Technical Memo

ECMWF
European Centre for Medium-Range Weather Forecasts

858

Reduced-resolution ocean configurations for efficient testing with the ECMWF coupled model

Christopher Roberts, Antje Weisheimer, Stephanie Johnson, Tim Stockdale, Magdalena Balmaseda, Philip Browne, Andrew Dawson, Martin Leutbecher, and Frederic Vitart
(Research Department)

January 2020

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our website under:
http://www.ecmwf.int/en/publications

Contact: library@ecmwf.int

Abstract

Since June 2018, all operational configurations of the Integrated Forecasting System (IFS) have been coupled to an ocean/sea-ice model, which has substantially increased the computational cost of research and development activities at ECMWF. The purpose of this memorandum is to assess whether some aspects of model development and testing at ECMWF can be undertaken with lower resolution ocean configurations without comprising the scientific integrity of the results. Although the higher resolution ocean configuration outperforms its lower resolution counterpart, there are many aspects of model testing and development for which the absolute model performance is less relevant than the difference (i.e. Δ) between two experiments. This study considers whether Δ estimated using a lower resolution ocean model (i.e. Δ_{LRO}) can be considered an appropriate proxy for Δ estimated using a higher resolution ocean model (i.e. Δ_{HRO}) for a variety of deterministic and probabilistic metrics at different lead times. In general, Δ_{LRO} is an extremely good proxy for Δ_{HRO} in medium-range forecasts (days 1-15), provided that ocean initial conditions used in lower and higher resolution systems are as consistent as possible. Δ_{LRO} is also good proxy for Δ_{HRO} in extended-range forecasts (weeks 1-4), particularly for metrics that are computed from forecast anomalies relative to a hindcast climatology. At extended-range lead times, we find a stronger sensitivity to changes in atmospheric resolution (50 km to 31 km) than changes in ocean resolution (100 km to 25 km). At seasonal (and longer) timescales, Δ_{LRO} is a useful proxy for Δ_{HRO} for some metrics (e.g. changes to the model climatology), but the approximation begins to break down due to a divergence of SST biases in lower and higher resolution systems. Finally, although lower resolution ocean configurations will continue to be a useful tool for research and development purposes at ECMWF, there will always be cases where it is necessary to use the highest resolution ocean model that is affordable.

1 Introduction

The importance of ocean-atmosphere coupling for numerical weather prediction has been emphasized in several recent studies (e.g Boisséson et al. 2012; Brassington et al. 2015; Mogensen et al. 2017). Accordingly, since June 2018, all operational configurations of the ECMWF Integrated Forecasting System (IFS) have been coupled to the eddy-permitting ORCA025 configuration of the NEMO ocean/sea-ice model, which has a horizontal resolution of ¼ degree (~25 km). However, the addition of ocean and sea-ice components to the IFS has substantially increased the computational cost of research and development activities. In the future, the adoption of even higher resolution ocean configurations could see further increases in the fractional cost of the ocean in the coupled IFS.

It is standard practice for new IFS developments to undergo initial testing at reduced atmospheric resolutions. However, it is less clear that this approach is applicable in the ocean due to the very different length scales associated with synoptic variability in the ocean and atmosphere. Lower-resolution atmospheric experiments at ECMWF commonly use the Tco199 or TL255 grid (i.e. 50-80 km grid resolution), which is more than sufficient to resolve the first baroclinic Rossby radius of deformation in the mid-latitude atmosphere (L_R ~ 1000 km). In contrast, the eddy-parameterized (~100 km grid spacing) and eddy-permitting (~25 km grid spacing) configurations of NEMO currently available at ECMWF cannot resolve the ocean mesoscale at all latitudes (L_R ~25 km in the midlatitudes). However, for some activities, the benefits of a more efficient coupled model may outweigh the penalty of reduced ocean resolution. For this reason, it is important to have a clear understanding of the processes, regions, and metrics that show a strong sensitivity to ocean resolution and the timescales at which differences become apparent.

Previous studies have demonstrated that increases to ocean model resolution can improve the representation of many oceanic processes, including mean surface temperature biases (Roberts et al. 2009; Marzocchi et al. 2015), the strength and position of western boundary currents (Kirtman et al. 2012; Marzocchi et al. 2015; Chassignet and Xu 2017), air-sea interaction (Roberts et al. 2016; Bryan et al. 2010), and mass, heat and freshwater budgets (e.g. Hewitt et al. 2016; Kirtman et al. 2012; Griffies et al. 2015). However, for many aspects of model testing and development, the absolute model performance is less relevant than the difference (i.e. Δ) between two experiments, where Δ is representative of a change in one or more summary statistics or forecast skill scores. A typical example of Δ is the collection of scores used to summarize the difference in forecast performance between two IFS cycles.

The primary purpose of this memorandum is to assess whether some aspects of model development and testing at ECMWF can be undertaken with lower resolution versions of the ocean/sea-ice model without comprising the scientific integrity of the results. In particular, we establish the circumstances under which Δ estimated using a lower resolution ocean model (i.e. Δ_{LRO}) can be considered an appropriate proxy for Δ estimated using a higher resolution ocean model (i.e. Δ_{HRO}).

We expect the validity of this approximation to depend on both forecast lead time and the metric (i.e. Δ) under consideration. In particular, metrics that are a non-linear function of the mean state may show a strong sensitivity to resolution at longer lead times. For example, the following equations illustrate that a change in squared bias will always be a function of the reference climatology:

$$\Delta MODEL = \langle G \rangle - \langle F \rangle$$

$$\Delta BIAS = (\langle G \rangle - \langle O \rangle) - (\langle F \rangle - \langle O \rangle) = \Delta MODEL \quad (1.1)$$

$$\Delta BIAS^2 = (\langle G \rangle - \langle O \rangle)^2 - (\langle F \rangle - \langle O \rangle)^2 = (\Delta MODEL)^2 + 2 \cdot \Delta MODEL \cdot (\langle F \rangle - \langle O \rangle)$$

where $<F>$ and $<G>$ represent model climatologies and $<O>$ is the observed climatology. Thus, even if the difference in the mean state between two cycles is very similar when tested at different resolutions (i.e. $\Delta MODEL_{HRO} = \Delta MODEL_{LRO}$), whether or not these changes represent an improvement will be sensitive to the climatological bias in the reference forecast. Therefore, a necessary (but not sufficient) condition for $\Delta BIAS^2_{LRO}$ to be a good approximation for $\Delta BIAS^2_{HRO}$ is for LRO and HRO systems to have sufficiently similar climatologies.

Although the sensitivity of model biases to changes in ocean resolution are discussed briefly in section 3, this memorandum is not intended to provide a systematic and process-based assessment of absolute model performance at different ocean resolutions. Instead, the interested reader is referred to Roberts et al. (2020), which provides a thorough review of resolution-sensitive mechanisms of ocean-atmosphere coupling and a detailed scientific evaluation of the timescale-dependent atmospheric response to increased ocean model resolution in the IFS.

The remainder of this memorandum is organized as follows. Section 2 describes the IFS configurations used in this study and the generation of consistent ocean and sea-ice initial conditions at different resolutions. Section 3 summarizes the impact of ocean resolution on the model climate at different lead times. Section 4 presents estimates of Δ_{LRO} and Δ_{HRO} for a variety of deterministic and probabilistic metrics in medium-range, extended-range, and seasonal forecasts. Section 5 provides a summary of our main conclusions and provides some guidance on the circumstances under which Δ_{LRO} can be considered an appropriate surrogate for Δ_{HRO}.

2 Methods

2.1 Model configurations

The ECMWF Integrated Forecasting System (IFS) is a global Earth system model that includes dynamic representations of the atmosphere, sea-ice, ocean, land surface, and ocean waves. An overview of the IFS coupled model is provided in Roberts et al. (2018) and further technical details can be found in the online documentation (ECMWF website 2019).

All IFS configurations discussed in this memorandum are coupled to version 3.4 of the Nucleus for European Models of the Ocean (NEMO) (Madec 2008) and version 2 of the Louvain-la-Neuve Sea-Ice Model (LIM2; Bouillon et al. 2009; Fichefet and Maqueda 1997). Higher resolution ocean configurations are labelled 'HRO' and use the ORCA025 tripolar grid, which has an eddy-permitting horizontal resolution of ¼ degree. Lower resolution ocean configurations are labelled 'LRO' and use the ORCA1 tripolar grid, which has a nominal horizontal resolution of ~ 1 degree, with meridional refinement to ~0.3 degrees near the equator. Both NEMO configurations are configured as described in Roberts et al. (2018). One of the significant differences between these configurations is the use of the Gent and Mcwilliams (1990) parameterization for the effect of mesoscale eddies with the ORCA1 grid, which is disabled when using the ORCA025 grid. Both configurations use the same vertical discretization, which consists of 75 z-levels and partial cells at the ocean floor.

Estimates of the fractional cost of the ocean in different configurations of the coupled IFS are summarized in Table 1. The total cost of coupled experiments with the Tco399L91 atmosphere (double precision, TSTEP=1200) is reduced by ~41 % when using the ORCA1 ocean model, which corresponds to a reduction in the fractional cost of the ocean from ~43 % (ORCA025) to ~4 % (ORCA1). The relative savings are even larger when running the IFS at lower atmospheric resolutions or at a reduced numerical precision (see table 1). However, these savings also depend on other details of the IFS setup, such as the number of vertical levels in the atmosphere and the specified time step. For example, coupled experiments with the Tco399L137 atmosphere (double precision, TSTEP=900) are ~27 % less expensive when using the ORCA1 ocean model, which corresponds to a reduction in the fractional cost of the ocean from ~28 % (ORCA025) to ~2 % (ORCA1).

Although this study emphasizes the comparison of ORCA1 and ORCA025 configurations of NEMO, the outlined approach can be considered a 'proof-of-concept' that will have relevance for future ocean configurations. For instance, future upgrades to high-performance computing may allow ECMWF to run an eddy-resolving model operationally, but with research and development experimentation limited to a 'workhorse' eddy-permitting configuration.

2.2 Ocean initial conditions.

This section describes the method used to ensure that initial conditions in ORCA1 and ORCA025 experiments are as consistent as possible. Ocean and sea-ice initial conditions for ORCA025 experiments, termed INI-HRO, are provided by the latest version of the ECMWF Ocean Reanalysis System (Zuo et al. 2019), which consists of a 5-member ensemble of eddy-permitting ocean/sea-ice analyses produced natively on the NEMO ORCA025 grid. Ocean and sea-ice initial conditions for ORCA1 experiments, termed INI-LRO, are generated via relaxation towards the higher resolution analysis. For each member of INI-HRO, an additional lower resolution experiment is run with identical

surface boundary conditions (including a -200 W/m 2 /K restoration towards observed SSTs) and additional relaxation terms that constrain 3D ocean temperature, 3D ocean salinity, sea ice thickness, and sea-ice concentration towards monthly mean values of INI-HRO. Relaxation timescales are 10 days for ocean variables and 5 days for sea-ice variables. This method provides balanced initial conditions suitable for initializing ORCA1 experiments that reproduce the high-frequency SST variability and lower-frequency sub surface variability of the INI-HRO ensemble of analyses. Roberts et al. (2020) provide further methodological details.

Table 1. Computational cost of the NEMO ocean model in different configurations of the coupled IFS. Estimated costs are derived from the timing statistics of a single 10 day forecast with IFS CY46R1.20190712.

ExpID	IFS precision	IFS config	Ocean grid	Total cost (SBUs)	Ocean cost (% of total)
hb4z	Double (64 bit)	Tco399L91, TSTEP=1200	ORCA025Z75, TSTEP=1200	3755	43 %
hb50	Double (64 bit)	Tco199L91, TSTEP=1800	ORCA025Z75, TSTEP=1200	1793	80 %
hb51	Single (32 bit)	Tco399L91, TSTEP=1200	ORCA025Z75, TSTEP=1200	2901	55 %
hb52	Single (32 bit)	Tco199L91, TSTEP=1800	ORCA025Z75, TSTEP=1200	1645	86 %
hb53	Double (64 bit)	Tco399L91, TSTEP=1200	ORCA1Z75, TSTEP=3600	2198	4 %
hb54	Double (64 bit)	Tco199L91, TSTEP=1800	ORCA1Z75, TSTEP=3600	400	11 %
hb55	Single (32 bit)	Tco399L91, TSTEP=1200	ORCA1Z75, TSTEP=3600	1347	6 %
hb56	Single (32 bit)	Tco199L91, TSTEP=1800	ORCA1Z75, TSTEP=3600	266	18 %

3 Impact of ocean resolution on SST biases

Figures 1 and 2 illustrate the impact of ocean model resolution on DJF sea surface temperature (SST) biases at different lead times in the North Atlantic and tropical oceans. The experimental data used to quantify these biases are summarized in table 2 and described in detail by Roberts et al. (2020). Importantly, the SST biases in INI-LRO and INI-HRO are extremely similar (figures 1a-b and 2a-b). This result demonstrates that the simple relaxation approach described in section 2 can adequately constrain the mean state of an eddy-parameterized ocean model using a higher resolution ocean analysis.

Table 2. Experiments used to evaluate SST biases at different lead times. Further details can be found in Roberts et al. (2020).

Name	Type	Members	Start dates	IFS cycle	Atm. grid	Ocean. grid	Reference
INI-HRO	Analysis	5				ORCA025_Z75, ~25 km	Zuo et al. (2019)
ENS-HRO	Forecast	5	Dec 1st, Jan 1st, Feb 1st	45r1	Tco319L91, ~31 km	ORCA025_Z75, ~25 km	Roberts et al. (2020)
SEAS-HRO	Forecast	5	Nov 1st	43r1 (SEAS5)	Tco319L91, ~31 km	ORCA025_Z75, ~25 km	Johnson et al. (2018)
INI-LRO	Analysis	5				ORCA1_Z75, ~100 km	Roberts et al. (2020)
ENS-LRO	Forecast	5	Dec 1st, Jan 1st, Feb 1st	45r1	Tco319L91, ~31 km	ORCA1_Z75, ~100 km	Roberts et al. (2020)
SEAS-LRO	Forecast	5	Nov 1st	43r1 (SEAS5)	Tco319L91, ~31 km	ORCA1_Z75, ~100 km	Roberts et al. (2020)

In the North Atlantic, both INI-HRO and INI-LRO have a notable warm bias that is associated with a northward displacement of the Gulf Stream (figure 1a-b). At a forecast lead time of 1-4 weeks, the SST biases in ENS-HRO and ENS-LRO configurations are very similar and in part inherited from the initial conditions (figure 1c-d). However, there is some evidence for a degradation in ENS-LRO relative to ENS-HRO, including a northward shift of the Gulf Stream and the development of a cold bias at the so-called northwest corner. The impact of ocean model resolution becomes more apparent at seasonal lead times (figure 1e-f). The North Atlantic SST biases in SEAS-HRO are similar to ENS-HRO, whereas SEAS-LRO exhibits more severe SST biases than ENS-LRO due to a more northerly Gulf Stream separation and further degradation in the northwest corner and Gulf Stream extension.

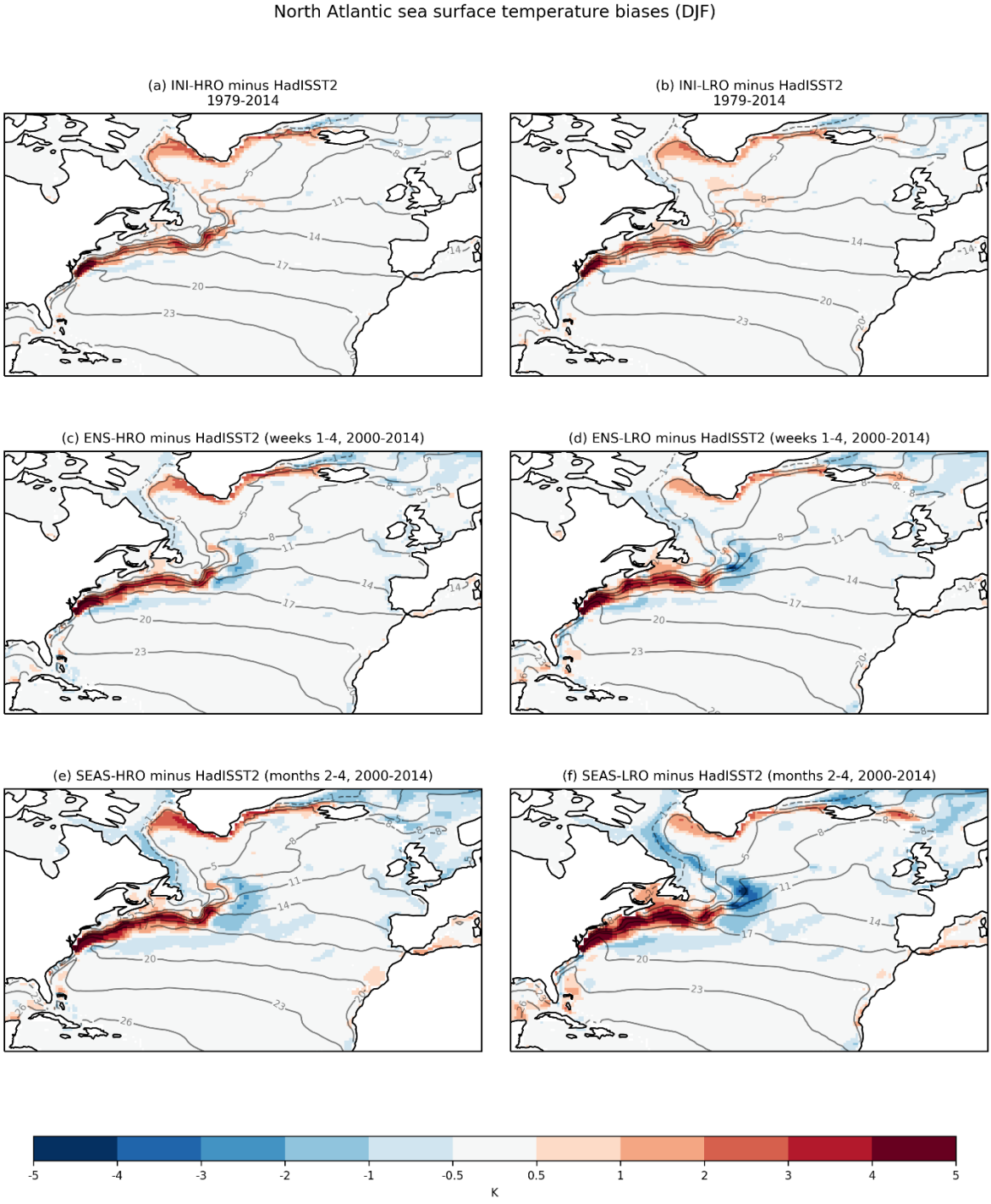

Figure 1. SST biases for the December-January-February (DJF) period relative to HadISST2 observations (Titchner and Rayner 2014). Contour lines indicate the model climatology in degC. Experimental details are summarized in table 2.

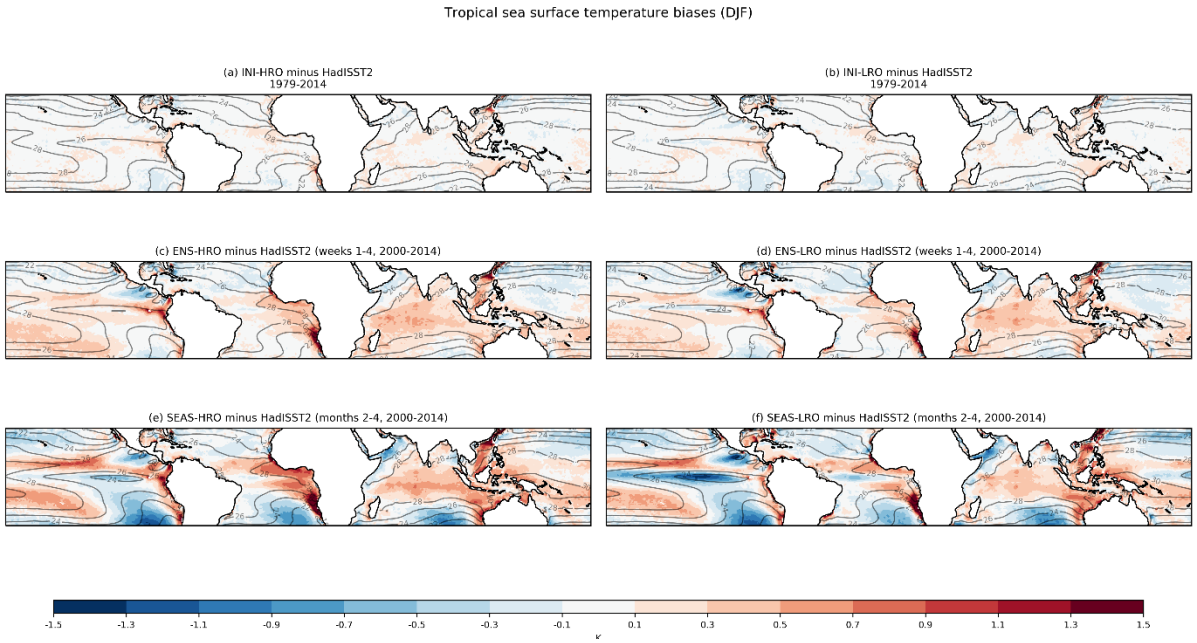

Figure 2. As figure 1, but for the tropical oceans.

Tropical SST biases are generally much smaller than those in the North Atlantic (note the different scales in figures 1 and 2). At lead times of 1-4 weeks, tropical SST biases in ENS-HRO and ENS-LRO are extremely similar to one another and are generally positive, particularly in the Indian Ocean and the regions bordering the maritime continent and the western coastlines of South America and Africa (figure 2c-d). At lead times of 2-4 months, SST biases are extremely similar in sub-tropical regions, but there is a marked degradation of an equatorial cold bias in the east Pacific in SEAS-LRO compared to SEAS-HRO (figure 2e-f).

At multidecadal timescales (not shown; see Roberts et al. 2018, 2020), it becomes clear that the ORCA025 and ORCA1 configurations of the coupled IFS have very different SST climatologies that arise due to fundamentally different representations of the ocean mesoscale. However, figures 1 and 2 demonstrate that SST biases at subseasonal lead times (i.e. 1-4 weeks) are relatively insensitive to the ocean model resolution of the forecast model, provided the ocean is initialized in a suitably consistent manner. However, more substantial differences between ORCA1 and ORCA025 configurations begin to emerge at seasonal lead times, particularly in the equatorial Pacific.

Further details on the impact of ocean resolution on the absolute performance of the IFS in the North Atlantic region are provided by Roberts et al. (2020). The main results from their study are summarized below:

- In general, mean biases are reduced in ORCA025 configurations relative to ORCA1 configurations and the magnitude of this impact is increased at longer lead times.
- Some aspects of air-sea interaction exhibit a clear improvement with increased ocean resolution at all lead times (weeks to decades).

- However, it is difficult to identify the impact of improved air-sea interaction and increased ocean eddy activity in the variability of the overlying atmosphere.
- Atmospheric blocking and the intensity of the storm track respond more strongly to mean biases than ocean eddy variability and thus have a larger response at longer lead times.
- Increased ocean resolution drives improvements to subseasonal predictability over Europe. This increase in skill comes from improvements to the Madden Julian Oscillation and its associated teleconnections rather than changes to air-sea interaction in the North Atlantic region.

4 Impact of ocean model resolution on the scientific assessment of IFS developments

4.1 Impact on medium-range forecasts (days 1-15)

This section compares the forecast performance of two recent IFS cycles (CY46R1 and CY47R1_v6) at lead times of 1-15 days. Figure 3 shows differences in the root mean square error (RMSE) of atmospheric temperature for deterministic forecasts with ORCA1 and ORCA025 configurations of IFS CY47R1_v6 (see table 3). As expected, the absolute performance of IFS CY47R1_v6 is significantly better when using NEMO ORCA025 ocean compared to the NEMO ORCA1 ocean.

Table 3. 10 day deterministic forecasts covering the period 2018/11/01 to 2019/03/15 (every 12h at 00 and 12 UTC). The total number of samples ranges from 250 to 269, depending on forecast lead time.

ExpID	IFS cycle	Atm. grid	Ocean. grid
h66e	CY46R1	Tco399L137, ~25 km	ORCA025_Z75, ~25 km
h9wr	CY46R1	Tco399L137, ~25 km	ORCA1_Z75, ~100 km
h8jj	CY47R1 (v6)	Tco399L137, ~25 km	ORCA025_Z75, ~25 km
h9xk	CY47R1 (v6)	Tco399L137, ~25 km	ORCA1_Z75, ~100 km

Figures 4 and 5 show differences between deterministic forecasts with CY46R1 and CY47R1_v6 estimated using ORCA1 ($\Delta RMSE_{LRO}$) and ORCA025 ($\Delta RMSE_{HRO}$) resolution configurations of the NEMO ocean model. From these plots it is clear that $\Delta RMSE_{LRO}$ is an extremely good approximation for $\Delta RMSE_{HRO}$ when considering geopotential height at lead times of 1-10 days. This result is consistent across other measures of forecast performance (e.g. anomaly correlation) and other atmospheric variables, including atmospheric temperature, relative humidity, and vector winds.

To complement our evaluation of deterministic forecasts, we also evaluate the impact of ocean resolution in medium-range ensemble forecasts (see table 4). Differences in ensemble forecast performance are quantified using the fair version of the continuous ranked probability score (FCRPS; Ferro et al. 2008; Ferro, 2014; Leutbecher, 2018):

(1.2)
$$FCRPS = \sum_{i=1}^{N} w_i \left(\frac{1}{M} \sum_{j=1}^{M} \left| x_i^j - y_i \right| - \frac{1}{2M(M-1)} \sum_{j=1}^{M} \sum_{k=1}^{M} \left| x_i^j - x_i^k \right| \right)$$

where x_i^j is member j of forecast i, y_i is the observed outcome, M is the number of ensemble members in each forecast, N is the total number of forecasts (all locations and start dates combined), and w_i is a weight for each forecast to account for variations in the area of grid-boxes as a function of latitude. For medium-range forecasts, FCRPS is computed from absolute forecast values without bias-correction. Figure 6 shows differences in FCRPS between CY47R1_v6 and CY46R1 estimated using the higher resolution ORCA025 ocean model (i.e. $\Delta FCRPS_{HRO}$) for a range of surface and atmospheric variables. Figure 7 is the companion plot computed using the lower resolution ORCA1 ocean model (i.e. $\Delta FCRPS_{LRO}$). These plots are extremely similar, though some differences can be identified under close inspection.

From this evaluation of deterministic and ensemble medium-range forecasts, we conclude that although the ORCA025 model outperforms the ORCA1 model in absolute terms, Δ_{LRO} can be an extremely good proxy for Δ_{HRO} at lead times of 1-15 days, provided that forecast systems are initialized in consistent manner.

Table 4. 15 day ensemble forecasts run for summer 2018 and winter 2018/19 seasons.

ExpID	IFS cycle	Forecast period	Members	Atm. grid	Ocean grid
h6mg	CY46R1	JJA 2018 (daily at 0 UTC)	8	Tco399L91, ~25 km	ORCA025_Z75, ~25 km
h6mf	CY46R1	NDJF 2018/19 (daily at 0 UTC)	8	Tco399L91, ~25 km	ORCA025_Z75, ~25 km
h9qa	CY46R1	JJA 2018 (daily at 0 UTC)	8	Tco399L91, ~25 km	ORCA1_Z75, ~100 km
h9qd	CY46R1	NDJF 2018/19 (daily at 0 UTC)	8	Tco399L91, ~25 km	ORCA1_Z75, ~100 km
h8k9	CY47R1 (v6)	JJA 2018 (daily at 0 UTC)	8	Tco399L91, ~25 km	ORCA025_Z75, ~25 km
h8k8	CY47R1 (v6)	NDJF 2018/19 (daily at 0 UTC)	8	Tco399L91, ~25 km	ORCA025_Z75, ~25 km
h9qc	CY47R1 (v6)	JJA 2018 (daily at 0 UTC)	8	Tco399L91, ~25 km	ORCA1_Z75, ~100 km
h9qb	CY47R1 (v6)	NDJF 2018/19 (daily at 0 UTC)	8	Tco399L91, ~25 km	ORCA1_Z75, ~100 km

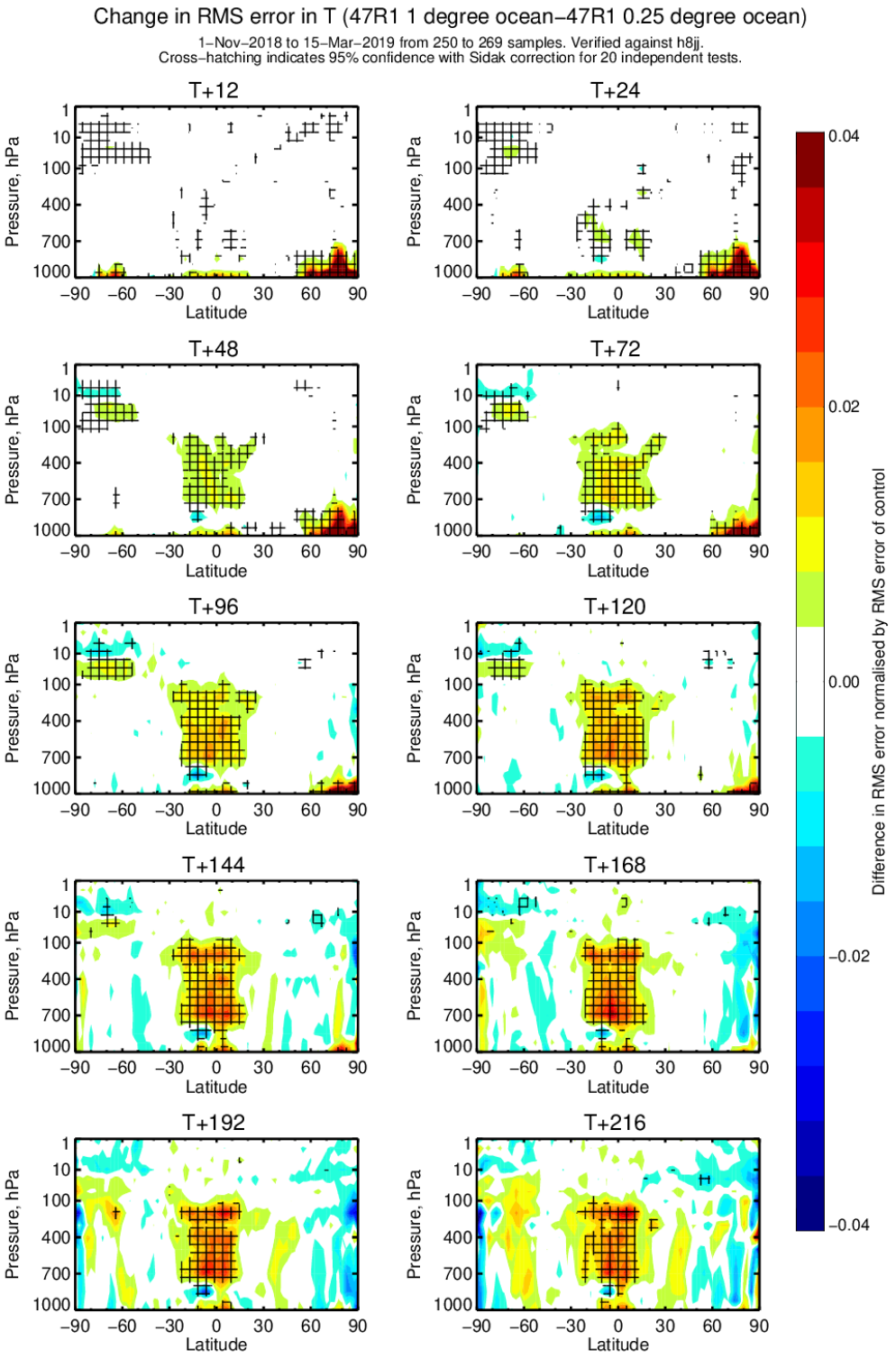

Figure 3. Difference in root mean square error of temperature in deterministic forecasts as a function of lead time (hours). Positive values correspond to larger errors in the lower resolution ORCA1 ocean configuration (h9xk) compared to the higher resolution ORCA025 ocean configuration (h8jj). Experimental details are provided in table 3.

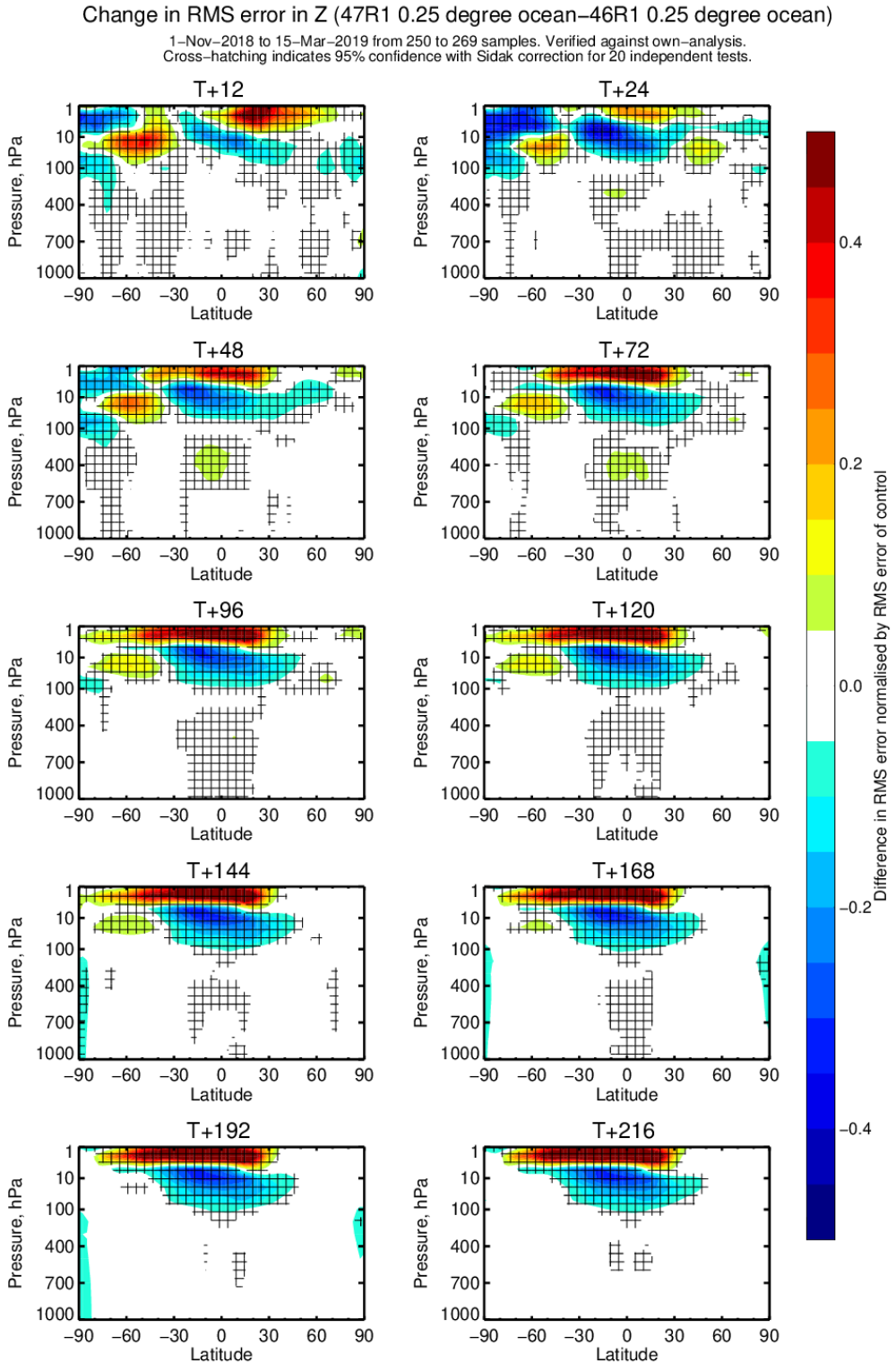

Figure 4. Difference in RMSE of geopotential height between CY47R1_v6 (h8jj) and CY46R1 (h66e) estimated using the higher resolution ORCA025 ocean model (i.e. $\Delta RMSE_{HRO}$). Negative values indicate improved performance in CY47R1_v6 relative to CY46R1. Experimental details are provided in table 3.

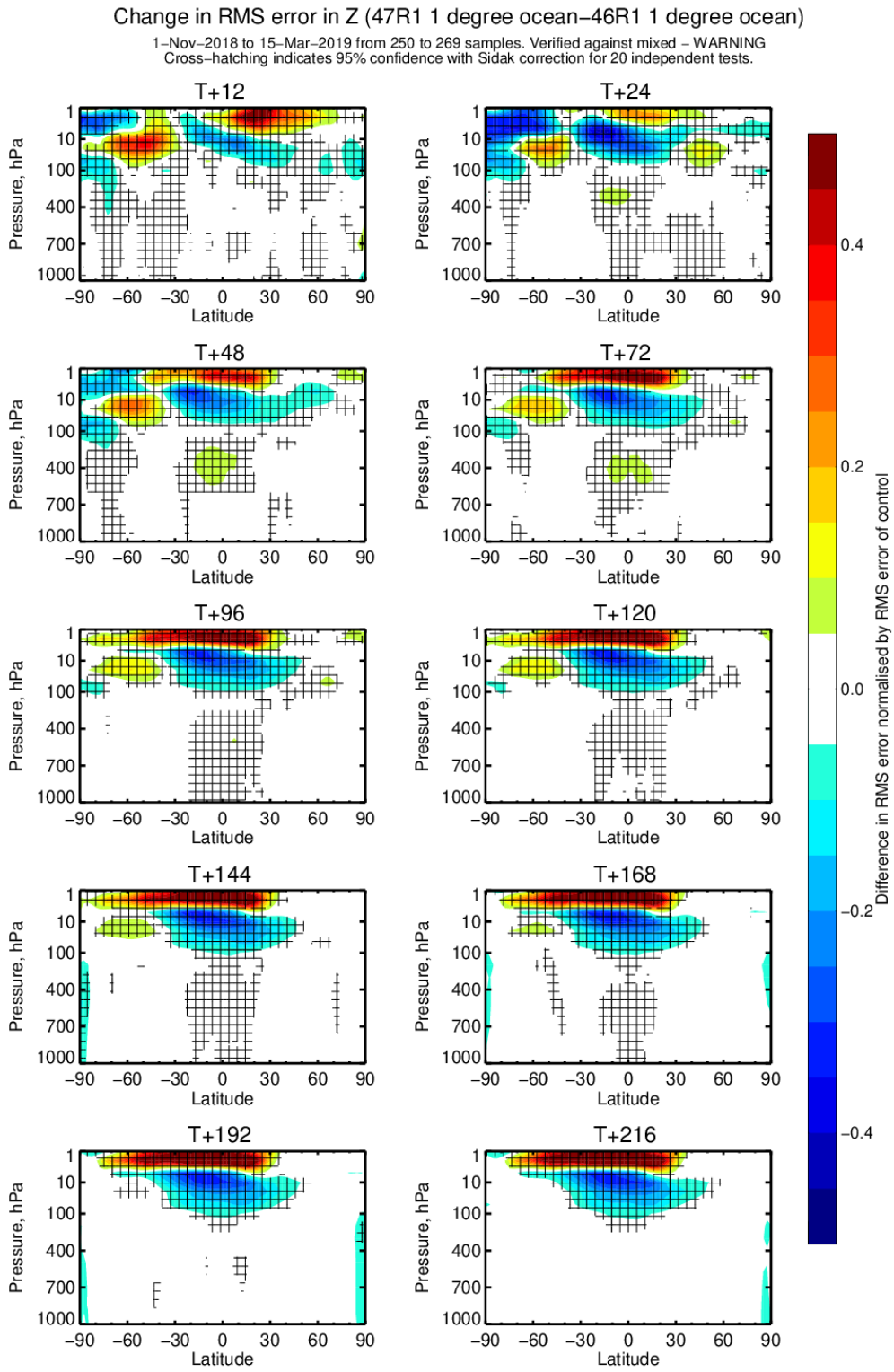

Figure 5. Difference in RMSE of geopotential height between CY47R1_v6 (h9xk) and CY46R1 (h9wr) estimated using the lower resolution ORCA1 ocean model (i.e. $\Delta RMSE_{LRO}$). Negative values indicate improved performance in CY47R1_v6 relative to CY46R1. Experimental details are provided in table 3.

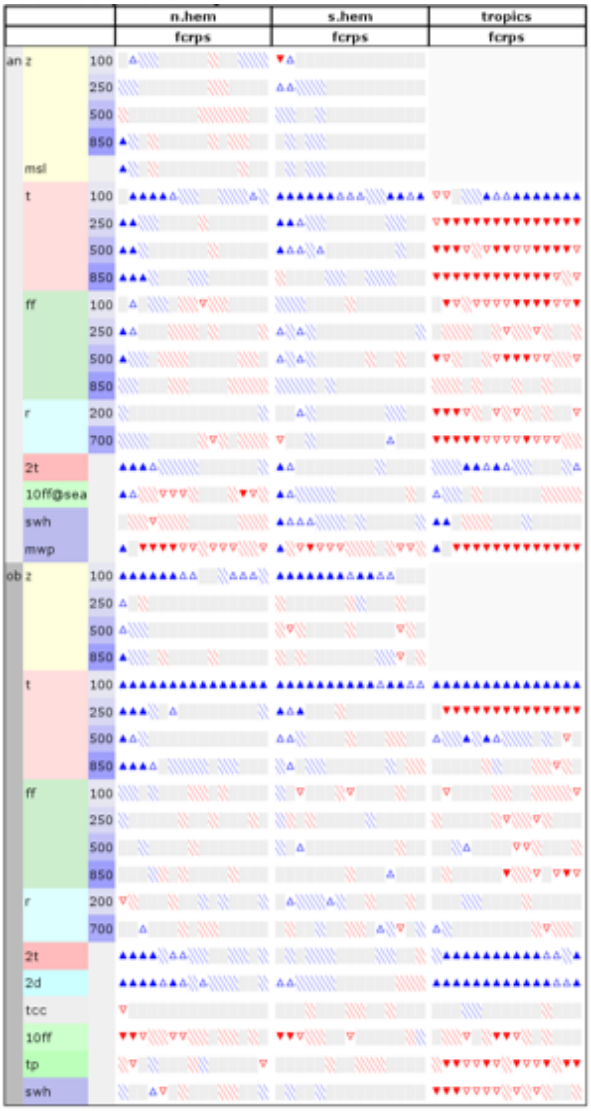

Figure 6. Scorecard summarizing differences in FCRPS between CY47R1_v6 (h8k8 and h8k9) and CY46R1 (h6mg and h6mf) estimated using the higher resolution ORCA025 ocean model (i.e. $\Delta FCRPS_{HRO}$). Scored variables include geopotential height at selected pressure levels (z), mean sea level pressure (msl), temperature at selected pressure levels (t), relative humidity at selected pressure levels (r), 2m temperature (2t), wind speed at selected pressure levels (ff), 10m wind speed (10ff), significant wave height (swh), mean wave period (mwp), 2m dewpoint temperature (2d), total cloud cover (tcc), and total precipitation (tp). Solid triangles correspond to changes that are 99.7% statistically significant and open triangles to changes that are 95% statistically significant. Positive impacts in CY47R1_v6 relative to CY46R1 (i.e. reduced FCRPS) are shown as blue triangles pointing upwards. Negative impacts are shown as red triangles pointing downwards. The three columns correspond to the northern hemisphere extratropics (20°N-90°N), the southern hemisphere extratropics (20°S-90°S), and the tropics (20°S-20°N). Each triangle corresponds to a different lead time, varying from 1 day (left) to 15 days (right). The top half of the scorecard shows verification against own analysis and the bottom half verification against radiosondes, SYNOP station data and buoys.

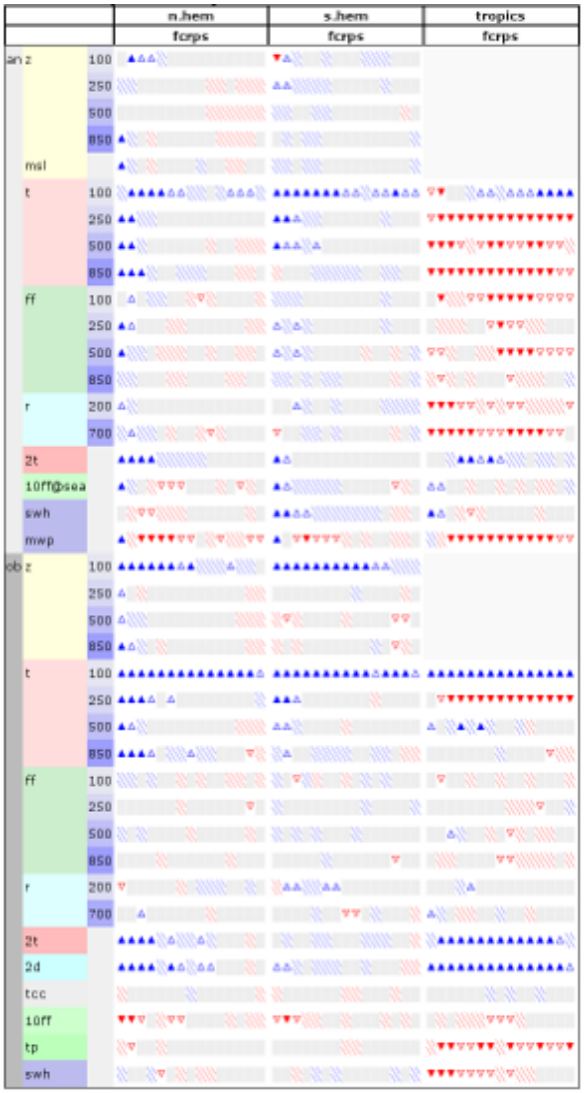

Figure 7. As figure 7 but comparing differences between CY47R1_v6 (h9qc and h9qb) and CY46R1 (h9qa and h9qd) estimated using the lower resolution ORCA1 ocean model (i.e. $\Delta FCRPS_{LRO}$).

4.2 Impact on extended-range forecasts (weeks 1-4)

This section compares the performance of IFS cycles CY45R1 (implemented June 2018) and CY43r3 (implemented July 2017) at lead times of weeks 1-4. Although these are different versions of the IFS to those considered in section 4.1, the approach of comparing differences between cycles at different resolutions is identical. In addition to comparing the impact of changes in ocean resolution, we also compare the Tco199 (~50 km) and Tco319 (~31 km) atmospheric grids to provide some context for the relative importance of ocean and atmospheric resolution. The experiments used for these comparisons are detailed in table 5.

Table 5. Extended-range ensemble forecasts run for the period 1989-2016.

ExpID	IFS cycle	Start dates	Members	Atm. grid	Ocean grid
gvnj	CY45R1	1st of each month	15	Tco319L91, ~31 km	ORCA025_Z75, ~25 km
h6ob	CY45R1	1st of each month	15	Tco199L91, ~50 km	ORCA025_Z75, ~25 km
h0cc	CY45R1	1st of each month	15	Tco199L91, ~50 km	ORCA1_Z75, ~100 km
gs67	CY43R3	1st of each month	15	Tco319L91, ~31 km	ORCA025_Z75, ~25 km
h6oc	CY43R3	1st of each month	15	Tco199L91, ~50 km	ORCA025_Z75, ~25 km
h0b2	CY43R3	1st of each month	15	Tco199L91, ~50 km	ORCA1_Z75, ~100 km

Extended-range forecast skill is quantified using the following skill score

$$FCRPSS = 1 - \frac{FCRPS}{FCRPS_{ref}} \tag{1.3}$$

where FCRPS is computed using forecast anomalies relative to a hindcast climatology (defined as a function of lead-time and start date) and $FCRPS_{ref}$ is a reference score derived using climatological forecasts. Figures 8 and 9 show differences in FCRPSS between extended-range forecasts with cycles CY45R1 and CY43r3 estimated using the Tco199 ORCA1 configuration ($\Delta FCRPSS_{LRO\text{-}LRA}$) and Tco199 ORCA025 configuration ($\Delta FCRPSS_{HRO\text{-}LRA}$). From figures 8 and 9 it is evident that $\Delta FCRPSS_{LRO\text{-}LRA}$ is a generally good approximation for $\Delta FCRPSS_{HRO\text{-}LRA}$ in extended-range forecasts, although some differences are evident at lead times of 3-4 weeks. This outcome is consistent with our results from section 4.1 and further evidence that it is possible to do useful testing and model development with the ORCA1 ocean model.

To provide some context to the differences between $\Delta FCRPSS_{LRO\text{-}LRA}$ and $\Delta FCRPSS_{HRO\text{-}LRA}$, we also evaluate the impact of a change in atmospheric resolution. Figures 9 and 10 show differences in FCRPSS estimated using the Tco199 ORCA025 configuration ($\Delta FCRPSS_{HRO\text{-}LRA}$) and Tco319 ORCA025 configuration ($FCRPSS\Delta_{HRO\text{-}HRA}$). Although there are strong similarities between the summary scorecards, the experiments run with the lower resolution Tco199 atmosphere generally give a more

positive impression of CY45R1 at lead times of 3-4 weeks. From this comparison we conclude that there is a stronger sensitivity to changes in atmospheric resolution (50 km to 31 km) than changes in ocean resolution (100 km to 25 km) when evaluating cycles at extended-range lead times.

The estimates of $\Delta FCRPSS_{LRO\text{-}LRA}$ and $\Delta FCRPSS_{HRO\text{-}LRA}$ presented in figures 8-10 are computed using forecast anomalies relative to a hindcast climatology, which means they are insensitive to changes in the mean state (except through non-linear interactions between the mean state and forecast variability). However, at longer lead times, we expect definitions of Δ that include a dependence on the mean state to be more sensitive to changes in ocean resolution due to the divergence of SST biases (see section 2). To evaluate the impact of ocean resolution on estimates of absolute bias in extended-range forecasts we compute the following score

$$BiasScore = 1 - \frac{\sum_i w_i \left| \langle F \rangle_i - \langle O \rangle_i \right|}{\sum_i w_i \left| \langle G \rangle_i - \langle O \rangle_i \right|} \tag{1.4}$$

where $\langle F \rangle_i$ and $\langle G \rangle_i$ represent forecast climatologies for a given location (i), $\langle O \rangle_i$ represents the observed climatology, and w_i is a weight to account for variations in grid-box area. A positive value indicates that the weighted average of absolute biases in forecast F are reduced compared to forecast G.

Estimates of *BiasScore* for cycle CY45R1 relative to CY43r3 are shown in figures 11 and 12 for extended-range forecasts initialized on March 1st. Despite some similarities, there are a number of differences when *BiasScore* is estimated using the using the Tco199 ORCA1 configuration ($BiasScore_{LRO\text{-}LRA}$) and Tco199 ORCA025 configuration ($BiasScore_{HRO\text{-}LRA}$). However, forecast biases are generally very similar in CY45R1 relative to CY43r3 so none of the individual *BiasScore* estimates are significant at the 99% level. Therefore, the difference between figures and 11 and 12 could be a result of sampling noise. We find similar results for other start dates (not shown).

To summarize, Δ_{LRO} can be a good proxy for Δ_{HRO} at lead times of 1-4 weeks, particularly for metrics that are computed from forecast anomalies relative to a hindcast climatology. However, this approximation is likely to become less accurate at longer lead times for metrics that are a non-linear function of the mean state, such as the *BiasScore* defined above.

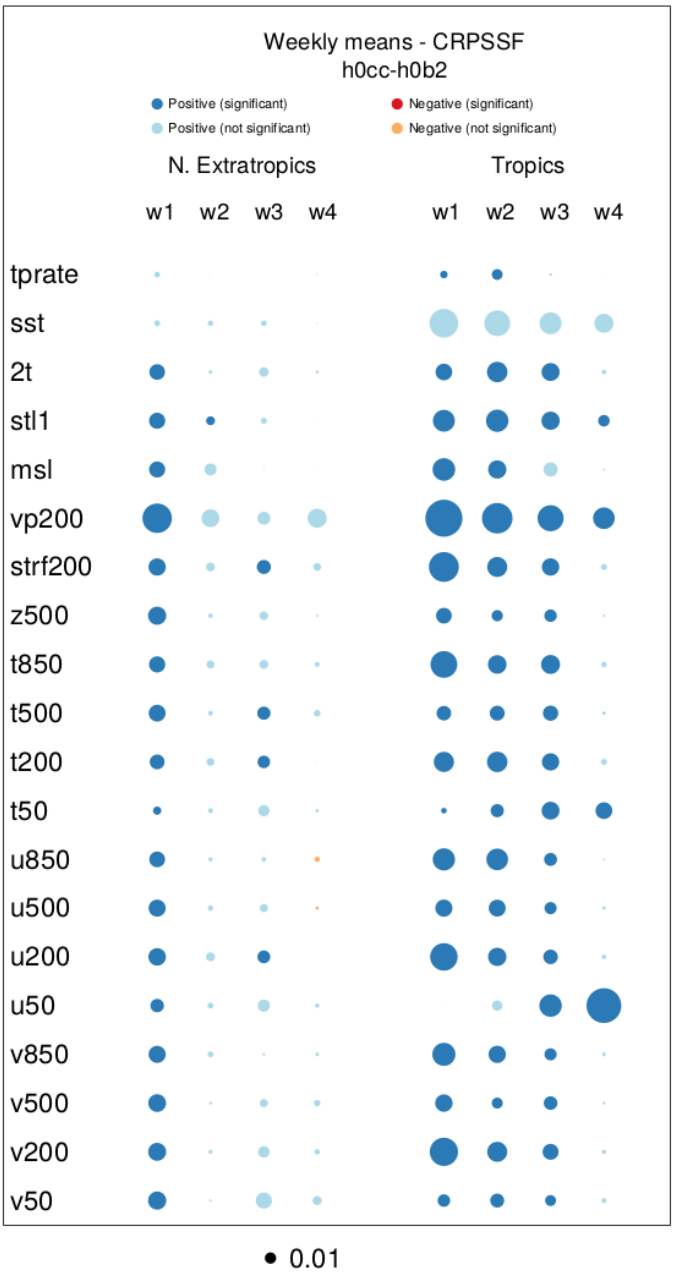

Figure 8. Scorecard summarizing differences in FCRPSS between CY45R1 (h0cc) and CY43R3 (h0b2) at lead times of 1-4 weeks estimated using the ORCA1 NEMO grid and Tco199 atmosphere (i.e. $\Delta FCRPSS_{LRO-LRA}$). Regions are defined as the northern extratropics (180°W-180°E, 30°N-90°N) and tropics (180°W-180°E, 30°S-30°N). Differences in skill are proportional to the area of the black dot (1 % change) and darker shading corresponds to differences that are determined to be significant at the 99% level. The variables shown are precipitation over land (tp), 2m air temperature over land (2t), surface temperature (stl1), sea surface temperature (sst), mean sea level pressure (msl), temperature at selected pressure levels (t), zonal wind at pressure levels (u), meridional wind at pressure levels (v), streamfunction at 200 hPa (sf200), velocity potential at 200 hPa (vp200), and geopotential height at 500 hPa (z500). Blue corresponds to improvements in CY45R1 relative to CY43R3.

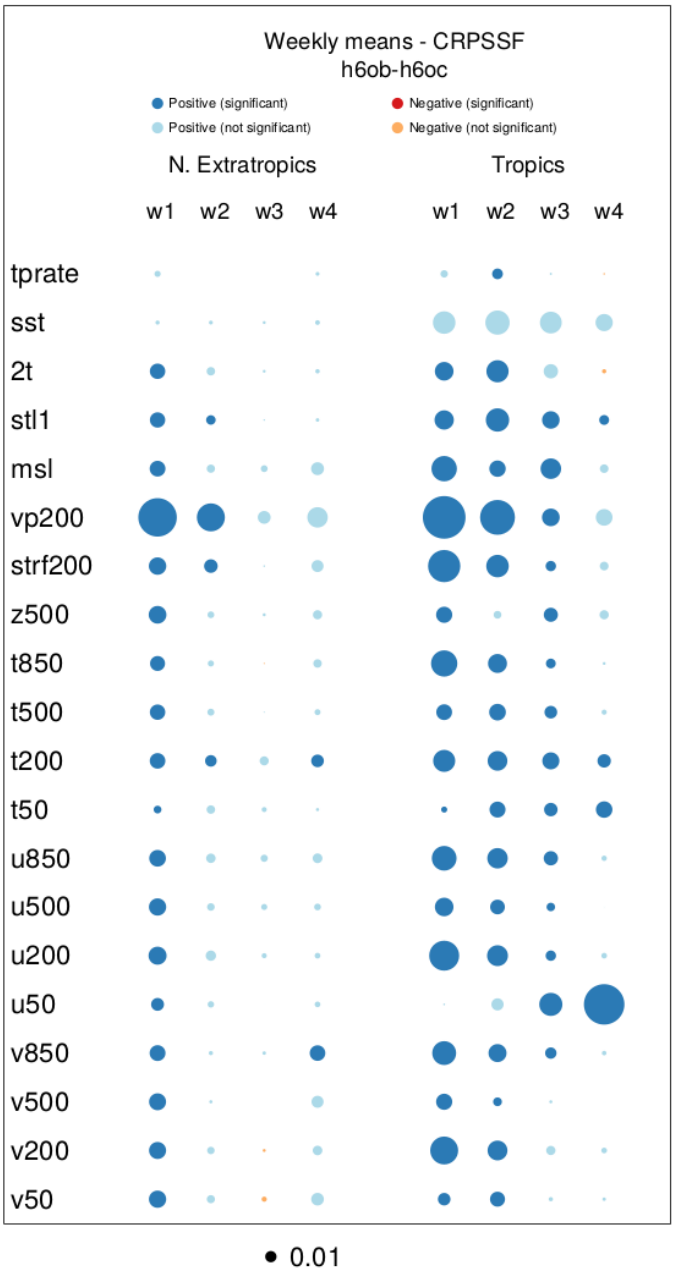

Figure 9. As figure 8, but summarizing differences in FCRPSS between CY45R1 (h6ob) and CY43R3 (h6oc) using the ORCA025 NEMO grid and Tco199 atmosphere (i.e. $\Delta FCRPSS_{HRO-LRA}$).

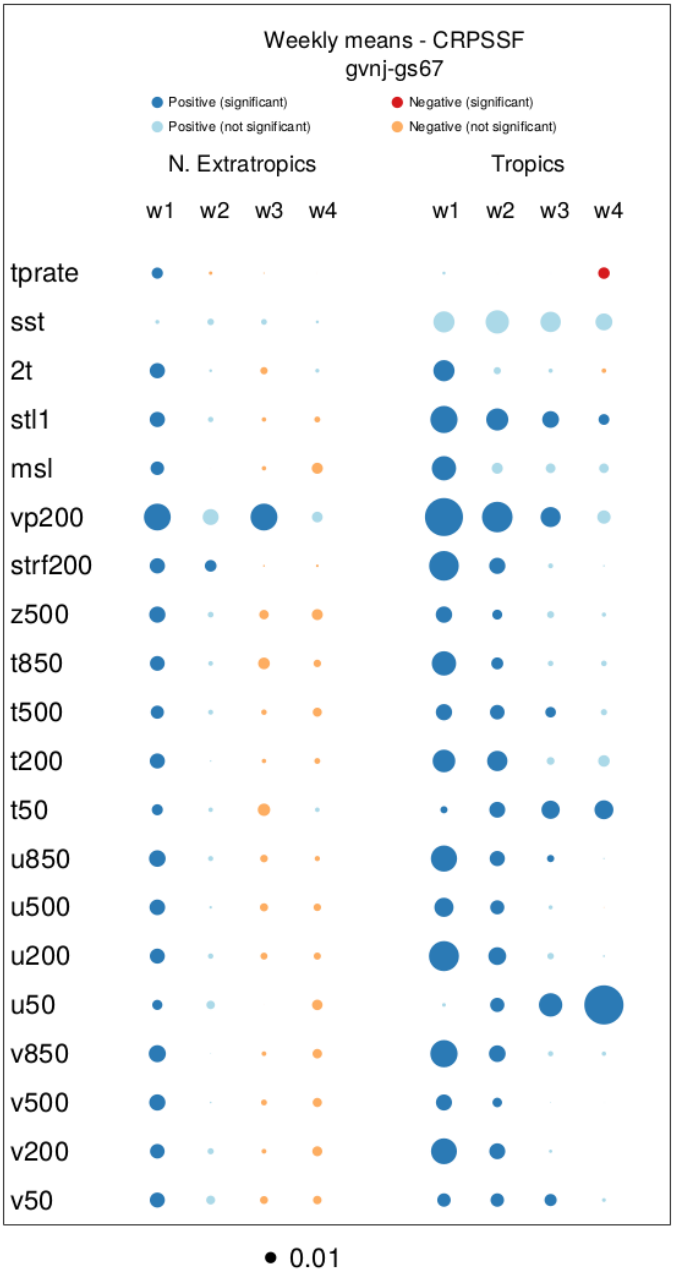

Figure 10. As figure 8, but summarizing differences in FCRPSS between CY45R1 (gvnj) and CY43R3 (gs67) using the ORCA025 NEMO grid and Tco319 atmosphere (i.e. $\Delta FCRPSS_{HRO-HRA}$).

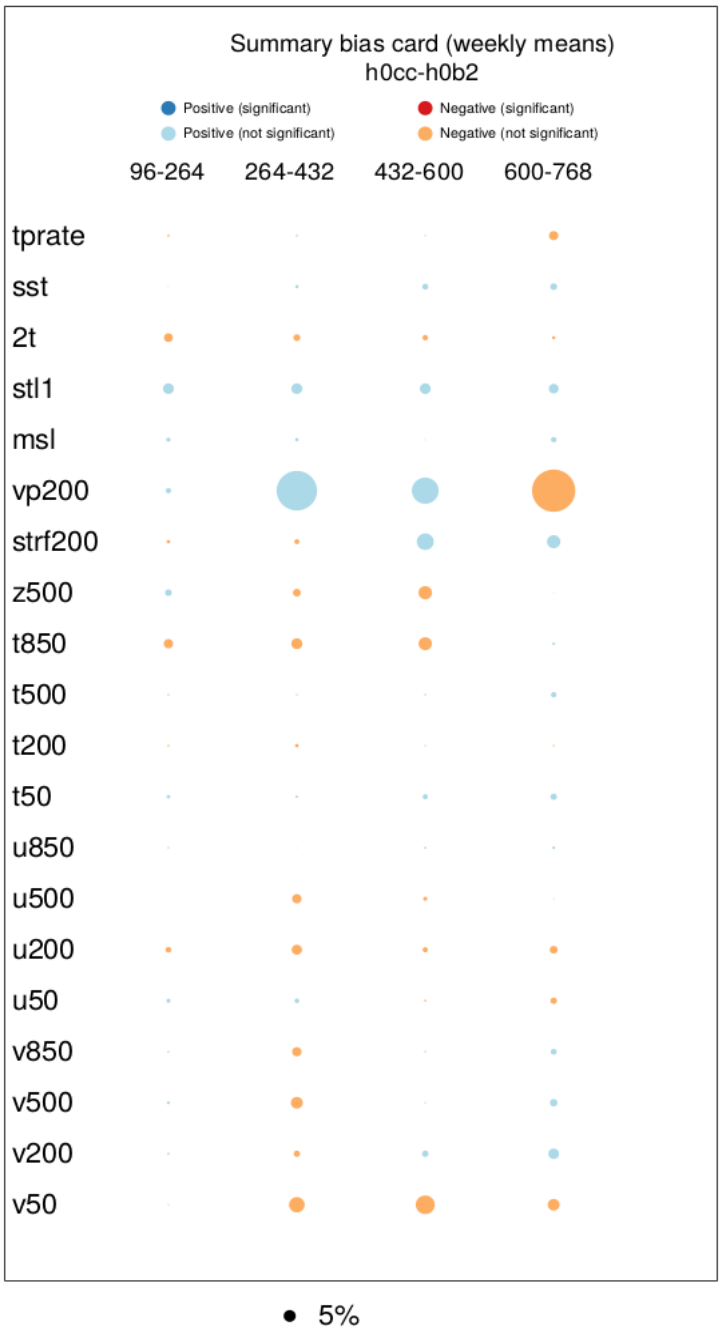

Figure 11. Estimates of BiasScore (equation 1.3) for CY45R1 (h0cc) relative to CY43R3 (h0b2) using the ORCA1 NEMO grid and Tco199 atmosphere (i.e. $BiasScore_{LRO-LRA}$). Scores are computed over the global domain using forecasts initialized on March 1st and are shown for different lead times (weekly means specified by start/end hours). Marker sizes are proportional to the area of the black dot (0.05 change) and darker shading corresponds to differences that are determined to be significant at the 99% level. Variable definitions are defined in the caption to figure 8.

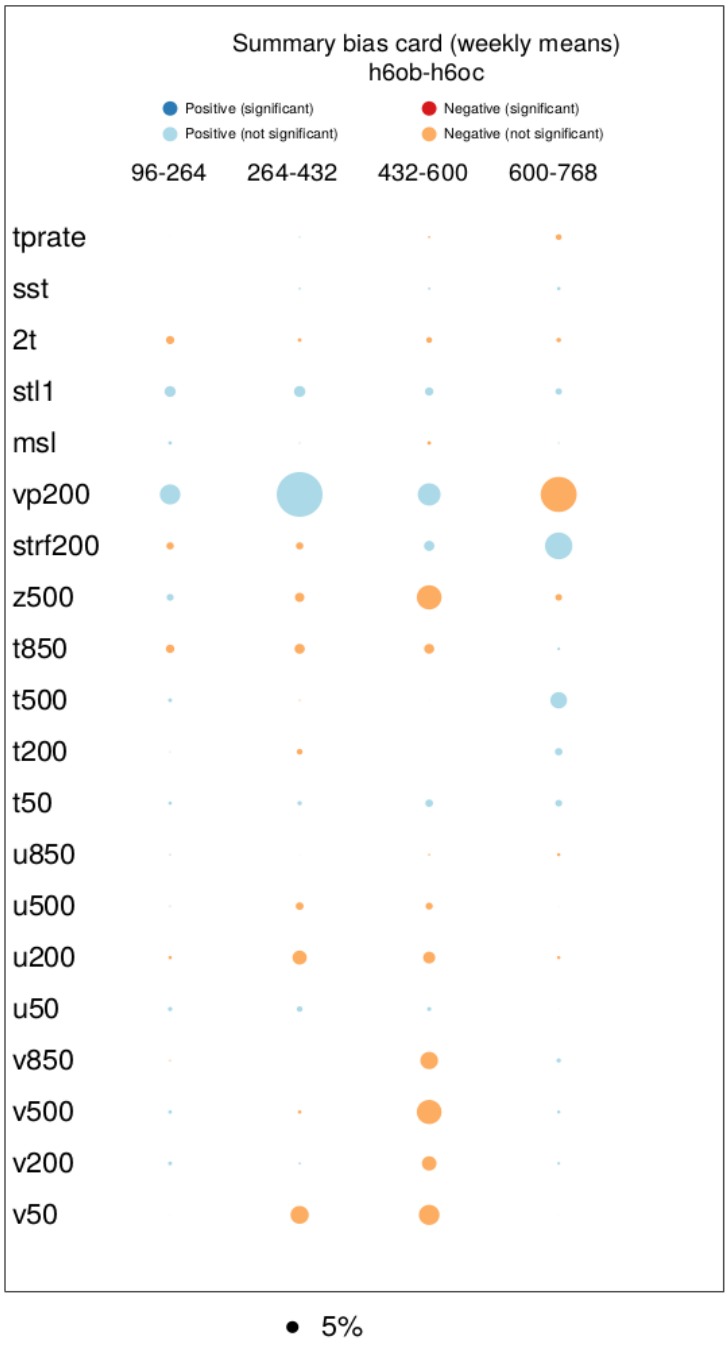

Figure 12. As figure 11 but comparing CY45R1 (h6ob) relative to CY43R3 (h6oc) using the ORCA025 NEMO grid and Tco199 atmosphere (i.e. BiasScore $_{HRO-LRA}$).

4.3 Impact on seasonal forecasts (months 1-7)

The latest version of the operational ECMWF seasonal forecast system is based on CY43R1 (SEAS5; Johnson et al. 2018). This section compares the seasonal performance of CY43R1 against CY45R1, a more recent cycle which introduced several changes to the IFS stochastic physics scheme that have a noticeable impact at seasonal lead times. We quantify this impact (i.e. Δ) using a higher-resolution configuration with a Tco319 atmosphere and ORCA025 ocean that reflects the operational system (i.e. $\Delta_{HRO-HRA}$) and a lower-resolution configuration with a Tco199 atmosphere and ORCA1 ocean (i.e. $\Delta_{LRO-LRA}$). The seasonal reforecast experiments discussed in this section are detailed in table 6.

Table 6. Seasonal reforecasts run for the period 1981-2016.

ExpID	IFS cycle	Start dates	Members	Atm. grid	Ocean grid
gws5	CY45R1	1st May, 1st Nov	25	Tco319L91, ~31 km	ORCA025_Z75, ~25 km
h269	CY45R1	1st May, 1st Nov	25	Tco199L91, ~50 km	ORCA1_Z75, ~100 km
SEAS5	CY43R1 (SEAS5)	1st May, 1st Nov	25	Tco319L91, ~31 km	ORCA025_Z75, ~25 km
gv6w	CY43R1 (SEAS5)	1st May, 1st Nov	25	Tco199L91, ~50 km	ORCA1_Z75, ~100 km

This section differs from sections 4.1 and 4.2 as we now compare the combined impact of changes to ocean and atmosphere resolution. However, the approach of comparing differences between cycles at different resolutions is otherwise identical. Figures 13-16 show differences in the mean state between CY43R1 and CY45R1 estimated using the higher-resolution (i.e. $\Delta MODEL_{HRO-HRA}$) and lower-resolution (i.e. $\Delta MODEL_{LRO-LRA}$) seasonal configurations. From these comparisons it is clear than $\Delta MODEL_{LRO-LRA}$ is a good approximation for $\Delta MODEL_{HRO-HRA}$ across a range of variables. However, whether or not these changes represent an improvement may be sensitive to the model mean state as discussed in section 1.

Figure 17 illustrates the evolution of SST biases in the equatorial Pacific (NINO3.4 region) in all four seasonal reforecast experiments. Although equatorial Pacific SSTs are degraded in the lower resolution ocean configurations. For this case, there is some consistency across resolutions such that lower and higher resolution forecasts initialized on Nov 1st have degraded (cooler) NINO3.4 SSTs in CY45R1 relative to CY43R1. Furthermore, both configurations have increased spread of NINO3.4 SSTs in CY45R1 relative to CY43R1 with very little change to the root mean square error of the bias-corrected ensemble mean (figure 18).

Figures 19 and 20 summarize differences in the regional weighted mean absolute error (MAE) of monthly means for other regions, variables, and seasons. MAE is defined as follows

$$MAE = \sum_{i=1}^{N} w_i \left| \overline{x}_i - o_i \right| \tag{1.5}$$

where N is the total number of forecasts (across all start dates, forecast months, and locations), w_i is a normalized weight to account for variations in the area of grid-boxes as a function of latitude, $\overline{x}_i$ is a monthly forecast ensemble mean, and o_i is the observed monthly mean. From this comparison it is evident that $\Delta MAE_{LRO-LRA}$ is generally a reasonable approximation for $\Delta MAE_{HRO-HRA}$ when the change is considered to be statistically significant. However, there are some cases where the sign and significance of ΔMAE is sensitive to horizontal resolution. For example, $\Delta MAE_{LRO-LRA}$ is positive (significant) whereas $\Delta MAE_{HRO-HRA}$ is negative (not significant) for MAM temperature at 500 hPa evaluated over the global domain. Other areas where the sign of ΔMAE is sensitive to resolution (though not significant) include DJF precipitation over the west Pacific and DJF/SON 2m temperature over Europe.

We evaluate the probabilistic performance of seasonal reforecasts at each grid-point using the continuous ranked probability score CRPS, which is defined as follows:

$$CRPS = \sum_{i=1}^{N} \int_{-\infty}^{\infty} \left[F_i(x) - O_i(x) \right]^2 dx \tag{1.6}$$

where $F_i(x)$ is the forecast cdf, $O_i(x)$ is the observed cdf, and N is the total number of start dates. As for the FCRPS computed in section 4.2, CRPS is computed using forecast anomalies relative to a hindcast climatology and skill scores are calculated following equation 1.3.

Figures 21-24 show differences in CRPSS between CY43R1 and CY45R1 for the DJF season estimated using the higher-resolution (i.e. $\Delta CRPSS_{HRO-HRA}$) and lower-resolution (i.e. $\Delta CRPSS_{LRO-LRA}$) seasonal configurations. The grid-point estimates of ΔCRPSS are generally rather noisy, with limited statistical significance. However, there are some statistically significant signals that are consistent across higher and lower resolution configurations. For example, there is a strong signal for improvements to SST and 2m temperature across the Southern Ocean in CY45R1 relative to 43R1 and, in this case, $\Delta CRPSS_{LRO-LRA}$ is a good approximation for $\Delta CRPSS_{HRO-HRA}$. However, we also identify the following cases where $\Delta CRPSS_{LRO-LRA}$ is a poor approximation for $\Delta CRPSS_{HRO-HRA}$: (i) the sign and significance of ΔCRPSS is inconsistent for SSTs for some locations along the sea-ice edge in the Arctic, (ii) $\Delta CRPSS_{LRO-LRA}$ is significant and positive for SSTs in the central equatorial Pacific, whereas $\Delta CRPSS_{HRO-HRA}$ is more neutral, and (iii) $\Delta CRPSS_{LRO-LRA}$ is significant and negative for mean sea level pressure over selected parts of Northern Africa, whereas $\Delta CRPSS_{HRO-HRA}$ is neutral.

To summarize, we have demonstrated using a range of variables that lower resolution seasonal configurations can be a good approximation for higher resolution seasonal configurations when estimating changes to the model climatology (i.e. $\Delta MODEL_{LRO-LRA}$ is a good approximation for $\Delta MODEL_{HRO-HRA}$). However, whether or not these changes represent an improvement may be sensitive to the model mean state, which is likely to be sensitive to ocean resolution at seasonal lead times (see section 2). Estimates of changes in forecast skill can also be sensitive to resolution at seasonal lead times. In particular, we identified several cases where $\Delta CRPSS_{LRO-LRA}$ was a poor analogue for $\Delta CRPSS_{HRO-HRA}$. Although we are unable to separate the relative role of ocean and atmosphere resolution

using the seasonal experiments described in table 6, the divergence of SST biases described in section 2 is consistent with Δ_{LRO} becoming a less valid approximation for Δ_{HRO} at seasonal and longer timescales.

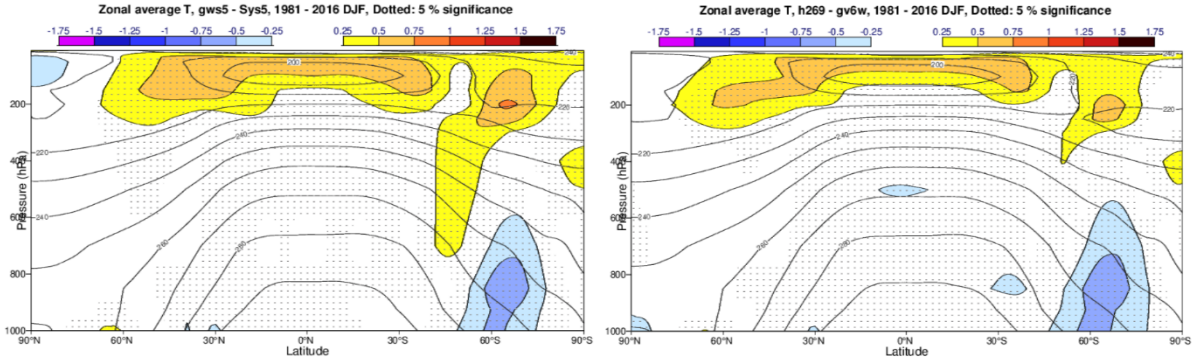

Figure 13. DJF difference in zonal mean temperature (T) between CY45R1 and CY43R1 estimated using higher-resolution ($\Delta MODEL_{HRO-HRA}$, left) and lower-resolution ($\Delta MODEL_{LRO-LRA}$, right) seasonal configurations. See table 6 for details of the reforecast experiments. Differences are displayed as CY45R1 minus CY43R1.

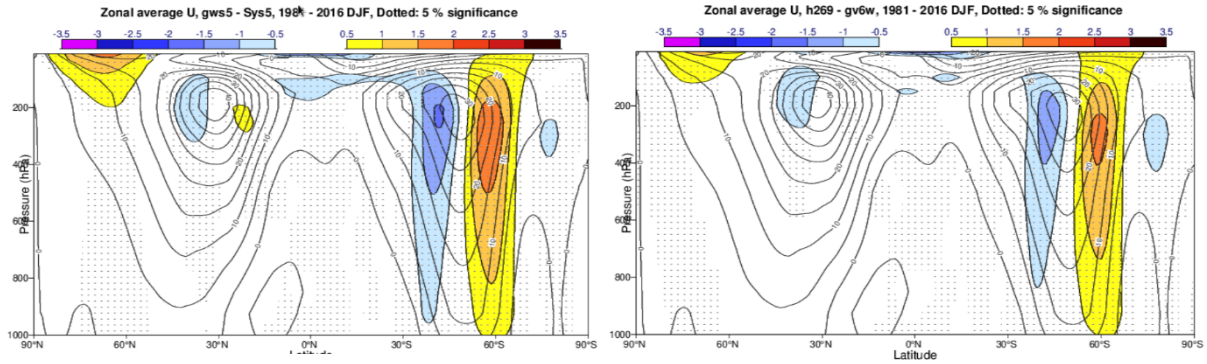

Figure 14. As figure 13, but for zonal mean zonal wind (U).

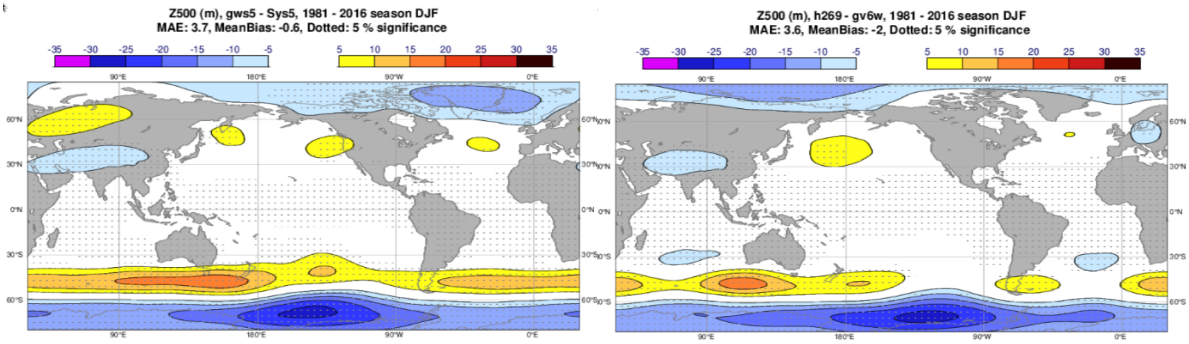

Figure 15. As figure 13, but for geopotential height at 500 hPa (Z500).

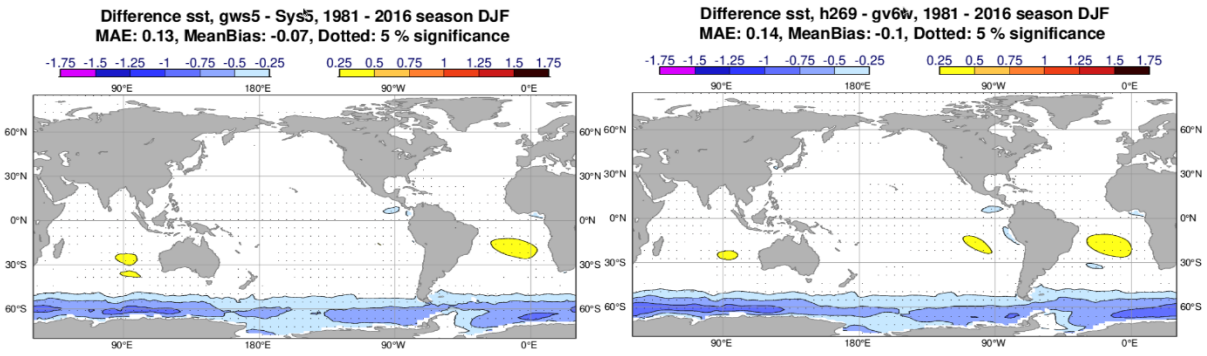

Figure 16. As figure 13, but for sea surface temperatures (SST).

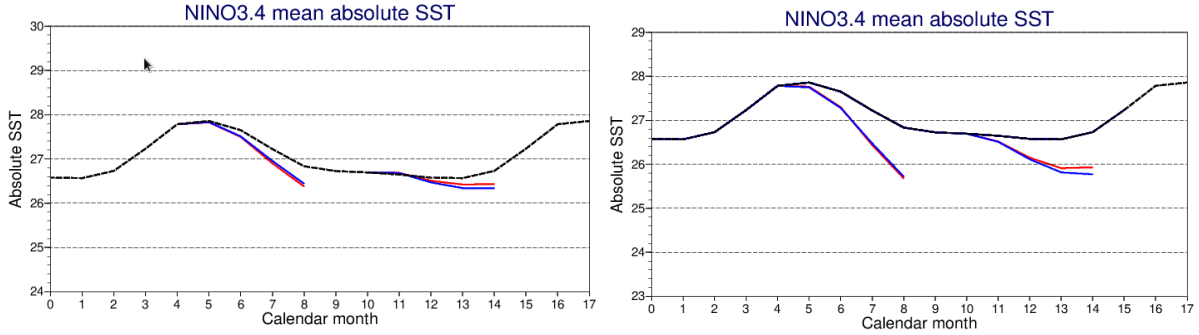

Figure 17. Sea surface temperature climatologies as a function of start date and lead time in the equatorial Pacific (NINO3.4 region) for HRO-HRA (left panel) and LRO-LRA (right panel) seasonal configurations (CY43R1 in red, CY45R1 in blue). The black line is the observed climatology.

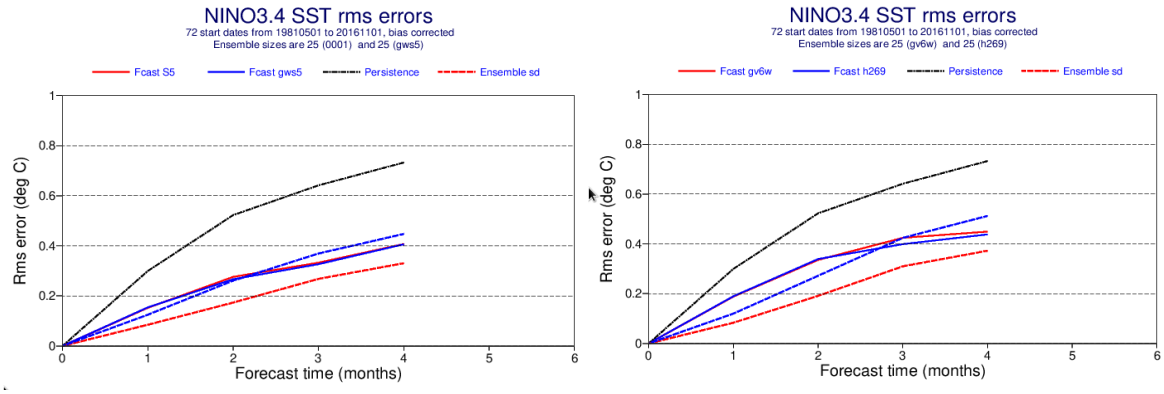

Figure 18. Root mean square error (solid lines) and ensemble standard deviation (dashed lines) of bias corrected sea-surface temperature anomalies as a function of lead time. Anomalies reflect averages over the NINO3.4 region for HRO-HRA (left panel) and LRO-LRA (right panel) seasonal configurations (CY43R1 in red, CY45R1 in blue). The black line is the RMS error of a persistence forecast.

Figure 19. Scorecard summarizing differences in MAE between CY45R1 (gws5) and CY43R1 (SEAS5) estimated using the ORCA025 NEMO grid and Tco319 atmosphere (i.e. $\Delta MAE_{HRO-HRA}$). Forecasts are initialized on Nov 1st (DJF, MAM) and May 1st (JJA, SON). The variables shown are sea surface temperature (sst), 2m air temperature (t2m), total precipitation (precip), mean sea level pressure (mslp), temperature at selected pressure levels (T), zonal wind at pressure levels (U), total cloud cover (TCC), net solar radiation at the top of the atmosphere (TSR), net thermal radiation at the top of the atmosphere (TTR), and total column water vapour (TCWV). Green corresponds to improvements in CY45R1 relative to CY43R1.

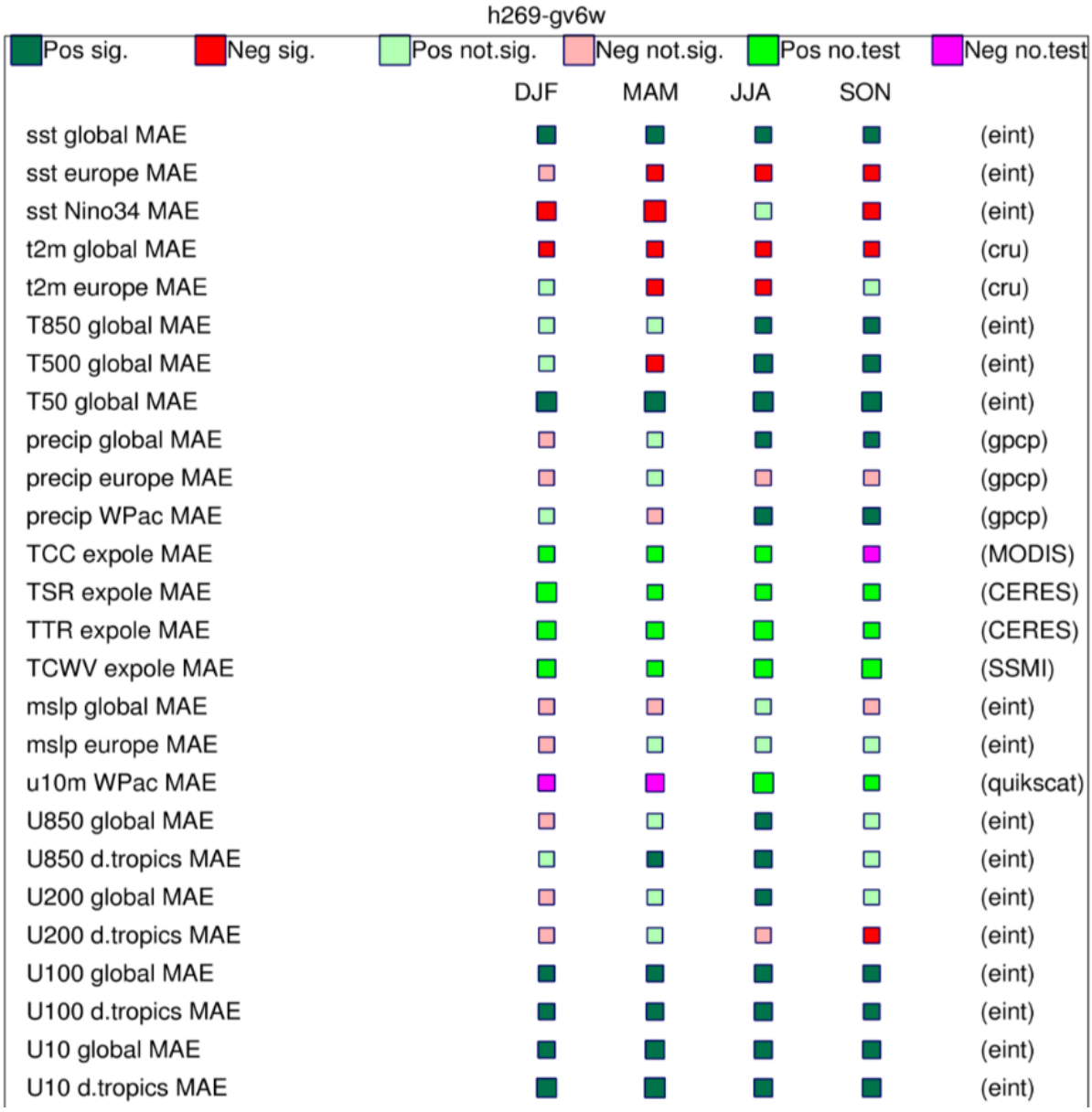

Figure 20. As figure 17,but estimated using the ORCA1 NEMO grid and Tco199 atmosphere (i.e. $\Delta MAE_{LRO-LRA}$).

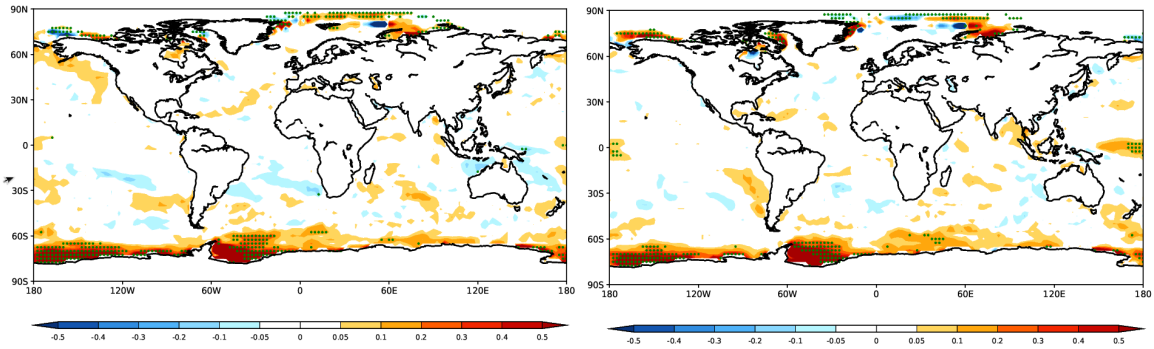

Figure 21. DJF difference in CRPSS of sea surface temperature between CY45R1 and CY43R1 estimated using higher-resolution ($\Delta CRPSS_{HRO-HRA}$, left) and lower-resolution ($\Delta CRPSS_{LRO-LRA}$, right) seasonal configurations. Stippling indicates that differences are considered statistically significant. Red values indicative of an improvement in CY45R1 relative to CY43R1. Statistics are calculated using 25 members for all reforecast experiments.

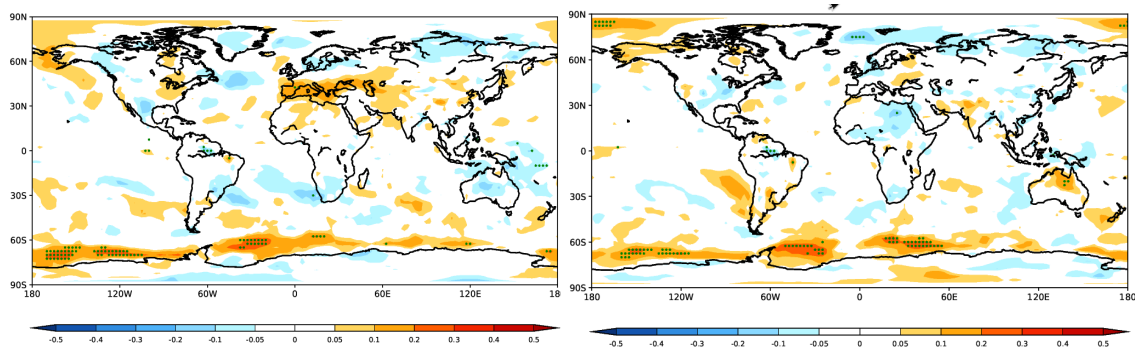

Figure 22. As figure 19, but for 2m temperature.

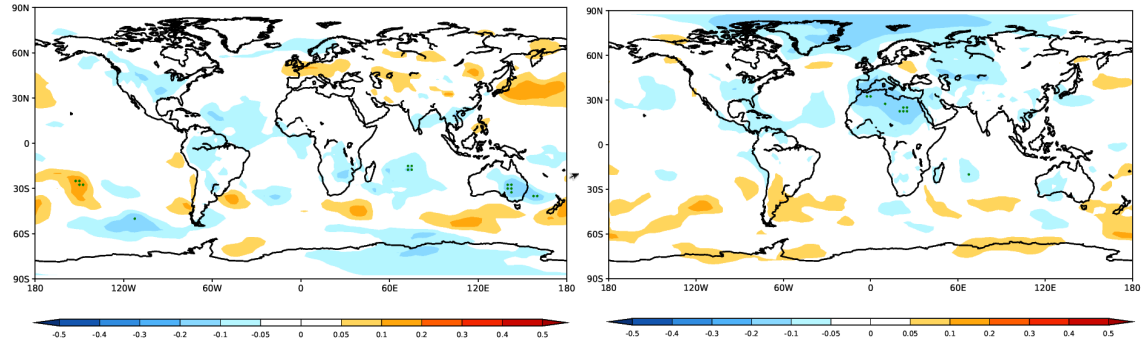

Figure 23. As figure 19, but for mean sea level pressure.

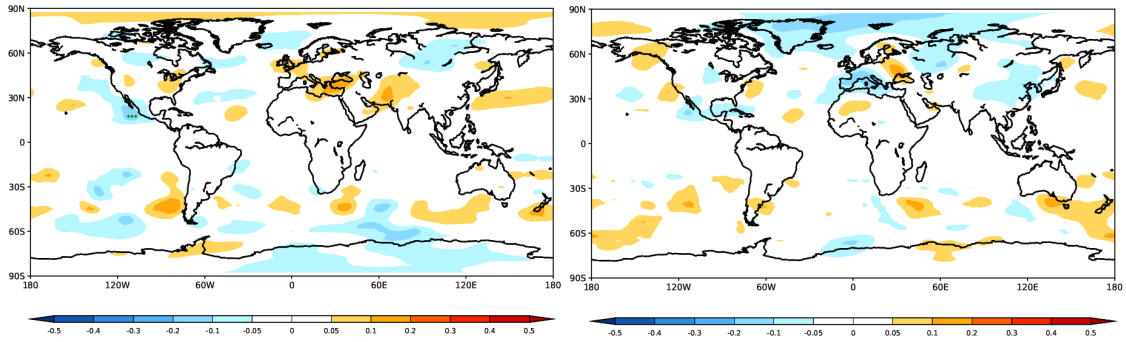

Figure 24. As figure 19, but for geopotential height at 500 hPa.

5 Summary and conclusions

The primary purpose of this memorandum was to assess whether some aspects of model development and testing at ECMWF can be undertaken with lower resolution versions of the ocean/sea-ice model without comprising the scientific integrity of the results. Although the higher resolution ocean outperforms its lower resolution counterpart, there are many aspects of model testing and development for which the absolute model performance is less relevant than the difference (i.e. Δ) between two experiments, where Δ is representative of a change in one or more summary statistics or forecast skill scores.

We have evaluated whether Δ estimated using a lower resolution ocean model (i.e. Δ_{LRO}) can be considered an appropriate proxy for Δ estimated using a higher resolution ocean model (i.e. Δ_{HRO}) for a variety of deterministic and probabilistic metrics at different lead times. Our main conclusions are summarized below:

- For medium-range forecasts (days 1-15), Δ_{LRO} is an extremely good proxy for Δ_{HRO} for both probabilistic and deterministic scores, provided that the ocean initial conditions used for lower and higher resolution systems are as consistent as possible.

- For extended-range forecasts (weeks 1-4), Δ_{LRO} is a good proxy for Δ_{HRO}, particularly for metrics that are computed from forecast anomalies relative to a hindcast climatology. At these lead times, we find a stronger sensitivity to changes in atmospheric resolution (50 km to 31 km) than changes in ocean resolution (100 km to 25 km) when evaluating the difference between two IFS cycles.

- At seasonal (and longer) timescales, we expect Δ_{LRO} to be less valid as an approximation for Δ_{HRO} as the SST biases in HRO and LRO configurations start to diverge. Nevertheless, we demonstrated that lower resolution seasonal configurations can make reliable estimates of changes to the model climatology for a range of variables. However, whether or not these changes represent an improvement will be sensitive to the model mean state. Estimates of changes in forecast skill can also be sensitive to resolution at seasonal lead times. In particular, we identified several cases where changes in CRPSS estimated using a lower resolution seasonal system were a poor analogue for changes estimated using a higher resolution configuration. However, we were unable to unambiguously attribute this difference to changes in ocean resolution as our seasonal experiments changed both ocean and atmosphere resolution at the same time.

Finally, this memorandum has shown that lower resolution configurations of the IFS based on the NEMO ORCA1 ocean model can still be a useful tool for research and model development purposes at ECMWF. However, there will still be cases where model developments and associated testing will require higher resolution configurations. In particular, studies into the role of sharp SST fronts and mesoscale ocean eddies for accurate simulation of air-sea exchange, interaction with atmospheric fronts, and the response of mid-latitude storm tracks and eddy-driven jets (Ma et al. 2015; Parfitt et al. 2016; Sheldon et al. 2017; Small et al. 2018) should use the highest resolution ocean model that is affordable.

References

Boisséson, E. D., Balmaseda, M. A., Vitart, F., & Mogensen, K. (2012). Impact of the sea surface temperature forcing on hindcasts of Madden-Julian Oscillation events using the ECMWF model. Ocean Science, 8(6), 1071-1084.

Bouillon, S., Maqueda, M. A. M., Legat, V., & Fichefet, T. (2009). An elastic–viscous–plastic sea ice model formulated on Arakawa B and C grids. Ocean Modelling, 27(3-4), 174-184.

Brassington, G. B., Martin, M. J., Tolman, H. L., Akella, S., Balmeseda, M., Chambers, C. R. S., ... & Laloyaux, P. (2015). Progress and challenges in short-to-medium-range coupled prediction. Journal of Operational Oceanography, 8(sup2), s239-s258.

Bryan, F. O., Tomas, R., Dennis, J. M., Chelton, D. B., Loeb, N. G., & McClean, J. L. (2010). Frontal scale air–sea interaction in high-resolution coupled climate models. Journal of Climate, 23(23), 6277-6291.

Chassignet, E. P., & Xu, X. (2017). Impact of horizontal resolution (1/12 to 1/50) on Gulf Stream separation, penetration, and variability. Journal of Physical Oceanography, 47(8), 1999-2021.

ECMWF website, 2019: [Online; accessed 02-12-2019], http://www.ecmwf.int.

Ferro, C. A., Richardson, D. S., & Weigel, A. P. (2008). On the effect of ensemble size on the discrete and continuous ranked probability scores. Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling, 15(1), 19-24.

Ferro, C. A. T. (2014). Fair scores for ensemble forecasts. Quarterly Journal of the Royal Meteorological Society, 140(683), 1917-1923.

Fichefet, T., & Maqueda, M. M. (1997). Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics. Journal of Geophysical Research: Oceans, 102(C6), 12609-12646.

Gent, P. R., & Mcwilliams, J. C. (1990). Isopycnal mixing in ocean circulation models. Journal of Physical Oceanography, 20(1), 150-155.

Griffies, S. M., Winton, M., Anderson, W. G., Benson, R., Delworth, T. L., Dufour, C. O., ... & Wittenberg, A. T. (2015). Impacts on ocean heat from transient mesoscale eddies in a hierarchy of climate models. Journal of Climate, 28(3), 952-977.

Hewitt, H. T., Bell, M. J., Chassignet, E. P., Czaja, A., Ferreira, D., Griffies, S. M., ... & Roberts, M. J. (2017). Will high-resolution global ocean models benefit coupled predictions on short-range to climate timescales?. Ocean Modelling, 120, 120-136.

Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., ... & Keeley, S. P. (2019). SEAS5: the new ECMWF seasonal forecast system. Geoscientific Model Development, 12(3), 1087-1117.

Kirtman, B. P., Bitz, C., Bryan, F., Collins, W., Dennis, J., Hearn, N., ... & Stan, C. (2012). Impact of ocean model resolution on CCSM climate simulations. Climate dynamics, 39(6), 1303-1328.

Leutbecher, M. (2018). Ensemble size: How suboptimal is less than infinity? Quarterly Journal of the Royal Meteorological Society.

Ma, X., Chang, P., Saravanan, R., Montuoro, R., Hsieh, J. S., Wu, D., ... & Jing, Z. (2015). Distant influence of Kuroshio eddies on North Pacific weather patterns? Scientific reports, 5, 17785.

Madec, G. (2008). NEMO ocean engine.

Marzocchi, A., Hirschi, J. J. M., Holliday, N. P., Cunningham, S. A., Blaker, A. T., & Coward, A. C. (2015). The North Atlantic subpolar circulation in an eddy-resolving global ocean model. Journal of Marine Systems, 142, 126-143.

Mogensen, K. S., Magnusson, L., & Bidlot, J. R. (2017). Tropical cyclone sensitivity to ocean coupling in the ECMWF coupled model. Journal of Geophysical Research: Oceans, 122(5), 4392-4412.

Parfitt, R., Czaja, A., Minobe, S., & Kuwano- Yoshida, A. (2016). The atmospheric frontal response to SST perturbations in the Gulf Stream region. Geophysical Research Letters, 43(5), 2299-2306.

Roberts, C. D., Vitart, F., Balmaseda, M. A., Molteni, F. (2020). The timescale-dependent response of the wintertime North Atlantic to increased ocean model resolution in a coupled forecast model, Journal of Climate, doi:10.1175/JCLI-D-19-0235.1

Roberts, C. D., Senan, R., Molteni, F., Boussetta, S., Mayer, M., & Keeley, S. P. (2018). Climate model configurations of the ECMWF Integrated Forecasting System (ECMWF-IFS cycle 43r1) for HighResMIP. Geoscientific model development, 11(9), 3681-3712.

Roberts, M. J., Hewitt, H. T., Hyder, P., Ferreira, D., Josey, S. A., Mizielinski, M., & Shelly, A. (2016). Impact of ocean resolution on coupled air- sea fluxes and large- scale climate. Geophysical Research Letters, 43(19), 10-430.

Roberts, M. J., Clayton, A., Demory, M. E., Donners, J., Vidale, P. L., Norton, W., ... & Slingo, J. (2009). Impact of resolution on the tropical Pacific circulation in a matrix of coupled models. Journal of Climate, 22(10), 2541-2556.

Sheldon, L., Czaja, A., Vannière, B., Morcrette, C., Sohet, B., Casado, M., & Smith, D. (2017). A 'warm path' for Gulf Stream–troposphere interactions. Tellus A: Dynamic Meteorology and Oceanography, 69(1), 1299397.

Small, R. J., Msadek, R., Kwon, Y. O., Booth, J. F., & Zarzycki, C. (2019). Atmosphere surface storm track response to resolved ocean mesoscale in two sets of global climate model experiments. Climate Dynamics, 52(3-4), 2067-2089.

Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., & Mayer, M. (2019). The ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: a description of the system and assessment. Ocean Science, 15(3), 779-808.