Supplementary Material

1 Ensemble Generation

Each of the ORA-OSES consists on an ensemble of 5 reanalysis, using the same perturbation method as in ORAS5 (Zuo et al 2019), which consist on applying perturbations to the surface fluxes and to the observations. The 5 ensemble members of the individual OSE experiments provide 5 different initial conditions for seasonal forecasts at any single time.

To create the 15 ensemble members of the FC-OSES we follow the same procedure as for SEAS5 (Johnson et al 2019). For each individual date, additional perturbations to the atmospheric initial conditions and atmospheric model are applied, in such a way the individual ensemble member of the ocean initial condition are given the same weight (e.g. in this case, a given ocean initial condition is used by 3 forecast ensemble member). The atmospheric initial condition perturbations consist on singular vectors and pre-computed ensemble from the operational ensemble of data assimilation (EDA). In addition, the uncertainty in the atmospheric model is taken into account by perturbing the model physics via a stochastic physic perturbation scheme. The total number of seasonal forecast integrations for a calendar month (say May or November) consist of 23 independent ensemble forecasts (initialized on that calendar month during the period 1993-2015), each of them comprising 15 ensemble coupled integrations up-to 7 months. This large number of ensembles is required to obtain significant and robust results, since signal to noise ratio in seasonal forecast is low, and the results exhibit a strong seasonal and lead-time dependence.

2 Perturbation Analysis

A linear multivariate perturbation analysis is conducted in order to explore how the ocean initial conditions perturbations translate into differences in the forecasts of SST over the tropical area. To increase the sample size and to include a range of temporal variations, we construct sets of initial δIni and forecast δFc perturbations by aggregating pairs of differences between pairs of experiments, each of them spanning the record 1993-2015 (i.e. 23 years). Only the tropical area 30S-30N is considered, at a resolution of 2x2 degrees. The analysis is conducted separately for the May and November starts. The main perturbation analysis is done using 3 pairs of experiments (NoInsitu-REF, NoArgo-REF, NoArgo-NoInsitu), yielding a sample of 69 perturbations. Two additional perturbation analysis were conducted: i) one analysis using all pairs, including NoInsAtl-REF, NoInsAtl-NoInsitu and NoInsAtl-NoArgo, to explore the role of the observing system in the Atlantic in the context of other basins; and ii) another analysis using only the pair NoInsAtl-REF, to explore the specific role of the observing system in the Atlantic basin.

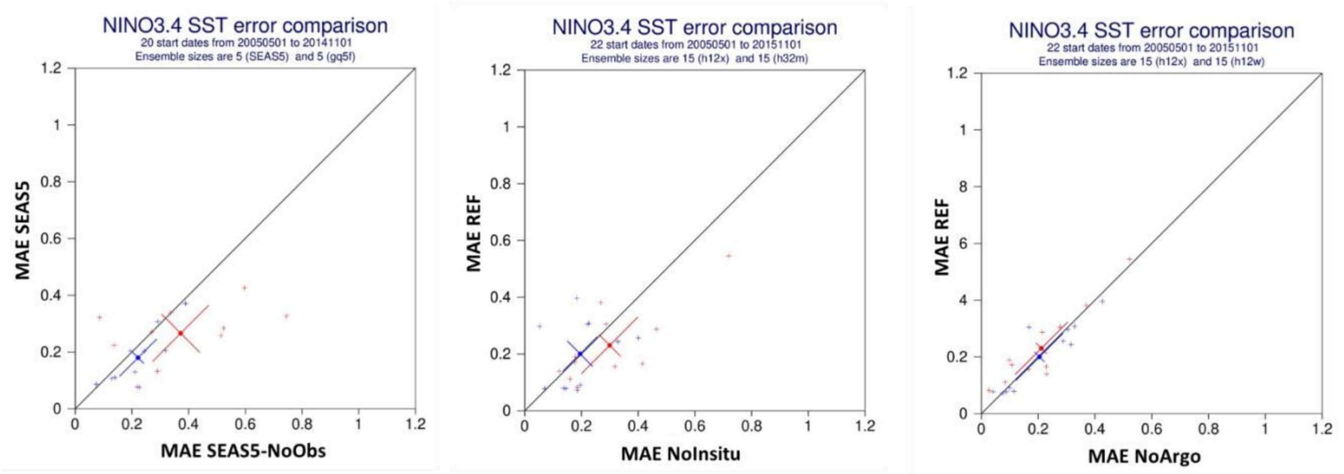
The inital perturbations δIni include different variables (***var****=[SST,OHC,D28I,MLD]*), and depend on latitude (*lat*), longitude (*lon*), and start date of the forecasts (*t0*, once a year, May or November starts). Since the samples are aggregated as described above, there are *t0*=69. As a proxy for the initial conditions of the forecasts for a given start date, we take the corresponding preceding month from the ORA experiments. Thus, we use ORA April monthly means as an approximation of the initial conditions for May starts, and October monthly means for the November starts. We can express the initial perturbations as δ*Ini(***var***,lon,lat,t0).*

The forecasts perturbations δ*Fc* are constructed in similar manner by collating pairs of differences from the ensemble mean of corresponding FC-OSES, and they depend on the forecast lead time, but the target variable includes only SST. We can express the forecast perturbation as δ*Fc(SST,lon,lat,t0,lead_time).* As lead time we chose two flavours: i) for the local correlation in analysis **Figure 4** and **Fig S7** we use lead_time=1,..7 months, e.g. monthly means for the individual months; ii) for the singular value decomposition in **Figure 5** we use lead_time=[month1, season1, season2], where month1 is the first month into the forecast, season1 is the average of months [2,3,4] and season2 is the average of months [5,6,7].

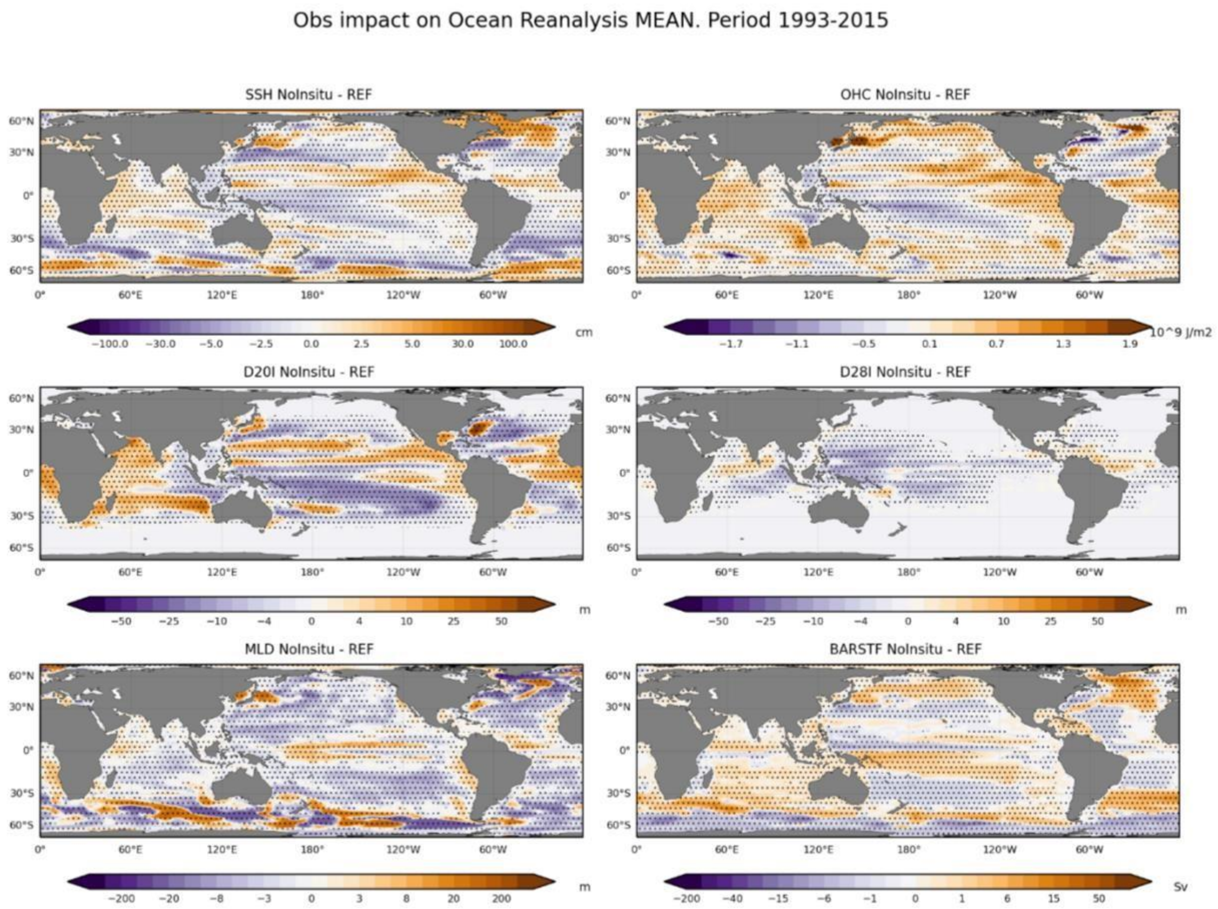
For the local correlation analysis we compute the local correlation in along t0 of δ*Ini* and δ*Fc,* independently for each grid point, lead time and variable. This yields a local correlation field *L_C(***var***,lon,lat,lead_time), where L_C(***var***=varx)* is the correlation of forecast SST with a single variable varx in the initial conditions. For the purpose of illustrating the results in **Figure 4** we average the local correlation values L_C across latitudes (top panels) or longitudes bands (bottom panels)

We also conduct a multivariate singular value decomposition (SVD) between the cross covariance matrix of δ*Ini* and δ*Fc.* The SVD of the covariance matrix finds pairs of modes which explain the maximize the explained joined covariance of the initial and final perturbations. Each mode is characterized by spatial patterns, explaining a given percentage of the total variance of the initial and forecast perturbations, and associated with time series for the initial and final perturbations that covary together, with a given correlation coefficient. To avoid giving unrealistic weights to different variables or forecast times, the individual fields are normalized by their global standard deviation before entering δ*Ini* and δ*Fc.* The normalization factors are later used to scale back the resulting SVD spatial patterns and convert them into physical units. **Figure 5** shows the spatial patterns of δ*Ini* and δ*Fc* and the associated time series, as well as the information on explained variance and correlation coefficient. The time series show the start dates aggregated across experiments. But it is easy to decompose and map them to specific years and experiment, as shown in **Figure S8**.

3 Supplementary Figures



Supplementary Figure S1. Impact of removing ocean observations in seasonal forecasts of Nino3.4 (5N-5S, 170W-120W) SST anomalies. The panels show the scatter diagrams of forecast Mean Absolute Error during the first 4 months of the forecasts, where the y-axis show the reference experiments (SEAS5 and REF), and the x-axis show the experiments withdrawing: left panel) all ocean observations in the case of SEAS5-NoObs, including altimeter, middle panel) all in-situ ocean observations; right panel) all Argo. Shown are the forecasts for the period 2005-2015 initialized in May (red) and November (blue). The mean and sample errors (95% significance level) are shown in bold. Points below the diagonal indicates that removing ocean observations have a detrimental impact on the skill of seasonal forecasts. The impact of ocean observations is more visible in forecasts initialized in May.



Obs impact on Ocean Reanalysis MEAN. Period 1993-2015

SSH NoInsitu - REF

OHC NoInsitu - REF

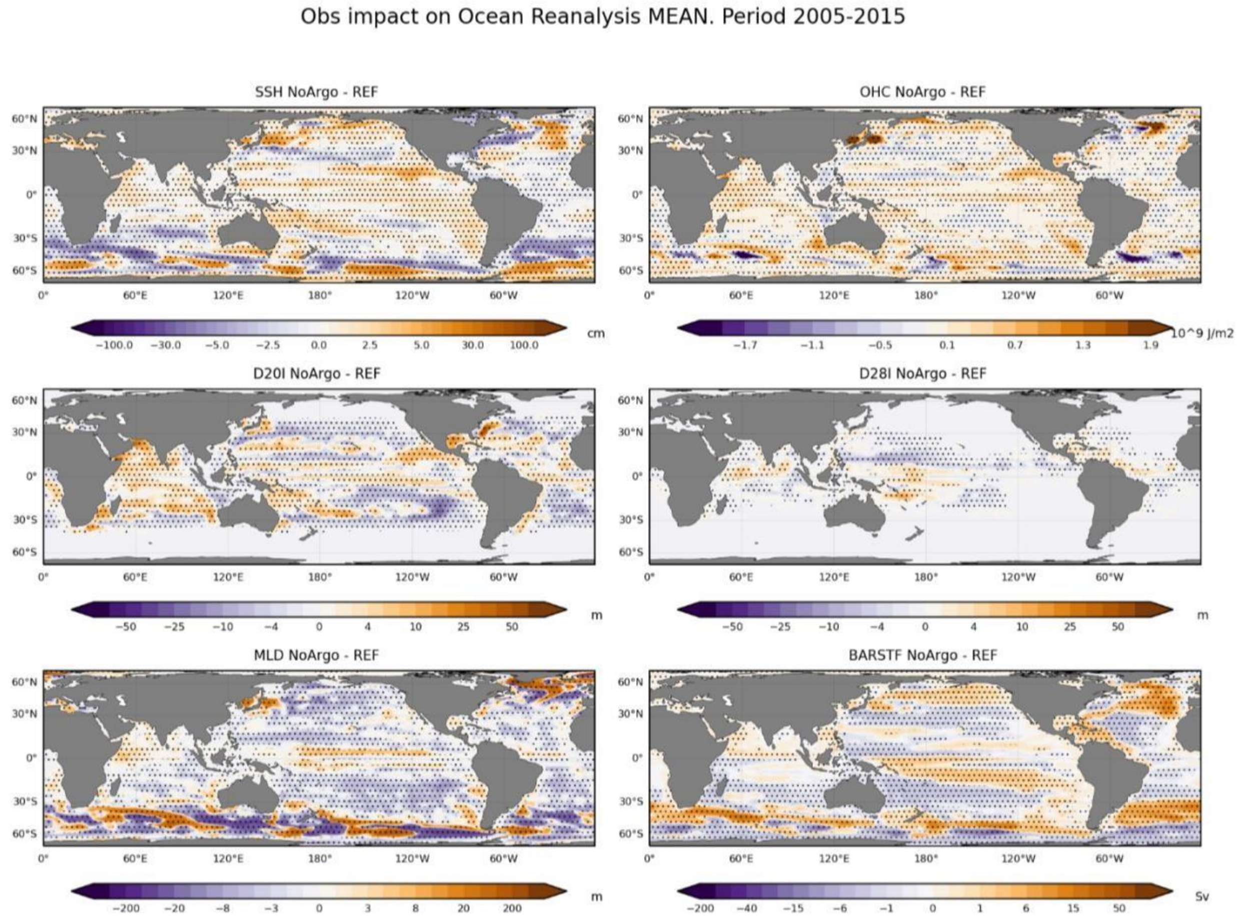
D20I NoInsitu - REF

D28I NoInsitu - REF

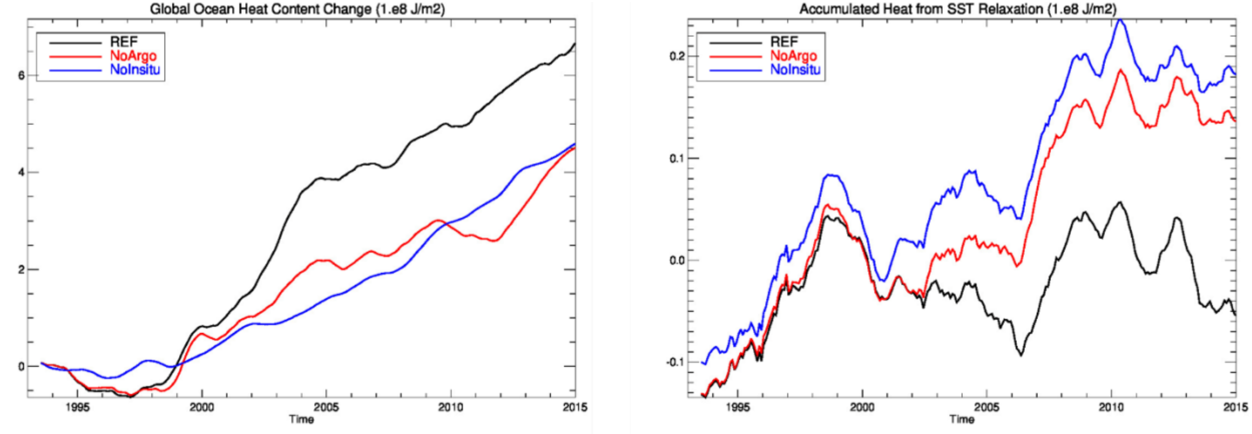
MLD NoInsitu - REF

BARSTF NoInsitu - REF

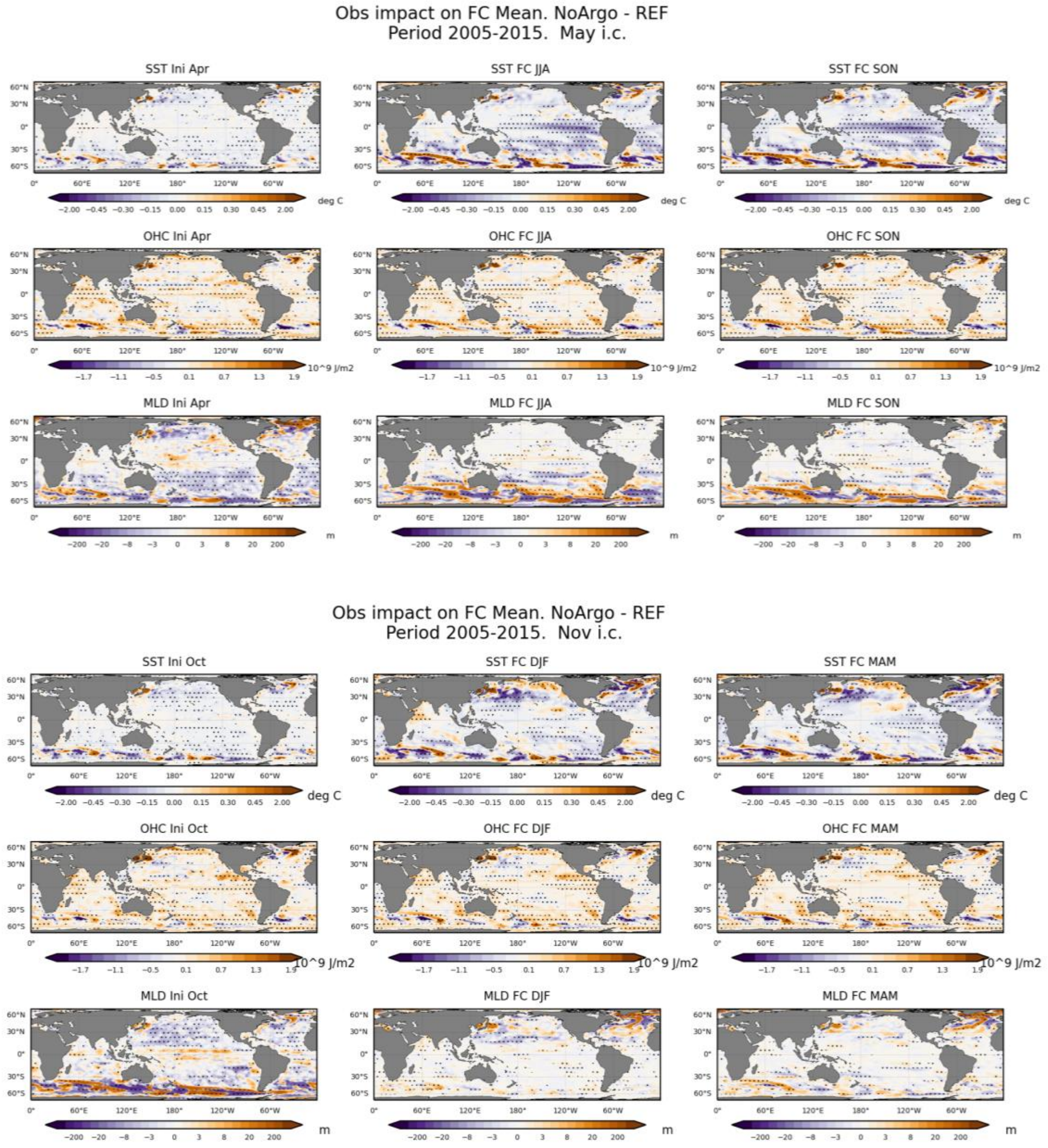
Supplementary Figure S2. Impact of removing all in-situ ocean observations on the 1993-2015 mean state of the ocean initial conditions, as measured by the differences between experiments NoInsitu – REF. Shown are differences in SSH, OHC, D20I, D28I, MLD and barotropic stream function (BARSTF). Dotted areas indicate where the differences are significant at the 90% level.



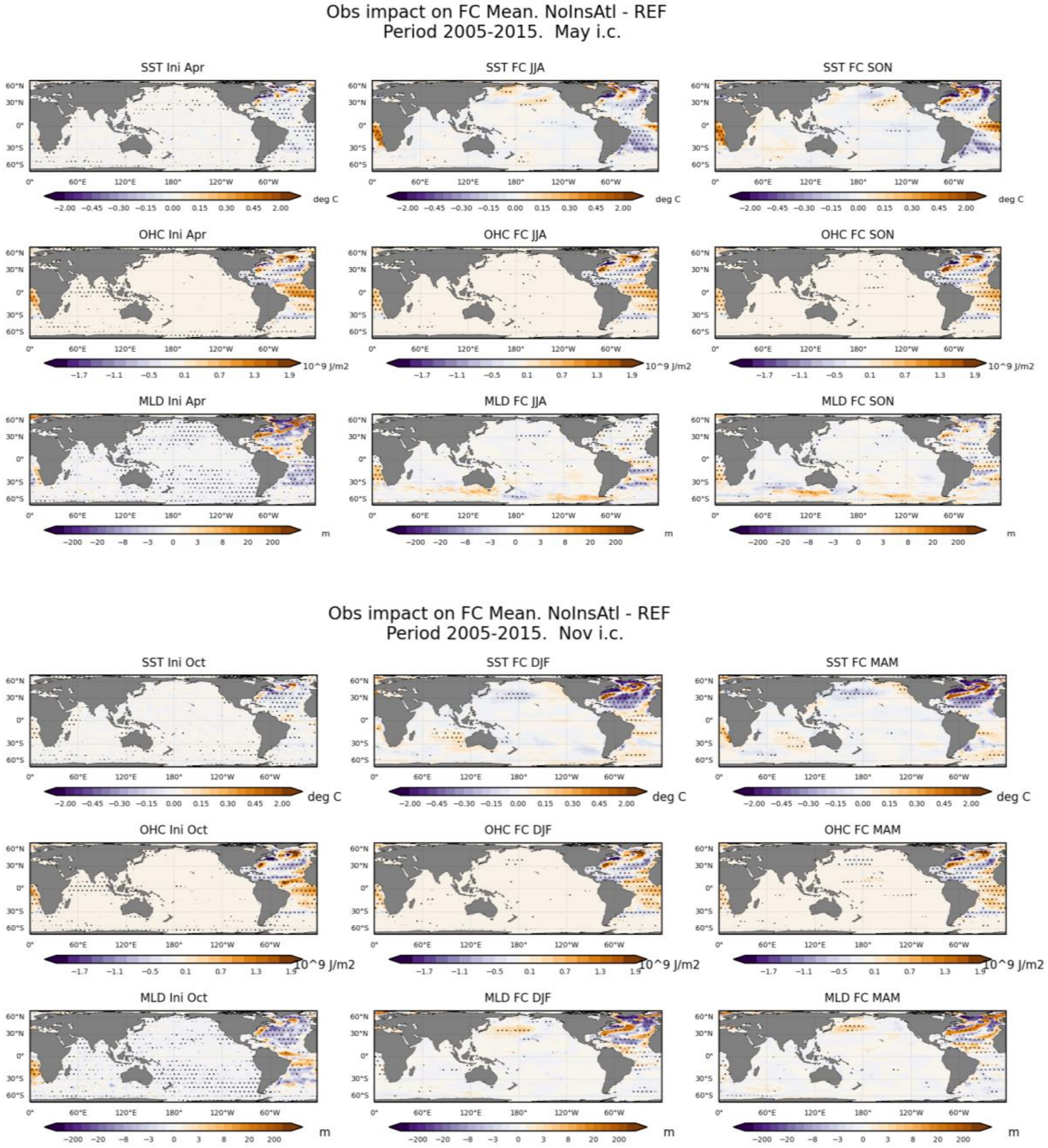
Supplementary Figure S3. Impact of removing the Argo ocean observations on the 2005-2015 mean state of the ocean initial conditions, as measured by the differences between experiments NoArgo – REF. Shown are differences in SSH, OHC, D20I, D28I, MLD and barotropic stream function (BARSTF). Dotted areas indicate where the differences are significant at the 90% level.



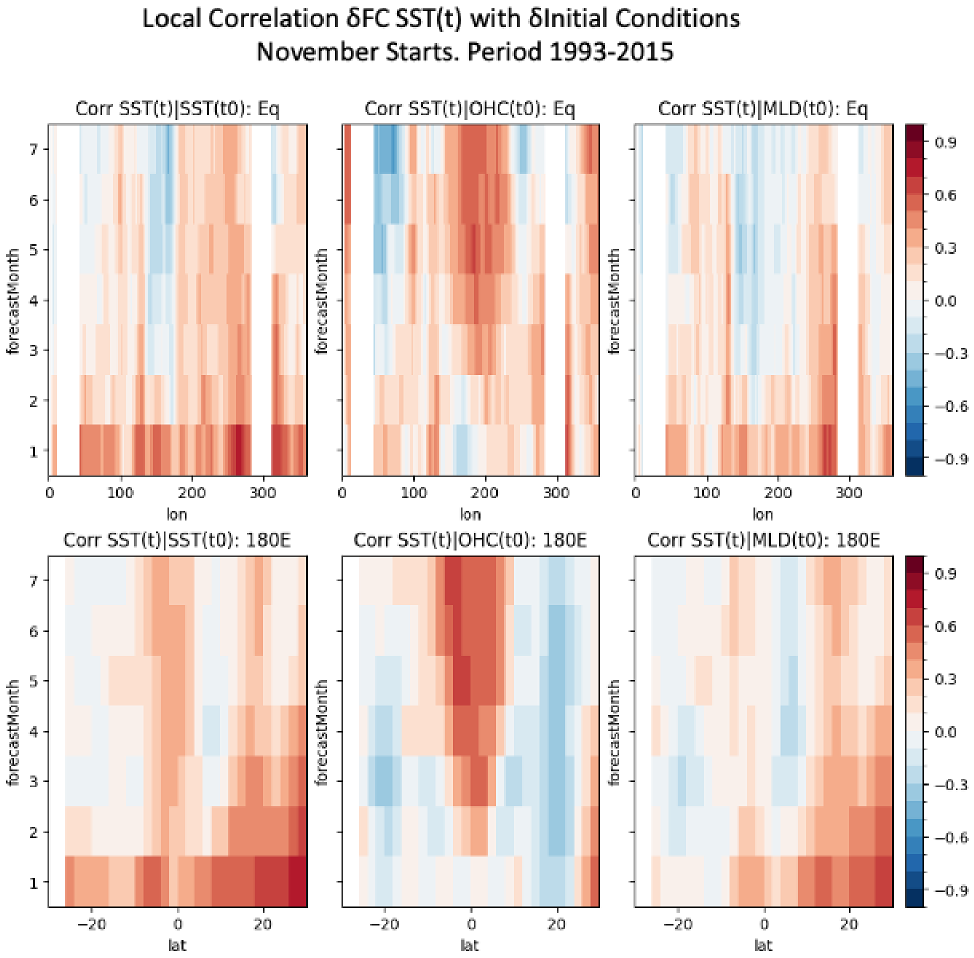
Supplementary Figure S4. Total ocean heat content change (left), and accumulated heat flux resulting from the SST relaxation (right) in the difference experiments. The removal of all the insitu or Argo observations results in lower estimations of ocean heat content, and they need stronger surface relaxation.



Supplementary Figure S5. Impact of Argo ocean observations in the mean state of seasonal forecasts of SST, OHC and MLD as measured by the experiments NoArgo – REF (top/middle/lower rows of each individual plate). Shown are the results for forecasts initialized in May (top plate) and November (bottom plate). The differences in the initial conditions are shown in the left column, and the forecasts for the first and second seasons are in the central and rightmost columns respectively. The dotted areas indicate where the differences are significant at the 90% significant level.

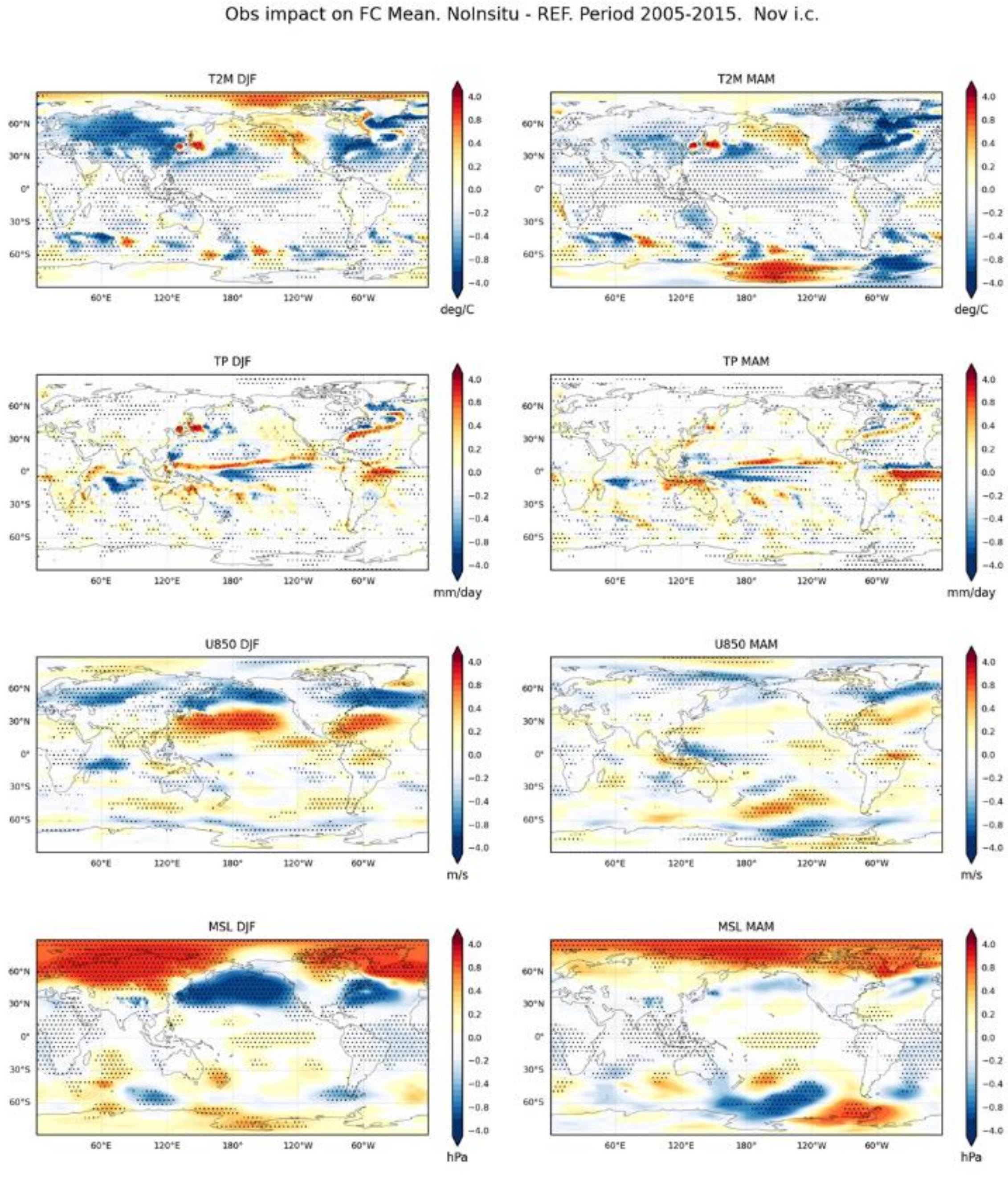


Supplementary Figure S6. Impact of Atlantic in-situ ocean observations in the mean state of seasonal forecasts of SST, OHC and MLD as measured by the experiments NoInsAtl – REF (top/middle/lower rows of each individual plate). Shown are the results for forecasts initialized in May (top plate) and November (bottom plate). The differences in the initial conditions are shown in the left column, and the forecasts for the first and second seasons are in the central and rightmost columns respectively. The dotted areas indicate where the differences are significant at the 90% significant level.



Supplementary Figure S7. Local correlation between the SST in δFc at different lead times the different variables in δIni averaged over the Equator (top) and at the date line for forecasts initialized in November. The local correlation has been averaged across the 5S-5N (top) and across all the longitudes (bottom).

Supplementary Figure S8. Time series of first (left) and 2nd (right) modes of the singular value decomposition between δFc and δIni. The time series are those shown in Fig 5 of the main text, but the generic sample time has been disaggregated into the contribution from the different OSE experiments.



Supplementary Figure S9. Impact of in-situ ocean observations in the 2005-2015 mean state of seasonal forecasts of atmospheric variables as measured by the differences between experiments NoInsitu and REF. shown are forecasts are initialized in November for the first and second seasons into the forecasts (left and right panels). The different rows correspond to different variables T2m, total precipitation (TP), U850 and mean sea level pressure (MSL). The dotted areas indicate that the differences in the mean are significant at the 90% level.