Project information	
Project full title	EuroSea: Improving and Integrating European Ocean Observing and Forecasting Systems for Sustainable use of the Oceans
Project acronym	EuroSea
Grant agreement number	862626
Project start date and duration	1 November 2019, 50 months
Project website	https://www.eurosea.eu

Deliverable information	
Deliverable number	D4.6
Deliverable title	**Skill assessment of ECV/EOV from seasonal forecast**
Description	Assess the seasonal forecast skill of selected ocean variables - SST, OHC300m, and SSH - from the ensemble of ECMWF and CMCC seasonal forecasts systems contributing to C3S
Work Package number	4
Work Package title	Data integration, Assimilation, and Forecasting
Lead beneficiary	ECMWF
Lead authors	Magdalena A. Balmaseda, Ronan McAdam
Contributors	Simona Masina, Retish Senan, Michael Mayer, Silvio Gualdi
Due date	31.10.2022
Submission date	28.10.2022
Resubmission date	18.08.2023 (revised version)
Comments	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 862626.

https://doi.org/10.3289/eurosea_d4.6

Table of contents

Executive summary

The availability of consistent and sufficiently long observational records of ocean variables has allowed, for the first time, the assessment of the spatial distribution of the skill of ocean variables from seasonal forecasts. Three ocean state-of-the-art records of Essential Ocean/Climate Variables (EOVs/ECVs) have been used as verification datasets: sea surface temperature (SST) and sea level anomaly (SLA) from the Copernicus Climate Change Service (C3S) and upper 300m ocean heat content (OHC) from the Copernicus Marine Environmental Service (CMEMS) Global ocean Reanalysis Ensemble Products (GREP). Seasonal means of these records for the period 1993-2016 have been used to quantify the spatial distribution of the skill, up to 2 seasons ahead, of the ensemble of ECMWF (European Centre for Medium-Range Weather Forecasts) and CMCC (Centro Euro-Mediterraneo sui Cambiamenti Climatici) seasonal forecasts contributing to C3S.

This report 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. We also go beyond the standard verification metrics to 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. The reasons for the trend errors are multifarious. 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 likely associated with the response of models to the radiative forcing, and further research is needed to pin them down. 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. A comparative assessment of the skill among variables is also presented. Therefore, 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.

1. Context within EuroSea: Quality assessment of ocean variables from C3S seasonal forecasts.

This deliverable is a contribution from Task 4.6, which deals with the quality assessment of ocean variables from the C3S seasonal. This task is part of EuroSea WP4 on Data Integration, Assimilation, and Forecasting.

Knowledge of forecast skill is a prerequisite for utilizing forecast information. Assessing the skill of ocean variables from seasonal forecasts has remained elusive due to the lack of verifying ocean datasets of sufficient quality and length. In this task, we aim at observable essential ocean/climate variables (EOVs/ECVs) to verify seasonal forecasts from two seasonal forecast systems (CMCC and ECMWF) contributing to the C3S seasonal multi-model product. The EOVs/ECVs are Ocean Heat Content (OHC) from the CMEMS GREP ensemble of ocean reanalyses, the SLA ECV distributed by the Copernicus Climate Change Service (C3S), and the new SST records from ESA-CCI (European Space Agency – Climate Change Initiative), also distributed by C3S. This task evaluates the spatial distribution of skill in these variables, with a particular focus on the skill

for user-relevant indicators. A previous deliverable (D4.3[1]) reported the preparations of data and software for the demonstrators of seasonal forecast of user-relevant ocean indicators and their probabilistic skill is reported in Deliverable D7.4. This deliverable reports the spatial distribution of the seasonal forecast of the ensemble mean and compares it with the persistence and climatological benchmarks.

2. Data

2.1. Selection of Essential Ocean/Climate Variables

Three observational records 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. The C3S global Sea Surface Temperature Reprocessed product (Merchant et al. 2019, Good et al 2019) is used to verify the SST. The ocean heat content (OHC) in the upper 300m is verified with the Global Ocean Reanalysis Ensemble Product (GREP, Storto et al 2019). The sea level data set used here is the Sea Surface Height (SSH) product distributed by CMEMS and C3S (Pujol et al.,2016 and Taburet et al.,2019). A full description of the data is provided in Appendix I.a.

For the purpose of verification, we use seasonal means of these records for the period 1993-2016. Table 1 below shows the naming convention for the seasons.

Table 1: Naming convention used for the seasonal verification

Season Number (s)	Season Name	Months Included
1	FMA	February, March, April
2	MJJ	May, June, July
3	ASO	August, September, October
4	NDJ	November, December, January

2.2. Ocean output from seasonal forecasts

The two forecast systems used here are the Seasonal Prediction System Version 3 from the Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC-SPS3), and the fifth generation Seasonal Forecasting System from the European Centre for Medium-Range Weather Forecasts (ECMWF-SEAS5). 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, but these are not yet publicly available, since they have not yet been verified. This is the gap that the current EuroSea activity tries to fill.

A full description of the model components was provided in Deliverable D.3 and for convenience is included in Appendix I.b. 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 reanalyses.

[1] https://doi.org/10.3289/eurosea_d4.3

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 two seasons. Thus, forecasts initialized in February will be verified for the FMA (lead 1) and MJJ (lead 2). 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. Table 2 below provides a summary of the naming conventions used in seasonal forecast verification.

Table 2: Summary and naming convention of the seasonal forecast records that will be verified

Fc Initial month (short name)	Initial season (s_0)	Lead 1 Season	Lead 2 Season
Feb (0201)	1	FMA	MJJ
May (0501)	2	MJJ	ASO
Aug (0801)	3	ASO	NDJ
Nov (1101)	4	NDJ	FMA

3. Verification method and calibration

The ability of a prediction system to forecast specific events at a given time is measured by a set of skill scores or metrics. In EuroSea we focus on the scores of the ensemble mean forecast anomalies. The use of the forecast anomalies comes from practical considerations: i) at seasonal time scales we are interested in predicting deviations concerning the seasonal cycle; ii) using forecast anomalies concerning their own model climate is a first-order calibration step since errors in the simulation of the mean seasonal cycle are removed.

3.1. Calculation of seasonal anomalies

Seasonal forecasts provide information on seasonal anomalies with reference to (w.r.t.) seasonal climatology. The computation of the seasonal anomalies in the verification datasets is straightforward, but we spell out the procedure to introduce the naming convention and the treatment of the time axis. The observations can be represented as time series ***z(t)*** of two-dimensional maps in longitude-latitude space. In this case, *t* represents seasonal means. As the statistics presented here are a function of the calendar season, it is convenient to represent the time as a function of the year (*y)* and calendar season (*s*), so *z(t)=z(y,s)*, where y=1… N_y is the year index, and s=1,…4, is the season index as indicated in Table 1, and *t=(y-1)*4+s*. In the case of EuroSea, the number of years N_y is 24, spanning the observed record 1993-2016.

The time mean $\langle z(s) \rangle$ and interannual variability $\sigma_z(s)$ of the observations are calculated for each calendar season by averaging over the number of years N_y. The calendar mean allows the computation of the interannual anomalies ***z'(y,s)*** and their interannual variability $\sigma_z(s)$ as shown below:

$$\langle \boldsymbol{z}(s) \rangle = \frac{1}{N_y} \sum_{y=1}^{N_y} \quad \boldsymbol{z}(y,s) \qquad (1.1)$$

$$\boldsymbol{z}'(y,s) = \boldsymbol{z}(y,s) - <\boldsymbol{z}(s)> \quad (1.2)$$

$$\boldsymbol{\sigma_z}(s) = \sqrt{\frac{1}{N_y-1}\sum_{y=1}^{N_y} (\boldsymbol{z}(y,s) - \langle \boldsymbol{z}(s) \rangle)^2} = \sqrt{\frac{1}{N_y-1}\sum_{y=1}^{N_y} (\boldsymbol{z}'(y,s))^2} \quad (1.3)$$

For simplicity, sometimes the dependency on calendar month is omitted, so the mean, interannual variability and anomalies will be denoted as **<z>** and $\boldsymbol{\sigma_z}$.

The interannual variability of the observations is equivalent to the mean square error (RMSE) of a forecast based on climatology. This value is used to benchmark the quality of the forecast.

The forecast seasonal anomalies are computed with respect to the model seasonal climatology, which depends on the forecast lead time, as described below.

We denote each of the individual ensemble members of seasonal reforecasts $\boldsymbol{x}^i(y,s_0,l)$, where super-index *i* refers to the ensemble member (*i=1,….M*); the pair (y,s_0) describes the initialization time (year and month) within the reforecast period, and *l* is the forecast lead time (*l=1,…2*), such that *t=(y-1)*4+s₀+/-l*. For each starting date and lead time, we compute the forecast ensemble mean $\overline{\boldsymbol{x}}(y,s_0,l)$, from which we derive the forecast seasonal climate $\langle \boldsymbol{x}_{\boldsymbol{s_0}}(l) \rangle$ and the ensemble mean interannual anomalies $\overline{\boldsymbol{x}}'(y,s_0,l)$:

$$\overline{\boldsymbol{x}}(y,s_0,l) = \frac{1}{M}\sum_i^M x^i(y,s_0,l) \quad (2.1)$$

$$\langle \boldsymbol{x}_{s_0}(l) \rangle = \langle \boldsymbol{x}(y,s_0,l) \rangle = \frac{1}{N_y}\sum_y^{N_y} \overline{\boldsymbol{x}}(y,s_0,l) \quad (2.2)$$

$$\overline{\boldsymbol{x}}'(y,s_0,l) = \overline{\boldsymbol{x}}_{t_0}(y,s_0,l) - \langle \boldsymbol{x}_{s_0}(l) \rangle \quad (2.3)$$

As can be seen in equation (2.1), the model climate is estimated as the average of all the ensemble forecasts starting in a given calendar month (e.g., all forecasts initialized on the first of May of the different years in the reforecast period). There are two important aspects in the above equation: i) the forecast anomalies are computed w.r.t the model seasonal climatology, rather than w.r.t the observed climatology; ii) The model climate depends then on the calendar month in which the forecasts were initialized, and changes with lead time, as the model drifts away from the observed climate.

As an example, Figure 1 shows the mean bias in forecasts of SST initialized in May from the ECMWF model, in the first and the second season. The amplitude of the model error is substantial. 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. The verification statistics used in what follows are therefore bias blind.

In the following, we will focus only on anomalies. Assessment of bias and variability in SST and OHC have been reported in McAdam et al 2022, and will not be further discussed here.

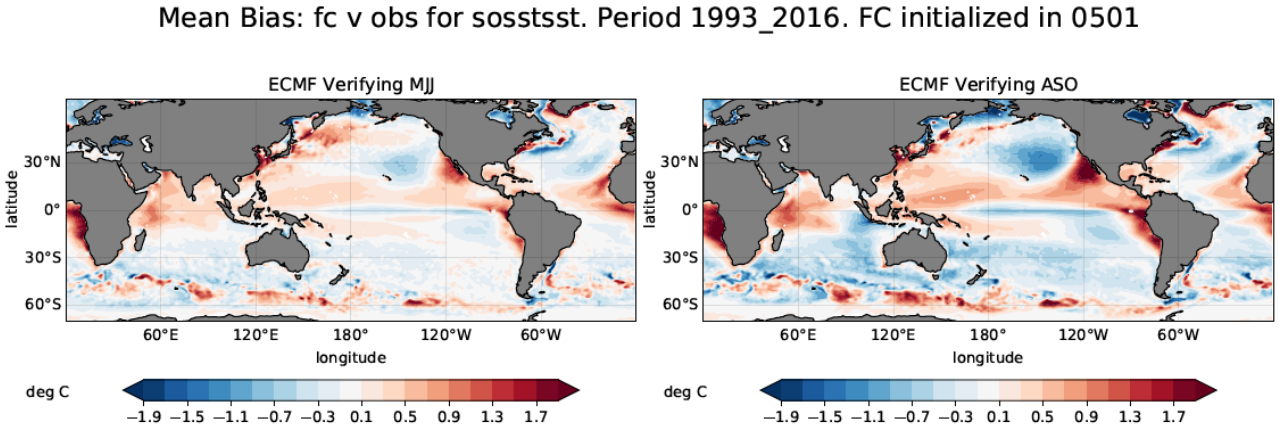


Figure 1. Differences between forecast and observation mean state for SST, in the ECMWF seasonal forecasts initialized in May and verifying in MJJ (lead 1) and ASO (lead 2).

3.2. Verification Statistics

Since the initial focus is to quantify the performance of the ensemble mean, we have chosen two different deterministic scores: anomaly correlation coefficient (ACC) and Mean Square Skill Score (MSSS). As for the model climate, the skill scores also depend on the forecast starting month and lead time, as shown below:

$$ACC(s_0, l) = \frac{{}^{1}/_{N_y} \sum_y \overline{\boldsymbol{x}'}(y,s_0,l) * \boldsymbol{z}'(y,s)}{\boldsymbol{\sigma}_{\bar{x}} \boldsymbol{\sigma}_{\boldsymbol{z}}} \quad (3.1)$$

$$MSSS(s_0, l) = 1. - \frac{RMSE_{fc}(s_0,l)}{RMSE_{clim}} = 1. - \frac{RMSE_{fc}}{\boldsymbol{\sigma}_{\boldsymbol{z}}} \quad (3.2)$$

$$RMSE(s_0, l) = \frac{1}{N_y} \sum_y (\overline{\boldsymbol{x}'}(y, s_0, l) - \boldsymbol{z}'(y,s))^2 \quad (3.3)$$

The MSSE 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.

3.3. 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 from equations (3), 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. 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: as equation (3) but removing the linear trend in the anomalies of both forecasts and observations. 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.

4. Skill without trend correction

Here we discuss the level of skill for the different variables, and how it compares to the persistence and climatological benchmarks as a function of lead time for selected initial dates. The skill for the different seasons and variables is presented in Appendix II.

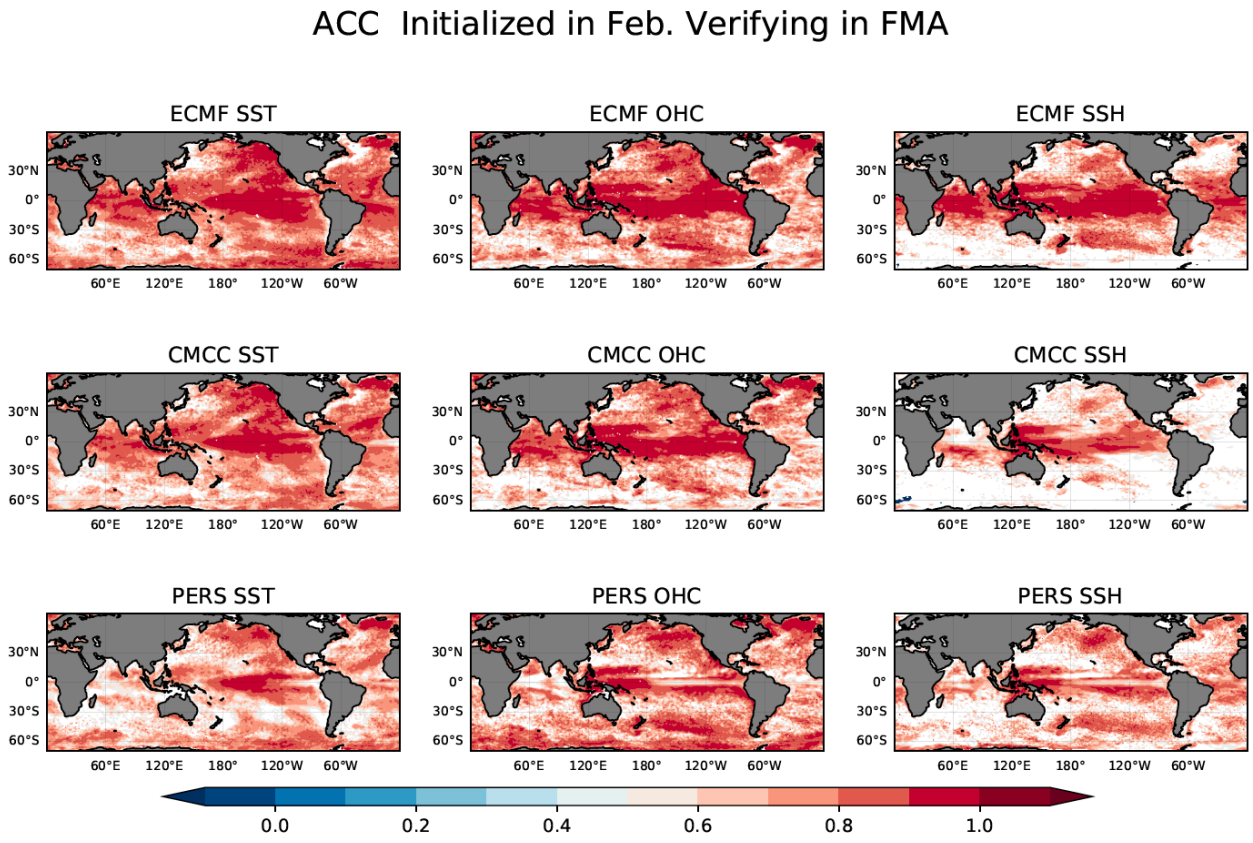


Figure 2. Anomaly Correlation Coefficient (ACC) of the seasonal forecasts initialized in February and verifying in FMA (1 season ahead) for SST (left column), OHC (middle) and SSH (right). Shown is the skill of the ECMWF (top) and CMCC (middle column) seasonal forecasting systems, as well as that of the persistence forecasts (bottom). Only correlation values with p-values < 0.05 are shown. Positive values indicate that the forecast skill is better than the climatological forecast.

The skill of the ECMWF and CMCC seasonal forecasting system in predicting the FMA season when initialized in February is shown in Figure 2, for the three different ocean variables. For reference, the skill of the persistence forecast is also shown. The persistence forecast is made by persisting the values of the observed anomalies (which for the forecast initialized in February, will be the January monthly mean). All the forecasts, including persistence, have a high level of skill in the first season. Nonetheless, even in the first season, the dynamical models appear 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 in section 6.

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. Therefore, 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 3. The dynamical models retain significant skill levels: for SST, ACC values larger than .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. This aspect is further discussed in section 4.1.

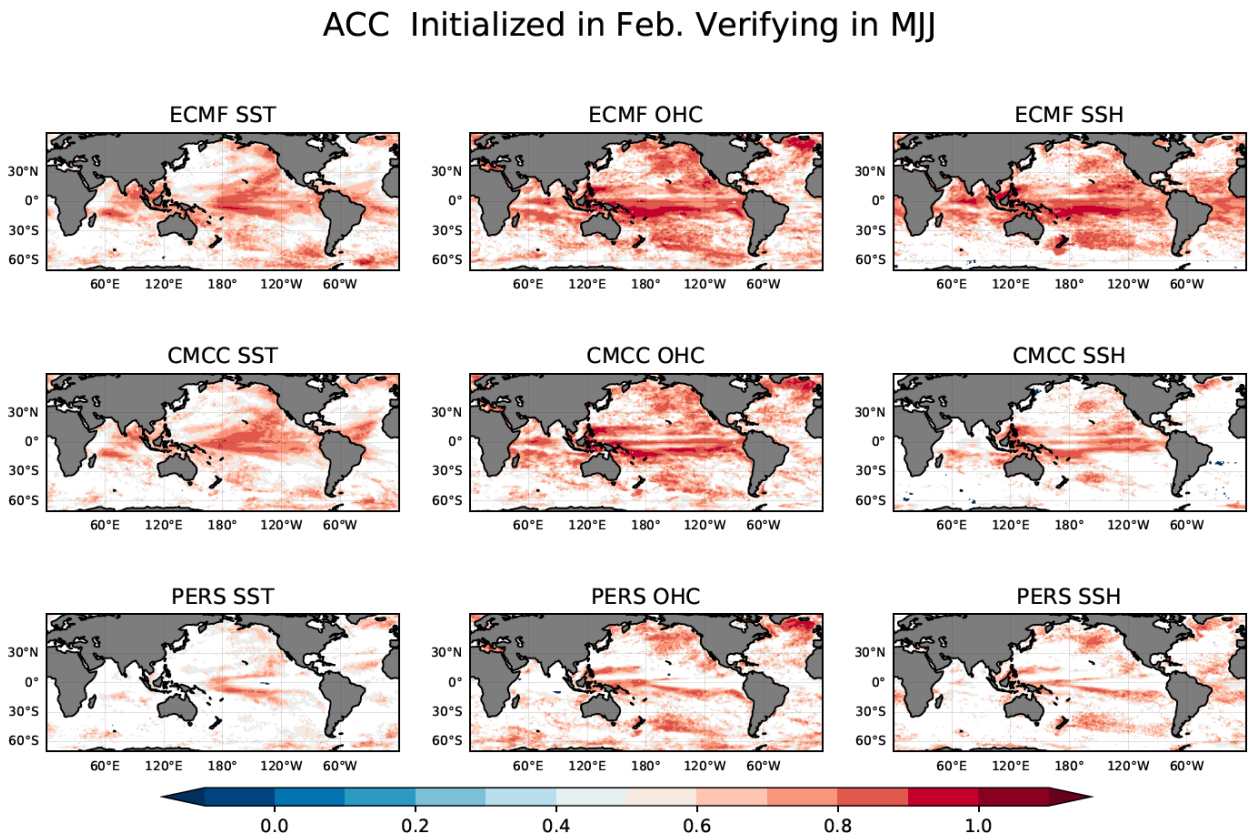


Figure 3. As Figure 2 but for the second season into the forecast.

The climatological forecast is another benchmark for the skill of seasonal forecasts, which is measured with the Mean Square Skill Score or MSSS (note that the ACC skill of the climatological forecast by definition is zero). The MSSS is sensitive to the amplitude of the anomalies in the ensemble mean, which decreases with lead time as the forecast ensemble spread decreases. A summary of the skill gained by the dynamical seasonal forecasts against persistence and climatology is given in Figure 4 and Figure 5, in terms of ACC differences w.r.t persistence and MSSS respectively. The figures show the skill gains in the three ocean variables for forecast initialized in May one and two seasons ahead. Positive values indicate that the seasonal forecasts are better than the benchmarks.

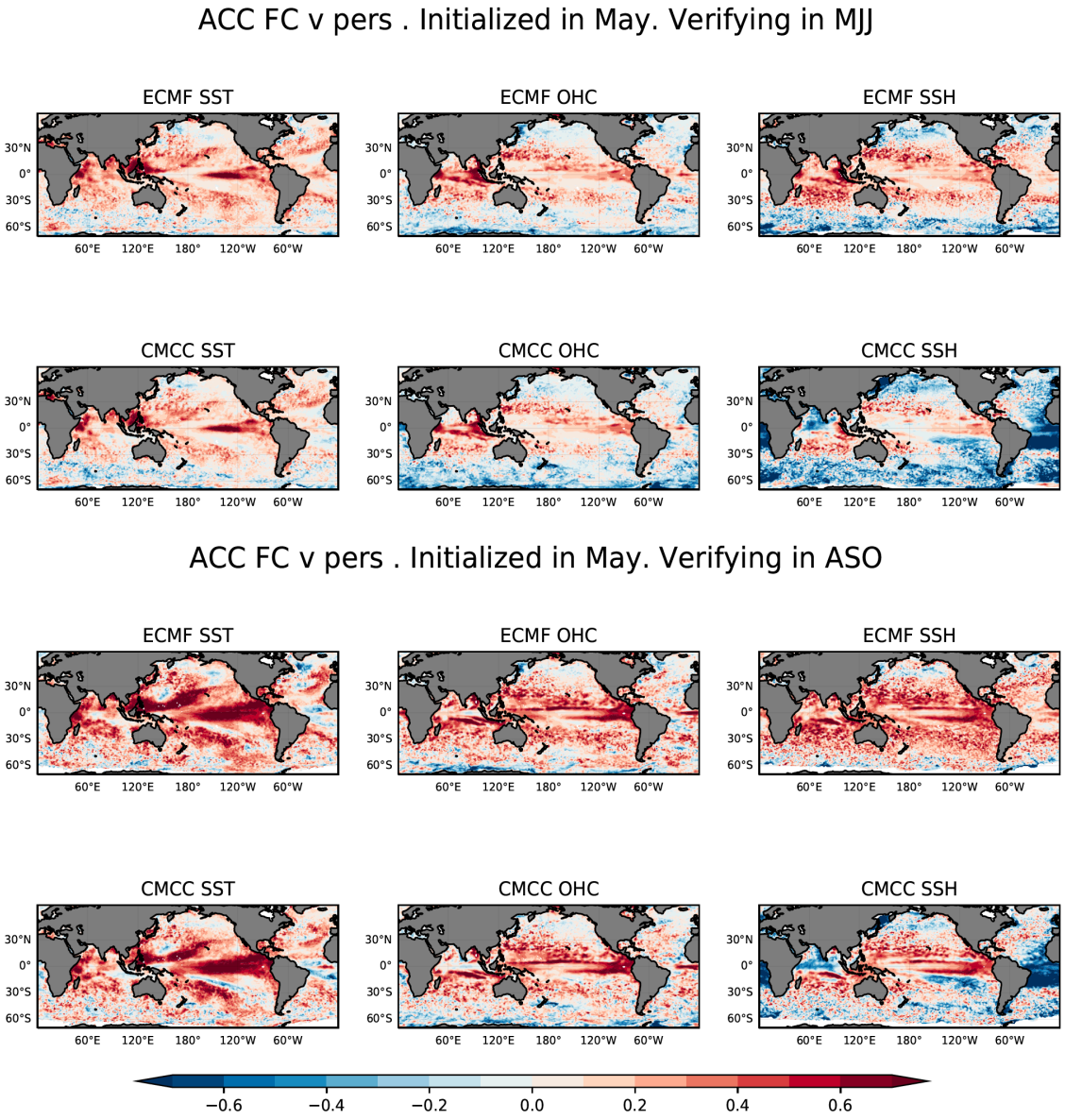


Figure 4. Summary of ACC skill differences of the dynamical seasonal forecasts against persistence, for the different variables for forecast initialized in May. Shown are the differences for the first and second seasons (top and bottom respectively). Positive values indicate that the dynamical seasonal forecasts are better than the persistence benchmark.

Figure 4 shows that the advantage against the persistence benchmark is already visible in the first season and increases further in the second season. The pattern of SST skill is indicative of the dynamical processes operating in the coupled model at these time scales, which are associated with tropical wave dynamics and are most noticeable in the Tropical Pacific, with the ENSO signature in the cold tongue region and Western Pacific clearly visible. The gains in the extratropics are likely related to the thermal memory of the mixed layer, but also as a result of the predictable variations in the atmospheric circulation (e.g., position and amplitude of anticyclones in summer will have an impact on the SST). An area of concern is the Equatorial Atlantic in the second season when the skill of both models is lower than persistence. This is attributed to

deficiencies in the forecast models or initialization, and highlights the potential for further skill gains that could be realized by improving the forecasting systems. The North Atlantic Subpolar Gyre in the ECMWF is another area where the model skill is lower than persistence. This has been attributed to a problem with the ocean initialization (Tietsche et al 2020), which is expected to be solved in the upcoming forecasting system SEAS6. There are also regions of decreased skill 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).

The patterns of ACC skill differences (or skill gains) between models and persistence are quite similar in OHC and SSH, indicative of the strong correlation between the two quantities, while exhibit visible differences with the patterns of SST differences. The pattern of skill gains in OHC and SSH bears the signature of the equatorial wave dynamics, confined to the Equatorial band and outside of which is rather homogeneous, has less meridional span than that of SST, and does not show a decrease of skill associated with atmospheric convergence zones apparent in the SST skill gain pattern. The different skill gain patterns between SST and OHC/SSH can be attributed to 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 in the atmospheric model can induce errors in the SST forecasts which are not so visible in the integrated ocean variables.

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.

Figure 5 shows the MSSS of forecasts initialized in May and verifying in MJJ and ASO (e.g. first and second seasons) for ECMWF, CMCC and persistence, and the three ocean variables of interest. Values larger than zero indicate that the forecast beat climatology. This is the case for the two dynamical forecasts almost everywhere, except for some regions. A notable exception is the North Atlantic subpolar gyre in the ECMWF system, and the Southern Ocean and the Subtropical Atlantic for the CMCC system. At these longer lead times, persistence is worse than climatology almost everywhere, except for some regions dominated by long-term trends and/or decadal variability, as will be discussed later.

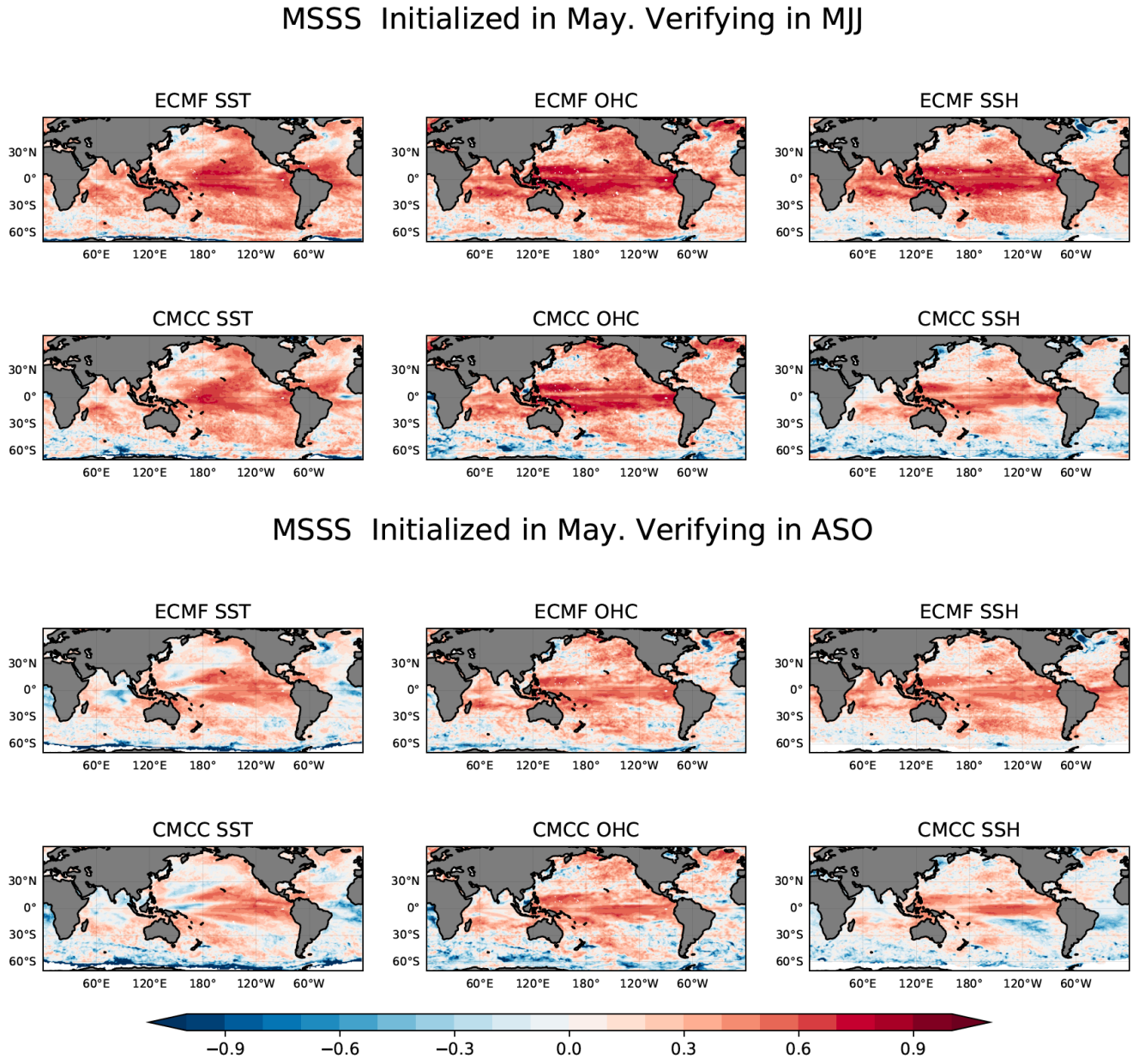


Figure 5. Summary of MSSS of the dynamical seasonal forecasts for the different variables for forecast initialized in May. Shown are the differences for the first and second seasons (top and bottom respectively). Positive values indicate that the dynamical seasonal forecasts are better than the climatological benchmark.

4.1. Comparative skill: OHC versus SST

Subsurface anomalies are thought to persist longer than surface anomalies, making subsurface seasonal predictions inherently easier. To test this, we compare the skills of OHC and SST persistence forecasts. The skill of a persistence model is used as a proxy for the persistence (duration) of anomalies in time (i.e., persistence skill is high if the anomalies in the validation dataset do not decay). In the interest of brevity, the anomaly correlation coefficient is used as the comparison skill measure. As expected, we find that OHC persistence correlation is higher than SST persistence correlation in 70% of the ocean surface area between

70S and 70N (in both lead seasons). However, this difference is statistically significant in only 20% of the ocean (in both lead seasons); if a p-value threshold of 0.1 is used instead of 0.05, the area covered only increases to 30%. It is possible, then, that OHC anomalies simply do not persist for significantly longer on the seasonal timescales than SST anomalies do. It is also possible that the common duration of each product (24 years) is too small of a dataset to detect significant differences between surface and subsurface anomaly persistence. Datasets stretching further back (or forward) in time would be required to confirm this point.

Nonetheless, there are regions where OHC is (significantly) more persistent than SST (on seasonal timescales), and they are the Equatorial Pacific, north-east Atlantic and parts of the North Pacific (Figure 6, top). In such regions, it is expected that the OHC re-forecast skill in our forecast systems would be higher than the SST re-forecast skill. This is shown to be true, as the improvements in OHC skill over SST skill in the dynamical systems (Figure 6, bottom) match the geographical extent and magnitude of the differences in persistence. In fact, at longer lead times the improvement of OHC forecasts over SST increases, again as a result of the longer persistence.

There are also many regions where OHC skill is less than or insignificantly different to SST skill in the dynamical models, particularly in the first season (Figure 6, bottom). The dynamically active ENSO region over the Central Equatorial Pacific stands out as the area where the dynamical model comparisons differ the most from the persistence forecast comparison. The widespread similarity of scores is a sign that SST dynamical forecasts either exploit the thermal and dynamical memory of the subsurface or benefit from improved initial conditions relative to the OHC (due to denser observations at the surface e.g., satellites). It is likely that both play a role, and the details of that role depend on the density of observations as well as the particular benefits created by these observations in particular regions. However, the differences (between OHC and SST in dynamical forecasts compared to persistence) indicate that, in places such as the equatorial band, SST forecasts from dynamical models implicitly include the predictive skill of OHC.

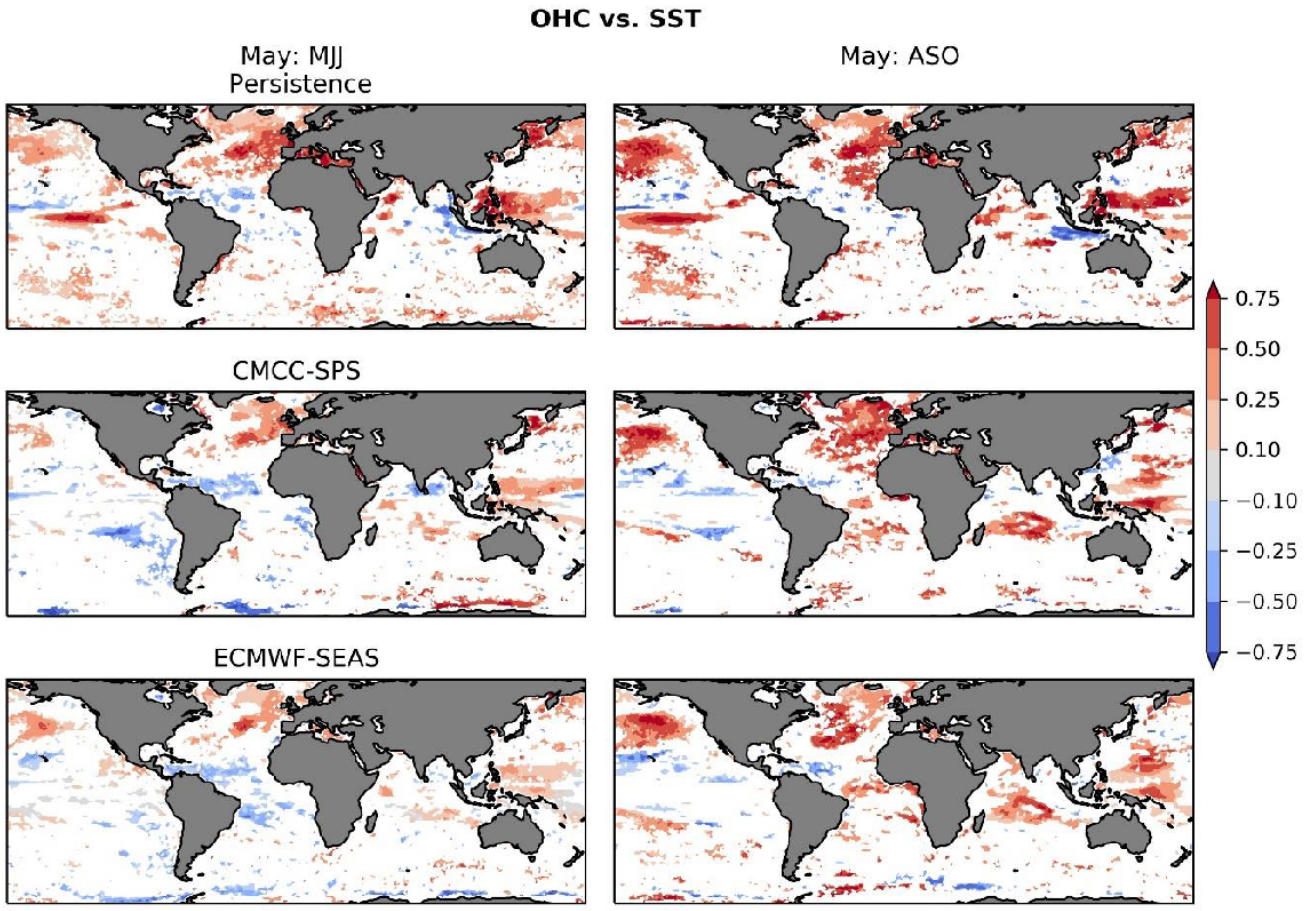


Figure 6. Difference in anomaly correlation coefficients for OHC 0-300 m and SST in three different models: Persistence (top), CMCC-SPS3 (middle) and ECMWF-SEAS5 (bottom). Positive values show where OHC skill is greater than SST skill in the corresponding model. Re-forecasts and the persistence model are initialised in May, and the seasonal averages of Lead 1 and Lead 2 seasons are shown. White regions indicate where differences in correlations are not statistically significant (p-value > 0.05).

5. Fidelity of trends

5.1. Linear trend in observational records

The 1993-2016 linear trend in the observational records of SST, SSH and OHC for the different seasons is shown in Figure 7. 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 marine heat waves, induced by persistent atmospheric anticyclones and deeper mixed layer (De Boisseson et al 2022). 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. Interestingly, the OHC also exhibits a negative trend over the North Atlantic subpolar gyre. The pattern of the trends is similar in all seasons, but the amplitude shows some seasonal variations (e.g., stronger extratropical SST warming in the respective summer hemisphere, and enhanced SSH trends in the northwestern Equatorial basins, likely related to the seasonal migration of the tropical gyre circulation). It is interesting that the clear trends in the tropical Atlantic in SSH and OHC do not have an obvious footprint in SST trend in that area.

5.2. 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 sake of brevity, we show only the fidelity of the forecast trends only for forecasts initialized in May. The results from other seasons can be seen in Appendix III.

Figure 8 (top 2 rows) shows the differences in linear trends for the MJJ season between seasonal forecasts initialized in May and observations. Differences are sizeable even at this short lead time. Most noticeable is the global difference in the SSH in the CMCC. This has been traced back to the fact that the ocean initial conditions in the CMCC system did not include the global changes in steric height. Aside from this global difference, both forecast systems exhibit also regional departures from the observational trends. The ECMWF overestimates 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 the ocean initial conditions, which could have spurious variability resulting from the changing observing system. For instance, equivalent seasonal forecasts initialized from an Ocean Observing System Experiment where the Argo observations were removed exhibited slightly different trend patterns (not shown). The eastern equatorial Pacific SST warming is also present in the CMCC seasonal forecasts, but with weaker amplitude, and it does not have a clear footprint in the OHC trends. This feature is present in most of the forecasts except for those initialized in November (see Appendix III). 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. Both ECMWF and CMCC favour warmer than observed SST in tropics at the expense of cooler than those in extratropics, although the footprint in OHC is not so consistent. The SST warming over warm pool regions is also overestimated in both forecast models. 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).

As the forecast progresses, the errors in the trends evolve, as can be seen in the 2 bottom rows of Figure 8, which shows the trend errors for forecasts initialized in May verifying in ASO. The excessive warming in the Eastern Equatorial Pacific is amplified in both models, now with a clear dynamical signature in OHC and SSH, which also show cooling in the Western Equatorial Pacific, consistent with the flattening of the thermocline. The errors in the trends have implications for the prediction of ENSO. Indeed, seasonal forecast models over-predicted the warming of El Nino in 2014-15 and are struggling to predict the three consecutive years of cool La Nina conditions since 2020.

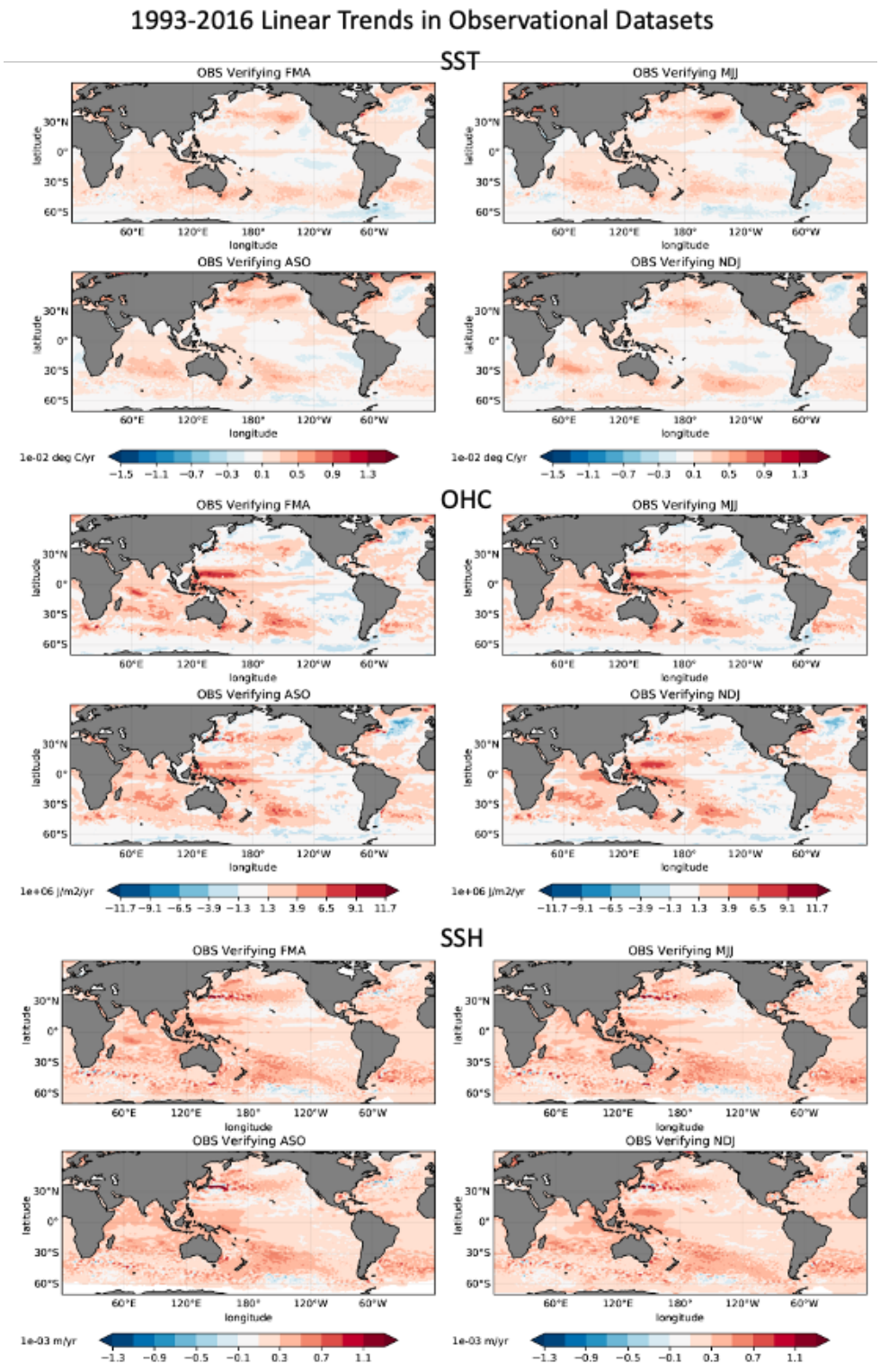

1993-2016 Linear Trends in Observational Datasets
SST
OBS Verifying FMA
OBS Verifying MJJ
OBS Verifying ASO
OBS Verifying NDJ
latitude
longitude
30°N
0°
30°S
60°S
60°E
120°E
180°
120°W
60°W
1e-02 deg C/yr
−1.5 −1.1 −0.7 −0.3 0.1 0.5 0.9 1.3
OHC
1e+06 J/m2/yr
−11.7 −9.1 −6.5 −3.9 −1.3 1.3 3.9 6.5 9.1 11.7
SSH
1e-03 m/yr
−1.3 −0.9 −0.5 −0.1 0.3 0.7 1.1

Figure 7. Linear trend in observational records for the period 1993-2016 for the different seasons in SST (4 upper panels, units are 0.01 deg C/year), OHC (middle 4 panels, units 10^6 J/m2/year) and SSH (4 lower panels, units mm/year).

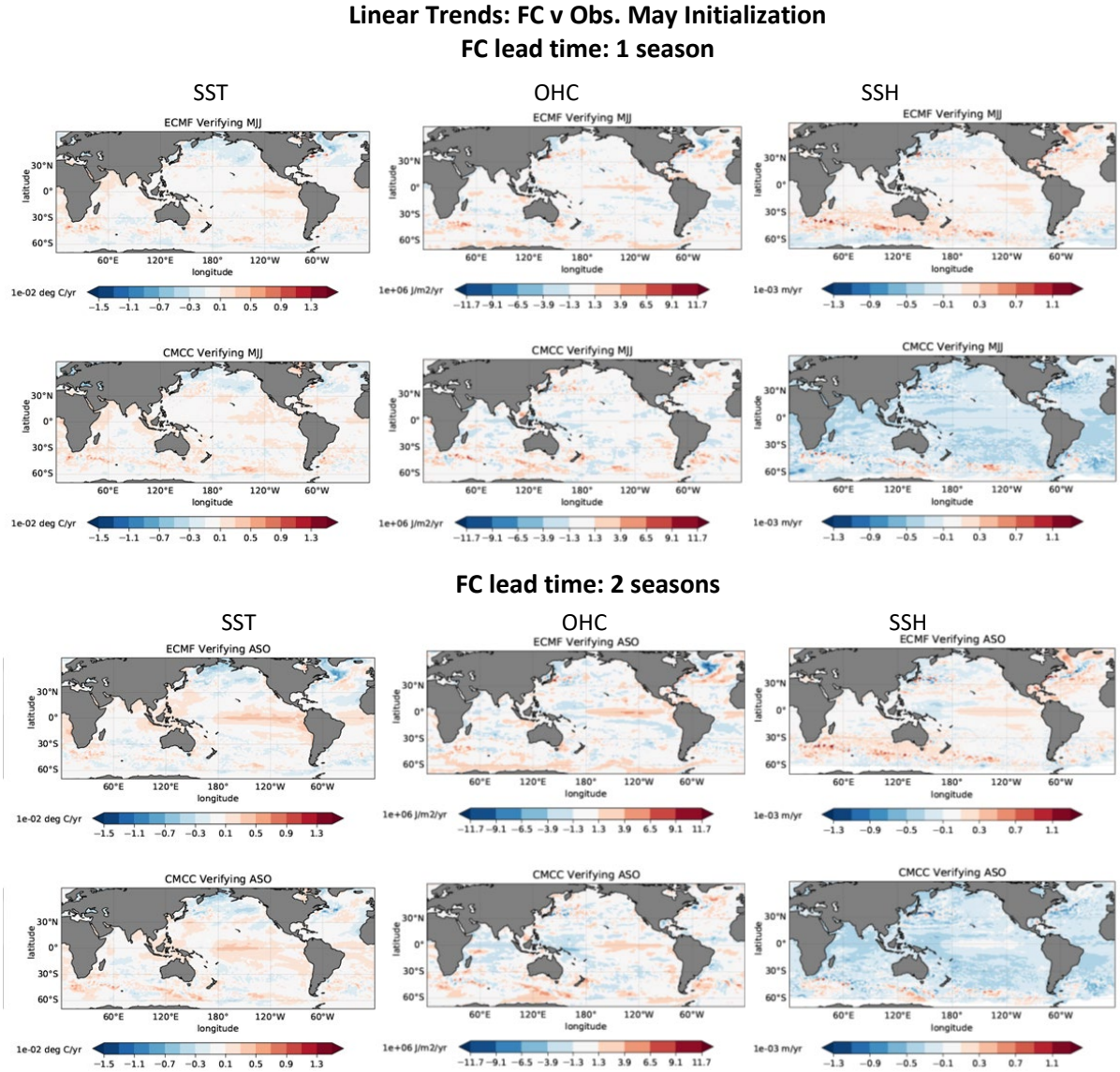


Figure 8. Differences in linear trends between seasonal forecasts and observations. The forecasts were initialized in May and verify in MJJ (top 2 rows) and ASO (lower 2 rows). Shown are trend differences in SST (left, units 0.01 degC/year), OHC (middle, units 10^6 J/m2/year), and SSH (right, units mm/year) for the ECMWF (top) and CMCC (bottom) seasonal forecasting systems.

6. Skill with trend correction

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. For the sake of brevity, we only use the ACC metric.

The skill gains obtained by the additional calibration step of correcting the linear trends are shown in Figure 9, which shows ACC differences between the trend-corrected versus the standard seasonal forecast initialized in May. Overall, positive values are seen for the 3 variables. In the first season, the impact in SST and OHC is confined to some regional areas around the sea-ice edge. But for the second season we see an impact on SST skill in wider areas of the northern extratropics, and noticeable in the Atlantic basin. Interestingly we also see a slightly negative contribution of the linear trend corrections in areas of the Indian Ocean and Pacific Warm Pool, which may be an indication of the trend not being so linear. For SSH the story is different: 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, consistent with the problem of the trends residing in the ocean initial conditions. The skill gains are especially high for the Atlantic basin, the Southeastern 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 SSH are not applied to the model. It is therefore not obvious why a global increase in sea level should have such a clear spatial structure.

It is of interest to evaluate the skill of the trend-corrected forecasts against persistence. This can be seen in Figure 10, which shows similar diagnostics as Figure 5 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. Equally, the underperformance in SST along the edges of the tropical atmospheric trade winds, which may be attributed to errors in the atmosphere, remains.

The contribution of the linear trends in the overall level of skill in seasonal forecasts appears in Figure 11, which shows 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. The impact of the trend in forecast skill is consistent in both forecasting systems.

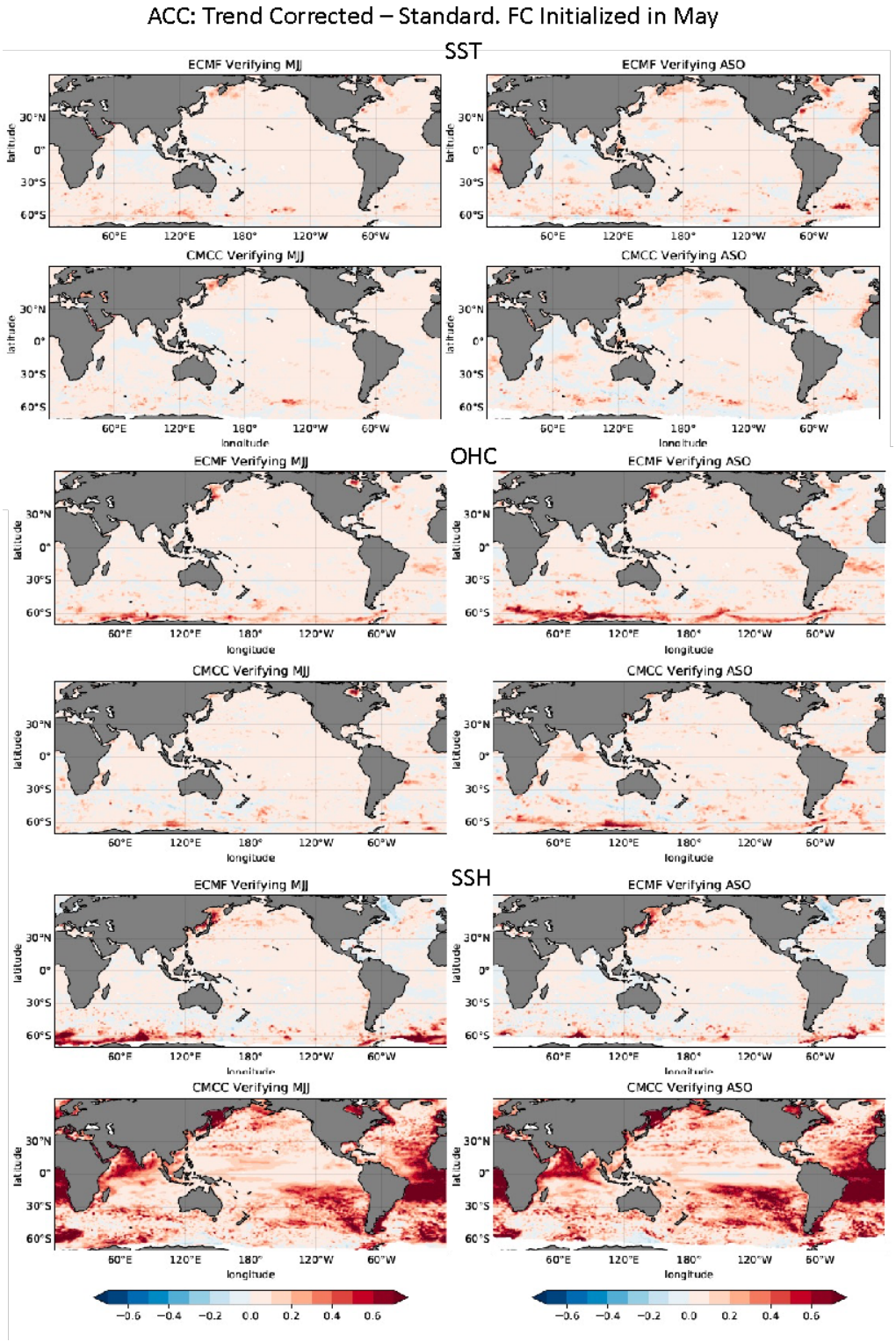


Figure 9. Differences in ACC skill between trend-corrected forecasts and those with standard calibration. Shown are the values for the ECMWF and CMCC seasonal forecasts initialized in May and verifying one and two seasons ahead (left and right columns), for SST (top 2 rows), OHC (central 2 rows) and SSH (lower 2 rows).

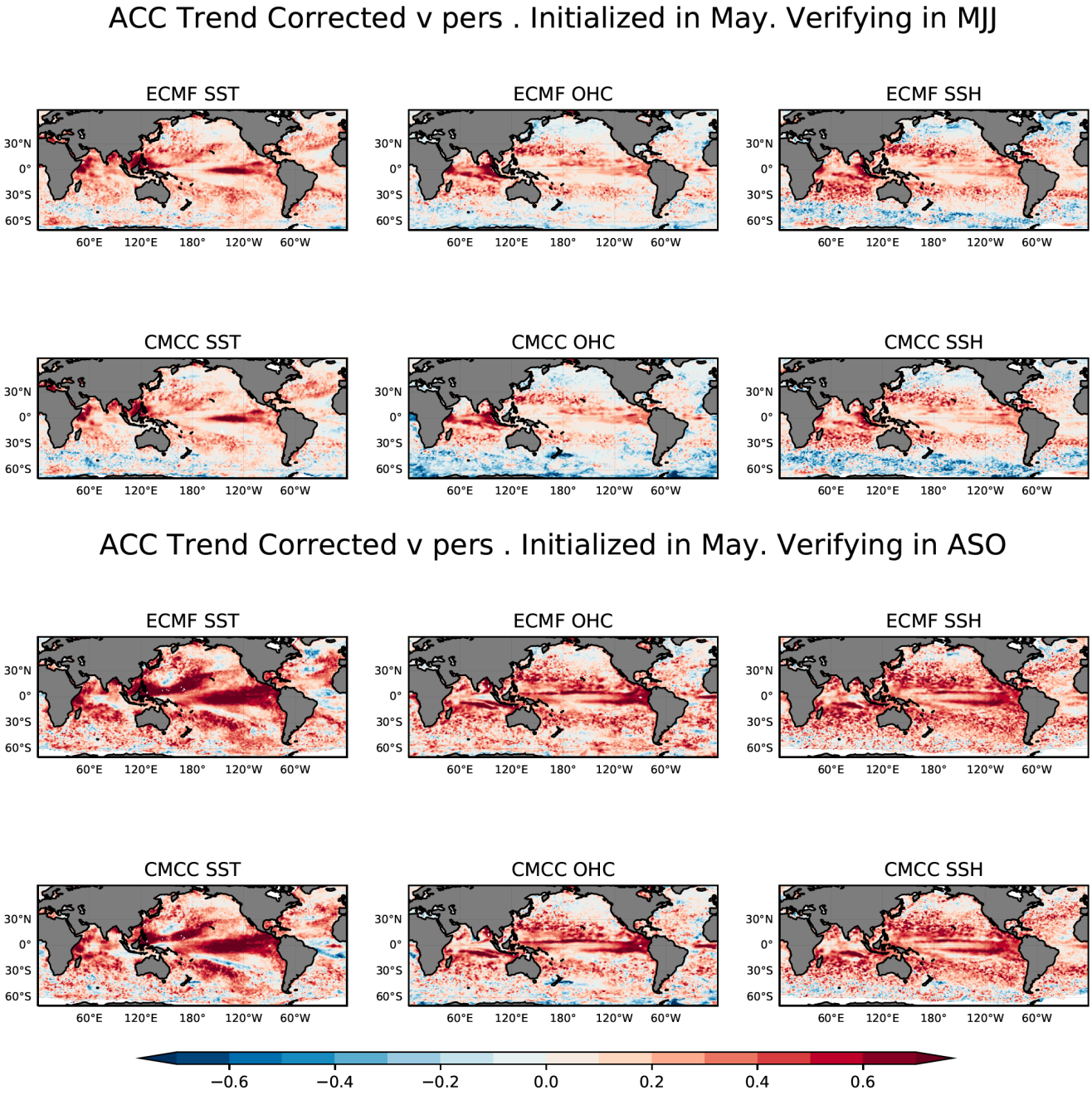


Figure 10. As Figure 4 but showing the difference in ACC between the Trend-Corrected forecasts and persistence.

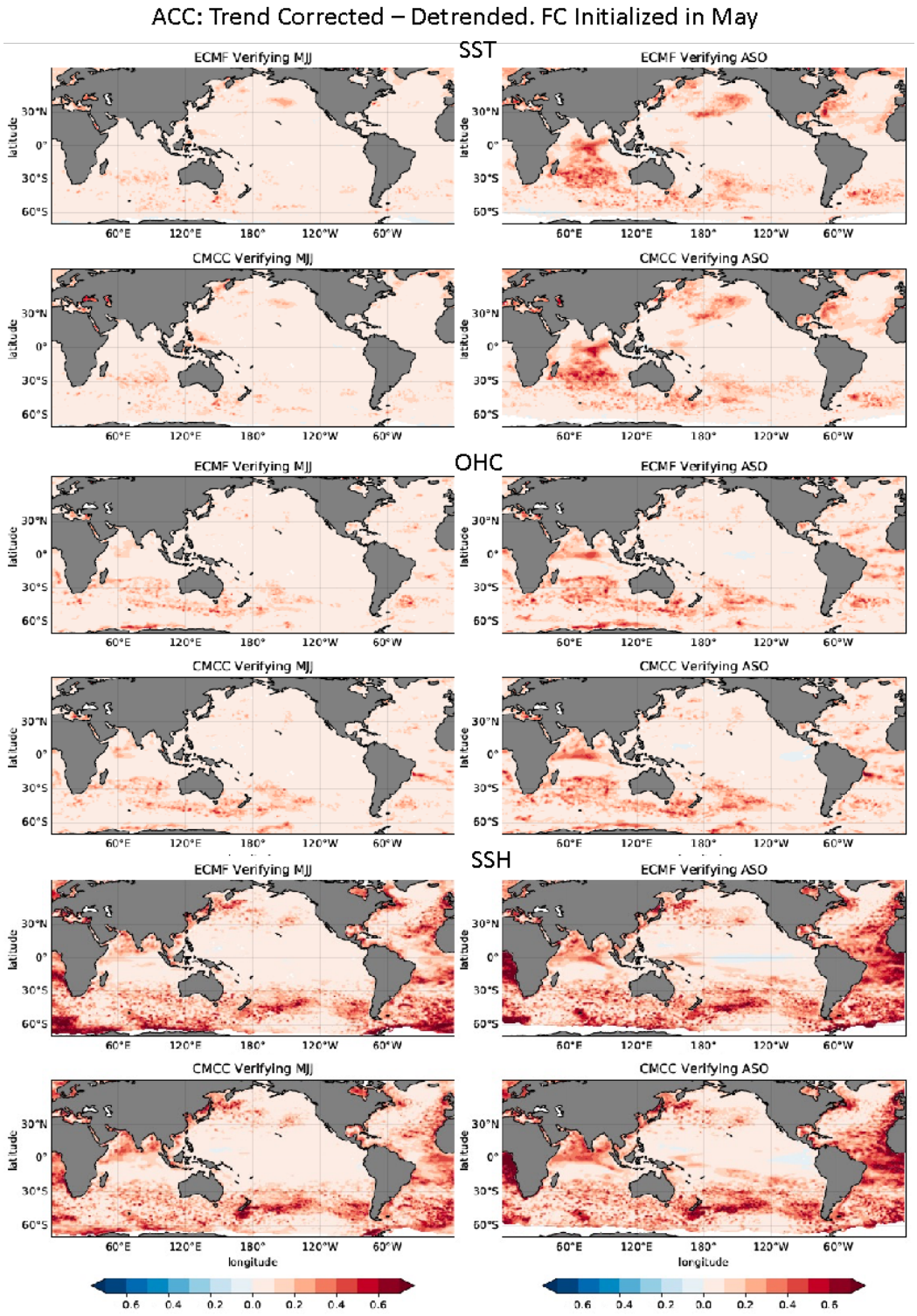


Figure 11. Contribution of the linear trend to the skill as measured by the differences between Trend Corrected and Detrended seasonal forecasts.

7. 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 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.
- The ability of the seasonal forecasts to capture the linear trends in observations have 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, although deficiencies of the models to persist the trends into the forecast cannot be ruled out at this stage. This is certainly the case for the SSH trends in the CMCC system.
- Additional linear trend correction calibration step can correct for deficiencies and improves the forecast skill further. The linear trend correction appears to contribute to the skill in several areas of the Atlantic basin.
- 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 scattered areas of the Atlantic basin.
- Comparison with the persistence and climatological benchmark indicates that there is still scope for further skill gains in the extratropical oceans (by improving the ocean initial conditions) and along the edges of the large atmospheric trade-wind system by improving the atmosphere models.

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.

Appendix I. Description of datasets

Selection of Essential Ocean/Climate Variables

ESA CCI Sea Surface Temperature

The ESA SST CCI (Climate Change Initiative) and C3S global Sea Surface Temperature Reprocessed product[2] provide gap-free maps of daily average SST at 20 cm depth at 0.05deg. x 0.05deg. horizontal grid resolution, using satellite data from the (A)ATSRs, SLSTR and the AVHRR series of sensors (Merchant et al., 2019). The ESA SST CCI and C3S level 4 analyses were produced by running the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system (Good et al., 2020) to provide a high resolution (1/20deg. - approx. 5km grid resolution) daily analysis of the daily average sea surface temperature (SST) at 20 cm depth for the global ocean. Only (A)ATSR, SLSTR and AVHRR satellite data processed by the ESA SST CCI and C3S projects were used, giving a stable product. The record covers the period 1982-2019.

Global ocean Reanalysis Ensemble Product (GREP)

GREP[3] is an ensemble of four global 3D ocean reanalysis products: C-GLORS v7 (CMCC: Storto and Masina, 2016), FOAM (Met Office UK: Blockley et al., 2014), GLORYS2V4 (Mercator: Garric et al., 2017) and ORAS5 (ECMWF: Zuo et al., 2019). All products are built on version 3 of NEMO and are provided from 1993 to 2019 on the native ORCA025 tri-polar curvilinear grid. There are 75 depth levels, 34 of which are shallower than 300 m. All use the same fluxes (CORE) and atmospheric forcing (ERA-I) (with the subtle exception being ORAS5, which also includes wave effects). All products assimilate similar data streams, typically ARGO, XBT temperature profiles and AVISO Sea Level Anomaly. However, the products all have diverse assimilation schemes, observation quality control, model parameters, spin-up and surface constraints (Storto et al., 2019). The GREP data used here covers the period 1993-2018.

Ocean reanalyses are a convenient choice for the task of global heat content validation because the ocean variables have coverage in space and time that is not matched by observations (Riser et al., 2016). Besides, ocean reanalyses integrate observational information with that of atmospheric reanalyses via a physical ocean model (Balmaseda et al., 2013). An ensemble of ocean reanalyses, such as GREP, is more powerful than a single standalone reanalysis; the ensemble product accounts for a range of uncertainties represented by the diverse inputs and methods used in each member. Storto et al. (2019) found the ensemble mean was a significant improvement on previous single-member versions of reanalyses, across a range of marine variables.

The correlation between individual GREP products and the ensemble mean is a way of estimating the robustness of the interannual variability in the GREP (Figure A1). Over most parts of the ocean tropics, there is a strong degree of consistency among the members of the GREP, with correlation values exceeding .9. The correlation is lowest over western boundary currents, Atlantic upwelling areas, and the Southern Ocean.

[2] https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-sea-surface-temperature-ensemble-product?tab=overview

[3] https://resources.marine.copernicus.eu/product-detail/GLOBAL_REANALYSIS_PHY_001_031/INFORMATION

Given the relatively large disagreement in the reanalyses in these regions, the forecast validation may be less reliable.

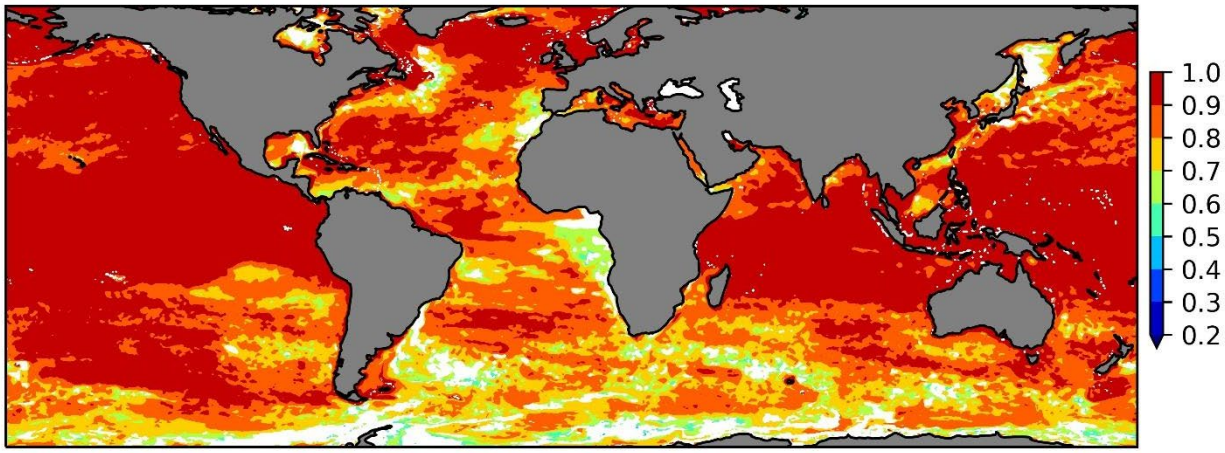


Figure A 1. GREP Temporal Correlation. Average correlation between the Ocean Heat Content 0-300 m of the GREP ensemble members and the GREP ensemble mean. Statistics are for the 1993-2016 period. From McAdam et al., Clim Dyn. 2022.

In this deliverable, the term "GREP" refers to the reanalyses' ensemble mean. As our forecast systems are initialised with either ORAS5 or C-GLORS, the OHC0-300 m validation dataset is not truly independent. However, as for the ESA CCI SST, the spatial and temporal coverage is unparalleled and necessary for a comparison of long-term data.

Sea Level Anomaly ESA-CCI

The sea level data set used here is based on the sea level Ocean Monitoring Indicators, produced by CMEMS[4], for which the C3S products[5] are used as input data. These C3S products are derived from the DUACS [6]delayed-time altimeter gridded maps of sea level anomalies based on a stable number of altimeters (two) in the satellite constellation. The altimeter satellite multi-mission gridded sea surface heights and derived variables are computed with respect to a twenty-year mean reference period (1993-2012). Up-to-date altimeter standards are used to estimate sea level anomalies. Contrary to near-real-time sea-level products, the stability and accuracy of the delayed-time products make them adapted to climate applications and ocean monitoring indicators. Details on the altimeter and processing algorithms are available in Pujol et al. (2016) and Taburet et al. (2019). The record covers the period 1993-2019.

Ocean output from seasonal forecasts

The two forecast systems used here are the Seasonal Prediction System Version 3 from the Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC-SPS3), and the fifth generation Seasonal Forecasting System from the European Centre for Medium-Range Weather Forecasts (ECMWF-SEAS5). Since 2018 both systems have been contributing to the Copernicus Climate Change Service (C3S), which makes seasonal forecasts of

[4] https://resources.marine.copernicus.eu/product-detail/SEALEVEL_GLO_PHY_CLIMATE_L4_MY_008_057/INFORMATION

[5] https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-sea-level-global?tab=overview

[6] https://duacs.cls.fr/

atmosphere and surface variables (precipitation, 2m-temperature) freely available online. These systems produce a forecast of ocean variables other than SST, but these are not yet publicly available, since they have not yet been verified. This is the gap that the current EuroSea activity tries to fill.

The model components of each system are detailed in Table A1. Both systems base their ocean model component on the eddy-permitting version 3.4 of NEMO, which has a horizontal resolution of 25 km at the equator. The ocean model grid is tripolar, introducing grid cell anisotropy north of 20°N towards the artificial poles over Canada and Siberia. The vertical resolution in ECMWF-SEAS5 is higher than in CMCC-SPS3; in the upper 300 m, there are 35 and 29 vertical levels in ECMWF-SEAS5 and CMCC-SPS3, respectively.

Table A1. Component, resolution, and initialisation details for CMCC-SPS3 and ECMWF-SEAS5 coupled forecast system

	ECMWF-SEAS5	CMCC-SPS3
Ensemble	51	40
Coupler	Single-Executable	CPL7
Atmosphere		
Model	IFS	CAM
Horizontal Resolution	36km	1°
Vertical Resolution (top)	91 levels (0.01 hPa)	45 levels (0.3 hPa)
Initialisation	ERA-Interim	ERA-Interim
Ocean		
Model	NEMO v.3.4	NEMO v3.4
Horizontal Resolution	0.25° tripolar grid	0.25° tripolar grid
Vertical Resolution	75 levels	50 levels
Initialisation	ORAS5	C-GLORS
Sea Ice	LIM2	CICE4
Waves	0.5°	N/A
Land	Embedded within IFS	CLM 4.5 1°
Rivers	N/A	River Transport Model (RTM)

Both systems use versions of their respective ocean reanalysis to create initial conditions. In CMCC-SPS3, the initial conditions are based on C-GLORS (Storto and Masina, 2016), while in ECMWF-SEAS5 they are based on ECMWF ORAS5 (Zuo et al., 2019). Both reanalyses use identical horizontal resolutions (0.25°), while the number of vertical levels is 75 and 50 for ECMWF-SEAS5 and CMCC-SPS3 respectively. Both use atmospheric forcing from ERA-Interim until 2016, and ECMWF's NWP analysis thereafter. Both systems used a variant of the CORE bulk formulation, although ORAS5 also includes wave forcing. Both systems assimilate temperature

and salinity profiles, and altimeter-derived sea-level anomalies, but the assimilation methods and observational datasets also differ. C-GLORS uses the 3D-variational data assimilation scheme OceanVar (Dobricic and Pinardi, 2008; Storto and Masina, 2016), while ORAS5 uses NEMOVAR. Thus, within the ocean component alone there are several factors which may contribute to differences in forecast output between the two systems.

The atmospheric model components have in common only the initial conditions (Table A1). The configuration of IFS in ECMWF-SEAS5 provides higher vertical and horizontal resolution than CAM in SPS3. CMCC-SPS3 uses the CPL7 coupler from the Community Earth System Model (CESM, Craig et al., 2012), while ECMWF-SEAS5 uses a single-executable (Mogensen et al., 2012). The coupling occurs every 90 minutes in CMCC-SPS3, and every 60 minutes for ECMWF-SEAS5, with both capturing diurnal cycles. In both, ocean and sea-ice models are tightly coupled (i.e., they share a horizontal grid). Meanwhile, the atmosphere and wave models provide fluxes of heat, momentum, freshwater and turbulent kinetic energy to the ocean and sea ice components, while the ocean and sea ice models provide SST, surface currents and sea-ice concentration in return.

Both ECMWF-SEAS5 and CMCC-SPS3 ensembles sample the uncertainty in the initial conditions of the land, ocean, and atmosphere (Table A1). The size of the ensemble (50 for ECMWF-SEAS5 and 40 for CMCC-SPS3) ensures a high signal-to-noise ratio in the ensemble mean. In CMCC-SPS3, the ensemble is made by combining various perturbations in each initial condition set: 10 perturbations of the atmospheric component, 4 of the ocean components, and 3 of the land-surface component. Then, 40 scenarios are picked from the possible 120. ECMWF-SEAS5 samples initial uncertainty in the ocean via the 5-member ensemble of ORAS5, and in the atmosphere via initial perturbations from the Ensemble of Data Assimilation (EDA) and singular vectors. It also applies stochastic physics perturbations to represent uncertainty arising from missing sub-scale processes. Further details of the ensemble generation are given in Johnson et al. (2019) and Sanna et al. (2017) for ECMWF-SEAS5 and CMCC-SPS3, respectively.

Appendix II. Skill maps for all initial dates without trend correction

Forecasts Initialized in February

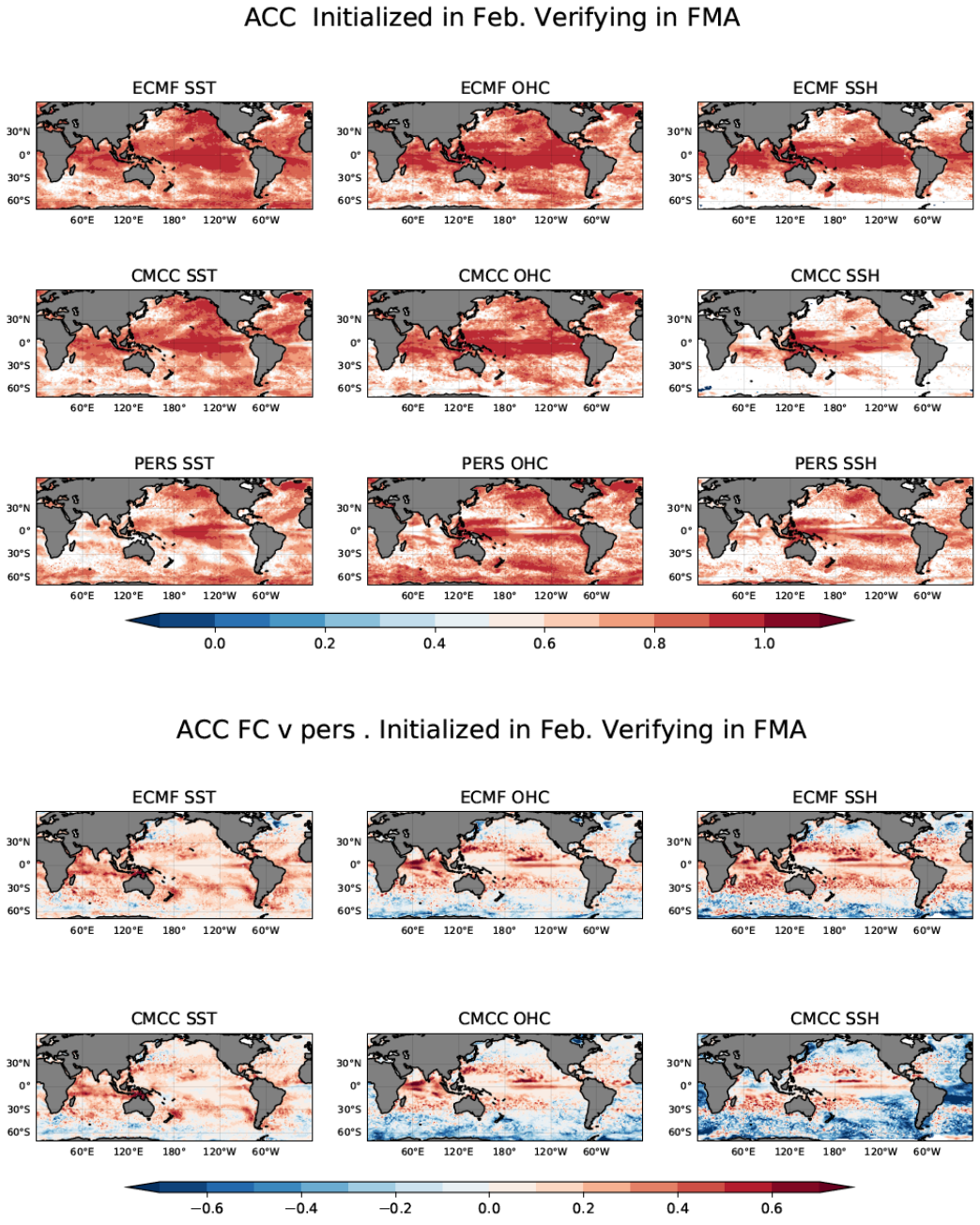
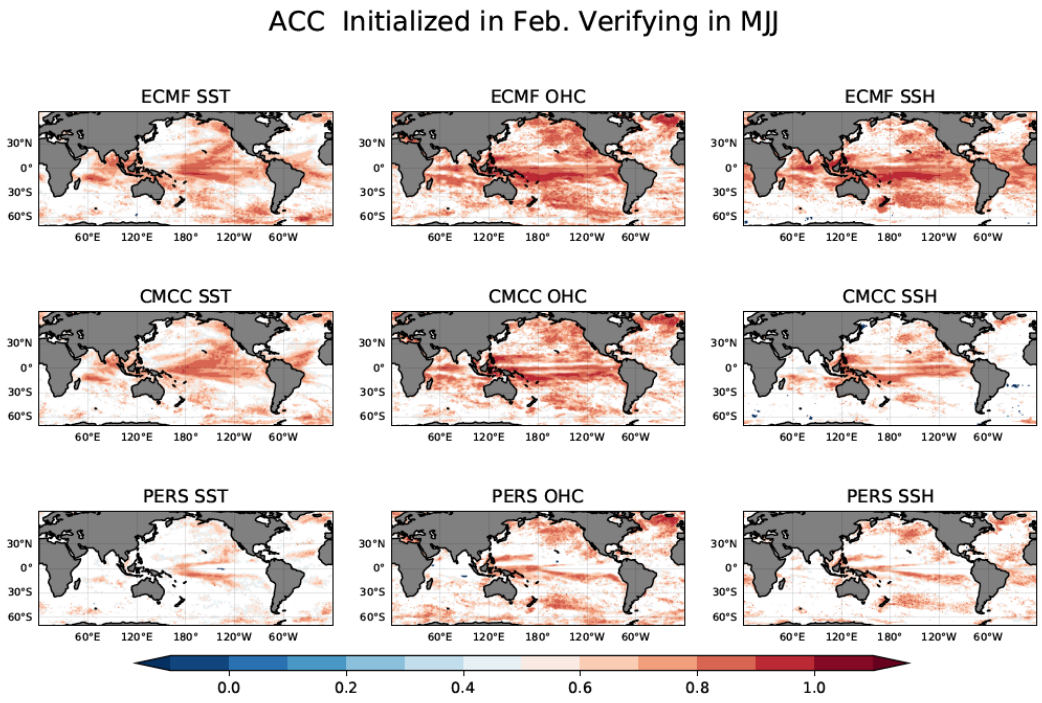


Figure A 2. Upper block) Anomaly Correlation Coefficient skill for the ECMWF (top), CMCC (middle) seasonal forecasts initialized in February and verifying in the first season (FMA), as well as the corresponding persistence forecast. Bottom block) Differences in ACC between the dynamical seasonal forecasts and persistence. Shown is the skill for SST (left), OHC (centre) and SSH (right).



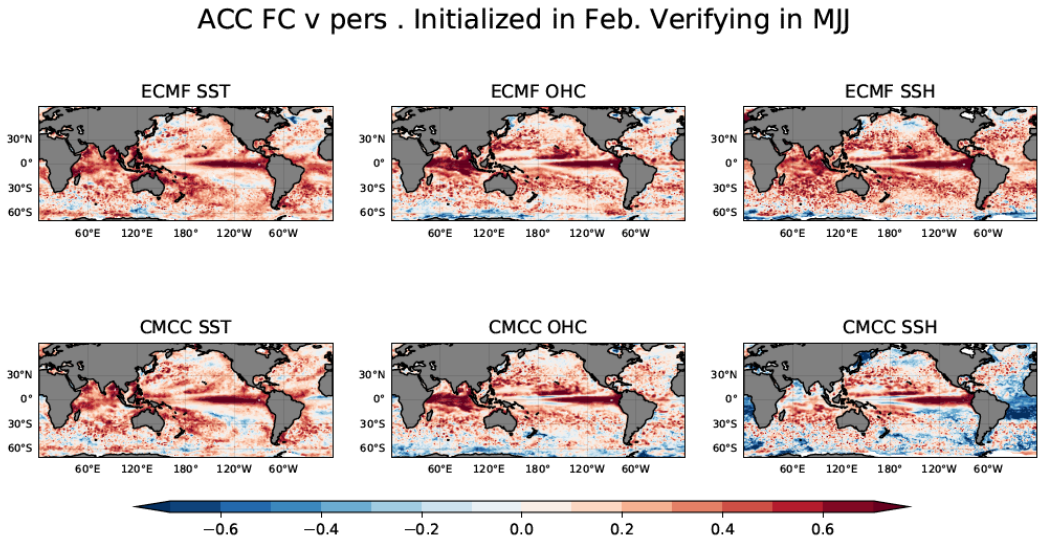
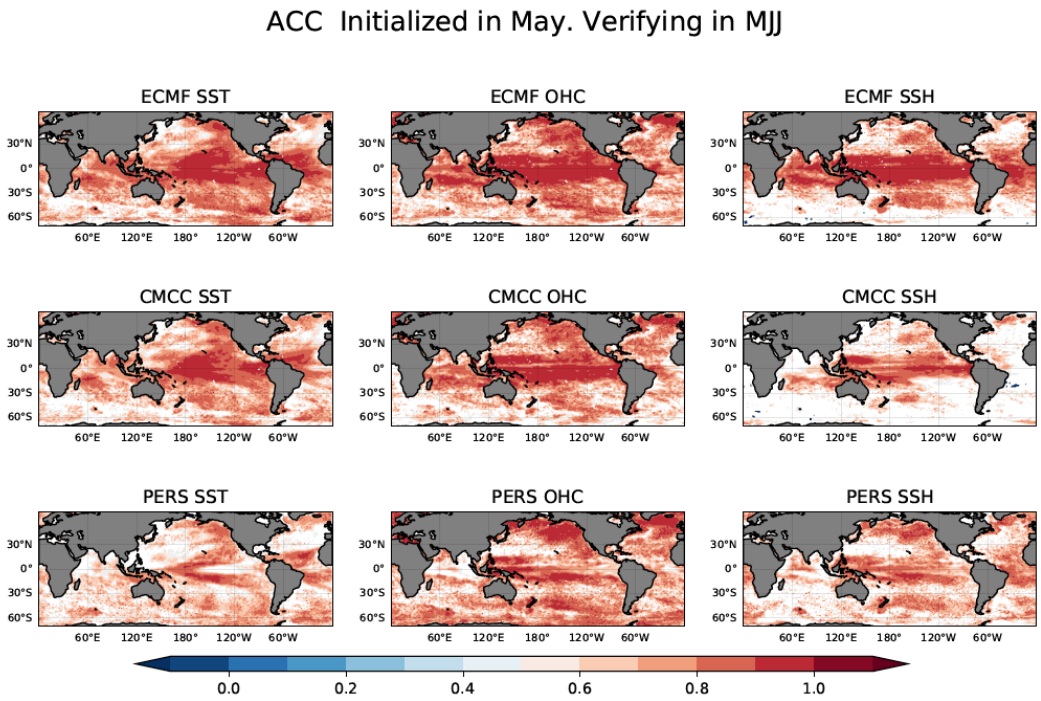


Figure A 3. As for Figure A 2 but for lead season 2.

Forecasts Initialized in May



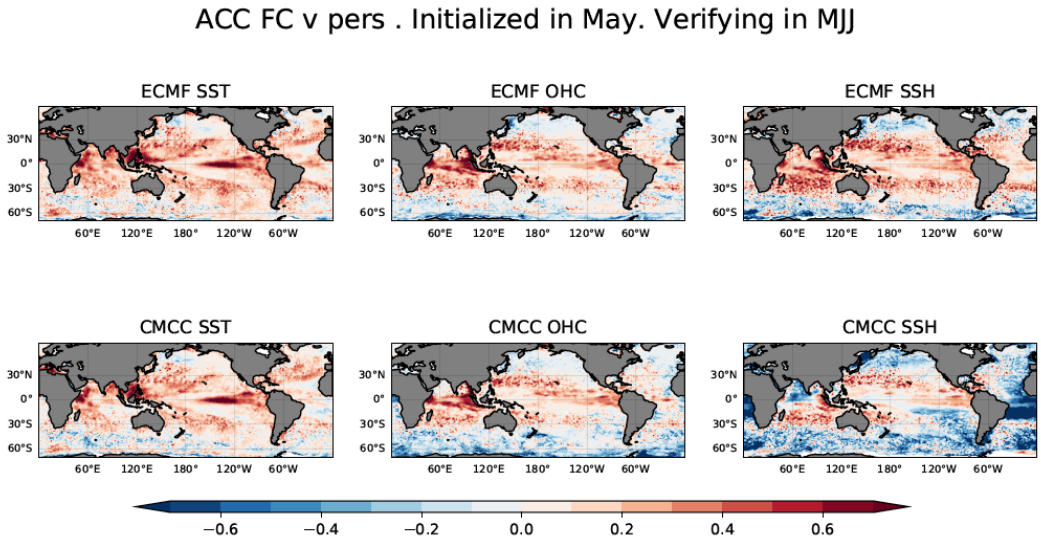
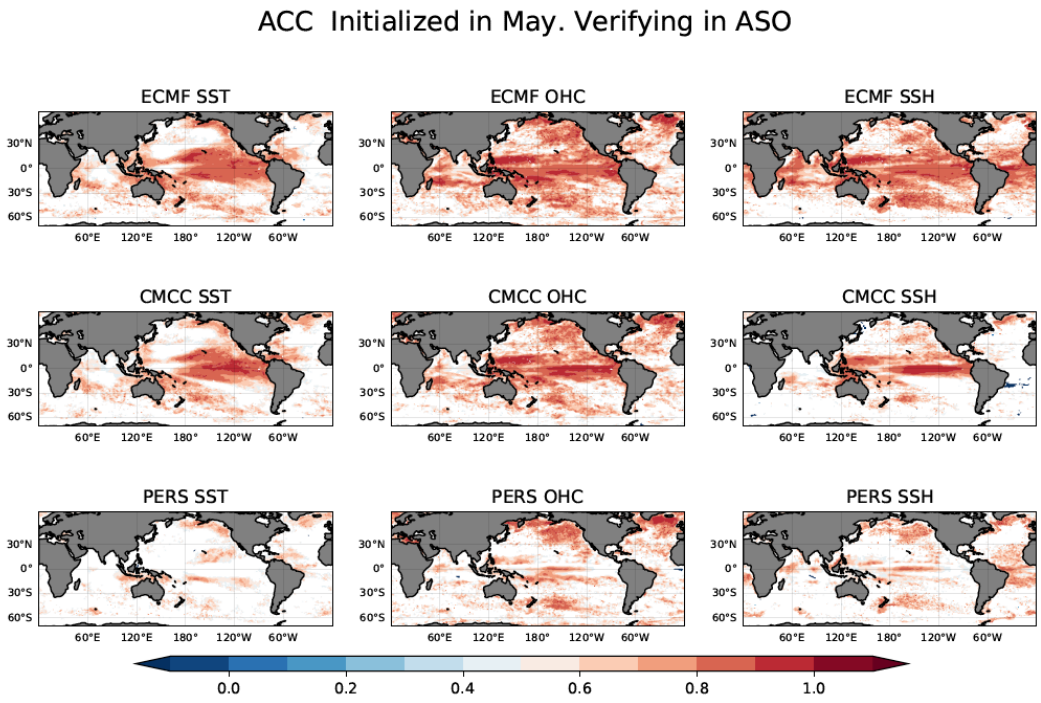


Figure A 3. As Figure A 2, for forecasts initialized in May at lead time 1 season



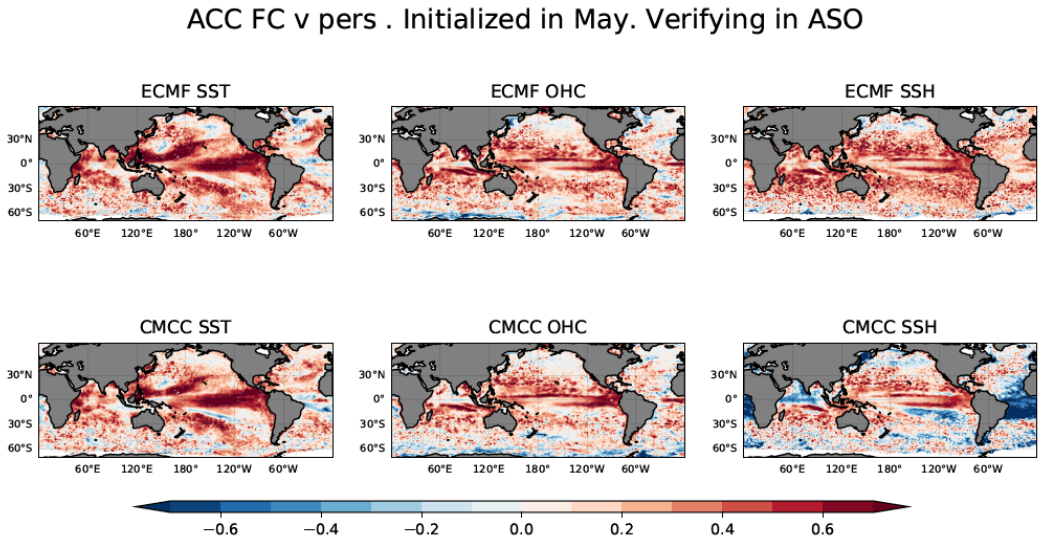


Figure A 4. As Figure A 2, for forecasts initialized in May at lead time 2 seasons.

Forecasts Initialized in August

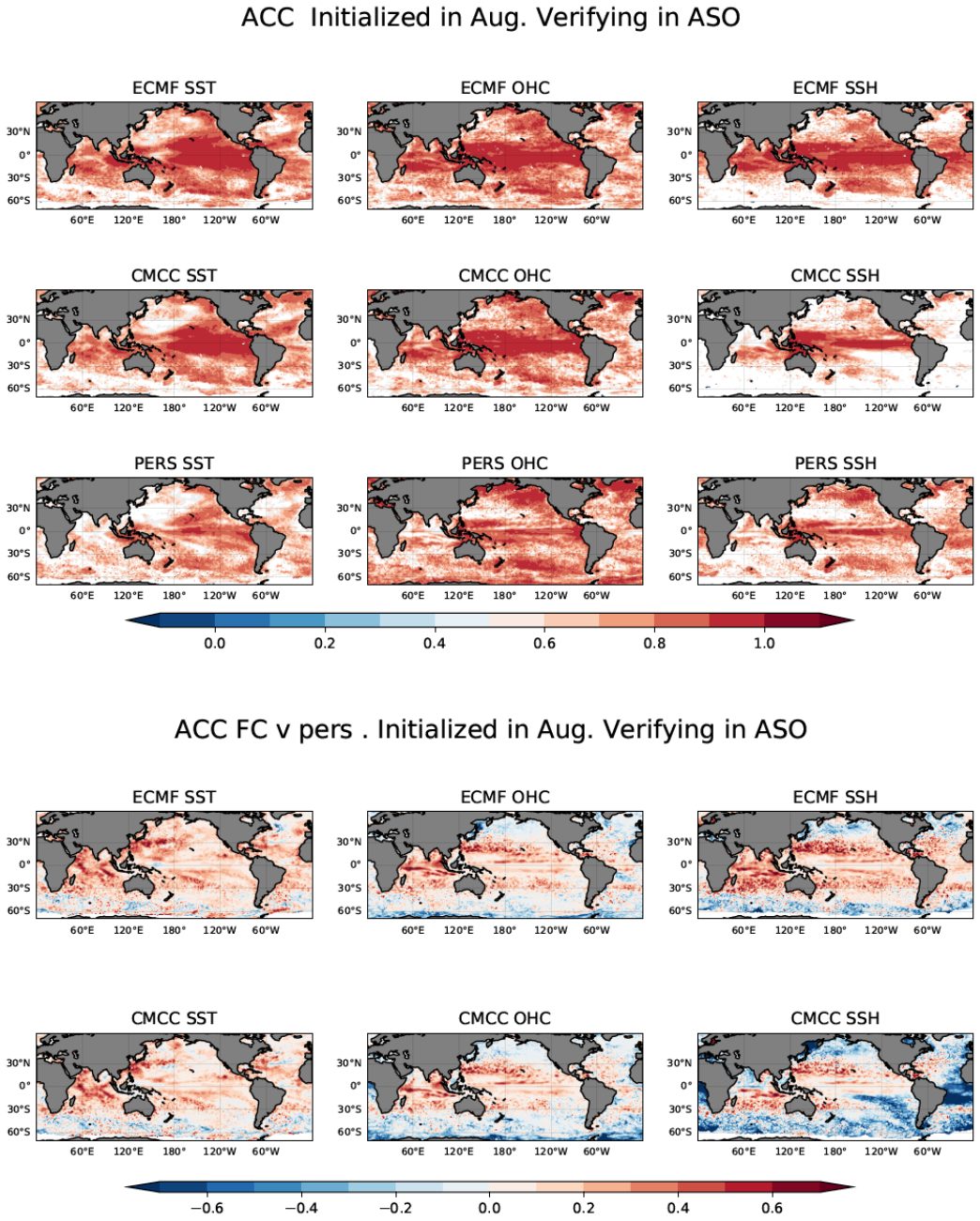
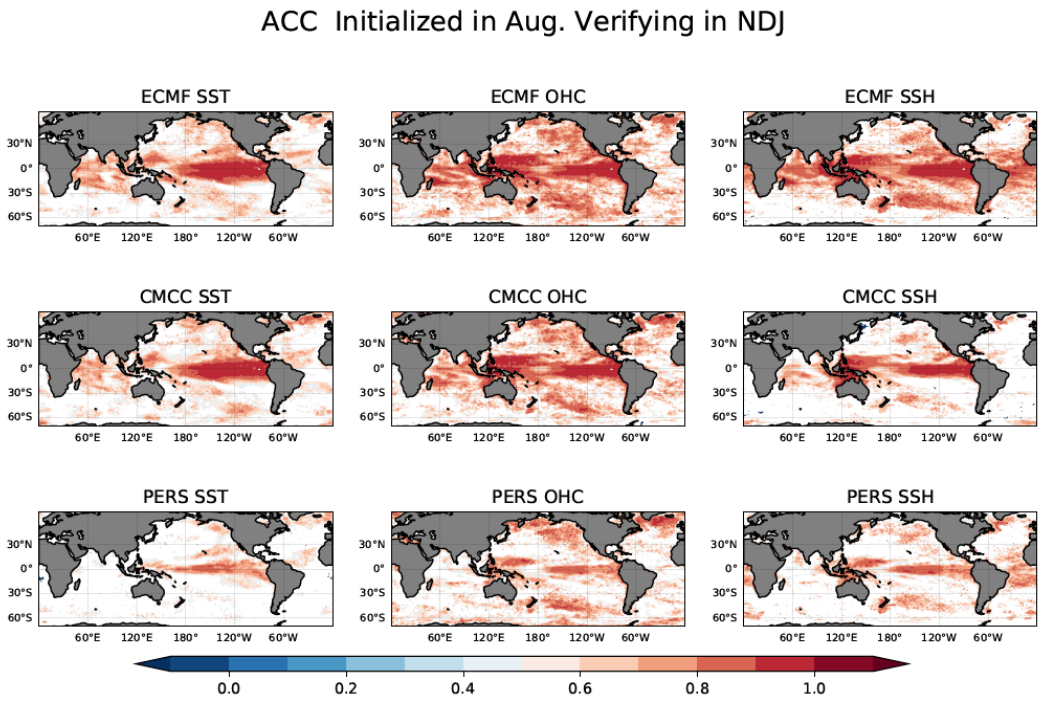


Figure A 5. As Figure A 2, for forecasts initialized in August and verifying at lead time 1 season



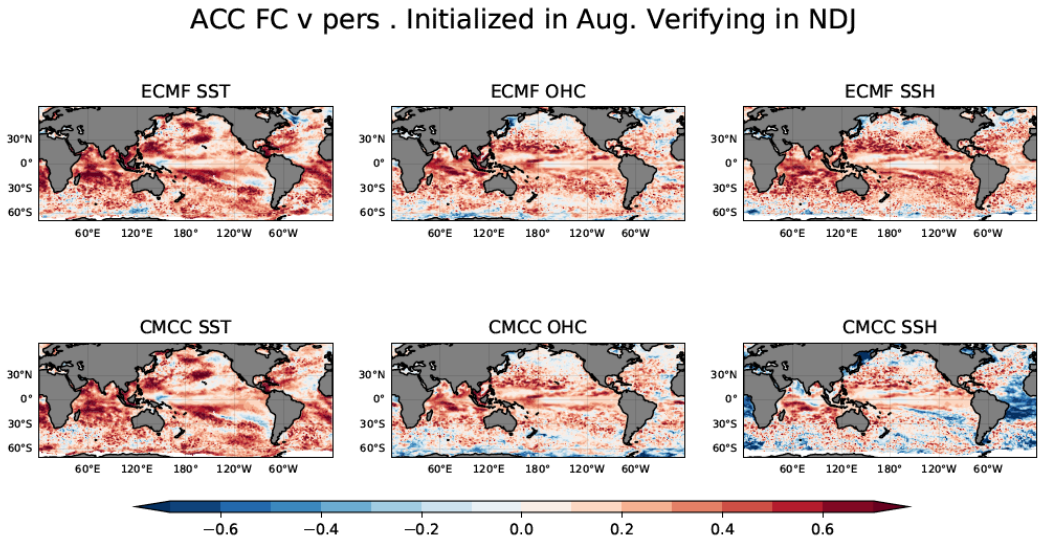


Figure A 6. As Figure A 2, for forecasts initialized in August and verifying at lead time 2 seasons.

Forecasts Initialized in November

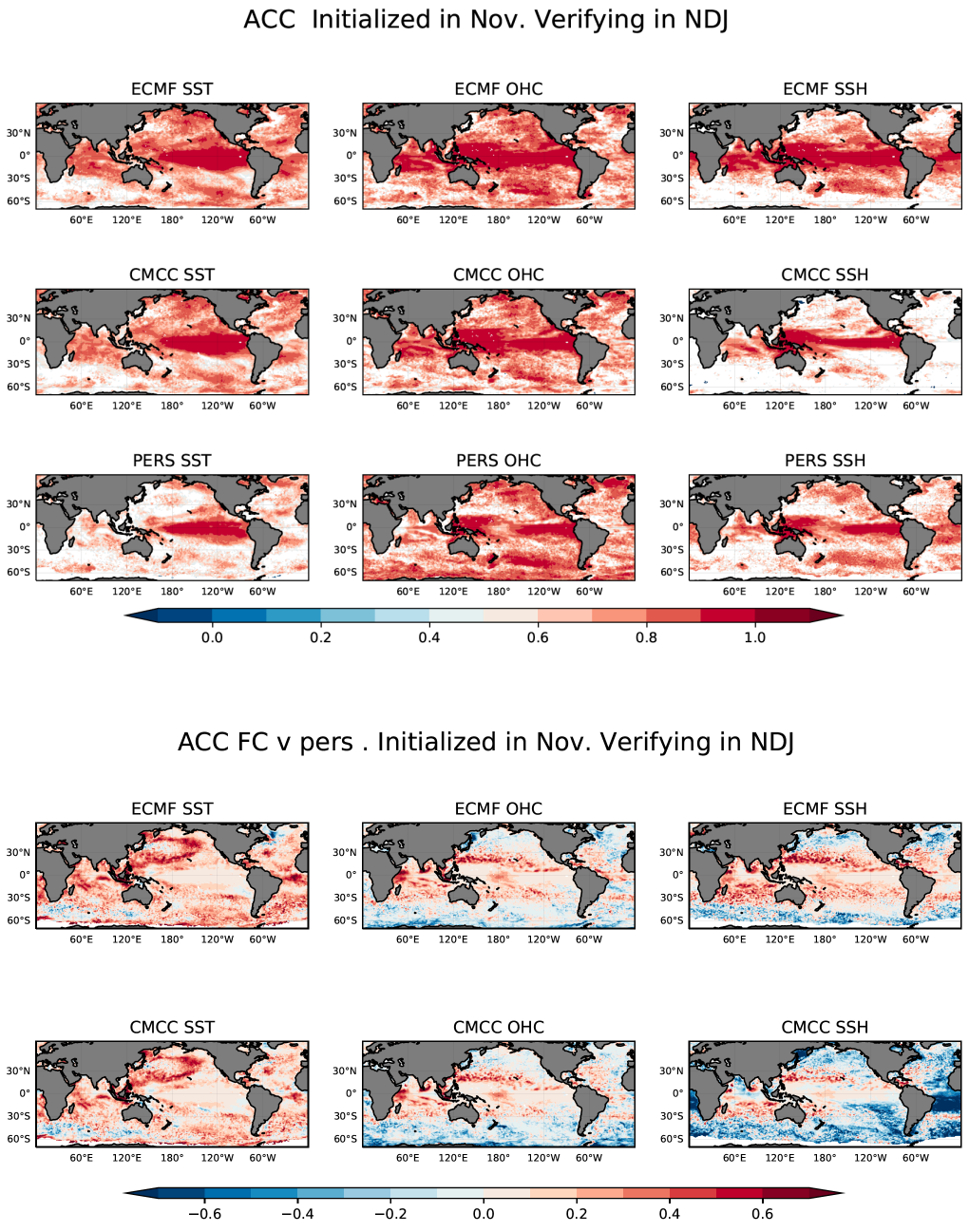


Figure A 8. As Figure A 2, for forecasts initialized in November and verifying at lead time 1 season

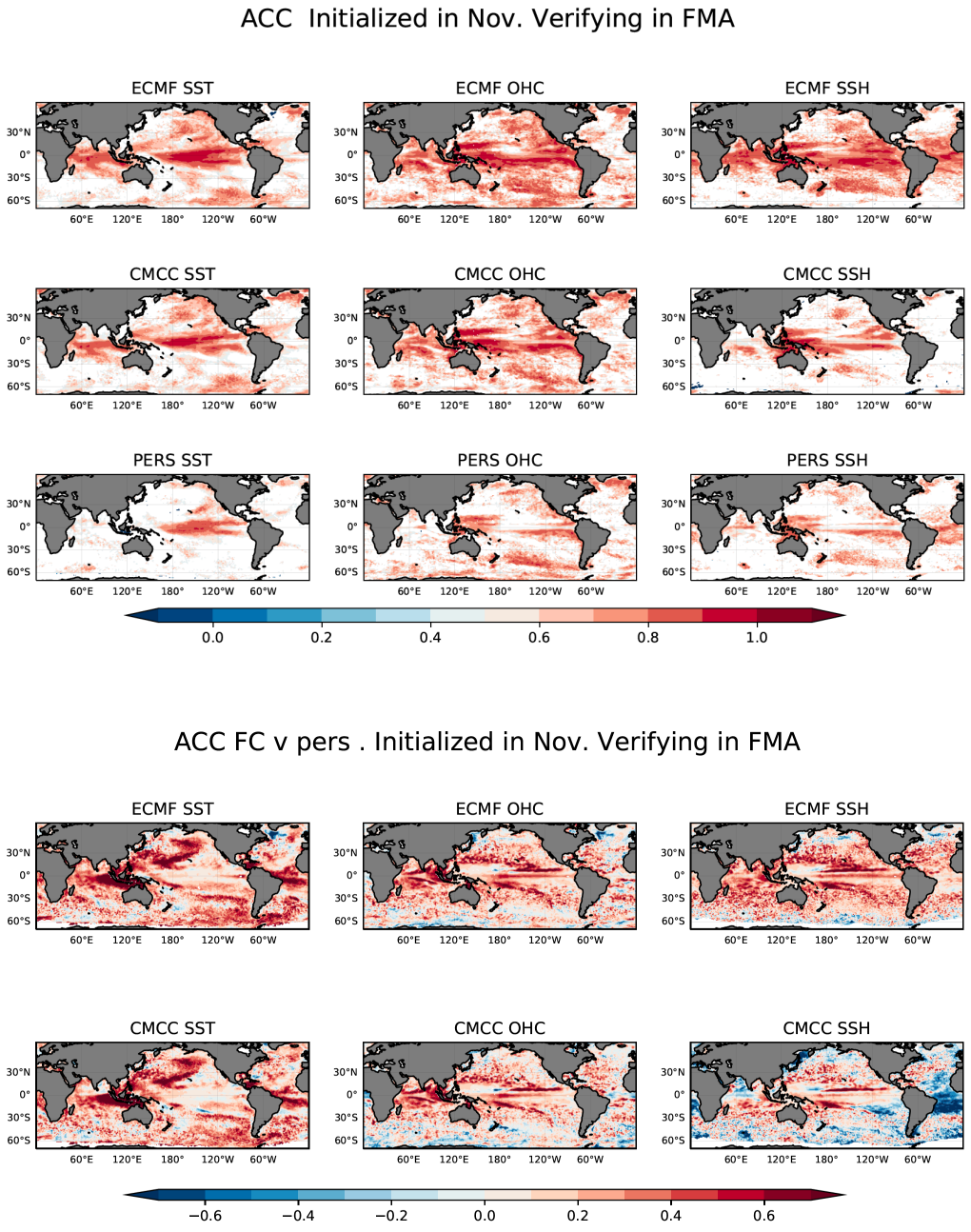


Figure A 9. As Figure A 2, for forecasts initialized in November and verifying at lead time 2 seasons.

Appendix III. Skill maps for all initial dates with trend correction

Forecasts initialized in February

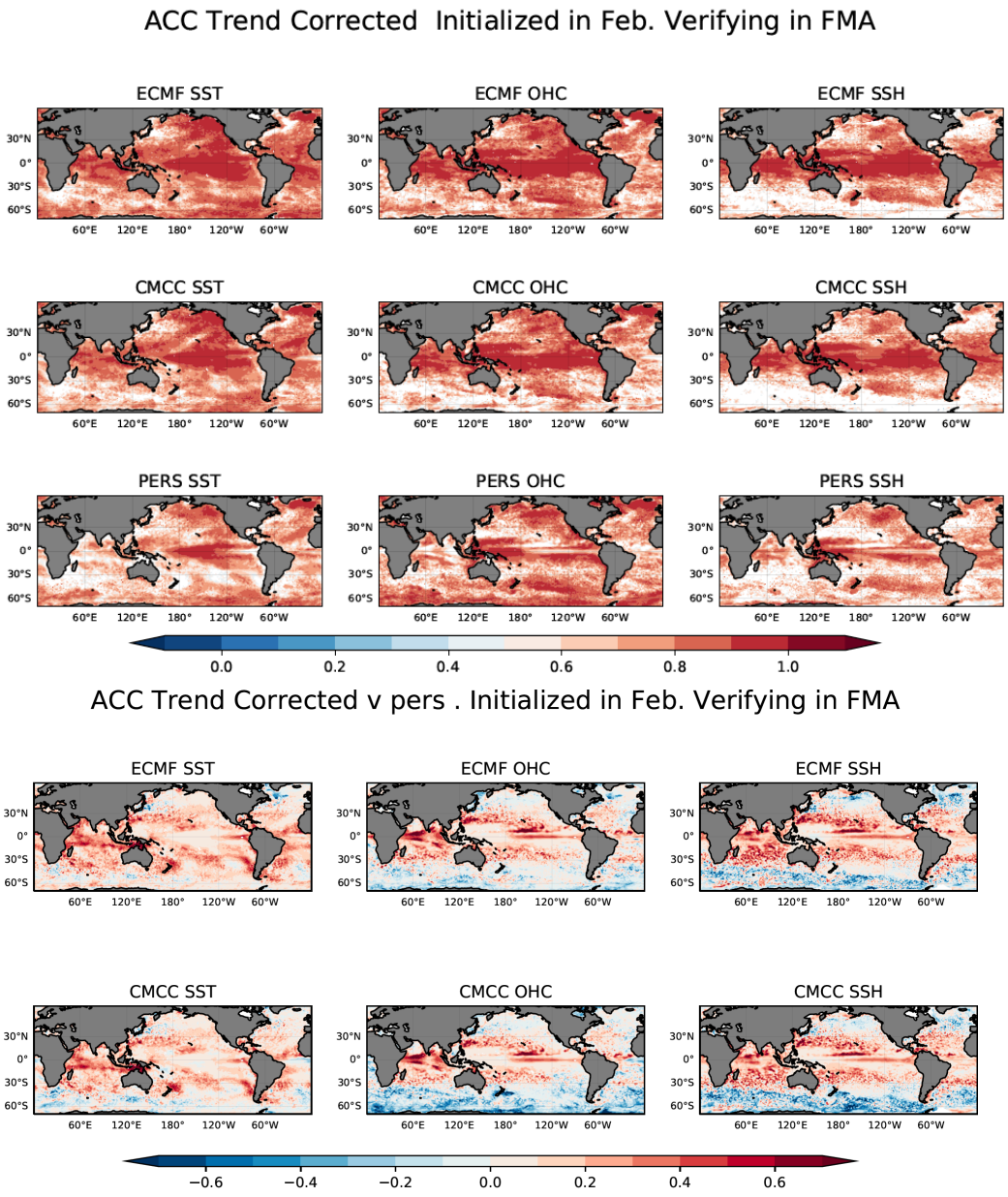


Figure A 7. As Figure A 2 but for trend corrected forecasts initialized in February and verifying at lead time 1 season.

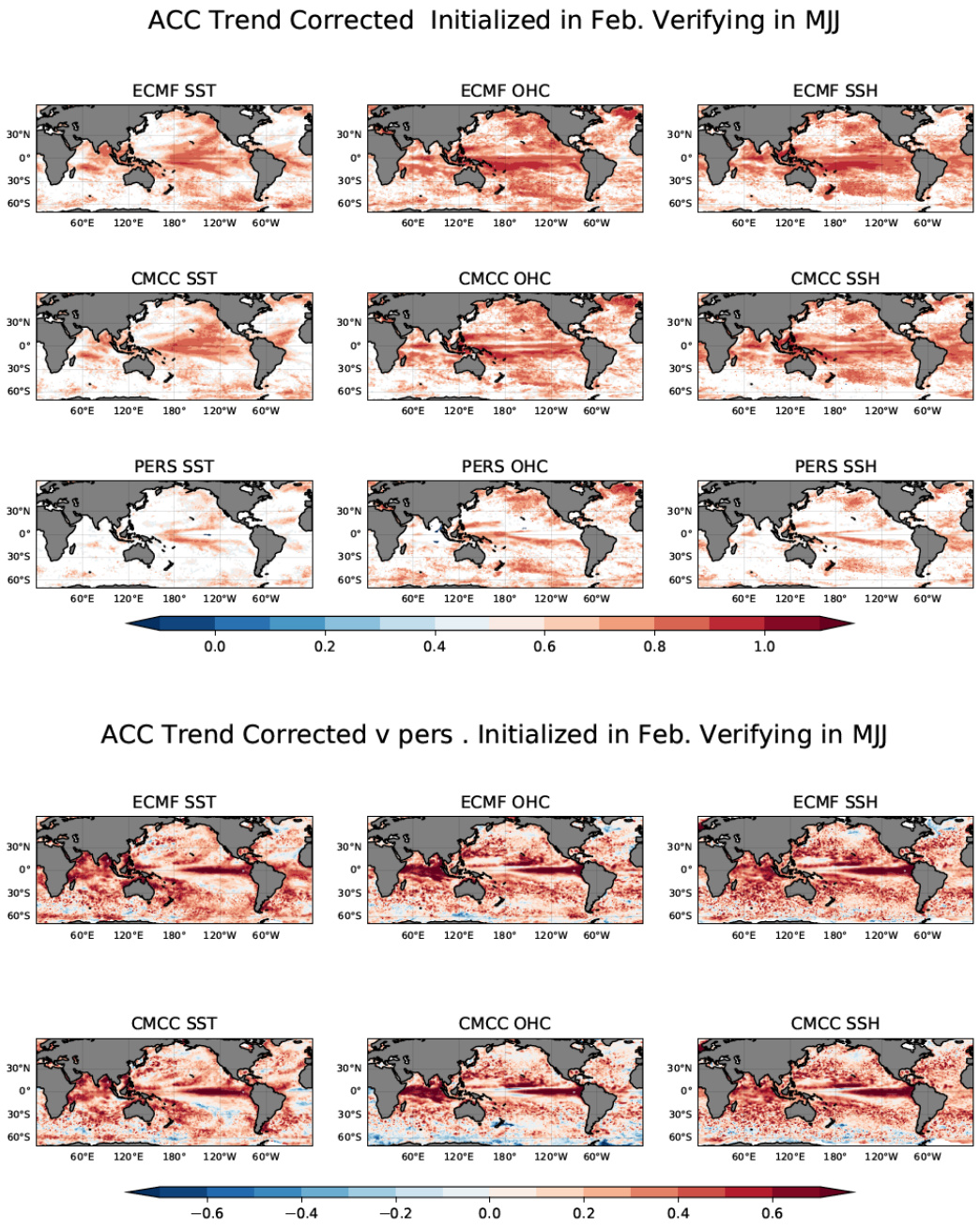


Figure A 8. As Figure A 2 but for trend corrected forecasts initialized in February and verifying at lead time 2 seasons.

Forecasts Initialized in May

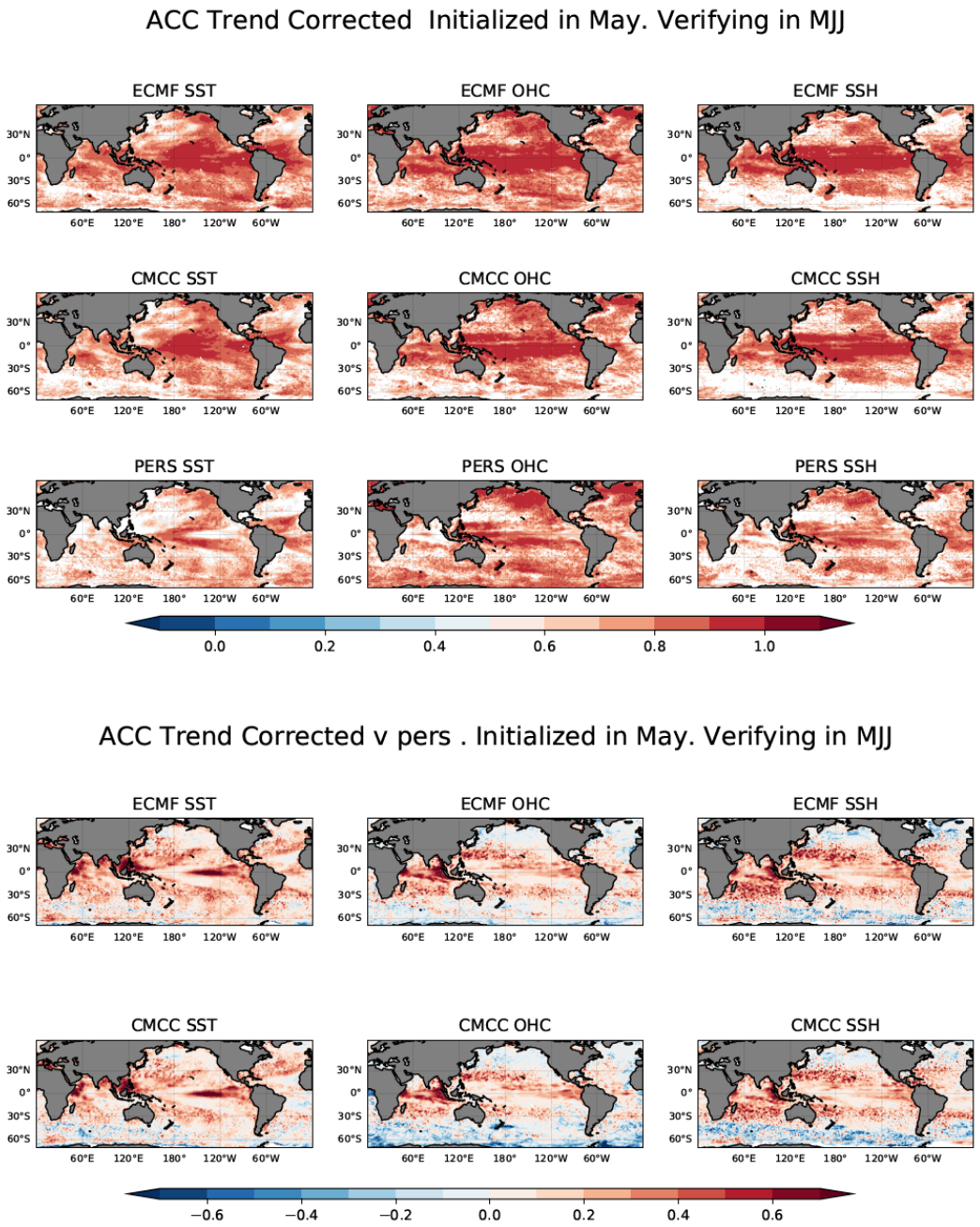
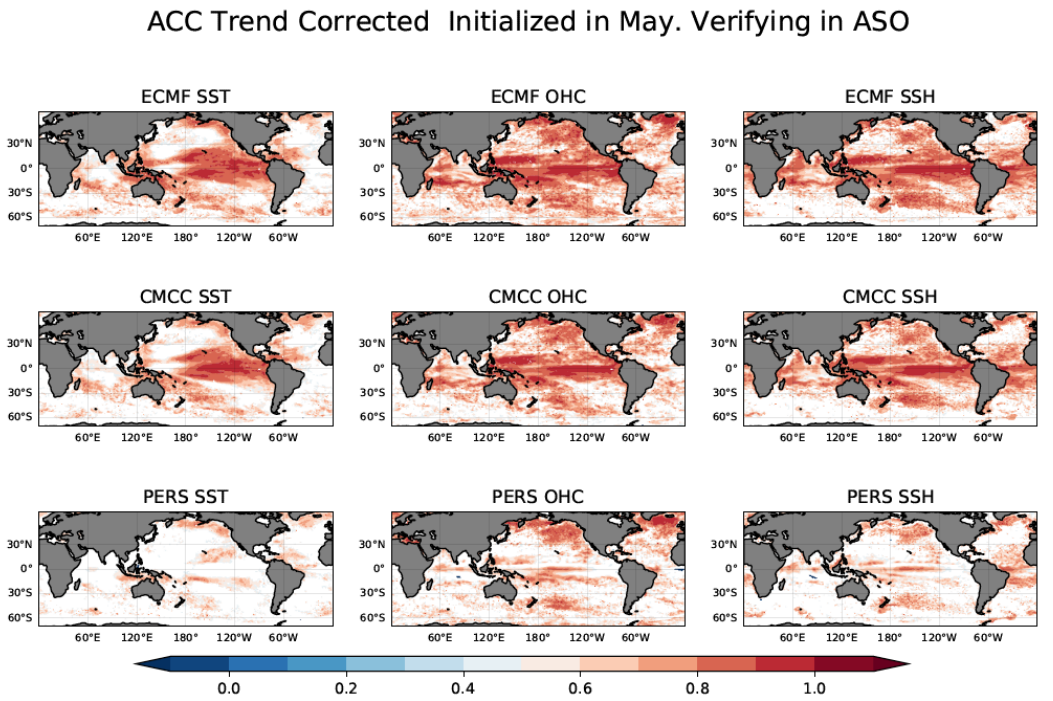


Figure A 9. As Figure A 2 but for trend corrected forecasts initialized in May and verifying at lead time 1 season.



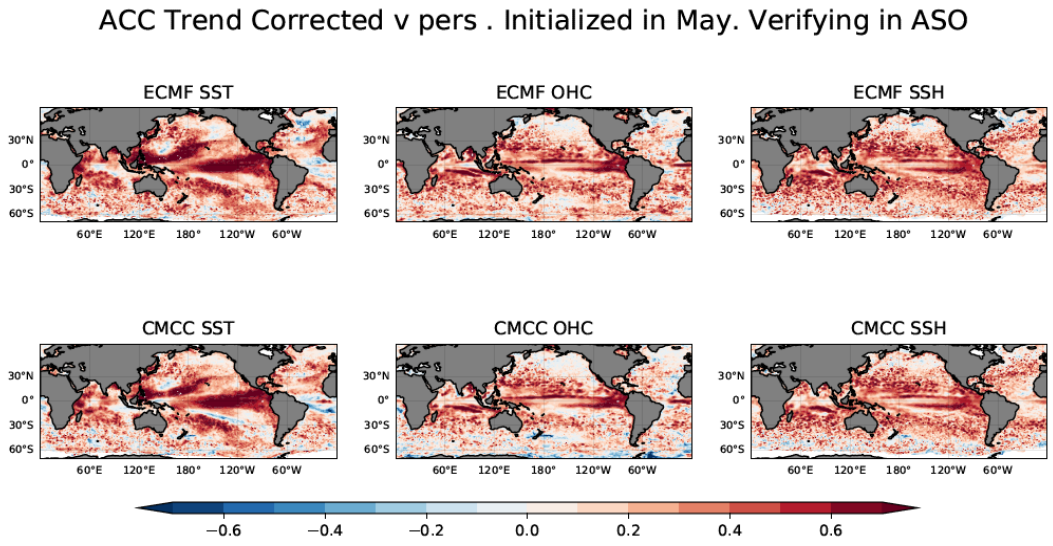


Figure A 10. As Figure A 2 but for trend corrected forecasts initialized in May and verifying at lead time 2 seasons.

Forecasts Initialized in August

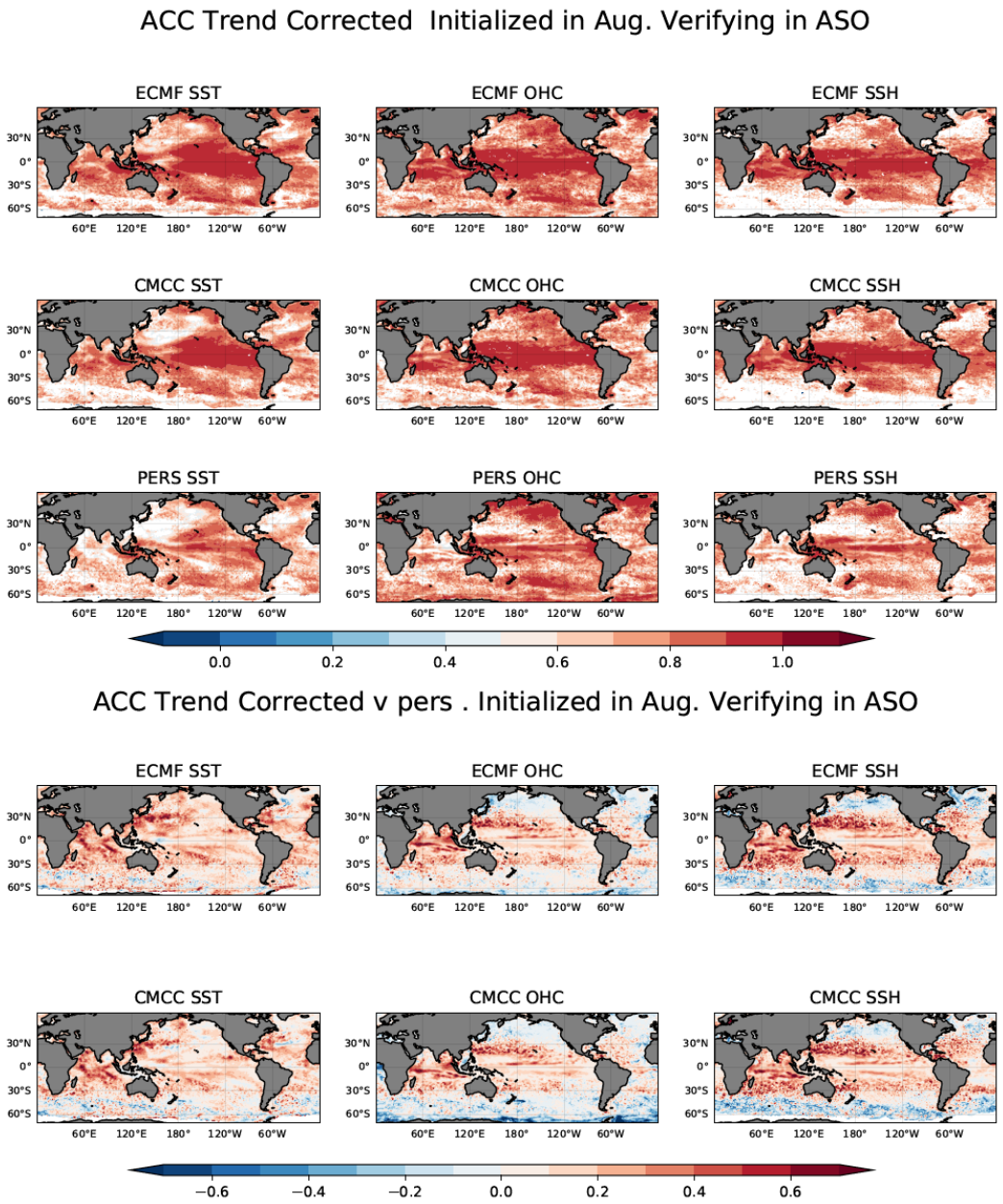
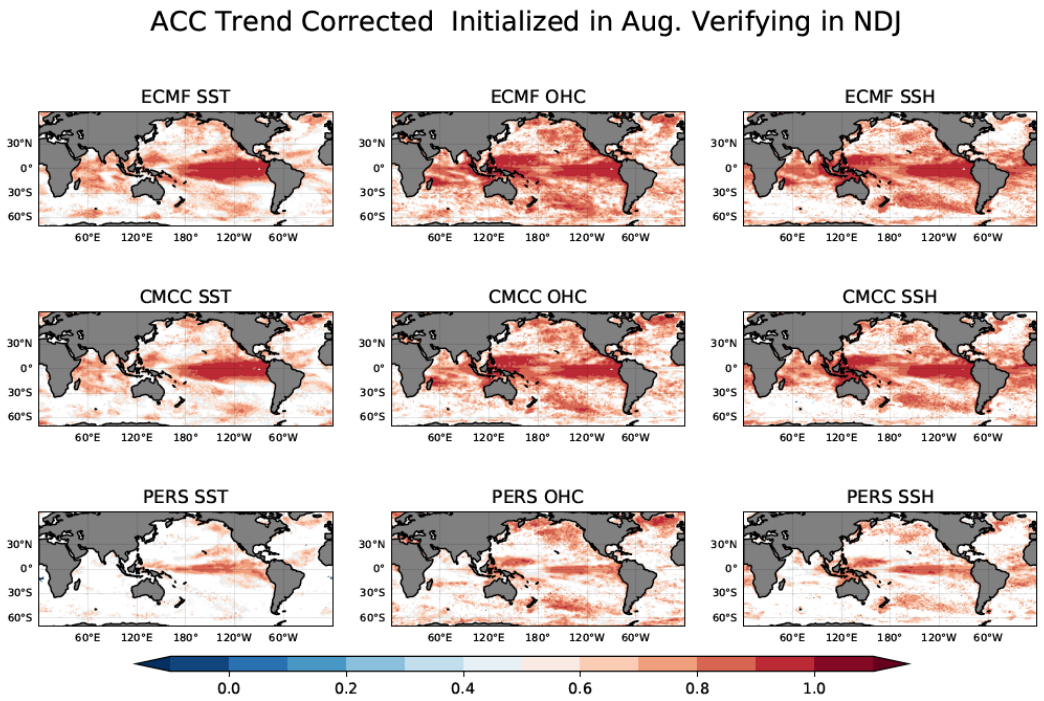


Figure A 11. As Figure A 2 but for trend corrected forecasts initialized in August and verifying at lead time 1 season.



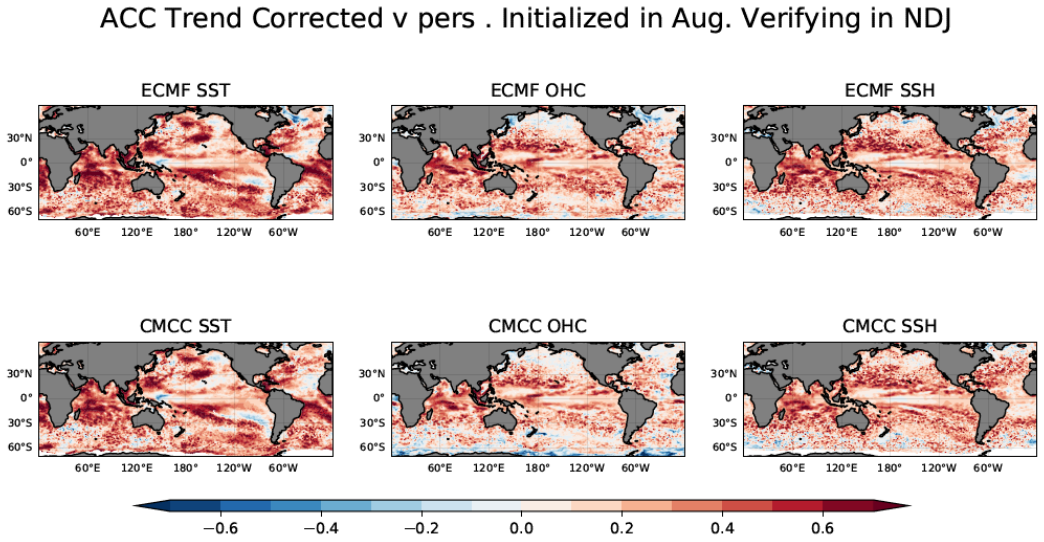


Figure A 12. As Figure A 2 but for trend corrected forecasts initialized in August and verifying at lead time 2 seasons.

Forecasts Initialized in November

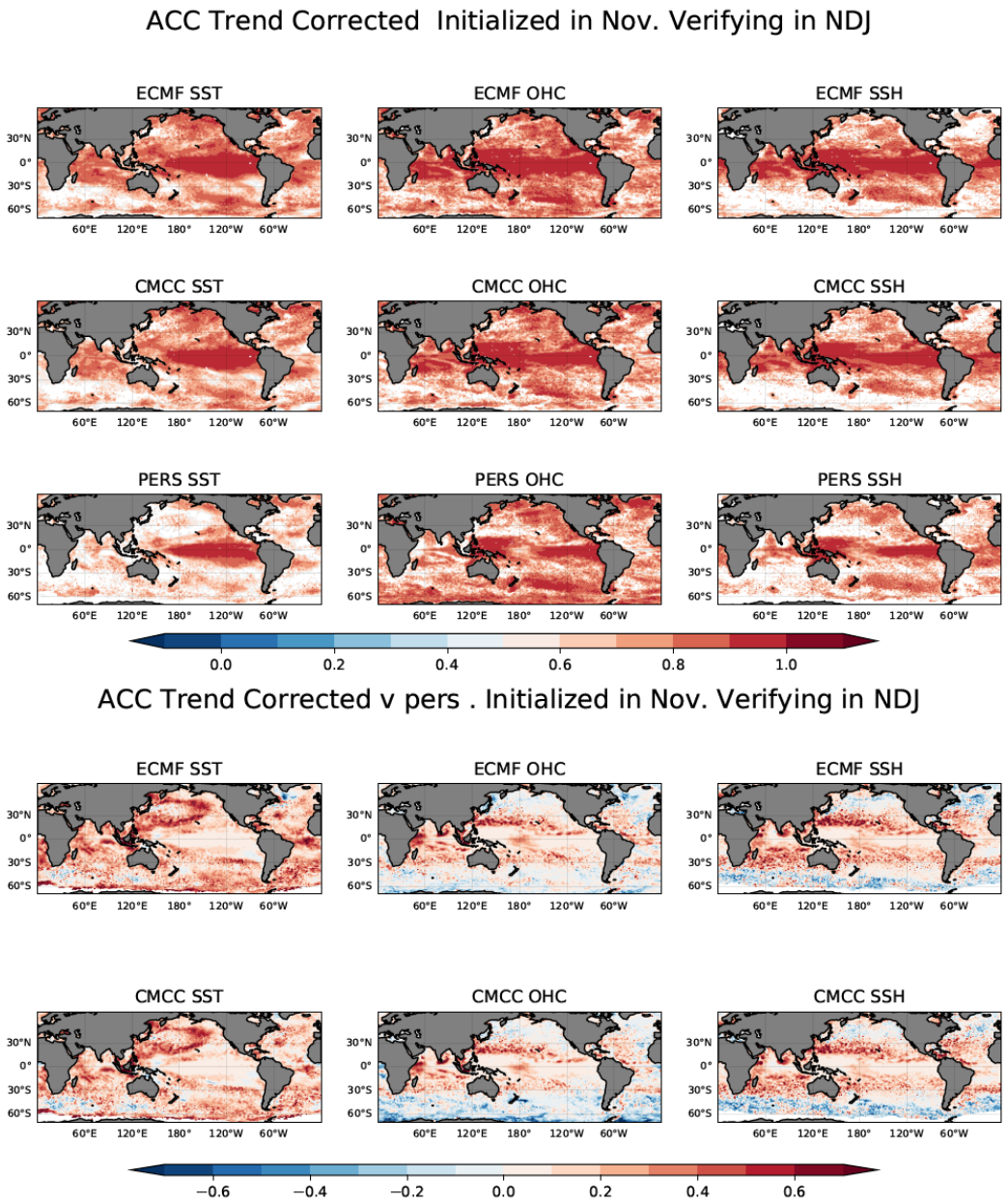


Figure A 13. As Figure A 2 but for trend corrected forecasts initialized in November and verifying at lead time 1 season.

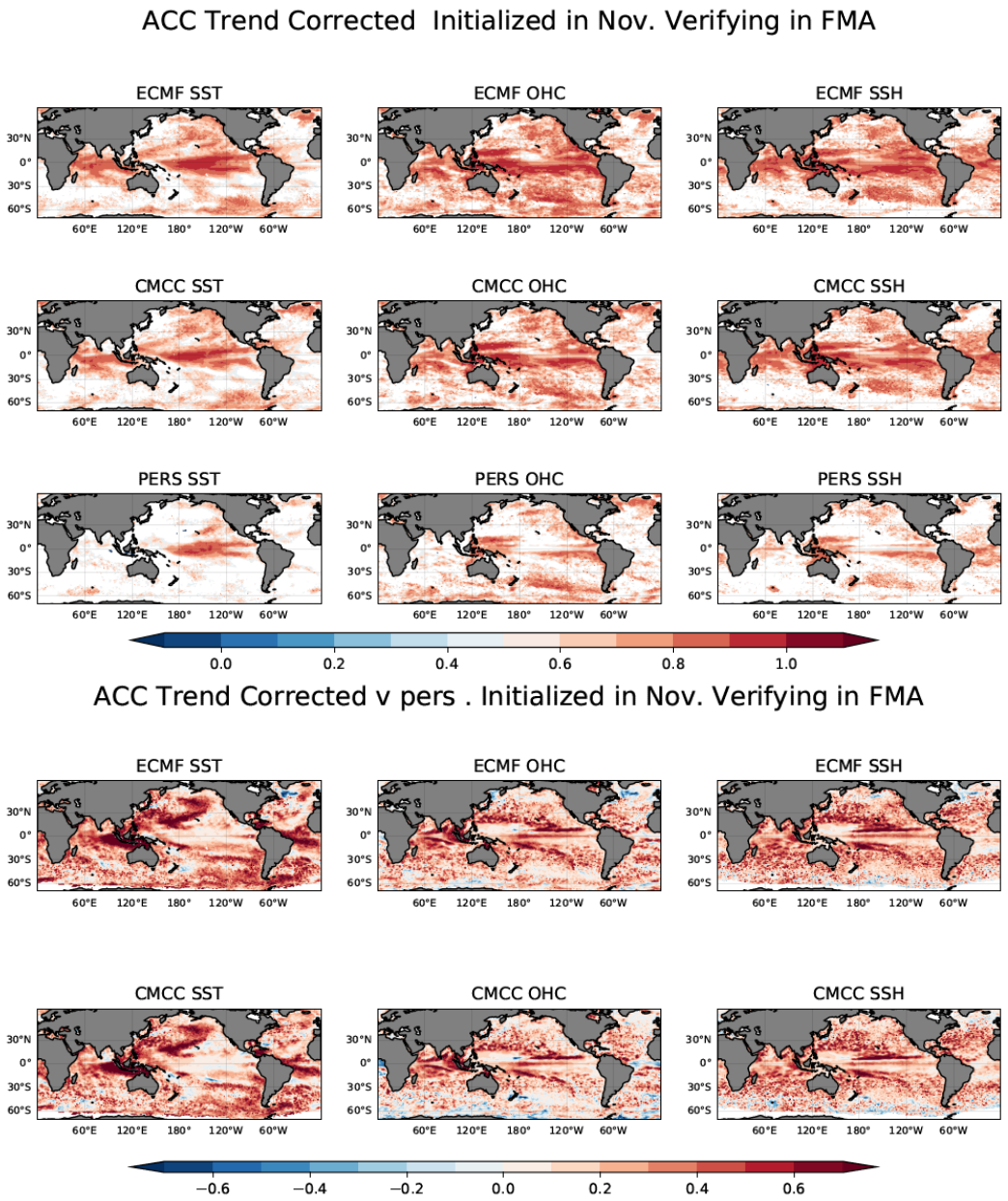


Figure A 14. As Figure A 1 but for trend corrected forecasts initialized in February and verifying at lead time 2 seasons.

Appendix IV. Trend Comparison for different fields and seasons

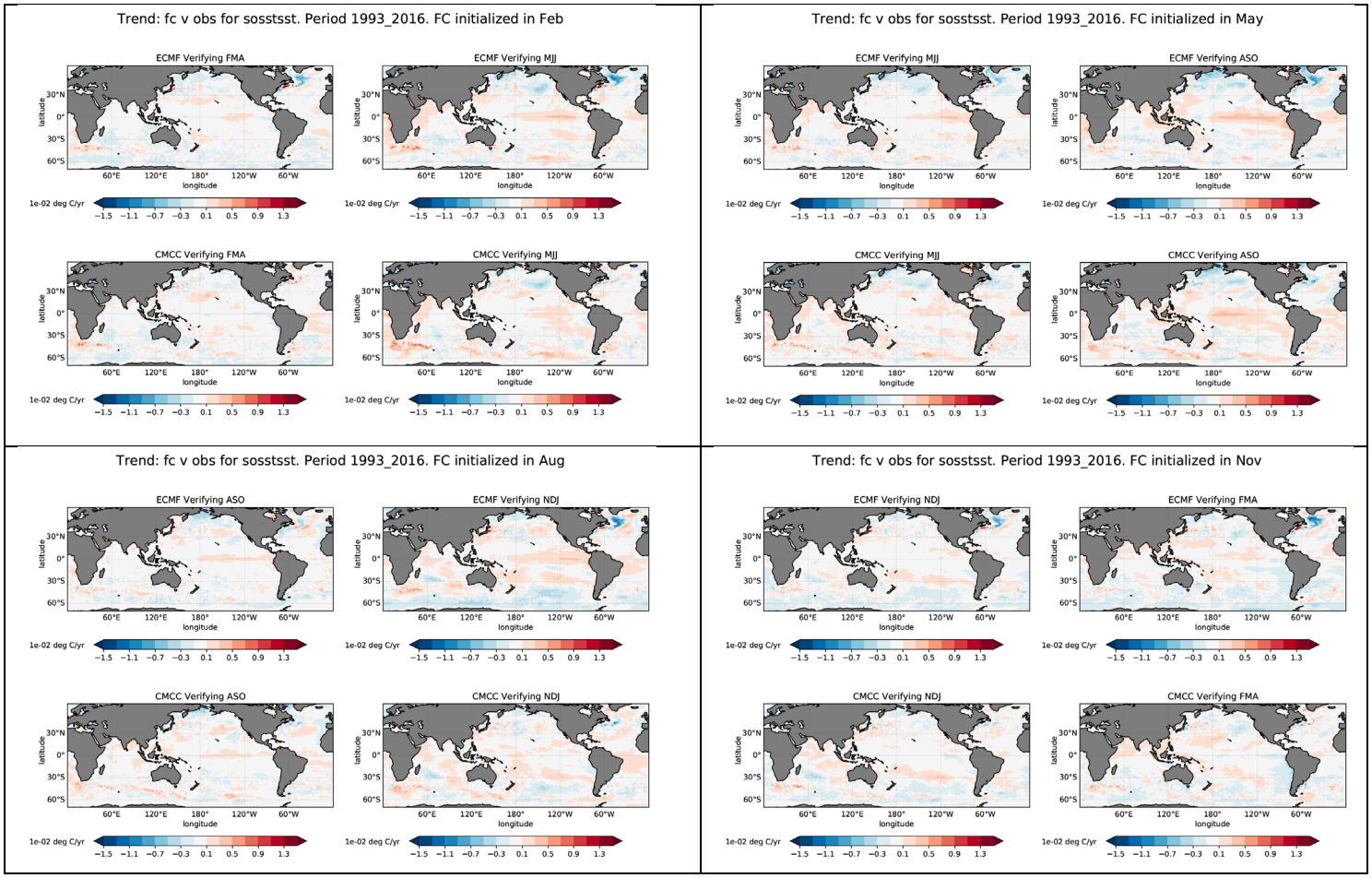


Figure A 15. Differences in linear trends for the period 1993-2016 in SST between seasonal forecasts and observations. Different blocks show results from different initialization times: Feb and May in top blocks (left/right), and Aug/Nov in the bottom blocks (left/right). Each block contains the differences for ECMWF (top panels) and CMCC (lower panels), verifying at the 1st and 2nd seasons.

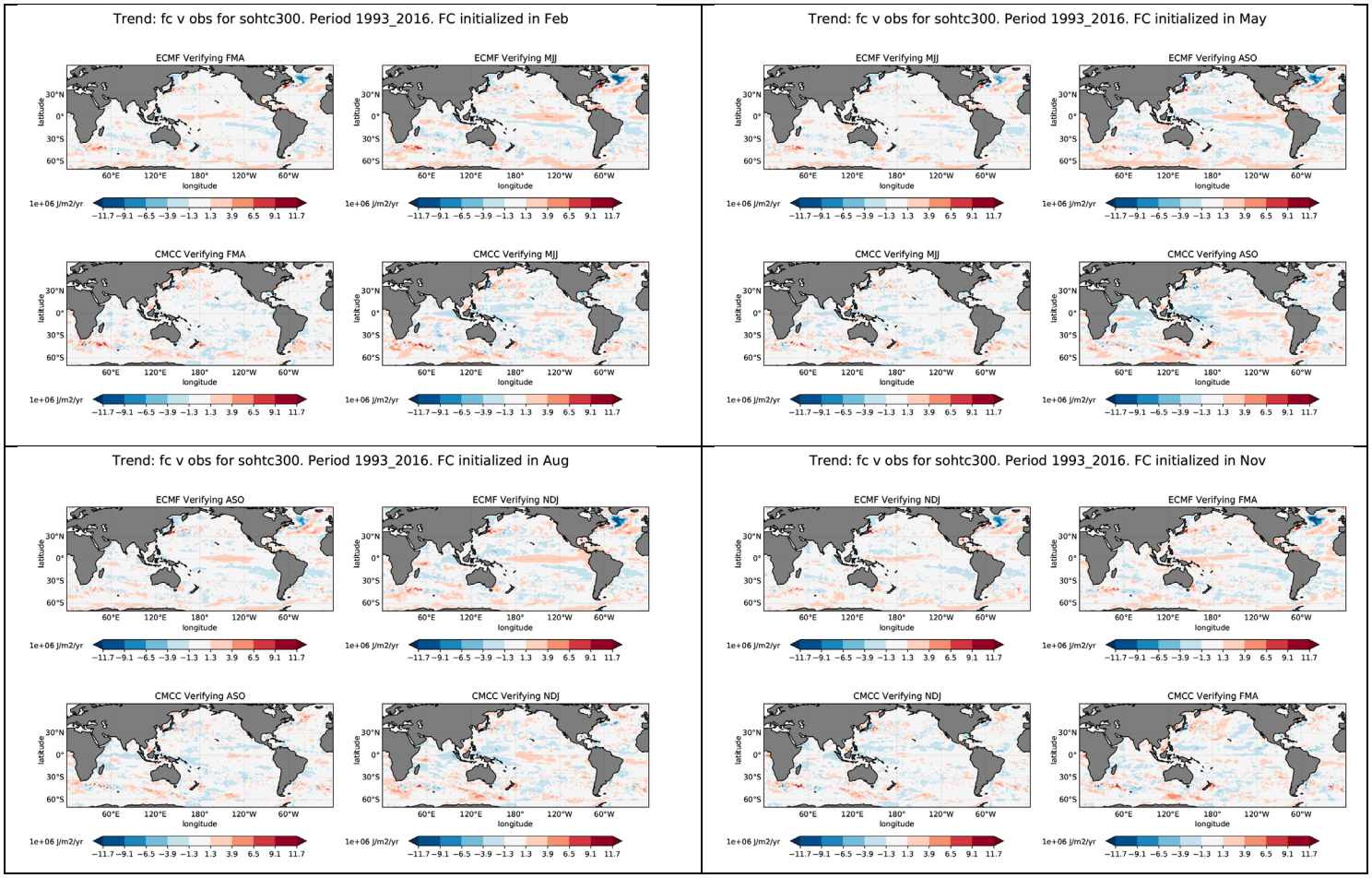


Figure A 16. As for Figure A 18, but for linear trends differences in forecasts of OHC.

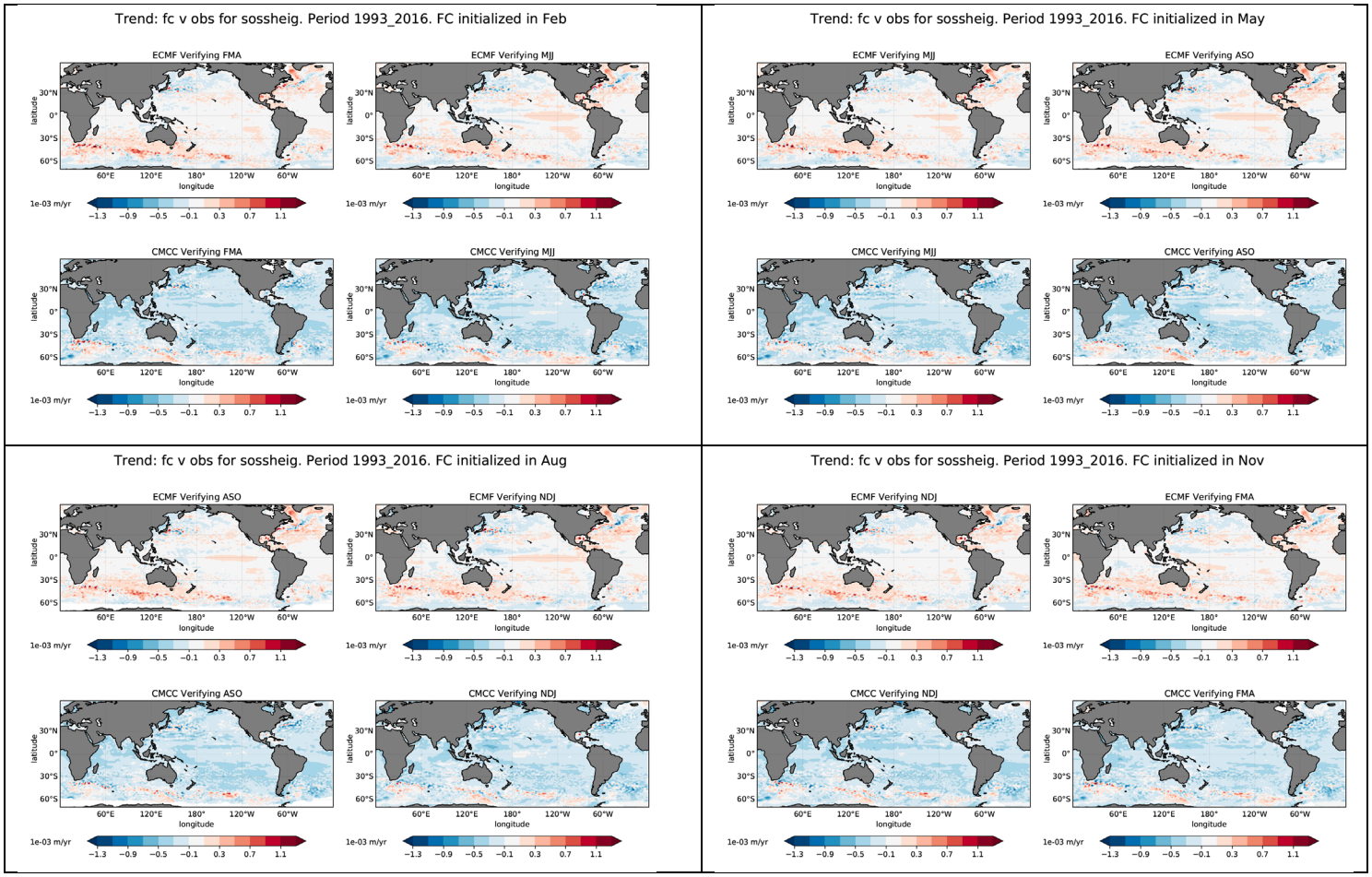


Figure A 17. As for Figure A 18, but for linear trends differences in forecasts of SSH.

References

Balmaseda, M. A.; Trenberth, K. E. & Källén, E., **2013**. Distinctive climate signals in reanalysis of global ocean heat content. *Geophysical Research Letters*, 40, 1754-1759

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

Craig, A. P.; Vertenstein, M. & Jacob, R., **2012**. A new flexible coupler for earth system modeling developed for CCSM4 and CESM1. *The International Journal of High Performance Computing Applications, 26*, 31-42

de Boisseson, E., Balmaseda, M., Mayer, M., Zuo, H., **2022**. Monitoring and predictions of the series of Marine Heatwave events impacting the Northeast Pacific in 2020, *Journal of Operational Oceanography*, Copernicus Ocean State Report, Issue 6, section 4.3

Dobricic, S. & Pinardi, N. **2008**. An oceanographic three-dimensional variational data assimilation scheme. Ocean modelling, Elsevier, 22, 89-105

Garric, G.; Parent, L.; Greiner, E.; Drévillon, M.; Hamon, M.; Lellouche, J.-M.; Régnier, C.; Desportes, C.; Le Galloudec, O.; Bricaud, C. & others, **2017**. Performance and quality assessment of the global ocean eddy-permitting physical reanalysis GLORYS2V4. *EGUGA*, 2017, 18776.

Good, S.A., **2020**. ESA Sea Surface Temperature Climate Change Initiative (SST_cci): GHRSST Multi-Product ensemble (GMPE), v2.0. Centre for Environmental Data Analysis, 05 August 2020.

Johnson, S.; Stockdale, T.; Ferranti, L.; Balmaseda, M.; Molteni, F.; Magnusson, L.; Tietsche, S.; Decremer, D.; Weisheimer, A.; Balsamo, G.; Keeley, S.; Mogensen, K.; Zuo, H. & Monge-Sanz, B. , **2019**. ECMWF-SEAS5: the new ECMWF seasonal forecast system**.** *Geoscientific Model Development, Geoscientific Model Development*, 12, 1087-1117.

L'Heureux M., M. K. Tippett and W. Wang, **2022**: Prediction Challenges from Errors in Tropical Pacific Sea Surface Temperature Trends. Front. Clim. Volume 4. https://doi.org/10.3389/fclim.2022.837483

McAdam, R.; Masina, S.; Balmaseda, M.; Gualdi, S.; Senan, R. & Mayer, M., **2022.** Seasonal forecast skill of upper-ocean heat content in coupled high-resolution systems. Climate Dynamics, 58(11), pp.3335-3350.

Merchant, C. J.; Embury, O.; Bulgin, C. E.; Block, T.; Corlett, G. K.; Fiedler, E.; Good, S. A.; Mittaz, J.; Rayner, N. A.; Berry, D.; Eastwood, S.; Taylor, M.; Tsushima, Y.; Waterfall, A.; Wilson, R. & Donlon, C., **2019**. Satellite-based time-series of sea-surface temperature since 1981 for climate applications. *Scientific Data*, 6, 223.

Mogensen, K.; Keeley, S. & Towers, P. , **2012**. Coupling of the NEMO and IFS models in a single executable. *ECMWF Techical Memorandum* 635

Pujol M.I., Y. Faugère, G. Taburet, S. Dupuy, C. Pelloquin, M. Ablain, and N. Picot. **2016**. DUACS DT2014: the new multi-mission altimeter data set reprocessed over 20 years. Ocean Sci., 12, 1067–1090, 2016 https://doi.org/10.5194/os-12-1067-2016

Riser, S. C.; Freeland, H. J.; Roemmich, D.; Wijffels, S.; Troisi, A.; Belbéoch, M.; Gilbert, D.; Xu, J.; Pouliquen, S.; Thresher, A. & others. **2016**. Fifteen years of ocean observations with the global Argo array. *Nature Climate Change*, 6, 145-153

Sanna, A.; A. Borrelli, P. A.; S. Materia, A. S. & S. Tibaldi, S. G**. 2017**. CMCC-SPS: The CMCC Seasonal Prediction System 3. Centro Euro-Mediterraneo sui Cambiamenti Climatici . *CMCC Tech. Rep. RP0285*, 61pp

Storto, A. & Masina, S., **2016**. C-GLORSv5: an improved multipurpose global ocean eddy-permitting physical reanalysis*. Earth System Science Data*, 8, 679-696

Storto, A.; Masina, S.; Simoncelli, S.; Iovino, D.; Cipollone, A.; Drevillon, M.; Drillet, Y.; von Schuckman, K.; Parent, L.; Garric, G.; Greiner, E.; Desportes, C.; Zuo, H.; Balmaseda, M. A. & Peterson, K. A. **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 Dynamics*, 53, 287-312

Taburet G., A.Sanchez-Roman, M. Ballarotta, M.I. Pujol, J.F Legeais, F. Fournier, Y. Faugere, and G. Dibarboure, **2019**. DUACS DT2018: 25 years of reprocessed sea level altimetry products. *Ocean Sci.*, 15, 1207–1224; https://doi.org/10.5194/os-15-1207-2019.

S. Tietsche, M. Balmaseda, H. Zuo, C. Roberts, M. Mayer, L. Ferranti, **2020**. The importance of North Atlantic Ocean transports for seasonal forecasts, Clim. Dyn. https://doi.org/10.1007/s00382-020-05364-6

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.