



OPEN ACCESS

EDITED BY

Peter R. Oke,
Oceans and Atmosphere (CSIRO), Australia

REVIEWED BY

Yonghong Yin,
Bureau of Meteorology, Australia
Liwei Jia,
Princeton University, United States

*CORRESPONDENCE

Magdalena Alonso Balmaseda
✉ Magdalena.balmaseda@ecmwf.int

RECEIVED 01 February 2024

ACCEPTED 30 April 2024

PUBLISHED 21 May 2024

CITATION

Balmaseda MA, McAdam R, Masina S,
Mayer M, Senan R, de Bosissésion E and
Gualdi S (2024) Skill assessment of seasonal
forecasts of ocean variables.
Front. Mar. Sci. 11:1380545.
doi: 10.3389/fmars.2024.1380545

COPYRIGHT

© 2024 Balmaseda, McAdam, Masina, Mayer,
Senan, de Bosissésion and Gualdi. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Skill assessment of seasonal forecasts of ocean variables

Magdalena Alonso Balmaseda^{1,2*}, Ronan McAdam³,
Simona Masina³, Michael Mayer^{1,2,4}, Retish Senan^{1,2},
Eric de Bosissésion^{1,2} and Silvio Gualdi³

¹Research Department, European Centre for Medium Range Weather Forecasts, Reading, United Kingdom, ²Research Department, European Centre for Medium Range Weather Forecasts, Bonn, Germany, ³Institute for Earth System Predictions, CMCC Foundation - Euro-Mediterranean Center on Climate Change, Bologna, Italy, ⁴Department of Meteorology and Geophysics, University of Vienna, Vienna, Austria

There is growing demand for seasonal forecast products for marine applications. The availability of consistent and sufficiently long observational records of ocean variables permits the assessment of the spatial distribution of the skill of ocean variables from seasonal forecasts. Here we use state-of-the-art temporal records of sea surface temperature (SST), sea surface height (SSH) and upper 300m ocean heat content (OHC) to quantify the distribution of skill, up to 2 seasons ahead, of two operational seasonal forecasting systems contributing to the seasonal multi-model of the Copernicus Climate Change Services (C3S). This study presents the spatial distribution of the skill of the seasonal forecast ensemble mean in terms of anomaly correlation and root mean square error and compares it to the persistence and climatological benchmarks. The comparative assessment of the skill among variables sheds light on sources/limits of predictability at seasonal time scales, as well as the nature of model errors. Beyond these standard verification metrics, we also evaluate the ability of the models to represent the observed long-term trends. Results show that long-term trends contribute to the skill of seasonal forecasts. Although the forecasts capture the long-term trends in general, some regional aspects remain challenging. Part of these errors can be attributed to specific aspects of the ocean initialization, but others, such as the overestimation of the warming in the Eastern Pacific are also influenced by model error. Skill gains can be obtained by improving the trend representation in future forecasting systems. In the meantime, a forecast calibration procedure that corrects the linear trends can produce substantial skill gains. The results show that calibrated seasonal forecasts beat both the climatological and persistence benchmark almost at every location for all initial dates and lead times. Results demonstrate the value of the seasonal forecasts for marine applications and highlight the importance of representing the decadal variability and trends in ocean heat content and sea level.

KEYWORDS

seasonal forecasts, skill, trend, essential climate/ocean variables, SST, sea level, ocean heat content

Introduction

Knowledge of forecast skill is a prerequisite for utilizing forecast information. Assessing the skill of ocean variables from seasonal forecasts other than sea surface temperature (SST) in a multi-model context has remained elusive due to the lack of verifying ocean datasets of sufficient quality and length. In recent years, the availability of longer observational records of surface ocean variables such as altimeter-derived sea level, and sea-ice concentration has allowed the verification of seasonal forecasts of these additional surface variables in a multi-model context. For instance, Long et al., 2021 and Widlansky et al., 2023 have used altimeter derived records of sea level anomalies to verify the ability of the multi-model seasonal forecasts to capture the variations of global and regional sea level. There have also been assessments of the skill of multi-model seasonal prediction of sea-ice (Guemas et al., 2016; Blanchard-Wrigglesworth et al., 2017). However, verification of sub-surface ocean variables in a multi-model context has remained challenging. Several studies have reported the potential predictability of upper ocean heat content (OHC) in the context of seasonal forecasts of El Niño-Southern Oscillation (ENSO) (e.g. Balmaseda et al., 1994; Sharmila et al., 2023, among others), but in these studies the forecast OHC was verified against own analysis. The recent availability of ensembles of ocean reanalyses has opened the possibility for independent verification of seasonal forecasts of the ocean subsurface. For instance, McAdam et al., 2022 have used the upper Ocean Heat Content from the Global Reanalyses Ensemble Product (GREP) for a comparative assessment of the forecast skill of seasonal forecasts between SST and upper Ocean heat content in a multi-model context.

Here we will compare the skill of seasonal forecast of SST, sea surface height (SSH) anomalies and OHC, with the expectation that the comparative assessment can shed some light on the sources of predictability and errors. One novel aspect of the current study is the assessment of the forecast models to represent the trends in the observational records, and to quantify how much the climate trend contributes to the skill and forecast errors. We also expect that the skill assessment of these ocean variables will encourage seasonal forecast providers to make the data publicly available, as it is currently done with the seasonal forecasts for the atmosphere.

In this study, we use these satellite observational records and ocean reanalyses to verify ocean variables from two seasonal forecast systems contributing to the C3S (Copernicus Climate Change Service) seasonal multi-model product. These systems are from ECMWF (European Centre for Medium-Range Weather Forecasts) and CMCC (Centro Euro-Mediterraneo sui Cambiamenti Climatici). The data and methods are described in section 2. The results are presented in Section 3, which provides a comparison of the spatial distribution of deterministic skill of SST, OHC and SSH, in terms of anomaly correlation and root mean square error, bench-marked against persistence and climatology. This section also investigates the ability of the seasonal forecasts to capture the recent linear trends present in observations, and it quantifies the contribution of the linear trend to the seasonal forecast skill. The lessons learned from the evaluation are summarized in section 4.

Data and methods

Data

Verification datasets

Three observation-based datasets have been identified as suitable for the verification of ocean variables from seasonal forecasts. Suitability criteria are based on the length of the available record (at least 1993–2016), and documentation on their uncertainty and temporal homogeneity. For the surface variable (SST and SSH) we use satellite-derived records from the ESA-CCI initiative: the global Sea Surface Temperature Reprocessed product (Merchant et al., 2019; Good, 2020), distributed by C3S; and the Sea Surface Height (SSH) product (Pujol et al., 2016 and Taburet et al., 2019), distributed by C3S and CMEMS (Copernicus Marine Environmental Monitoring Service). Since satellite information is insufficient to constrain the ocean subsurface, the ocean heat content (OHC) in the upper 300m is verified with the Global Ocean Reanalysis Ensemble Product (GREP, Storto et al., 2019), distributed by CMEMS, which is constrained by *in-situ* observations. A full description of the data is provided in section S1.1 of [Supplementary Material](#). For the purpose of verification, we use seasonal means of these records for the period 1993–2016.

Seasonal forecasts ocean data

The two forecast systems used here are the Seasonal Prediction System Version 3 from the Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC-SPS3, Sanna et al. (2017)), and the fifth generation Seasonal Forecasting System from the European Centre for Medium-Range Weather Forecasts (ECMWF-SEAS5, Johnson et al. (2019)). Since 2018 both systems have been contributing to the Copernicus Climate Change Service (C3S), which makes seasonal forecasts of atmosphere and surface variables (precipitation, 2m-temperature) freely available online. These systems produce a forecast of ocean variables other than SST, although the ocean variables are not yet publicly available for the multi-model.

A full description of the models is included in Section S1.2 of [Supplementary Material](#). Suffice to say that both systems base their ocean model component on the eddy-permitting version 3.4 of NEMO (Nucleus for European Modelling of the Ocean), which has a horizontal resolution of 25 km at the equator, and they are initialized from ocean reanalyses: CMCC-SPS3 is initialized from C-GLORS (Storto and Masina, 2016), while ECMWF-SEAS5 is initialized from ORAS5 (Zuo et al., 2019).

Seasonal means of SST, SSH and OHC have been gathered from a set of retrospective seasonal forecasts (re-forecasts) from the two models. The re-forecast dataset comprises 96 independent initial dates, spanning the 1993–2016 period, initialized 4 times per year, with starting dates on the 1st of February, May, August and November. This 1993–2016 period was chosen, so it is the same as that used for the C3S seasonal multi-model product. The forecast range is 6 months, and for the purpose of map verification, we split it into first and second season. Thus, forecasts initialized in February will be verified for the FMA (lead 1) and MJJ (lead 2)

seasons. For an individual date, the forecast from each system comprises 40 ensemble members, which are averaged to estimate the ensemble mean. The forecast data are stratified by initial and verifying calendar date (or by initial calendar date and lead time). For instance, lead 1 (2) forecasts initialized in May will comprise the MJJ (ASO) season for the period 1993–2016, which can then be verified against the corresponding MJJ (ASO) values of the observational record. For the sake of brevity, in the following we discuss the skill statistics averaged over all starting months, unless explicitly stated.

Verification methods

The ability of a prediction system to forecast specific events at a given time is measured by a set of skill scores or metrics. Here we focus on the scores of the ensemble mean forecast anomalies. The forecast seasonal anomalies are computed with respect to the model seasonal climatology, which depends on the forecast lead time. By subtracting the model climate from the individual forecasts, we effectively remove the forecast bias, which is the first order correction in the calibration of seasonal forecasts (Stockdale, 1997). For the purpose of deterministic verification, only the anomalies of the ensemble mean will be used (see description of method in [Supplementary Material 1.3.1](#)). The verification statistics used in what follows are therefore bias blind. An assessment of bias and variability in SST and OHC have been reported in [McAdam et al., 2022](#), and it will not be further discussed here.

Since the initial focus is to quantify the performance of the ensemble mean, we have chosen two different deterministic scores (Wilks, 2011): anomaly correlation coefficient (ACC) and Mean Square Skill Score (MSSS). The mathematical expression of these skill metrics is given in [Supplementary Section 1.3.2](#). The MSSS compares the root mean square error (RMSE) of the forecast with that of the climatology, which is the simplest model for a seasonal forecast to be compared to. The climatological benchmark is therefore already built in the definitions of ACC and MSSS: positive values in these scores imply that the forecast is more skillful than climatology. We also benchmark against persistence (of the observed anomaly at the time of the initialization), which is the second simplest statistical forecast model after climatology.

Treatment of linear trends

In addition to the interannual variability, in a changing climate, it is also important to evaluate the ability of the forecast model to capture the linear trend present in observations. Since the trend contributes to the temporal variability, it can potentially impact the forecast skill: it can enhance it if models are successful in capturing the linear trend, or it can deteriorate it if the model errors prevent the correct representation of the observed trends. In addition to the standard ACC and MSSS statistics, we compute two additional sets of statistics:

- Trend-corrected (Tc) ACC and MSSS, where the linear trend of the forecast is corrected with that of observations. This is done by linearly detrending the forecasts first (see

below), and then adding the observational linear trend. Comparison of these statistics with the standard ones gives an idea of the gains in skill by this additional calibration step, and it illustrates the potential gains that could be obtained if the models were able to represent the trends adequately.

- Detrended (D) ACC and MSSS, where the linear trend has been removed from the anomalies of both forecasts and observations before computing the ACC and MSSS. Differences between the Detrended with the Trend-corrected statistics quantify the amount of the skill due simply to the presence of trends in the observations.

Results

Skill without trend correction

The skill of the ECMWF and CMCC seasonal forecasting systems in predicting the first season is shown in [Figure 1](#), for the three different ocean variables. For reference, the skill of the persistence forecast is also shown. All the forecasts, including persistence, have a high (significant) level of skill in the first season. Nonetheless, even in the first season, the dynamical models are more skillful than persistence in the wider tropics, where ocean dynamics are faster. A notable exception is the skill of SSH in the CMCC model, which in the tropical Atlantic is less than persistence. It has been found that the underrepresentation of the sea level trends stems from the ocean initial conditions; we will return to this point later.

The advantage of the dynamical seasonal forecasts w.r.t persistence is more obvious in the predictions for the second season, as can be seen in [Figure 2](#). As expected, the overall level of skill decreases as the forecast lead time increases, but this decline is faster for persistence than for the dynamical seasonal forecasts. The dynamical models retain significant skill levels: for SST, ACC values larger than 0.6 are seen over the wider tropics, along the coast of North-West America and some areas of the Southern Ocean and Northern Seas. The skill levels at longer forecast leads for OHC and SSH are also higher than for SST, as expected from the larger memory of the deeper water column.

[Figure 3](#) shows that the skill gains of the dynamical models against the persistence benchmark, already visible in the first season, increases further in the second season. The pattern skill gains in SST (left panels) is indicative of the dynamical processes operating in the coupled model at these time scales, which persistence is not able to represent. For instance, the signature of the tropical wave dynamics is noticeable in the Tropical Pacific, with the ENSO signature in the cold tongue region clearly visible. The skill gains in the extratropics are likely related to the thermal memory of the mixed layer, but may also be a result of the predictable variations in the atmospheric circulation, which persistence will not be able to account for. The comparison of skill between OHC and SST provides additional criteria to attribute the predictability gains to the thermal memory or to predictable atmospheric circulation. For instance, during the

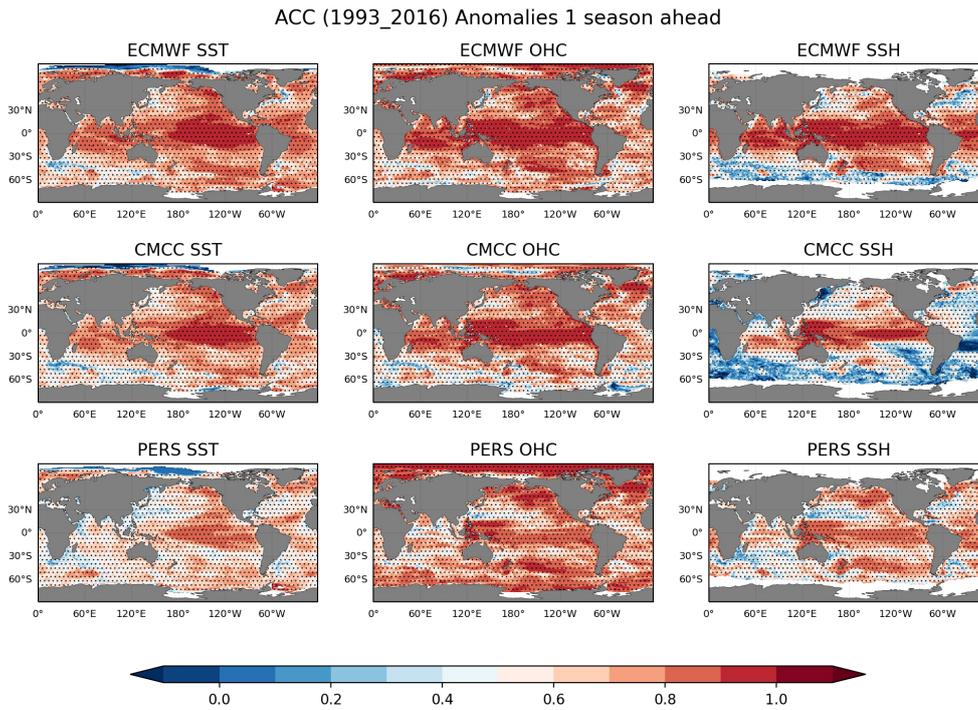


FIGURE 1
 Anomaly Correlation Coefficient (ACC) of the seasonal forecast anomalies in the first season for all initialization times for SST (left column), OHC (middle) and SSH (right). Shown is the skill of the ECMWF (top) and CMCC (middle row) seasonal forecasting systems, as well as that of the persistence forecasts (bottom). Correlation values with p-values < 0.05 are shown as dotted areas. Positive values indicate that the forecast skill is better than the climatological forecast.

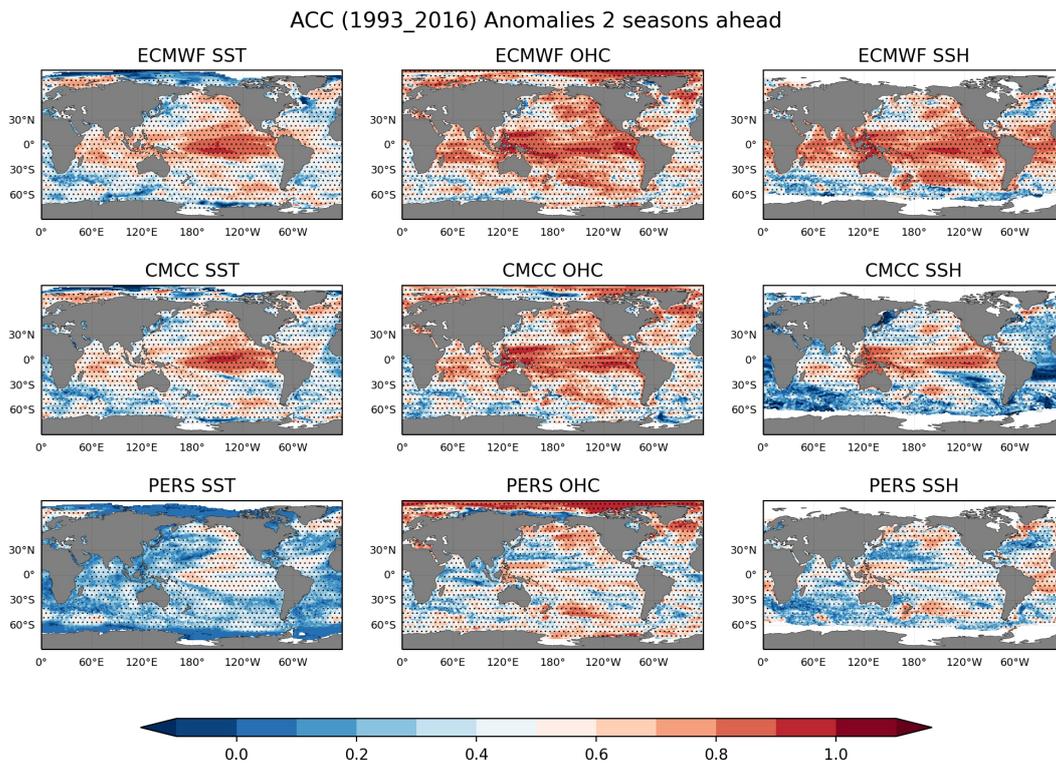
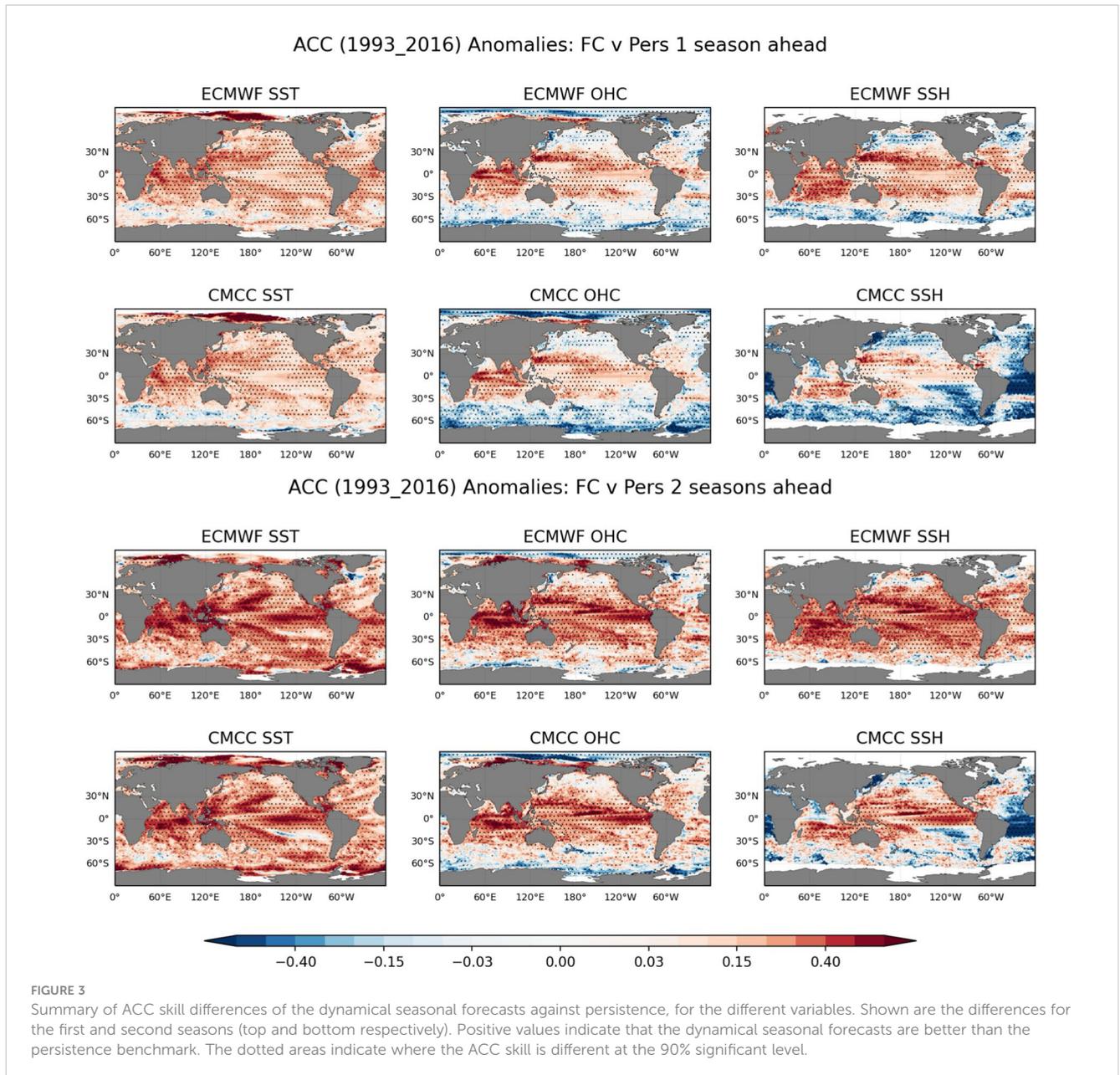


FIGURE 2
 As Figure 1 but for the second season into the forecast.



boreal winter, the predictable atmospheric teleconnection patterns such as the PNA (Wallace and Gutzler, 1981) will have more impact on SST predictability than on OHC. The ACC over the North-Eastern Pacific is indeed stronger for SST than for OHC for forecasts of the first season verifying in boreal winter (see Supplementary Figure S1).

An area of concern is the North Atlantic Subpolar Gyre in the ECMWF, where the model skill is lower than persistence. This has been attributed to a problem with the ocean initialization (Tietsche et al., 2020). In the second season, there are also regions of decreased skill gains along the boundaries of the atmospheric convergence zones, which is symptomatic of errors in the spatial patterns of anomalies produced by atmospheric models (e.g., meridional extension of the Hadley Circulation). This feature is

more noticeable in forecasts initialized from May (Supplementary Figure S2 in Supplementary Material).

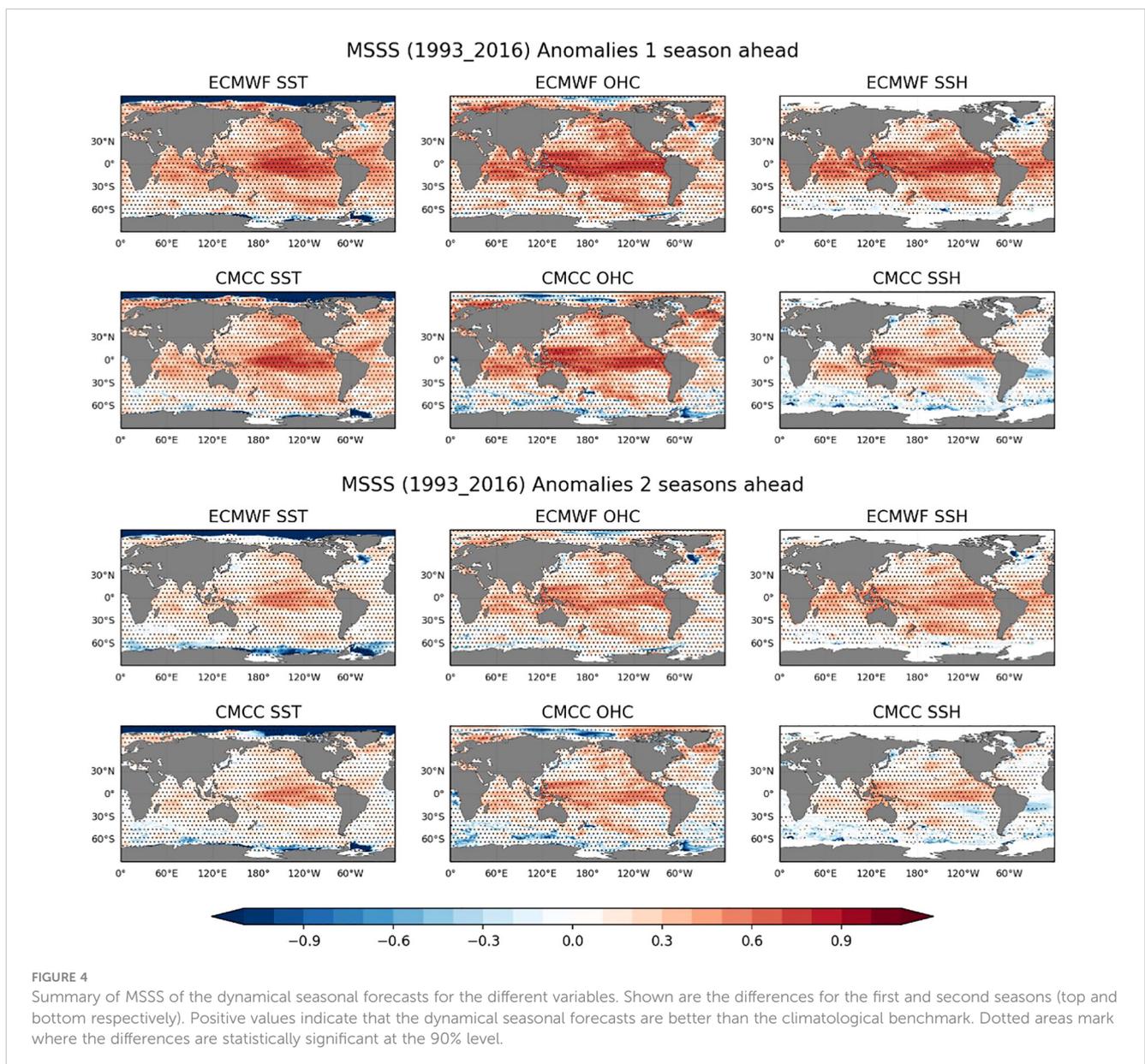
The patterns of ACC skill gains of models versus persistence are quite similar in OHC and SSH, indicative of the strong correlation between the two quantities, while they exhibit visible differences with the patterns of SST skill gains. The pattern of skill gains in OHC and SSH bears the signature of the equatorial wave dynamics, more confined to the Equatorial band than SST. Outside the Equatorial band is rather homogeneous, having less meridional span than that of SST, and for the second season the decrease of skill associated with atmospheric convergence zones is less apparent than in the SST skill gain pattern. Along the Western Boundary Currents, the OHC/SSH have less skill than persistence, a feature associated with the insufficient resolution of the ocean component, as discussed by

Feng et al. (2024). The different skill gain patterns between SST and OHC/SSH can be attributed to different factors: i) the SST forecast in dynamical models benefit from the memory of the OHC in the initial conditions, which will be included in the persistence of OHC itself but not in the persistence of SST; ii) the impact of the atmospheric component, which is stronger in SST than in the subsurface. Thus, the predictable atmospheric component enhances the skill of SST in the tropics and mid-latitudes; conversely, errors (or unpredictable variability) in the atmospheric model can induce errors (or unpredictable variability) in the SST forecasts which are not so visible in the integrated ocean variables. This seems to be the case in localized areas of the Western Pacific, where the skill in SST is lower than in OHC and SSH (Figures 1, 2).

We also note that the model skill for OHC and SSH is lower than persistence at high latitudes, especially during the first season of the forecast. This is suggestive of potential problems with the ocean initial conditions, which should be the target of developments in future data

assimilation systems. We see again the low skill of the CMCC model for SSH, which is related to the trend in the initial conditions and will be discussed later. But luckily, this is an error that does not affect the structure of the water column substantially and does not manifest in SST or OHC. Over some coastal areas of the eastern tropical Atlantic Equatorial Atlantic, the models prediction of OHC in the first season is lower than persistence. This is partially attributed to deficiencies in ocean initialization (McAdam et al., 2022), and it highlights the potential for further skill gains that could be realized by improving the forecasting systems.

Figure 4 shows the MSSS of forecasts at the first and second seasons, for ECMWF and CMCC, and the three ocean variables of interest. Values larger than zero indicate that the forecast beats climatology. This is the case for the two dynamical forecasts over most of the ocean. Notable exceptions include the North Atlantic subpolar gyre in the ECMWF system, and the Southern Ocean and the Subtropical Atlantic for the CMCC system.



Fidelity of the trends in seasonal forecasts

Linear trends in observational datasets

The 1993–2016 linear trend in the observational records of SST, SSH and OHC for the different seasons is shown in Figure 5. Positive trend values are visible in all the fields, albeit with different patterns. The SST field shows positive trends in the Indian Ocean, and warm pools, in the Equatorial cold tongue and in the extratropics, especially summer hemisphere, e.g. North Eastern Pacific in boreal summer and South Western Pacific in austral summer. The former coincides with the location of recent long-lasting marine heat waves, induced by persistent atmospheric anticyclones and deeper mixed layer (de Boissésou et al., 2022; de Boissésou and Balmaseda, 2024). There are also warming SST trends in the north/south subtropical Atlantic, and in the European Northern Seas. SSH shows positive values everywhere, with enhanced amplitude in the Western Pacific north of the Equator, Tropical Indian and Atlantic Ocean and Western Boundary Currents. The OHC trends resemble those in SSH, but they show a stronger footprint of changes in ocean circulation (e.g., dipole in the tropical Pacific associated with strengthening of the trades). They also reflect the deepening of the mixed layer in the summer extratropics, consistent with trends in SST. Both SST and OHC exhibit a negative trend over the North Atlantic subpolar gyre, consistent with the findings of Li et al. (2022). Interestingly, this negative trend is not visible in SSH which suggests that other factors

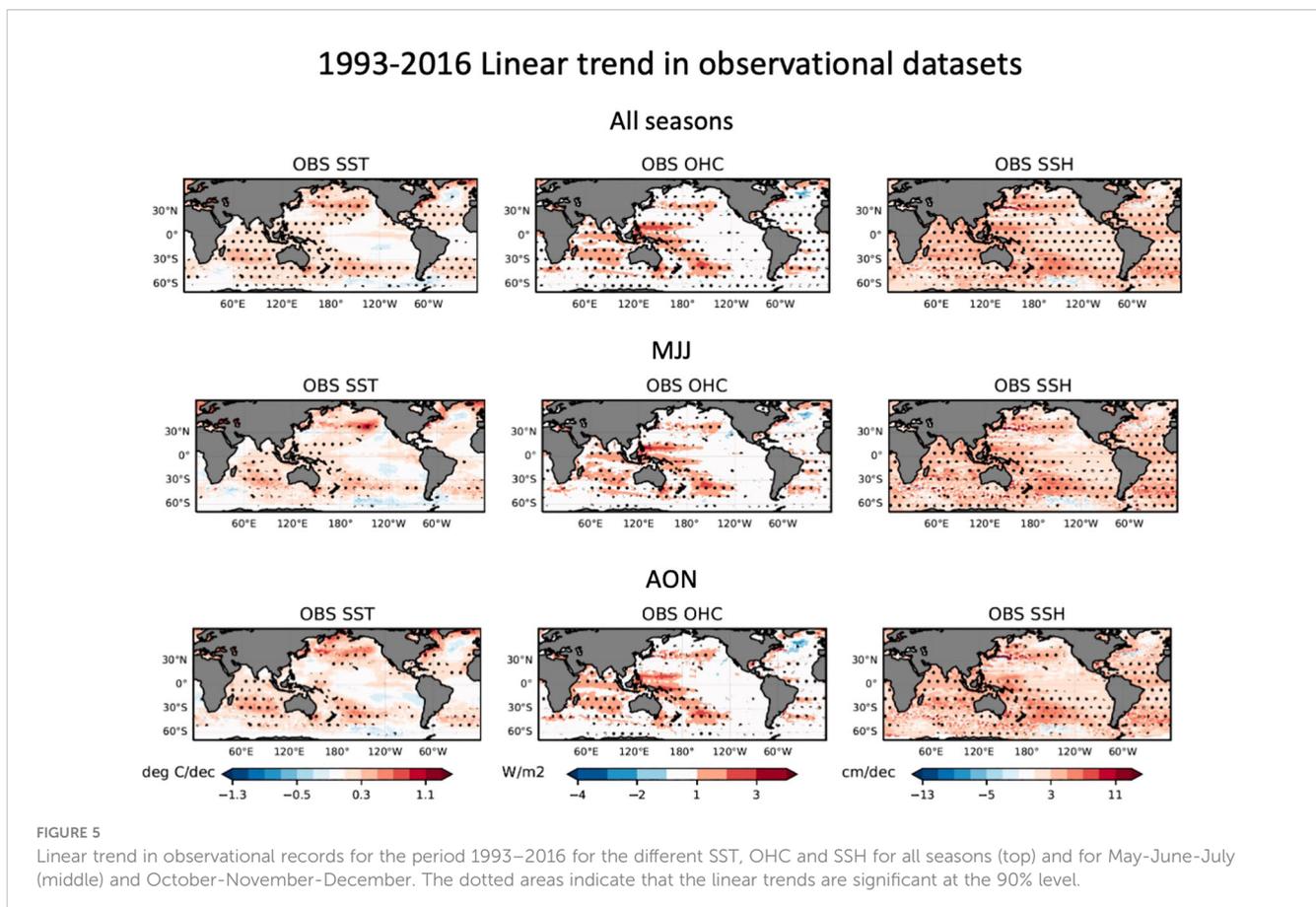
are at play (such as the deep ocean or variations in salinity). It is also interesting that the positive trends in the tropical South Atlantic in SSH and OHC do not have an obvious footprint in SST trend in that area.

In general, the spatial pattern of the trends is similar in all seasons, but the amplitude of the trends shows some seasonal variations, with stronger extratropical SST warming in the respective boreal summer (middle panels of Figure 5), and enhanced SSH and OHC trends in the western Equatorial Pacific in boreal autumn (August–October–November, bottom panels), likely related to the strengthening of the trade winds over the Pacific (de Boissésou et al., 2014).

Linear trends in seasonal forecasts

We now evaluate the ability of the forecasts to capture observational trends. Any discrepancy between forecasts and observation could be attributed to model errors or errors in the initial conditions. For the purpose of illustrating the main messages, we focus on show only the fidelity of the trends for the forecasts initialized in May. The results for the forecast trends averaged over February, May, August and November initial dates are shown in Supplementary Figure S3.

Figure 6 (top 2 rows) shows the differences in linear trends for the first season between seasonal forecasts and observations. Differences are sizeable even at this short lead time. Most noticeable is the global difference in the SSH in the CMCC forecast. This has been traced

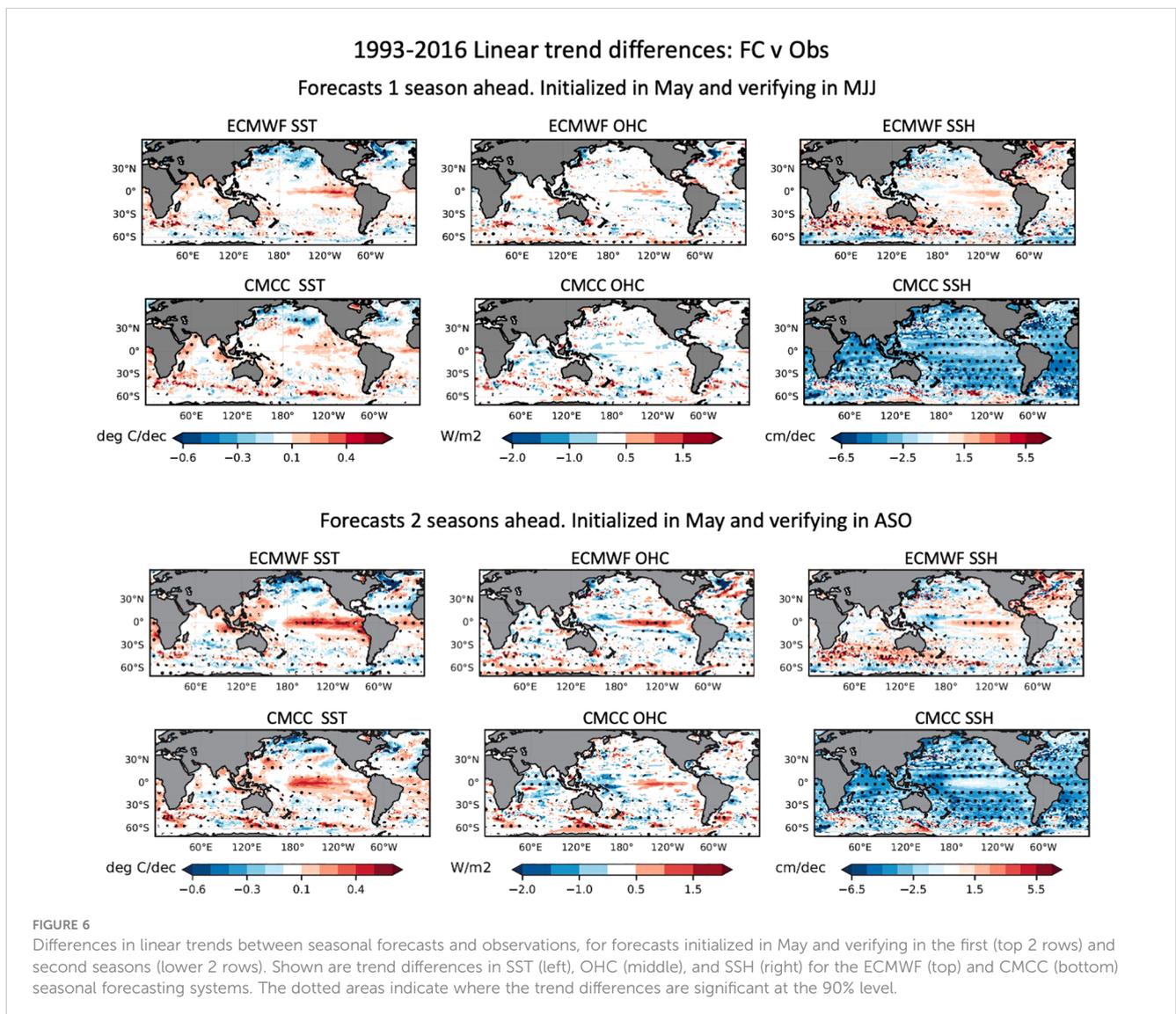


back to the fact that the ocean initial conditions of SSH in the CMCC system did not include explicitly the global changes in steric height mean values, which current generation of models do not represent due to the Boussinesq approximation. In ORAS5 (used to initialize the ECMWF forecasts) this is diagnosed and added to the SSH (Balmaseda et al., 2013; Zuo et al., 2019). Aside from this global difference, both forecast systems exhibit also regional departures from the observational trends, which are amplified as the in the forecasts for the second season (bottom 2 rows), which are largely similar when considering all initial forecast days (Supplementary Figure S3).

The ECMWF forecasts overestimate the warming trend in the Eastern Equatorial Pacific, a signature that manifests in all three considered variables. Further investigations point towards a sensitivity of this trend to both ocean initial conditions and atmospheric model. Two additional experiments were conducted with the ECMWF system, replacing the atmospheric model and the ocean initial conditions with more up-to-date versions one at a time. The changes in the ocean version were only related with the data assimilation and forcing fields, maintaining the same version of the

ocean model. The SST trends over the Equatorial Pacific trends were reduced in these new experiments. In region Nino3.4, for forecasts initialized in May and verifying in ASO, the SST trends went from 0.45 °C/decade in SEAS5, to 0.4 °C/decade with the new atmosphere model version, and to 0.3 °C/decade when both atmosphere and ocean initial conditions changed. The eastern equatorial Pacific SST warming is also present in the CMCC seasonal forecasts, but with weaker amplitude. We note that errors in seasonal forecasts of SST trends are common to other models (L'Heureux et al., 2022), and it is an emerging research topic in the scientific community. The errors in the trends are likely to have implications for the prediction of ENSO. For instance, seasonal forecast models over-predicted the warming of El Nino in 2014–15 (Mayer and Balmaseda, 2021) and struggled to predict the duration of the prolonged La Nina conditions during 2020–2022.

The CMCC and ECMWF seasonal forecasts tend to overestimate SST warming over the Western Indian Ocean and Bay of Bengal, and over the tropical Atlantic. We note that this overestimation of in the surface is not mirrored by the OHC. Both ECMWF and CMCC



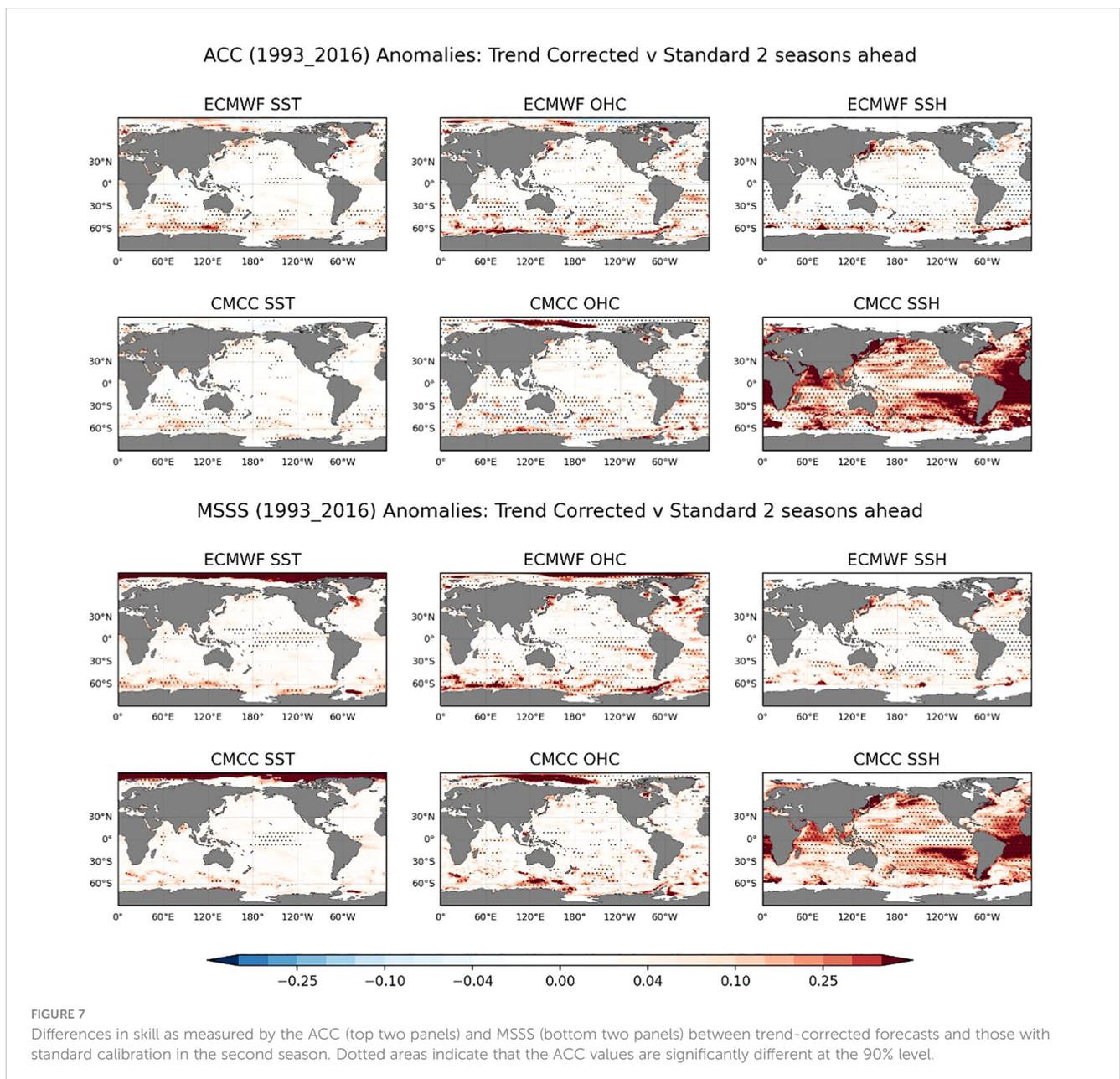
underestimate the surface warming trends at high latitudes during the summer. Both ECMWF and CMCC underestimate the surface warming trends at high latitudes during the summer hemisphere (Figure 6 bottom 2 panels). The ECMWF model produces a cooler than observed trend in the North Atlantic Subpolar Gyre, more visible in the OHC, but also SST. This is believed to be related to the overestimation of the decadal variability of the AMOC in the ECMWF ocean initial conditions reported by [Tietsche et al. \(2020\)](#), and it is present in all seasons ([Supplementary Figure S3](#)).

Impact of trends on errors and skill

The errors in the seasonal forecast linear trends could be easily removed by correcting the linear trend, an additional calibration

step that is not currently carried out when using or assessing seasonal forecast skill. Equally, if the linear trend is sizeable, it will influence the interannual variability and its potential predictability. Here we quantify the impact on skill of correcting the linear trend, by comparing the skill of trend corrected versus standard calibration. We can also measure the contribution of the linear trend to the skill by comparing the skill of trend-corrected versus detrended forecasts.

The skill gains obtained by the additional calibration step of correcting the linear trends are shown in [Figure 7](#), which shows the differences in ACC (top two panels) and MSSS (bottom two panels) between the trend-corrected versus the standard calibration for the second season. Overall, positive values are seen for the 3 variables. The equivalent results for the first season are shown in [Supplementary Figure S4](#). In the first season, the impact of trend



correction is clearly visible in the CMCC forecasts of SSH, and on forecasts of OHC and SST for both systems in MSSS. For ACC the differences in the first season are clearly visible in SSH, and for OHC and SST over some areas close to the sea-ice edge and mid-latitudes. The trend correction improves the SST skill in several areas, most noticeable over the Southern Indian Ocean, and the high latitudes. There are also small but significant SST skill gains over the Equatorial Central Pacific and Eastern Atlantic, which is more noticeable in MSSS. Significant gains in OHC skill are widespread across the different ocean basins. We also note that over the Arctic and North Atlantic subpolar gyre the linear trend correction slightly deteriorates the skill in the ECMWF forecasts of OHC, suggesting the presence of non-linear trends.

The impact of linear trend correction on SSH deserves special attention. While for the ECMWF the influence of trend correction is similar in SSH to the other variables, in the CMCC system the trend correction has a sizeable positive impact in the wider ocean from the early stages in the forecast (Supplementary Figure S4), consistent with the problem of the trends residing in the ocean initial conditions. The skill gains are especially high for the Atlantic basin, the South-Eastern Pacific, the Northern Indian Ocean, and the Western Boundary currents. The problem in the CMCC ocean initial conditions has been identified, as it is related to the fact that the global trends in the steric component of SSH are not applied to the model. It is however not obvious why a global increase in sea level should have such a clear spatial structure. We note over the regions where the trend correction has such a pronounced impact on the SSH seasonal forecasts, the trend correction also has a significant (although small) impact on the ECMWF seasonal forecasts of SSH, and in both systems the same regions the trend correction has impact on the skill of OHC.

It is of interest to evaluate the ACC skill of the trend-corrected forecasts against persistence. This can be seen in the top panels of Figure 8, which show similar diagnostics as Figure 3 but for the additional calibration step. A simple linear trend correction solves the problem with the skill of SSH in the CMCC system. In addition, some areas with poor skill originally, such as OHC in the tropical Atlantic and mid-latitudes, are improved. However, the trend correction does not solve the underperformance in the first season of OHC and SSH at high latitudes, indicating that more work is needed to improve the ocean initialization in these areas. The prediction of OHC over the Arctic in dynamical seasonal forecasts is still poorer than persistence. The trend correction does not improve the underperformance in SST along the edges of the tropical atmospheric trade winds either, which was visible for individual seasons and which may be attributed to errors in the atmosphere (not shown). To verify where the dynamical seasonal forecasts still have an advantage over persistence when predicting the interannual variability only, the lower panels of Figure 8 show the comparison of the detrended dynamical forecasts versus the detrended persistence. In this case the linear trend has been removed from forecasts, persistence and verifying observation. The detrended dynamical forecasts maintain their skill advantage in the wider tropics for the three variables considered. As for the trend corrected forecasts, the most striking difference of the detrended forecasts in comparison with Figure 3 is the skill for

SSH of CMCC system over the Atlantic, which now is superior to the detrended persistence. These results confirm that the erroneous forecast trends can mask the skill in predicting the interannual variability.

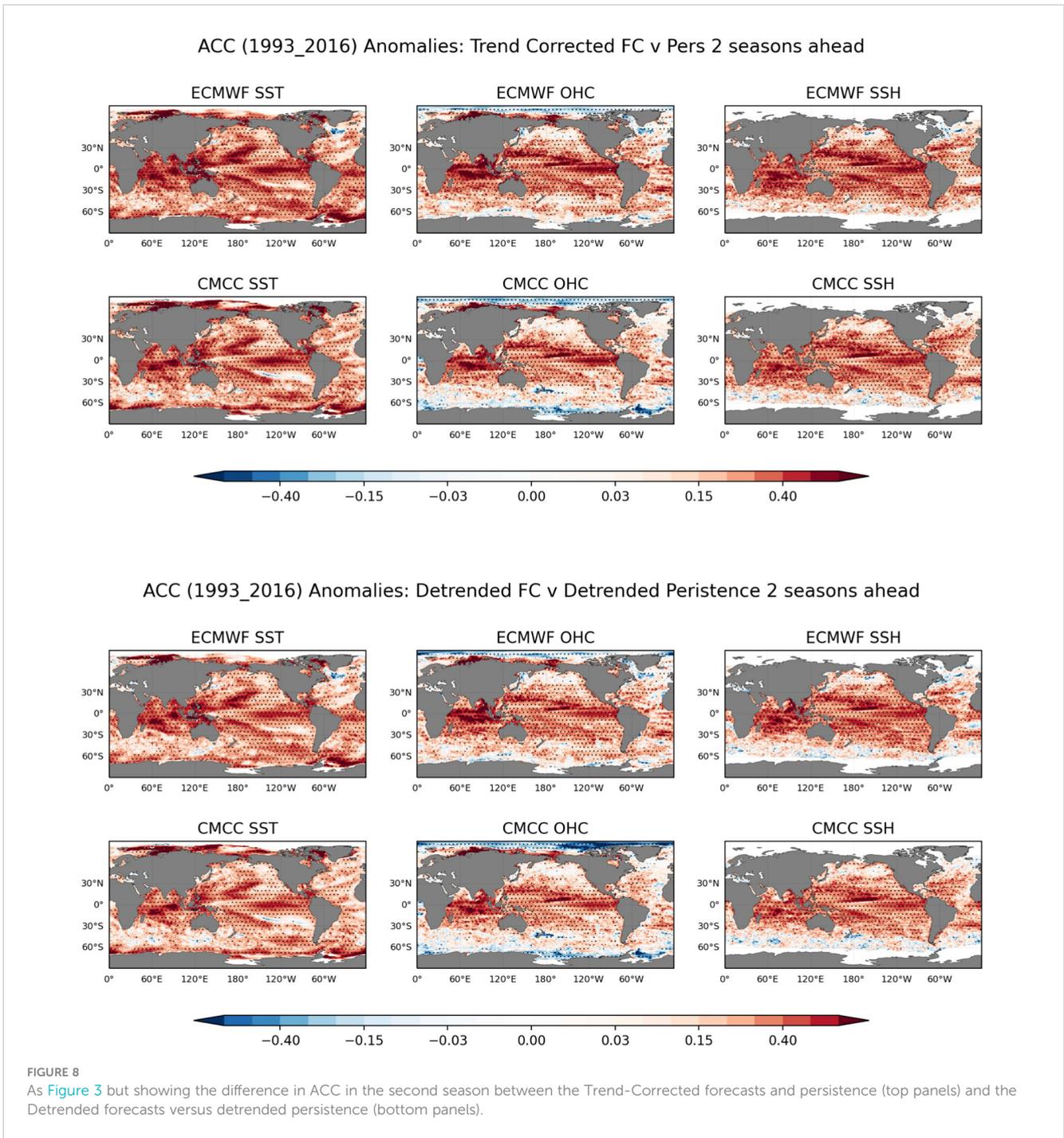
The contribution of the linear trends to the overall level of ACC skill in seasonal forecasts is displayed in the top panels of Figure 9, which show the ACC differences between trend-corrected and de-trended forecasts (the latter verified against de-trended observations). In SSH and OHC the linear trends contribute to the skill in the Tropical Indian Ocean and extratropical Pacific. The impact is also seen across the Atlantic basin. The impact of the trends is stronger in predictions of SSH, notably over the Atlantic basin and the Southern Ocean, as discussed previously. The trends also contribute to the predictability of SST over the extratropical oceans, and notably over the European-Arctic area. The impact of the trend in forecast skill is consistent in both forecasting systems. The bottom panels of Figure 9 show the ACC differences between the detrended forecasts and the standard calibration. Regions with negative values indicate where the linear trends contribute to the ACC skill. Conversely, positive values indicate where the erroneous trends in the forecasting systems are detrimental for the skill.

Summary and conclusions

Selected ocean variables (SST, OHC and SSH) from the ensemble of ECMWF and CMCC seasonal forecasts contributing to C3S have been verified against independent observational records. The observational records chosen are the state-of-the-art datasets of Essential Ocean/Climate Variables (EOVs/ECVs). These are monthly SST and SLA from the Copernicus Climate Change Service (C3S) and OHC from Copernicus Marine Environmental Service (CMEMS) Global Ensemble of ocean Reanalyses Products (GREP).

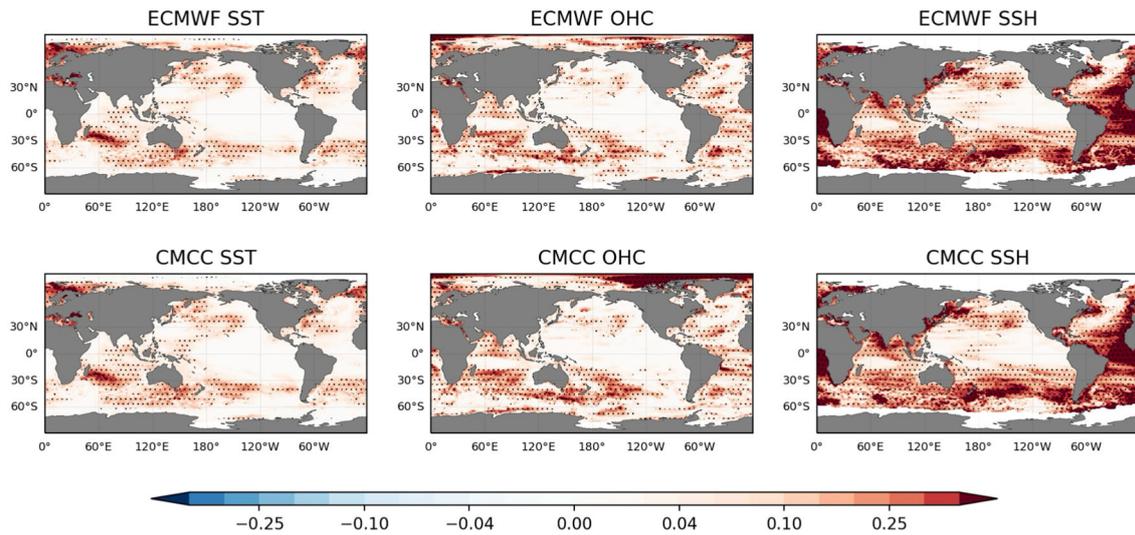
The C3S seasonal forecasts dataset comprises probabilistic forecasts initialized four times per year during the period 1993–2016. Each individual forecast consists of 40 ensemble members, integrated for up to 6 months. The forecast and observational data have been stratified in seasonal means for which the anomaly correlation and mean square skill score metrics have been derived. The forecast performance has been benchmarked against two statistical forecasts, namely persistence and climatology. The fidelity of the linear trends in the forecasts has also been evaluated, as well as the contribution of the observed trend to the seasonal forecast skill. From the analysis of the results, we obtain the following conclusions:

- Skill of seasonal forecast for 3 variables outperforms that of persistence and climatology in most regions in the first and the second season over the tropics. There is still scope for further skill gains in the extratropical oceans, where the persistence forecast beats the dynamical models. This is more noticeable in forecasts of OHC and SSH, and therefore it is expected that the improvements of the ocean initial conditions can contribute to further skill gains.



- Differences among the variables in the spatial distribution of skill are indicative of processes contributing to predictability. For example, over the tropical Atlantic and North-Eastern Pacific, the higher skill in SST forecasts than in OHC or SSH, is likely the consequence of the additional predictability arising from the remote effect of ENSO. Conversely, the lower skill of SST over the Western Pacific Warm pool is probably related with unpredictable atmospheric processes interfering with the predictable signal in the ocean subsurface.
- The ability of the seasonal forecasts to capture the linear trends in observations has been evaluated. Results show that some aspects of the observed linear trends are not well captured by seasonal forecasts. This includes overestimation of the warming in the tropics (warm-pool regions, and Equatorial Pacific cold tongue) and under-estimation of mid-latitude warming. These deficiencies are visible early in the forecast and appear to be associated with the trends in the ocean initial conditions. This is certainly the case for the SSH trends in the CMCC system. However, deficiencies in the forecasts models cannot be ruled out.

ACC (1993_2016) Anomalies: Trend Corrected v Detrended 2 seasons ahead



ACC (1993_2016) Anomalies: Detrended v Standard 2 seasons ahead

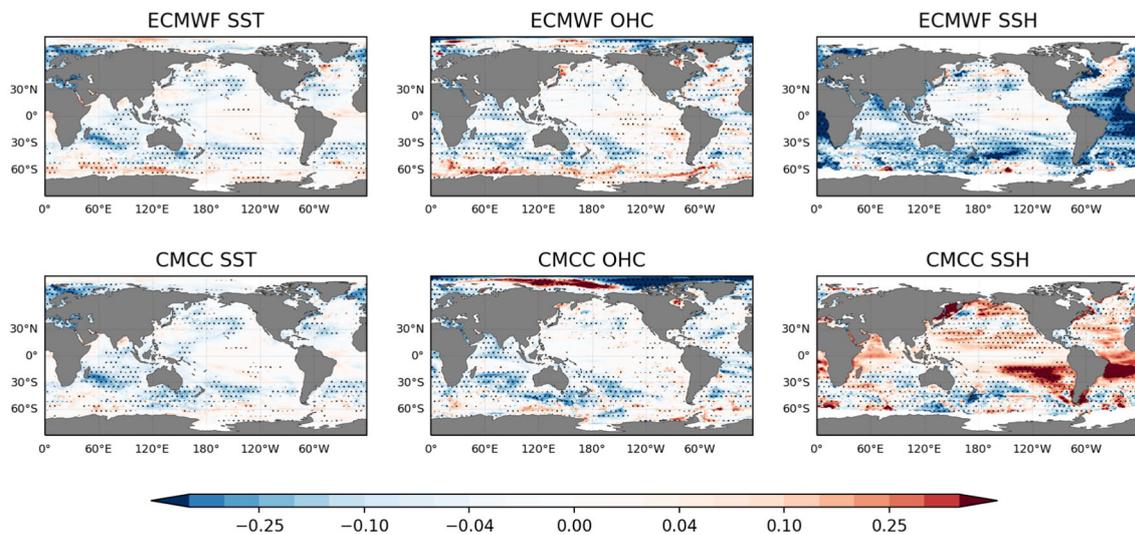


FIGURE 9

Contribution of the linear trend to the skill as measured by the differences between the anomaly correlation of the Trend Corrected and Detrended seasonal forecasts (top panels) and by the differences between the Detrended and standard seasonal forecasts (bottom panels). Dotted area indicate that the correlation differences are significant at the 90% level.

- Additional linear trend correction calibration step corrects some of these deficiencies and improves the forecast skill further. The linear trend correction appears to contribute to the skill in several areas of the Atlantic basin. However, it does not improve the forecast skill over the Arctic, suggesting that in these regions the trends may not be linear.
- The contribution of the linear trend to the skill has been quantified, and it is shown that this contribution is sizeable for SSH in the Atlantic and Southern Ocean and is also

visible in SST and OHC in the Indian Ocean, mid-latitudes, and areas of the Atlantic basin.

Results also highlight the importance of representing the decadal variability and trends in ocean heat content and sea level in the initial conditions. This is a non-negligible challenge for the ocean data assimilation systems used in the production of ocean initial conditions. The representation of decadal variability and trends is essential for decadal forecasts and climate projections. Therefore, the results from the seasonal forecasts

are also very relevant for the efforts on decadal variability and climate projections.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://cds.climate.copernicus.eu/cdsapp#!/dataset/seasonal-monthly-ocean?tab=overview>.

Author contributions

MB: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. RM: Data curation, Investigation, Writing – review & editing. SM: Funding acquisition, Investigation, Resources, Supervision, Writing – review & editing. MM: Investigation, Writing – review & editing. RS: Writing – review & editing. ED: Investigation, Writing – review & editing. SG: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was partially funded by the European Union's Horizon 2020 research and innovation program under grant agreement No 862626, as part of the EuroSea project. We also thank C3S_330 tender and the MEDSCOPE project (MEDSCOPE is part of

References

- Balmaseda, M. A., Anderson, D. L. T., and Davey, M. K. (1994). ENSO prediction using a dynamical ocean model coupled to statistical atmospheres. *Tellus* 46A, 497–511. doi: 10.1034/j.1600-0870.1994.00012.x
- Balmaseda, M. A., Mogensen, K., and Weaver, A. T. (2013). Evaluation of the ECMWF ocean reanalysis system ORAS4. *Q.J.R. Meteorol. Soc.* 139, 1132–1161. doi: 10.1002/qj.2063
- Blanchard-Wrigglesworth, E., Barthélemy, A., Chevallier, M., Cullather, R., Fučkar, N., Massonnet, F., et al. (2017). Multi-model seasonal forecast of Arctic sea-ice: Forecast uncertainty at pan-Arctic and regional scales. *Climate Dynam.* 49, 1399–1410. doi: 10.1007/s00382-016-3388-9
- de Boissésou, E., and Balmaseda, M. (2024). Predictability of Marine Heatwaves: assessment based on the ECMWF seasonal forecast system. *Ocean Sci.* 20, 265–278. doi: 10.5194/os-20-265-2024
- de Boissésou, E., Balmaseda, M. A., Abdalla, S., Källén, E., and Janssen, P. A. E. M. (2014). How robust is the recent strengthening of the Tropical Pacific trade winds? *Geophys. Res. Lett.* 41, 4398–4405. doi: 10.1002/2014GL060257
- de Boissésou, E., Balmaseda, M., Mayer, M., and Zuo, H. (2022). Section 4.3 of Copernicus Ocean State Report, issue 6. *J. Operat. Oceanogr.* 15, 1–220. doi: 10.1080/1755876X.2022.2095169
- Feng, X., Widlansky, M., Balmaseda, M. A., Zuo, H., Spillman, C., Smith, G., et al. (2024). Improved capabilities of global ocean reanalyses for analyzing sea level variability near the Atlantic and Gulf of Mexico coastal U.S. *Front. Mar. Sci. Sec. Coast. Ocean Processes* 11. doi: 10.3389/fmars.2024.1338626
- Good, S. A. (2020). ESA Sea Surface Temperature Climate Change Initiative (SST_cci): GHRSSST Multi-Product ensemble (GMPE), v2.0. *Centre for Environmental Data Analysis, 05 August 2020*. doi: 10.5285/a963d9415bb74247830f8704f825aa90
- Guemas, V., Blanchard-Wrigglesworth, E., Chevallier, M., Day, J. J., Déqué, M., Doblas-Reyes, F. J., et al. (2016). A review on Arctic sea-ice predictability and prediction on seasonal to decadal time-scales. *Q.J.R. Meteorol. Soc.* 142, 546–561. doi: 10.1002/qj.2401
- Johnson, S., Stockdale, T., Ferranti, L., Balmaseda, M., Molteni, F., Magnusson, L., et al. (2019). ECMWF-SEAS5: the new ECMWF seasonal forecast system. *Geosci. Model. Develop. Geosci. Model. Dev.* 12, 1087–1117.
- L'Heureux, M., Tippett, M. K., and Wang, W. (2022). Prediction challenges from errors in tropical Pacific sea surface temperature trends. *Front. Clim.* 4. doi: 10.3389/fclim.2022.837483
- Li, L., Lozier, M. S., and Li, F. (2022). Century-long cooling trend in subpolar North Atlantic forced by atmosphere: an alternative explanation. *Clim. Dyn.* 58, 2249–2267. doi: 10.1007/s00382-021-06003-4
- Long, X., Widlansky, M. J., Spillman, C. M., Kumar, A., Balmaseda, M., Thompson, P. R., et al. (2021). Seasonal forecasting skill of sea-level anomalies in a multi-model prediction framework. *J. Geophys. Res.: Oceans* 126, e2020JC017060. doi: 10.1029/2020JC017060
- Mayer, M., and Balmaseda, M. (2021). Indian Ocean impact on ENSO evolution 2014–2016 in a set of seasonal forecasting experiments. *Clim. Dyn.* 56, 2631–2649. doi: 10.1007/s00382-020-05607-6

ERA4CS, an ERANET initiated by JPI Climate, with co-funding by the European Union (grant agreement number 690462)).

Acknowledgments

The content of this manuscript has been presented in part of deliverable D4.6 for the EU project EuroSea: https://doi.org/10.3289/eurosea_d4.6_v2.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmars.2024.1380545/full#supplementary-material>

- McAdam, R., Masina, S., Balmaseda, M., Gualdi, S., Senan, R., and Mayer, M. (2022). Seasonal forecast skill of upper-ocean heat content in coupled high-resolution systems. *Climate Dynam.* 58, 3335–3350.
- Merchant, C. J., Embury, O., Bulgin, C. E., Block, T., Corlett, G. K., Fiedler, E., et al. (2019). Satellite-based time-series of sea-surface temperature since 1981 for climate applications. *Sci. Data* 6, 223.
- Pujol, M. I., Faugère, Y., Taburet, G., Dupuy, S., Pelloquin, C., Ablain, M., et al. (2016). DUACS DT2014: the new multi-mission altimeter data set reprocessed over 20 years. *Ocean Sci.* 12, 1067–1090. doi: 10.5194/os-12-1067-2016
- Sanna, A., Borrelli, A., Athanasiadis, P., Materia, S., Storto, A., Tibaldi, S., et al. (2017). CMCC-SPS: the CMCC seasonal prediction system 3. *Centro Euro-Mediterraneo sui Cambiamenti Climatici*. CMCC Tech. Rep. RP0285, 61pp.
- Sharmila, S., Hendon, H., Alves, O., Weisheimer, A., and Balmaseda, M. (2023). Contrasting el Niño–la Niña predictability and prediction skill in 2-year reforecasts of the twentieth century. *J. Climate* 36, 1269–1285. doi: 10.1175/JCLI-D-22-0028.1
- Stockdale, T. N. (1997). Coupled ocean–atmosphere forecasts in the presence of climate drift. *Mon. Wea. Rev.* 125, 809–818. doi: 10.1175/1520-0493(1997)125<0809:COAFIT>2.0.CO;2
- Storto, A., and Masina, S. (2016). C-GLORSv5: an improved multipurpose global ocean eddy-permitting physical reanalysis. *Earth Sys. Sci. Data* 8, 679–696.
- Storto, A., Masina, S., Simoncelli, S., Iovino, D., Cipollone, A., Drevillon, M., et al. (2019). The added value of the multi-system spread information for ocean heat content and steric sea level investigations in the CMEMS GREP ensemble reanalysis product. *Climate Dynam.* 53, 287–312.
- Taburet, G., Sanchez-Roman, A., Ballarotta, M., Pujol, M. I., Legeais, J. F., Fournier, F., et al. (2019). DUACS DT2018: 25 years of reprocessed sea level altimetry products. *Ocean Sci.* 15, 1207–1224. doi: 10.5194/os-15-1207-2019
- Tietsche, S., Balmaseda, M., Zuo, H., Roberts, C., Mayer, M., and Ferranti, L. (2020). The importance of North Atlantic Ocean transports for seasonal forecasts. *Clim. Dyn.* 55, 1995–2011. doi: 10.1007/s00382-020-05364-6
- Wallace, J. M., and Gutzler, D. S. (1981). Teleconnections in the geopotential height field during the northern hemisphere winter. *Mon. Wea. Rev.* 109, 784–812.
- Widlansky, M. J., Long, X., Balmaseda, M. A., Spillman, C. M., Smith, G., Zuo, H., et al. (2023). Quantifying the benefits of altimetry assimilation in seasonal forecasts of the upper ocean. *J. Geophys. Res.: Oceans* 128, e2022JC019342. doi: 10.1029/2022JC019342
- Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* Vol. 100 (Elsevier: Academic press).
- Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., and Mayer, M. (2019). The ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: a description of the system and assessment. *Ocean Sci.* 15, 779–808. doi: 10.5194/os-15-779-2019