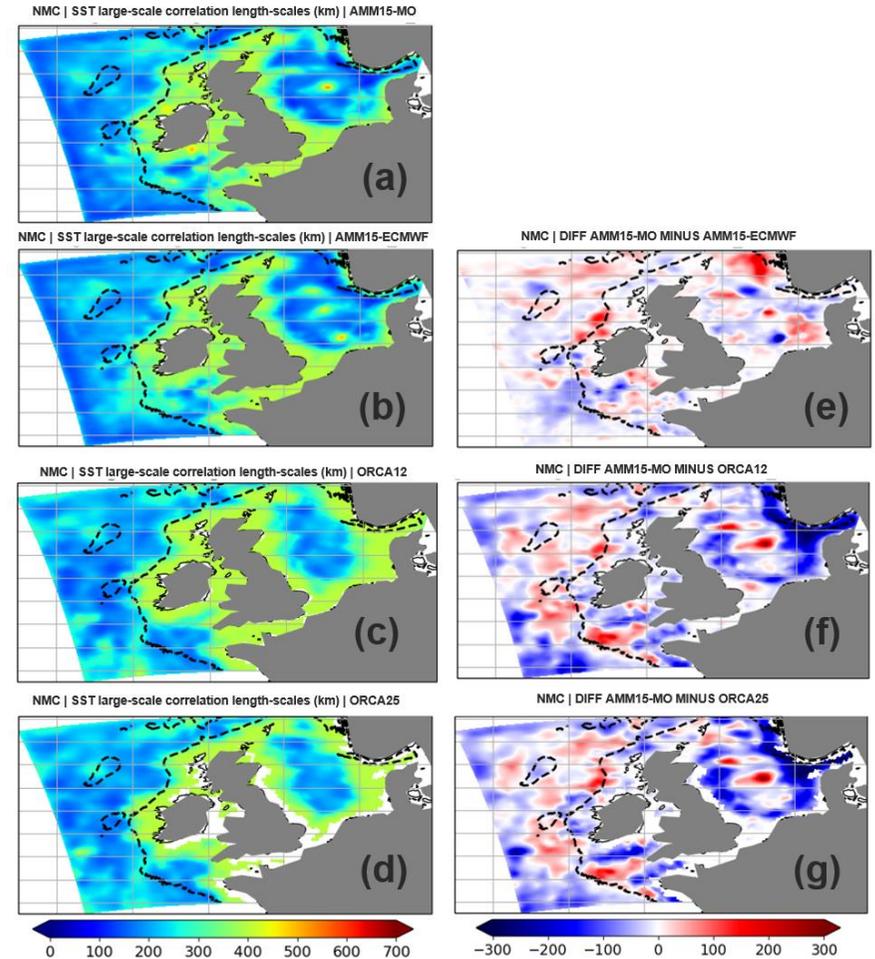


Figure 27: 201701-201805 large-scale SST correlation length-scales (km) estimated by the NMC method for (a) AMM15-MO, (b) AMM15-ECMWF, (c) ORCA12 and (d) ORCA25. The differences between AMM15-MO and the other model configurations are shown in (e-g).

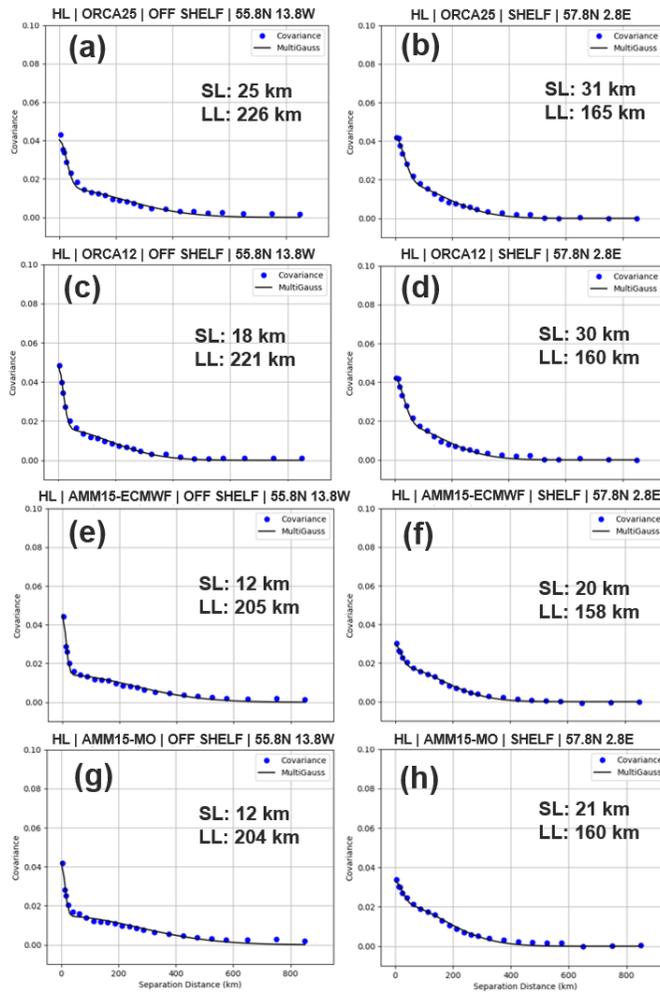


Figure 28: (a-b) ORCA25, (c-d) ORCA12, (e-f) AMM15-ECMWF and (g-h) AMM15-MO SST function fitting for two points located off, and on the shelf, respectively. The blue dots represent the SST error covariances ( $K^2$ ) estimated from the HL method as a function of the separation distance, whereas the black line corresponds to the Multi-Gaussian function fit.

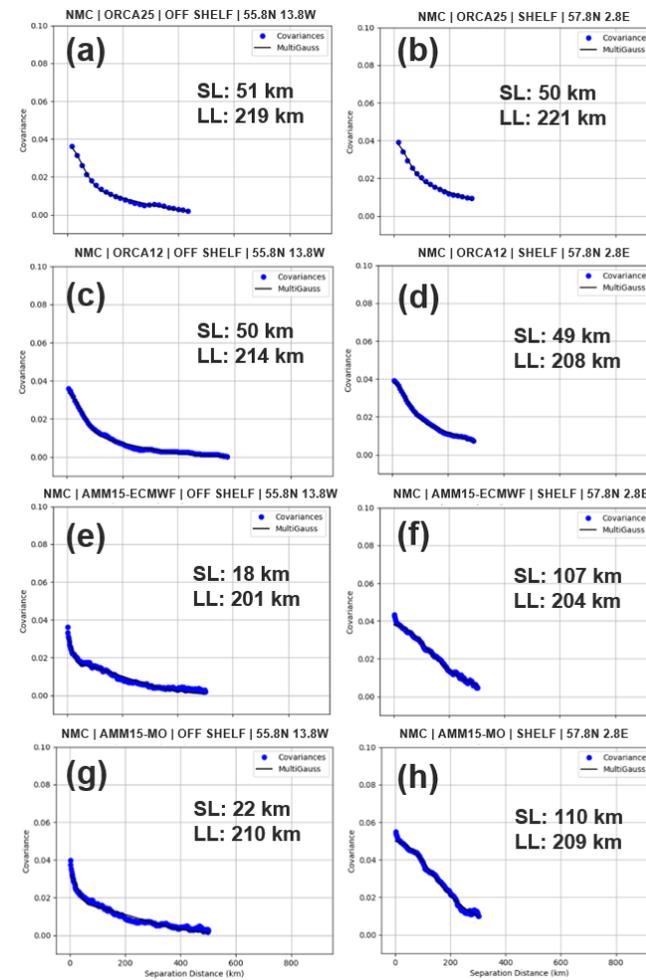


Figure 29: (a-b) ORCA25, (c-d) ORCA12, (e-f) AMM15-ECMWF and (g-h) AMM15-MO SST function fitting for two points located off, and on the shelf, respectively. The blue dots represent the SST error covariances ( $K^2$ ) estimated from the NMC method as a function of the separation distance, whereas the black line corresponds to the Multi-Gaussian function fit.

## 5. Discussion and conclusions

In this work error covariances given by the HL and NMC methods for SST were compared between model configurations with distinct horizontal resolutions and assimilation configurations. The comparison involved runs from two global systems - ORCA12 and ORCA025 - where the seasonal aspects of the SST error covariances and their difference between the models were evaluated on a global scale. Similar comparisons of the error covariances, but without including seasonal estimates, were conducted in the NWS shelf region with the addition of two high-resolution AMM15 systems with different atmospheric forcing.

A clear seasonality of the SST forecast errors is found in both ORCA025 and ORCA12, with larger errors in the summer hemisphere. This seasonal pattern of the SST forecast errors was also reported by Robert-Jones et al. (2016) which applied the HL method to in-situ and Along Track Scanning Radiometer (AATSR) innovations from a long-term SST reanalysis. Their explanation to the seasonality of the SST errors is directly linked to the MLD. Less mixing and a shallower MLD in summer months make the SSTs more sensitive to errors in the surface fluxes, leading to an increased variability of the SSTs. Interestingly, Robert-Jones et al. (2016) did find seasonal variations in the SST forecast errors, but not in their length-scales. Here, the seasonal variations also extend to the small length-scales of the errors which consistently become larger as the forecast errors increase during summer. However, in agreement with Robert-Jones et al. (2016), there is no clear seasonal variation for the long length-scales of the SST forecast errors.

It is also worth noting that the magnitudes of the SST forecast errors are generally similar between the HL and NMC methods on a global scale. Both techniques pick up consistent seasonal patterns and they provide similar magnitudes of the errors within boundary current regions. They also show that ORCA12 has larger small-scale forecast errors compared to ORCA025 in regions of high SST gradients, despite the fact that ORCA12 has smaller length-scales there. Further analysis demonstrated that most of the increase in the small-scale errors for ORCA12 occurs within 25 km of separation distance in both NMC and HL methods. Therefore, it is evident that ORCA12 resolves smaller SST scales than ORCA025, but the assimilation cannot properly constrain the large SST gradients at those scales. This is consistent with the work of Aguiar et al. (2021) who found that ORCA12 performs generally better than ORCA025 if double penalty effects are removed.

There are several possible reasons which may contribute to the poor performance of the SST assimilation in constraining small-scale ocean features in ORCA12. One could be the current mismatch in resolution between SLA and SST satellite observations that are assimilated. While SST satellite observations are very high resolution and can constrain very fine scales in the analysis, SLA satellite observations are not as high resolution as the SSTs. As a result, the SLA assimilation is unlikely to constrain the same mesoscale features in the analysis. This in turn might contribute to the inability of the SST assimilation to properly constrain fine scales

in ORCA12 forecasts, resulting in increased small-scale forecast errors. New SLA satellite data with much higher resolution will become available soon with the Surface Water and Ocean Topography (SWOT) mission (Morrow et al., 2019). Recent findings using the SWOT simulator, such as the ones from D'Addezio et al. (2019), identified that the addition of SWOT to their current assimilation system can provide gains of several kilometres in predicting ocean fine-scale structures. Assimilation experiments using the SWOT simulator are already being conducted in the AMM15 configuration at the Met Office and future plans also include the assimilation of SWOT data in ORCA12.

Another possible explanation for limitations of ORCA12 in reproducing small-scale structures is the fact that the assimilation in ORCA12 is performed on the ORCA025 grid. Therefore, ORCA12 assimilation increments do not have the same resolution as the model itself. Future work will aim to improve ORCA12 forecast results by performing the assimilation on the ORCA12 grid. So, the same error covariance analyses can be repeated to check whether this improvement contributes to a better representation of fine-scale features in ORCA12 forecasts.

In contrast to ORCA12, the benefits of increasing the model resolution to adequately represent small-scale features is seen in the high-resolution AMM15 systems with both NMC and HL results. The AMM15 configurations give reduced small-scale SST errors relative to the global systems on the shelf, due to the inclusion of shelf processes, and there are also clear improvements in the deeper part of the NWS domain. Comparing the small-scale background errors between ORCA12 and AMM15-MO within the off-shelf (on-shelf) region, there is a reduction from 0.18 K (0.19 K) to 0.14 K (0.16 K) given by the NMC results, whereas the HL results show a decrease from 0.20 K (0.20 K) to 0.17 K (0.18 K). A few exceptions to this occur particularly near the coast where tide-induced SST gradients might increase the AMM15 forecast errors, perhaps indicating that tidal fronts need to be better constrained by the assimilation. For the AMM15, the reduction of the small-scale forecast errors is accompanied by an overall decrease of their short length-scales, mainly in the deep part of the domain. The variability of the short length-scales in AMM15 is also much greater than the global configurations, such as on the shelf and following the sharp bathymetric slope of the shelf break. This is particularly visible in the NMC results which provide additional fine-scale information of the error covariances. Unlike its small-scale counterpart, there was no clear impact on the reduction of the SST large-scale errors in AMM15 configurations. This is expected since an increase in model resolution should mainly improve the representation of small-scale structures in the forecasts.

The NMC method also identified that in particular regions a third component of the errors with even shorter length-scales might be needed to fully represent the forecast error characteristics in the AMM15 runs. A detailed investigation of these finer error scales in AMM15 will be conducted as future work. These AMM15 results are relevant because they clearly demonstrate the ability of very high-resolution systems to capture much greater variability and finer structures, not only in the shelf but also in the off-shelf part of the domain. Therefore, the findings here reinforce the large potential for future higher-resolution global models and support the effort of operational centres to implement them (Iovino et al., 2016; Barton et al.,

2020; Aguiar et al., 2021).

Although most of the findings are well explained in this report, other results are not so obvious and further investigation is necessary. For example, the impact of the surface forcing on the representation of the AMM15 forecast errors is unclear with different results given by the HL and NMC methods. For instance, the large-scale errors seen by the HL method are generally larger over most of the on-shelf region for the AMM15 experiment using the Met Office fluxes compared with the one forced using ECMWF fluxes, whereas the opposite is seen for the NMC method. These discrepancies also extend to the long length-scales. The Met Office forcing gives much longer length-scales than the ECMWF forcing with the HL method, whereas this is not captured by the NMC technique.

These discrepancies between HL and NMC results are also seen for ORCA025 and ORCA12 length-scales on the global domain. Apart from the regions of high SST gradients and boundary currents, the short length-scale results disagree between both methods. As expected, the HL method gives smaller scales for ORCA12 almost everywhere in the globe compared to ORCA025, whereas the NMC results surprisingly show the opposite, particularly in the subtropical and tropical regions. As well as in the short length-scales, ORCA12 shows increased long length-scales relative to ORCA025 given by the NMC method, whereas this is not obvious for the HL results. Due to a three-fold increase in horizontal resolution, it is quite surprising and unexpected to see the NMC results indicating longer length-scales for ORCA12 than for ORCA025 almost everywhere apart from the boundary current regions.

A limitation of the NMC method in properly reproducing the length-scales of the errors is related to the way in which the error covariances are calculated for this method (see Bannister et al., 2008). Usually, the NMC method uses longer (48 h and 24 h) forecast differences which allow the assimilation information to spread out further, directly reflecting on the NMC length-scales. Since the HL technique computes the model error covariance directly, based on the observation-minus-forecast data, and not as a proxy of the forecast differences, one would expect this method to be more reliable than the NMC method in regions where the HL assumptions are valid. For instance, the NMC statistics often require adjustment to compensate for the use of forecasts that are temporally longer than those used for the a-priori state (e.g. Derber and Bouttier, 1999; Ingleby, 2001). In this report, the forecast length of 24 h is chosen to avoid any contamination of the error covariance statistics in the forecast initialisation due to assimilation shocks. However, previous works found beneficial the use of forecast lengths closer to the background state, for example 36 h and 12 h forecasts or even 30 h and 6 h forecasts (e.g. Berre et al., 2000; Berre et al., 2006). As future work, sensitivity tests can be conducted using shorter forecast lengths to evaluate whether these changes have impacts on the NMC length-scales.

Overall, this work has produced useful insights into the effect of the resolution of the model and data assimilation systems on the forecast error characteristics for SST. The spatial patterns of the forecast errors and their seasonal variations have been explored for the global systems. The impact of different atmospheric forcing on the magnitude and scales of the SST forecast errors in a high-resolution system has also been described. The work has

demonstrated the positive impact of high resolution 1.5 km systems, which include representation of tidal processes, on reducing the magnitude and scales of the short-range SST forecast errors, both in shelf-seas and in the deeper ocean. In addition to the impact of the resolution on the forecast errors, the forecast error covariances themselves can be valuable for many other applications. The forecast error estimates produced here will be useful in further identifying areas where the global and regional forecasting systems can be improved, either through model or data assimilation developments. Furthermore, these new spatially and seasonally varying estimates of the error covariances will also be used to update the background error parameters in the data assimilation system, which is expected to reduce the forecast errors of the model configurations assessed in this report.

## References

Aguiar, AB, Water, J, Price, M, Inverarity, G, Pequignet, C, Maksymczuk, J, Smout-Day, K, Martin, M, Bell, M, King, R, While, J, Lea, D, Siddorn, J. The new Met Office global ocean forecasting system at 1/12<sup>th</sup> degree resolution. In preparation for submission to Q.J.R. Meteorol. Soc.

Badin, G., Williams, R. G., Holt, J. T., and Fernand, L. J. (2009), Are mesoscale eddies in shelf seas formed by baroclinic instability of tidal fronts? *J. Geophys. Res.*, 114, C10021, doi:10.1029/2009JC005340.

Bannister, R.N. (2008), A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances. *Q.J.R. Meteorol. Soc.*, 134: 1951-1970.

Barton, N., Metzger, E. J., Reynolds, C. A., Ruston, B., Rowley, C., Smedstad, O. M., et al. (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*, 7, e2020EA001199. <https://doi.org/10.1029/2020EA001199>

Berre, L. (2000). Estimation of synoptic and mesoscale forecast error covariances in a limited area model. *Mon. Weather Rev.* 128, 644-667.

Berre L., Stefanescu S. E., Pereira M. B. (2006). The representation of the analysis effect in three error simulation techniques. *Tellus* 58A, 196-209.

Berg, P. and Poulsen, J.W., 2012. Technical Report 12-11 Implementation details for HBM. Danish Meteorological Institute, Ministry for Climate and Energy.

Blockley, E.W., Martin, M.J., McLaren, A.J., Ryan, A.G., Waters, J., Lea, D.J., Mirouze, I., Peterson, K.A., Sellar, A. and Storkey, D., 2014. Recent development of the Met Office operational ocean forecasting system: an overview and assessment of the new Global FOAM forecasts. *Geoscientific Model Development*, 7(6).

Bloom, S.C., Takacs, L.L., Da Silva, A.M. and Ledvina, D., 1996. Data assimilation using incremental analysis updates. *Monthly Weather Review*, 124(6), pp.1256-1271.

Cooper, M. and Haines, K., 1996. Altimetric assimilation with water property conservation. *Journal of Geophysical Research: Oceans*, 101(C1), pp.1059-1077.

Craig, A., Valcke, S. and Coquart, L., 2017. Development and performance of a new version of the OASIS coupler, OASIS3-MCT\_3.0. *Geoscientific Model Development*, 10(9).

D'Addezio, J. M., Smith, S., Jacobs, G. A., Helber, R. W., Rowley, C., Souopgui, I., Carrier, M. J., 2019. Quantifying wavelengths constrained by simulated SWOT observations in a submesoscale resolving ocean analysis/forecasting system. *Ocean Model.* 135, 40-55. <http://dx.doi.org/10.1016/j.ocemod.2019.02.001>.

Davidson F, Alvera-Azcárate A, Barth A, Brassington GB, Chassignet EP, Clementi E, De Mey-Frémaux P, Divakaran P, Harris C, Hernandez F, Hogan P, Hole LR, Holt J, Liu G, Lu Y, Lorente P, Maksymczuk J, Martin M, Mehra A, Melsom A, Mo H, Moore A, Oddo P, Pascual A, Pequignet A-C, Kourafalou V, Ryan A, Siddorn J, Smith G, Spindler D, Spindler T, Stanev EV, Staneva J, Storto A, Tanajura C, Vinayachandran PN, Wan L, Wang H, Zhang Y, Zhu X and Zu Z (2019) Synergies in Operational Oceanography: The Intrinsic Need for Sustained Ocean Observations. *Front. Mar. Sci.* 6:450. doi: 10.3389/fmars.2019.00450.

Derber J., Bouttier F, 1999. A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus*, 51, pp.195-221.

Flather, R.A., 1981. Results from a model of the northeast Atlantic relating to the Norwegian Coastal Current. *The Norwegian coastal current*, 2, pp.427-458.

Good, S.A., Martin, M.J. and Rayner, N.A., 2013. EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, 118(12), pp.6704-6716.

Graham, J.A., O'Dea, E., Holt, J., Polton, J., Hewitt, H.T., Furner, R., Guihou, K., Brereton, A., Arnold, A., Wakelin, S. and Castillo Sanchez, J.M., 2018. AMM15: a new high-resolution NEMO configuration for operational simulation of the European north-west shelf. *Geoscientific Model Development*, 11(2), pp.681-696.

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

Hollingsworth, A. and Lönnberg, P., 1986. The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus A*, 38(2), pp.111-136.

Hunke, E.C., Lipscomb, W.H., Turner, A.K., Jeffery, N. and Elliott, S., 2010. CICE: the Los Alamos Sea Ice Model Documentation and Software User's Manual Version 4.1 LA-CC-06-012. *T-3 Fluid Dynamics Group, Los Alamos National Laboratory*, 675, p.500.

Ingleby N. B., 2001. The statistical structure of forecast errors and its representation in the Met Office global three-dimensional variational data assimilation system. *Q. J. R. Meteorol. Soc.*, 127, 209-231.

Iovino, D., Masina, S., Storto, A., Cipollone, A., and Stepanov, V. N.: A 1/16° eddy simulation of the global NEMO sea-ice–ocean system, *Geosci. Model Dev.*, 9, 2665–2684, <https://doi.org/10.5194/gmd-9-2665-2016>, 2016.

Jackson, J. C., Dubois, C., Forget, G., Haines, K., Harrison, M., Iovino, D., Kohl, A., Mignac, D., Mashina, S., Peterson, K. A., Piecuch, C. G., Roberts, C., Robson, J., Storto, A., Toyoda, T., Valdivieso, M., Wilson, C., Wang, Y., and Zuo, H.: The mean state and variability of the north Atlantic circulation: a perspective from ocean reanalyses, *J. Geophys. Res.-Ocean.*, 124, 9141-9170, <https://doi.org/10.1029/2019JC015210>, 2019.

King, R.R., While, J., Martin, M.J., Lea, D.J., Lemieux-Dudon, B., Waters, J. and O'Dea, E., 2018. Improving the initialisation of the Met Office operational shelf-seas model. *Ocean Modelling*, 130, pp.1-14.

Lea, D.J., Drecourt, J.P., Haines, K. and Martin, M.J., 2008. Ocean altimeter assimilation with observational-and model-bias correction. *Quarterly Journal of the Royal Meteorological Society*, 134(636), pp.1761-1774.

Lellouche, J.-M., Greiner, E., Le Galloudec, O., Garric, G., Regnier, C., Drevillon, M., Benkiran, M., Testut, C.-E., Bourdalle-Badie, R., Gasparin, F., Hernandez, O., Levier, B., Drillet, Y., Remy, E., and Le Traon, P.-Y.: Recent updates to the Copernicus Marine Service global ocean monitoring and forecasting real-time  $1/12^\circ$  high-resolution system, *Ocean Sci.*, 14, 1093–1126, <https://doi.org/10.5194/os-14-1093-2018>, 2018.

Lewis, H.W., Castillo Sanchez, J.M., Siddorn, J., King, R.R., Tonani, M., Saulter, A., Sykes, P., Pequignet, A.C., Weedon, G.P., Palmer, T. and Staneva, J., 2019. Can wave coupling improve operational regional ocean forecasts for the north-west European Shelf?. *Ocean Science*, 15(3), pp.669-690.

Madec, G., 2016. NEMO ocean engine, version 3.6 stable. *Note du Pole de modelisation de l'Institut Pierre-Simon Laplace*, (27).

Martin, M.J., Hines, A. and Bell, M.J., 2007. Data assimilation in the FOAM operational short-range ocean forecasting system: a description of the scheme and its impact. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 133(625), pp.981-995.

Mirouze, I. and Weaver, A.T., 2010. Representation of correlation functions in variational assimilation using an implicit diffusion operator. *Quarterly Journal of the Royal Meteorological Society*, 136(651), pp.1421-1443.

Mirouze, I., Blockley, E.W., Lea, D.J., Martin, M.J. and Bell, M.J., 2016. A multiple length scale correlation operator for ocean data assimilation. *Tellus A: Dynamic Meteorology and Oceanography*, 68(1), p.29744.

Mogensen, K., Balmaseda, M.A., Weaver, A.T., Martin, M. and Vidard, A., 2009. NEMOVAR: A variational data assimilation system for the NEMO ocean model. *ECMWF newsletter*, 120, pp.17-22.

Mogensen, K., Balmaseda, M.A. and Weaver, A., 2012. The NEMOVAR ocean data assimilation system as implemented in the ECMWF ocean analysis for System 4.

Moore AM, Martin MJ, Akella S, Arango HG, Balmaseda M, Bertino L, Ciavatta S, Cornuelle B, Cummings J, Frolov S, Lermusiaux P, Oddo P, Oke PR, Storto A, Teruzzi A, Vidard A and Weaver AT (2019). Synthesis of Ocean Observations Using Data Assimilation for Operational, Real-Time and Reanalysis Systems: A More Complete Picture of the State of the Ocean. *Front. Mar. Sci.* 6:90. doi: 10.3389/fmars.2019.00090.

Morrow, R., Fu, L. L., Arduin, F., Benkiran, M., Chapron, B., Cosme, E., D'Ovidio, F., Farrar, J. T., Gille, S. T., Lapeyre, G., Le Traon, P. Y., Pascual, A., Ponte, A., Qiu, B., Rasle, N., Ubelmann, C., Wang, J., and Zaron, E. D.: Global observations of fine-scale ocean surface topography with the Surface Water and Ocean Topography (SWOT) Mission, *Frontiers in Marine Science*, 6, 232, <https://doi.org/10.3389/fmars.2019.00232>, 2019.

O'Dea, E., Furner, R., Wakelin, S., Siddorn, J., While, J., Sykes, P., King, R., Holt, J. and Hewitt, H., 2017. The CO5 configuration of the 7 km Atlantic Margin Model: large-scale biases and sensitivity to forcing, physics options and vertical resolution. *Geoscientific Model Development*, 10(8), p.2947.

Parrish, D.F. and Derber, J.C., 1992. The National Meteorological Center's spectral statistical-interpolation analysis system. *Monthly Weather Review*, 120(8), pp.1747-1763.

Penny SG, Akella S, Balmaseda MA, Browne P, Carton JA, Chevallier M, Counillon F, Domingues C, Frolov S, Heimbach P, Hogan P, Hoteit I, Iovino D, Laloyaux P, Martin MJ, Masina S, Moore AM, de Rosnay P, Schepers D, Sloyan BM, Storto A, Subramanian A, Nam S, Vitart F, Yang C, Fujii Y, Zuo H, O'Kane T, Sandery P, Moore T and Chapman CC (2019). Observational Needs for Improving Ocean and Coupled Reanalysis, S2S Prediction, and Decadal Prediction. *Front. Mar. Sci.* 6:391. doi: 10.3389/fmars.2019.00391

Ridley, J. K., Blockley, E. W., Keen, A. B., Rae, J. G. L., West, A. E., and Schroeder, D.: The sea ice model component of HadGEM3-GC3.1, *Geosci. Model Dev.*, 11, 713–723, <https://doi.org/10.5194/gmd-11-713-2018>, 2018.

Roberts-Jones, J., K. Bovis, M.J. Martin, A. McLaren, 2016. Estimating background error covariance parameters and assessing their impact in the OSTIA system. *Rem. Sens. of Environment*, 176, 117-138. doi:10.1016/j.rse.2015.12.006

Saulter, A., Bunney, C., Li, J.G. and Palmer, T., 2017, September. Process and resolution impacts on UK coastal wave predictions from operational global-regional wave models. In *Proceedings of the 15th International Workshop on Wave Hindcasting and Forecasting and 6th Coastal Hazard Symposium, Liverpool, UK* (pp. 10-15).

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., and Sinha, B.: UK Global Ocean GO6 and GO7: a traceable hierarchy of model resolutions, *Geosci. Model Dev.*, 11, 3187–3213, <https://doi.org/10.5194/gmd-11-3187-2018>, 2018.

Storkey, D., Blockley, E.W., Furner, R., Guiavarc'h, C., Lea, D., Martin, M.J., Barciela, R.M., Hines, A., Hyder, P. and Siddorn, J.R., 2010. Forecasting the ocean state using NEMO: The new FOAM system. *Journal of operational oceanography*, 3(1), pp.3-15.

Tonani, M., Sykes, P., King, R.R., McConnell, N., Péquignet, A.C., O'Dea, E., Graham, J.A., Polton, J. and Siddorn, J., 2019. The impact of a new high-resolution ocean model on the Met Office North-West European Shelf forecasting system. *Ocean Science*, 15, pp.1133-1158.

Troccoli, A. and Haines, K., 1999. Use of the temperature–salinity relation in a data assimilation context. *Journal of Atmospheric and Oceanic Technology*, 16(12), pp.2011-2025.

Waters, J., Lea, D.J., Martin, M.J., Mirouze, I., Weaver, A. and While, J., 2015. Implementing a variational data assimilation system in an operational  $\frac{1}{4}$  degree global ocean model. *Quarterly Journal of the Royal Meteorological Society*, 141(687), pp.333-349.

Weaver, A.T., Deltel, C., Machu, É., Ricci, S. and Daget, N., 2005. A multivariate balance operator for variational ocean data assimilation. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 131(613), pp.3605-3625.

Weaver, A.T. and Mirouze, I. (2013), On the diffusion equation and its application to isotropic and anisotropic correlation modelling in variational assimilation. *Q.J.R. Meteorol. Soc.*, 139: 242-260. <https://doi.org/10.1002/qj.1955>

While, J. and Martin, M.J., 2019. Variational bias correction of satellite sea-surface temperature data incorporating observations of the bias. *Quarterly Journal of the Royal Meteorological Society*, 145(723), pp.2733-2754.

Met Office  
FitzRoy Road  
Exeter  
Devon  
EX1 3PB  
United Kingdom