Relating model bias and prediction skill in the equatorial Atlantic

François Counillon[1,2] · Noel Keenlyside[1,2] · Thomas Toniazzo[3] · Shunya Koseki[2] · Teferi Demissie[3,4] · Ingo Bethke[2] · Yiguo Wang[1]

Abstract

We investigate the impact of large climatological biases in the tropical Atlantic on reanalysis and seasonal prediction performance using the Norwegian Climate Prediction Model (NorCPM) in a standard and an anomaly coupled configuration. Anomaly coupling corrects the climatological surface wind and sea surface temperature (SST) fields exchanged between oceanic and atmospheric models, and thereby significantly reduces the climatological model biases of precipitation and SST. NorCPM combines the Norwegian Earth system model with the ensemble Kalman filter and assimilates SST and hydrographic profiles. We perform a reanalysis for the period 1980–2010 and a set of seasonal predictions for the period 1985–2010 with both model configurations. Anomaly coupling improves the accuracy and the reliability of the reanalysis in the tropical Atlantic, because the corrected model enables a dynamical reconstruction that satisfies better the observations and their uncertainty. Anomaly coupling also enhances seasonal prediction skill in the equatorial Atlantic to the level of the best models of the North American multi-model ensemble, while the standard model is among the worst. However, anomaly coupling slightly damps the amplitude of Atlantic Niño and Niña events. The skill enhancements achieved by anomaly coupling are largest for forecast started from August and February. There is strong spring predictability barrier, with little skill in predicting conditions in June. The anomaly coupled system show some skill in predicting the secondary Atlantic Niño-II SST variability that peaks in November–December from August 1st.

Keywords Equatorial Atlantic · Seasonal prediction · Model bias · Data assimilation

1 Introduction

The Atlantic Niño and the weaker Atlantic Niño-II dominate interannual variability in the equatorial Atlantic (Zebiak 1993; Keenlyside and Latif 2007; Ding et al. 2010;

✉ François Counillon
francois.counillon@nersc.no

1   Nansen Environmental and Remote Sensing Center and Bjerknes Centre for Climate Research, Bergen, Norway

2   Geophysical Institute, University of Bergen and Bjerknes Centre for Climate Research, Bergen, Norway

3   NORCE Norwegian Research Centre, Bjerknes Centre for Climate Research, P.O. Box 22, 5838 Bergen, Norway

4   CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS), East Africa, Addis Ababa, Ethiopia

Lübbecke et al. 2018; Okumura and Xie 2006). These phenomena peak respectively in boreal summer and November–December. They both exhibit dynamics similar to the much stronger El Niño Southern Oscillation (ENSO), but coupled ocean–atmosphere interactions are less pronounced (Jansen et al. 2009; Lübbecke and McPhaden 2013) and positive and negatives events are more symmetric (Lübbecke and McPhaden 2017). Like ENSO, they also have important climatic impacts (Okumura and Xie 2006; Lübbecke et al. 2018; Foltz et al. 2019; Losada et al. 2010).

Unlike with ENSO, little skill has been demonstrated in predicting equatorial Atlantic variability (Stockdale et al. 2006, 2011; Richter et al. 2018; Barreiro et al. 2005). The low predictability can be partly explained by the strong seasonal modulation of coupled ocean–atmosphere interactions in this basin and the greater influence of largely stochastic wind variability (Richter et al. 2014a; Richter and Doi 2019; Nnamchi et al. subm.). Furthermore, thermodynamic ocean–atmosphere interaction and a range of other mechanisms contribute to equatorial Atlantic variability (Lübbecke

et al. 2018; Nnamchi et al. 2015; Brandt et al. 2011; Richter et al. 2013); and most of these mechanisms may be less predicable than ENSO like dynamics.

Model error is also a potential cause for poor prediction skill. In particular, the tropical Atlantic biases in state-of-the-art models are much larger than the amplitude of the interannual variability (Richter et al. 2014b). They are characterised by a large warm SST bias (up to 8 °C in the southeast tropical Atlantic), a too deep thermocline in the eastern Atlantic (Richter et al. 2014b), a cyclonic surface wind bias in the Angola–Benguela Frontal Zone and a southward shift of the intertropical convergence zone (ITCZ) (Richter et al. 2012). These biases cause an underestimation of the thermocline feedback (Deppenmeier et al. 2016) and an overestimation of thermodynamic ocean–atmosphere interaction (Jouanno et al. 2017); and reducing them enhances the simulation of dynamical ocean–atmosphere interaction and Atlantic Niño variability (Ding et al. 2015a, b; Dippe et al. 2018; Harlaß et al. 2018). Furthermore, Dippe et al. (2019) show that reducing this bias improves the prediction of Atlantic Niño variability, although prediction skill remains poor—i.e. only beating marginally persistence for May start and beating persistence by 0.1 for August start up to 4 lead month.

Unfortunately, improvements in simulating tropical Atlantic climate have been only moderate during the past 20 years (Davey et al. 2002; Richter et al. 2012; Toniazzo and Woolnough 2014). There is a greater understanding of the causes of tropical Atlantic biases: The equatorial Atlantic SST biases are linked to too weak trade winds in boreal spring (Richter and Xie 2008; Wahl et al. 2011; Voldoire et al. 2019); whereas southeastern tropical Atlantic bias has been related to the misrepresentation of (1) the persistent low level cloud cover (Zuidema et al. 2016), (2) along shore coastal winds (Xu et al. 2014a; Koseki et al. 2018) and (3) coastal upwelling (Xu et al. 2014b). Several studies have shown that increasing atmospheric model resolution can reduce tropical Atlantic biases, especially in the southeastern Atlantic (Small et al. 2015; Milinski et al. 2016; Harlaß et al. 2018), whereas de la Vara et al. (2020) have shown improvements from increasing oceanic model resolution. While detecting the cause of the bias and formulating a solution that is computationally tractable is most desirable, the remaining bias even in the computationally demanding systems (e.g., Harlaß et al. 2018) is still large (about 2–3 °C). Thus, the tropical Atlantic warm bias is likely to remain a problem for seasonal prediction for a while to come.

Flux correction and anomaly coupling are alternate approaches that mitigate climatological biases and their impacts, and thereby enhance predictions. In flux correction a fixed numerical correction is added to the exchanged fluxes (Sausen et al. 1988), while in anomaly coupling the variables used to compute air–sea fluxes are corrected

(Kirtman et al. 1997). Although neither approach guarantees an improved simulation of variability, anomaly coupling has been shown effective in improving the simulation and prediction of equatorial Atlantic interannual variability (Ding et al. 2015a; Dippe et al. 2018, 2019). A drawback with both approaches is that estimating effective numerical corrections is difficult, because coupled feedbacks in the tropics can strongly amplify errors (Neelin and Dijkstra 1995). This is particularly problematic when the corrections are estimated from uncoupled ocean general circulation model (GCM) and atmospheric GCM simulations. To overcome this, Toniazzo and Koseki (2018) developed an iterative approach to estimate the anomaly coupling corrections from a coupled GCM simulation. They applied their approach to the Norwegian Earth system model (NorESM; Bentsen et al. 2012), which has tropical Atlantic biases comparable to those of the models from the Coupled Model Intercomparison Project 5 (CMIP5) (Koseki et al. 2018; Toniazzo and Koseki 2018). Their technique alleviated the SST and precipitation biases in the tropical Atlantic (and elsewhere) without damping too much the variability.

In this study, we aim to use Toniazzo and Koseki (2018) improved technique to assess whether current model biases limit seasonal prediction skill in the equatorial Atlantic. This work is based on the Norwegian Climate Prediction Model (NorCPM, Counillon et al. 2014, 2016), which combines the NorESM and the ensemble Kalman filter (EnKF, Evensen 2003) data assimilation method. We will use NorESM in its standard and anomaly coupled configurations. NorCPM aims to provide long-term reanalyses and seasonal-to-decadal climate predictions. NorCPM demonstrated good skill in controlling the upper ocean heat content in the equatorial and north Pacific, the north Atlantic subpolar gyre region and the Nordic Seas seas by assimilating surface temperature anomalies (SSTAs, Counillon et al. 2016). In Wang et al. (2019) NorCPM with assimilation of SST reaches skill comparable to the top-performing prediction systems of the North American Multi Model ensemble (NMME, Kirtman et al. 2014) in most regions, but performance in the tropical Atlantic—were the model has large biases—were found to be poor. One can thus expect anomaly coupling to be particularly beneficial there, by correcting the model bias, improving the representation of the dynamics and enhancing the prediction skill.

This paper is organised as follow. Section 2 presents the Norwegian climate prediction model, in its standard and anomaly coupled configurations. We then assess the impact of the anomaly coupling on the accuracy of a reanalysis for the period 1980–2010 (Sect. 4), and on the skill of retrospective hindcasts of Atlantic Niño variability (Sect. 5).

2 Norwegian climate prediction system and experimental design

2.1 The Norwegian Earth system model

The NorESM (Bentsen et al. 2012) is a global fully coupled model for climate simulations. It is based on the Community Earth System Model version 1.0.3 (CESM1, Vertenstein et al. 2012), a successor to the Community Climate System Model version 4 (CCSM4, Gent et al. 2011). Unlike in the Community Atmosphere Model (CAM4, Neale et al. 2010), the atmospheric component (CAM4-OSLO) features advanced aerosol chemistry schemes (Kirkevåg et al. 2013). The ocean component (Bentsen et al. 2012) is an updated version of the isopycnal coordinate ocean model MICOM (Bleck et al. 1992). This study uses the medium-resolution NorESM1-ME (Tjiputra et al. 2013), which has the capability to be fully emission driven and has contributed output to CMIP5. External forcings used here comply with CMIP5's historical experiment (see Bentsen et al. 2012 for details). The atmosphere and land components are configured on a 2° finite-volume grid that has a latitude longitude resolution of $1.9° \times 2.5°$. The atmosphere component uses 26 hybrid sigma-pressure levels with a model top at approximately 3 hPa. The horizontal resolution of the ocean and sea-ice model is approximately 1°, but is enhanced in the meridional direction about the equator and enhanced in both zonal and meridional directions at high latitudes. The ocean uses 51 isopycnal layers and two layers for representing the bulk mixed layer with time-evolving thicknesses and densities. We refer to this configuration as CTRL. The NorESM CTRL has climatological SST, precipitation, and wind biases in the tropical Atlantic that are typical of CMIP5 models (Koseki et al. 2018).

2.2 The anomaly coupled version of NorESM

The anomaly coupling technique developed by Toniazzo and Koseki (2018) allows to substantially reduce such biases, through correcting the SST and wind stress exchanged between the ocean and atmosphere models at each coupling steps. The corrections in the anomaly coupling vary in space and time according to a monthly climatology. They compensate for biases that arise when the coupled surface fluxes of energy and momentum act to cause systematic departures of local SSTs from the observed surface climatology. The correction thus represents a compensation for the difference between the sea-surface climatology of the observations and that of the (standard) model simulations. In this sense the model thus becomes *anomaly coupled* (hereafter "ACPL") i.e., the ocean model receives wind stress consisting of atmosphere model simulated anomalies added to the observed climatology—and likewise for SST received by the atmosphere. The corrections are calculated in coupled mode according to an iterative scheme that allows to account for the effect of coupled feedbacks. For details of this procedure, including the target climatology, the reference observation data set, and a validation of the resulting simulated climatology, see Toniazzo and Koseki (2018). Note that we do not use Toniazzo and Koseki (2018)'s proposed global energy conservation scheme. The global energy imbalances are globally uniform and within $\pm 1\,W/m^2$ that is generally small compared to regional physical flux anomalies. While this may become important for inter-annual or slower modes of variability, we do not expect its inclusion to substantially alter seasonal predictions and the findings of our paper. Furthermore, we did not identify strong regional climate drifts in our seasonal predictions system (not shown).

The impact of anomaly coupling on SST and precipitation in our model configuration for the period 1980–2010 is now depicted (Fig. 1). For precipitation we calculate the bias against the Global Precipitation Climatology Project monthly precipitation (GPCP) Version 2.3 dataset (Adler et al. 2003). In the tropical Atlantic anomaly coupling dramatically reduces the SST and precipitation biases. Although the scope of the paper focuses is in the tropical Atlantic, it is worth mentioning that the SST bias in the tropical Pacific is not fully eliminated by anomaly coupling (Fig. S1 for a global assessment). The remaining bias can be related to the fact that (1) the bias of the individual components (here in the ocean model MICOM) is not reduced by the anomaly coupling (2) the flux correction term was trained with constant external forcing, while we are now verifying the performance with transient forcing.

Although anomaly coupling corrects the mean bias for our region of interest, it does not guarantee an improved simulation of variability. We illustrate this by presenting the standard deviation of equatorial Atlantic SST anomalies as a function of calendar month from observations and free running simulations with the standard and anomaly coupled configurations of NorESM (Fig. 2). The seasonality of equatorial Atlantic SST variability provides insights into the model's performance in capturing the timing of Atlantic Niño variability. The standard version of NorESM shows a marked seasonal peak in variability along the equator in July while in observations the peak is reached a month earlier and is weaker. With ACPL, the maximum occurs in June as in the observation but is a little weaker and displaced to the east. In the standard NorESM the strong SST variability persist until the end of the year while in the observation it decays rapidly during
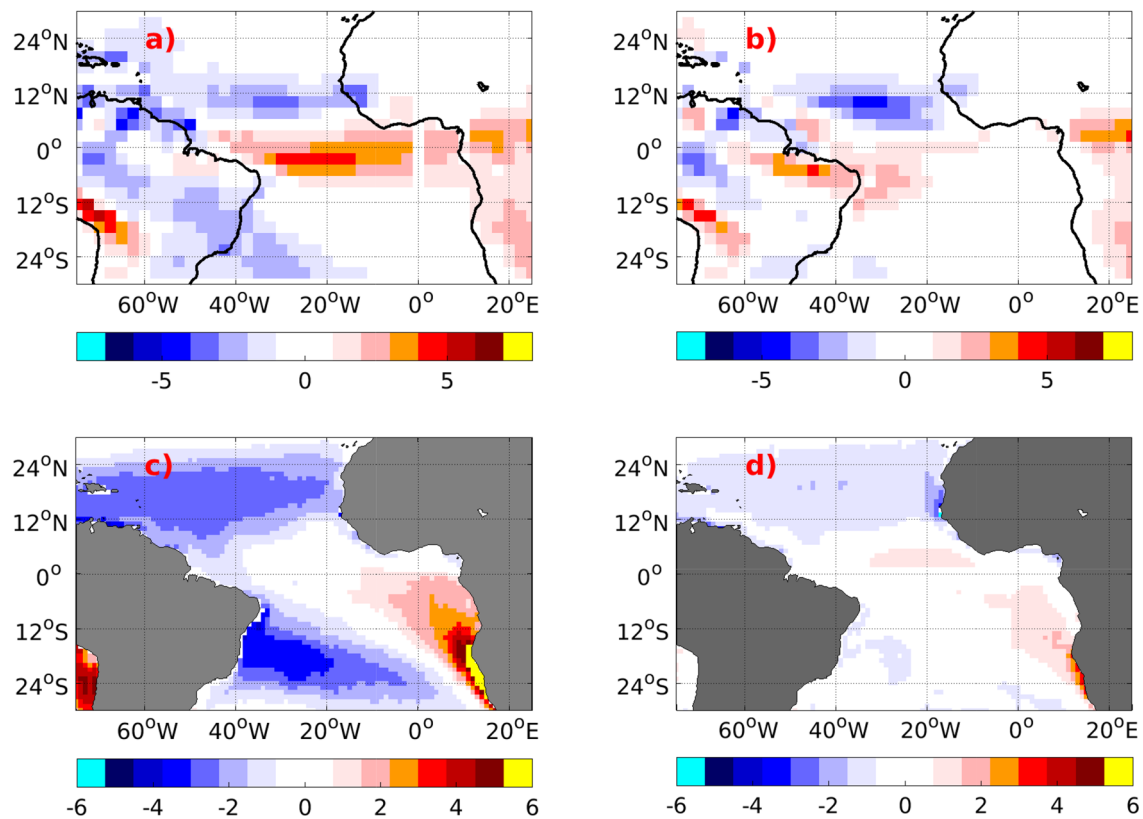
Fig. 1 The first row is the precipitation bias (in mm per day) of the mean of a 5 ensemble member free running versions (i.e. without assimilation) of CTRL (**a**) and ACPL (**b**) calculated against GPCP V2.3 for the period 1980–2010. The second row (panels **c** and **d**) is the same for SST against HadISST2
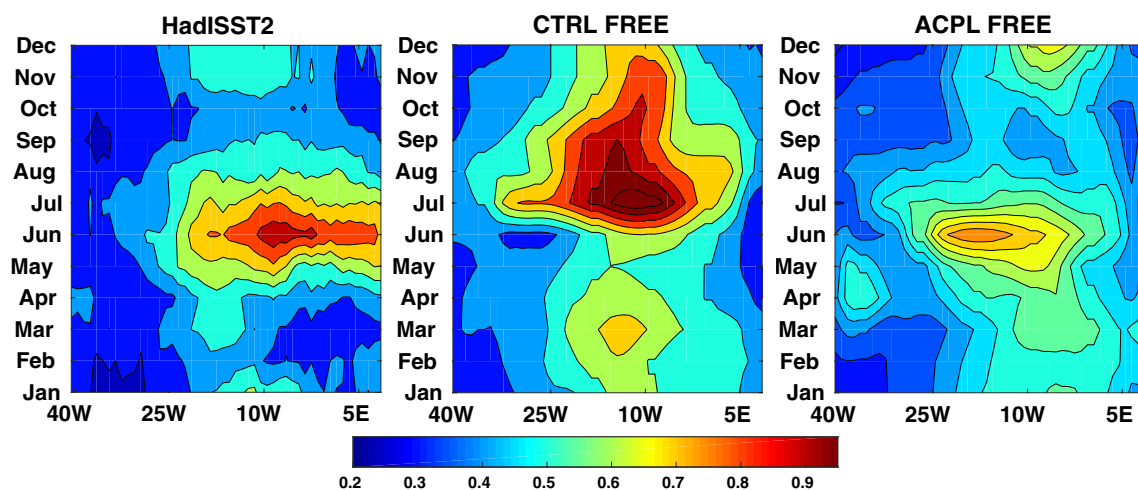

Fig. 2 The annual cycle of SST standard deviation along the equator in the observation (HadISST2) and in free running versions of CTRL and ACPL. The simulated standard deviations are estimated for the period 1980–2010 using five members for both CTRL and ACPL

September before showing a second weaker peak during November–December associated with the Atlantic Niño-II phenomena (Okumura and Xie 2006). Again, the ACPL model better captures this secondary maximum in SST variability, although it is slightly too far to the east compared to the observations and a little too strong. Overall, the anomaly coupled model simulates more realistic variability compared to the standard version of NorESM.

2.3 Reanalyses and seasonal hindcasts experiment description

The Norwegian Climate prediction model (Counillon et al. 2014, 2016; Wang et al. 2019) assimilates observations with the EnKF. Here, we assimilate SST and hydrographic profiles as described in Counillon et al. (2014, 2016) and Wang et al. (2017). The SST data is from the HadISST2 data set (HadISST2.1.0.0, Rayner et al., personal communication) that provides an ensemble of reconstruction of SST for the period 1850–2010. We assimilate the ensemble mean and use the ensemble spread, which varies in space and time, to quantify the accuracy of the observational data set, as is needed for the data assimilation. The hydrographic profiles are from EN4.1.1 (Gouretski and Reseghetti 2010). The observation error is estimated as described in Karspeck (2016) and the localisation radius varies with latitudes as described in Wang et al. (2017). The entire ocean state vector is updated; that is we update the temperature, salinity, and layer thickness of all vertical layers. We employ the aggregation method (Wang et al. 2016) for layer thickness. It is a cost-efficient modification of the linear analysis update in data assimilation (DA) for physically constrained variables, which ensures that the analysis satisfies physical bounds without changing the expected mean of the update and thus avoid introducing a drift.

Two 30 member reanalyses are performed with CTRL and ACPL configurations for the period 1980–2010. The initial conditions for both reanalyses are branched in 1980 from the two corresponding ensemble historical simulations. These two ensemble historical simulations are produced by selecting random initial condition from a stable preindustrial simulations and integrating the ensemble from 1850 to 2005 using CMIP5 historical forcings and there after the RCP8.5 is used (Taylor et al. 2012). CTRL historical ensemble simulations of 30 members have been performed for a previous study (Counillon et al. 2016, data set available on https://doi.org/10.11582/2019.00035). As producing such initial conditions is computationally expensive, ACPL historical simulation is performed differently: using the anomaly coupled configuration with fixed monthly climatological correction, we perform a 5 ensemble member simulation from 1850 to 2010; the other 25 ensemble members are spawned from these 5 members, by perturbing the SST of the initial condition in January 1970 with 0.1 °C spatially white noise and then integrating these 25 members to 1980. The resulting ensemble spread in the near surface and intermediate ocean depths of the 30 member ACPL simulation has comparable amplitude to that of CTRL (not shown); and thus it should satisfy a well balanced EnKF (i.e. error versus ensemble spread) for seasonal prediction.

We perform anomaly assimilation; that is the climatological monthly mean for the observations and the model are removed before comparing the two. This option was preferred over full field assimilation for several reasons. First handling model bias in DA is a real challenge because it is designed to correct random, zero-mean errors, i.e. the model and observations are assumed (erroneously) to be unbiased. As a result, the analysis state with full field assimilation will still include part of the bias (Dee 2005) and yields a too strong reduction of ensemble spread (Anderson 2001). Second, full field assimilation can produce a large shock as the model often drifts back rapidly to its own climatology (or attractor). Third, when models are attracted to their biased climatology, full field assimilation will cause recurrent corrections of the model bias, and yield a transfer of bias to the non-observed variables (via the multivariate updates). All of these can lead to a slow degradation of the performance of the data assimilation system during the analysis period (Dee 2005). For anomaly assimilation we used the climatological reference period 1980–2010, a period that is sufficiently long for sampling the variability of the tropical Atlantic and during which there is enough data to estimate the observed climatology accurately. Note that for the hydrographic profiles (Gouretski and Reseghetti 2010), the climatological mean is calculated from the EN4 objective analysis (Good et al. 2013), because profile data are too sparse and heterogeneously distributed to estimate a trustworthy climatology. The monthly climatological mean of the model is estimated from the historical ensemble for the period 1980–2010. For CTRL it is calculated from 30 members, while for ACPL it is calculated from only 5 members.

We assimilate data every month and only update the ocean component. The other components (atmosphere, sea ice and land) adjust dynamically between the assimilation cycle. In the reanalysis and hindcasts, we apply the same external forcing as in the historical simulations. We assess the prediction skill based on hindcasts (i.e. retrospective predictions). Seasonal hindcasts start on the 15th of January, April, July and October each year during 1985–2010. In total, there are 104 hindcasts (26 years with 4 hindcasts per year). Each hindcast runs 9 realisations (ensemble members) for 13 months. Initial conditions are taken from the first 9 members of the 30 ensemble member reanalyses. Note that this choice has no influence on the results, because with the EnKF all members are equally likely.

3 Validation data sets and metrics

For validation of the reanalysis, we use the independent sea surface height (SSH) altimetry data and the EN4 objective analysis (Good et al. 2013). Monthly sea surface height (SSH) anomalies (i.e., sea level anomalies) are computed from the Global ARMOR3D L4 Reprocessed dataset (http://marine.copernicus.eu, available from 1993 to present)

produced by Ssalto/Duacs and distributed by AVISO with support from the Centre National D'Études Spatiales. The product is gridded to a resolution of 1° for comparisons with our coarser model data.

The EN4 objective analysis is not independent from our reanalysis because it uses the same raw observations (i.e. the EN4 hydrographic profiles, Gouretski and Reseghetti 2010) to construct the 4-dimensional reconstruction. However, this comparison is of interest because objective analysis and model reanalysis are different by construction. Objective analysis provides a 4D interpolation of the observations without dynamical constraints and reverts to climatology when no data are available—as every monthly estimate minimises the error locally in space and time based on the available observations. Model reanalysis on the contrary use available observations to correct the state of a dynamical model and provide a dynamical reconstruction (Murphy 1993). An advantage of reanalyses is that they propagate dynamically the improvements of sparse observations to the unobserved regions (Storto et al. 2019), but in a region where the biases are very large (such as in the tropical Atlantic) the growth of the model error within the assimilation cycle may become quantitatively larger than the benefit. The DA settings used in NorCPM favour dynamical consistency at the expenses of accuracy. Namely, (1) we use anomaly assimilation to minimise dynamical adjustment post assimilation; (2) we add a representativity error term to the observation error to ensure that the ensemble DA system remains reliable (i.e. ensemble spread matches the error of the ensemble mean); and (3) we use the k-factor formulation (Sakov et al. 2012) in which observational error is artificially inflated if the assimilation pushes the update beyond two times the ensemble spread.

For validating the hindcasts, we focus on the performance in predicting the ATL3 index (SST anomalies averaged over the region 20°W to 0° and 3°S and 3°N), which is a good indicator of the Atlantic Niño (Zebiak 1993). We have also analysed the performance of hindcasts for the ATL2 SST index (15°W–5°W, 3°S–3°N), an indicator of the Atlantic Niño II that occurs during November–December (Okumura and Xie 2006). The area of both indices is marked on Fig. 5. We benchmark the performance of the two predictions systems against the persistence forecast and hindcasts from the North American Multimodel Ensemble (Kirtman et al. 2014, https://www.earthsystemgrid.org/search.html?Project=NMME). Note that for comparing the skill of NorCPM with NMME, we use NOAA OI-SST V2 (Reynolds et al. 2002), which differs from the data set used for assimilation (HadISST2), making the comparison fairer. The validation was also repeated with other data set HADISSTV1.1 (Rayner et al. 2003) with similar results (not shown).

The NMME is a multi-model seasonal forecasting system that consists of several coupled climate models from US and

Table 1 Brief description of NorCPM and 13 NMME models used in this paper

Model	Institution	Hindcast period	Ensemble size	Lead times (months)	References
NorCPM	Bjerknes Centre for Climate Research	1985–2010	9	1–12	Counillon et al. (2016)
CMC1-CanCM3	Canadian Meteorological Centre (CMC)	1981–2010	10	1–12	Merryfield et al. (2013)
CMC1-CanCM4	CMC	1981–2010	10	1–12	Merryfield et al. (2013)
COLA-RSMAS-CCSM3	National Center for Atmospheric Research (NCAR)	1982–2017	6	1–12	Kirtman and Min (2009)
COLA-RSMAS-CCSM4	NCAR	1982–2017	10	1–12	Kirtman et al. (in preparation)
GFDL-CM2p1	Geophysical Fluid Dynamics Laboratory (GFDL)	1982–2012	10	1–12	Delworth et al. (2006) and Zhang et al. (2007)
GFDL-CM2p1-aer04	GFDL	1982–2017	10	1–12	Delworth et al. (2006) and Zhang et al. (2007)
GFDL-CM2p5-FLOR-A06	GFDL	1980–2017	12	1–12	Vecchi et al. (2014)
GFDL-CM2p5-FLOR-B01	GFDL	1980–2017	12	1–12	Vecchi et al. (2014)
IRI-ECHAM4p5-Anomaly	International Research Institute for Climate and Society (IRI)	1982–2012	12	1–8	DeWitt (2005)
IRI-ECHAM4p5-Direct	IRI	1982–2012	12	1–8	DeWitt (2005)
NASA-GMAO	National Aeronautics and Space Administration (NASA)	1981–2012	10	1–9	Vernieres et al. (2012)
NASA-GMAO-062012	NASA	1981–2017	12	1–9	Vernieres et al. (2012)
NCEP-CFSv2	National Centers for Environmental Prediction (NCEP)	1982–2010	24	1–10	Saha et al. (2014)

Canadian modelling centres. We select 13 NMME systems that provide SST hindcasts from 1985 to 2010 (Table 1). All NMME hindcasts start on the first day of each month and are 8–12 months long. The NorCPM hindcasts start 15 days earlier than the NMME hindcasts, but make use of monthly average to perform a centered analysis. As an example, our hindcasts starting on the 15th of April have assimilated monthly average April data and is compared to the hindcasts of NMME starting on the first of May; it will be referred to as May hindcast in the following for simplicity. The ensemble size ranges from 6 to 24 among the NMME models. Here, we use the first 9 ensemble members of each NMME model (except for CCSM3 that only provides 6 ensemble members) to have a comparable ensemble size to NorCPM. The NMME hindcast data are provided as monthly means with a horizontal resolution of $1° \times 1°$.

The performance of the model reanalyses and hindcasts are assessed by calculating anomaly correlation coefficient (ACC) and root mean square error (RMSE). Statistical significance is estimated by a Students t-test at the significance level of 5%. The degrees of freedom are estimated from the auto-correlation (equation 8.7, p.149 Von Storch and Zwiers 1999). We also assess the reliability of our ensemble prediction systems for the ATL3 index. Reliability refers to the property of the ensemble spread to match the accuracy of the ensemble mean (Eq. 1). There are advanced formulation for checking the reliability of an ensemble system using a probabilistic Attributes Diagrams (Corti et al. 2012), but here we will only assess the reliability of the total variance. When calculating the RMSE with imperfect observations, one should account for observational error variance in the reliability budget analysis (e.g., Sakov et al. 2012; Rodwell et al. 2016) and ensure that the error variance of the ensemble mean matches the sum of the variance of the ensemble (σ_{ens}) and the observation error variance (σ_{obs}). For this purpose, we use the HadISST2, because it provides a more accurate estimate of the observation error than the OI-SST V2, and it is the product assimilated in ACPL and CTRL reanalyses and thus is the data set for which the reliability relation should be satisfied:

$$\text{RMSE} = \sqrt{\sigma_{obs}^2 + \sigma_{ens}^2}. \tag{1}$$

4 Reanalysis

Here we investigate the accuracy of the reanalyses performed with CTRL and ACPL configurations of NorCPM that are used to initialise the hindcasts (Sect. 5) for the period 1985–2010. The correlations in the tropical Atlantic are substantially higher than in CTRL (Fig. 3). It should be noted that both reanalyses are able to constrain monthly

variability of upper 200 m ocean heat content over much of the globe (Fig. S2) and correlations are mostly very similar exception made of our area of interest. There are also some difference in the tropical Pacific: ACPL yields an improved representation of the variability in the western part, but causes a slight degradation in the central part, where the bias is amplified (see Figure S1).

The improvements in the tropical Atlantic from anomaly coupling are related to Atlantic Niño variability (Fig. 3). In particular, the signature of the Atlantic Niño is absent in the performance of CTRL reanalysis for upper ocean heat content, while it is marked by correlations exceeding 0.5 in ACPL. The discrepancies in correlation coefficient are also clearly visible in the reanalysis performance for subsurface temperature variability at the equator and for SSH. The tropical Atlantic can be approximated as a 1.5 layer system, (i.e., an active less dense layer over a much thicker and denser inactive layer), and SSH variations are closely related to the thickness of the upper layer and thermocline-depth variations (Wyrtki and Kendall 1967; Rebert et al. 1985). The pattern of correlation coefficient improvement in SSH in the tropical Atlantic and in temperature at 150–200 m depth in the eastern equatorial Atlantic are consistent with a better representation of the thermocline variations linked to Atlantic Niño variability in ACPL compared to CTRL. Correlation with the upper 200 m ocean salt content is also improved in ACPL (not shown). These results suggest that by improving the dynamical representation in our model, we are able to make more efficient use of the observations to provide a dynamical reconstruction. In the supplementary material, one can see that similar results are found for RMSE (Fig. S4), where we present the difference of correlation and RMSE to better highlight the improvements.

In the last row of Fig. 3, we present the seasonal variability of the averaged correlation and RMSE of the temperature in the upper 200 m of the ocean in the equatorial band. A more detailed verification can be found in the supplementary with the spatial and vertical correlation at the starting months of the four seasonal hindcasts (see Figure S5). ACPL reanalysis performs better than CTRL reanalysis from May to October, with the greatest improvements in August and September. The improvements coincide with ACPL's better representation of equatorial SST variability, while in CTRL there is excessive variability from July to November (Fig. 2). ACPL reanalysis performs slightly better in January and December, coinciding with the better representation of the secondary maximum in equatorial SST variability in ACPL. There are no improvements in February to April, when the simulated variability is weak in both models, in reasonable agreement with the observations.

Fig. 3 First row shows the pointwise de-seasoned (with the mean seasonal cycle removed) correlation of monthly 200 m ocean heat content of EN4 objective analysis with CTRL (**a**) and ACPL (**b**). The second row (**c, d**) shows the temperature correlation in the vertical along the tropical Atlantic band (between 3°S and 3°N), the third row (**e, f**) the same for SSH correlation. Correlations that are not significant (at significance level of 5%) are masked. The last row shows the averaged correlation and RMSE of temperature within the top 200 m in the equatorial band over calendar months (between 3°S and 3°N)
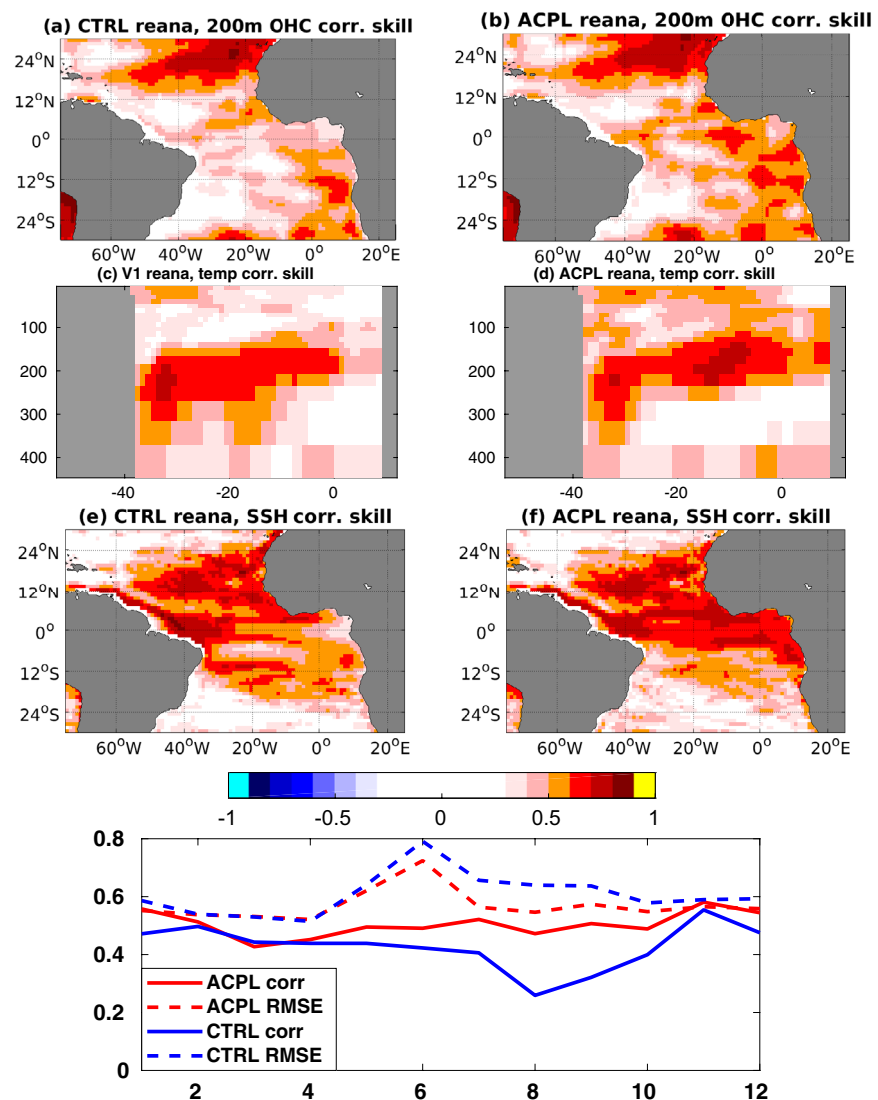
5 Hindcasts

The performance of ACPL and CTRL hindcasts in predicting the ATL3 SST index is assessed in terms of anomaly correlation and RMSE, and their skill is compared against that of the NMME forecasting systems (Fig. 4). Considering all four start dates (Fig. 4a), CTRL performs very poorly and is less skilful than the NMME systems and most of the time than persistence. ACPL improves on CTRL. Up to forecast lead month 3, ACPL shows skill comparable to the NMME system, though it does not beat persistence. From forecast lead month 4 onward, ACPL beats persistence and is as skilful as the best models of the NMME system. The improvements seem at least partially explained by the improved accuracy of the initial SST conditions (denoted with a square on the y-axes in Fig. 4 and calculated from the monthly averaged of the reanalyses). However, the skill of all the models is low, and no model has a correlation skill above 0.5 after month lead month 4. Next we investigate how the skill varies with start month.

For hindcasts started from February, ACPL achieves a correlation skill larger than 0.5 at 4 months lead time (i.e., in May), a skill that is better than all NMME systems. CTRL and NNME hindcasts do not beat persistence during the first 4 months. Thereafter the skill of all systems is low. Interestingly, the performance of the SST initial condition (marked with the squares) is indiscernible between CTRL and ACPL, but the accuracy of the initial 200 m heat content is improved in ACPL (see Fig. 3 and Fig. S5). A good subsurface initialisation can remerge to the surface when the shoaling of the thermocline occurs. The evolution of skill in ACPL may also be related to the behaviour of the model during the hindcast integration. The strong drop in skill of ACPL in June coincide with the boreal summer peak

Fig. 4 The first row shows the skill in predicting ATL3 SST in terms of **a** correlation and **b** RMSE for the different prediction systems considering all start dates. CTRL is indicated by the blue solid line, ACPL with the red solid line, persistence by the dashed black, and the different models comprising the NMME system by the other thin coloured lines. The following rows show the skill of hindcasts started in **c**, **d** February, **e**, **f** May, **g**, **h** August and **i**, **j** November. The squares on the y-axes indicate the accuracy of the monthly averaged ATL3 index in ACPL and CTRL reanalyses (i.e. lead month 0). The skill is calculated using NOAA OISST V2. A circle is added for CTRL and ACPL when correlations are significant at a 5% significance level. Legend for the NMME system is shown in Fig. S3
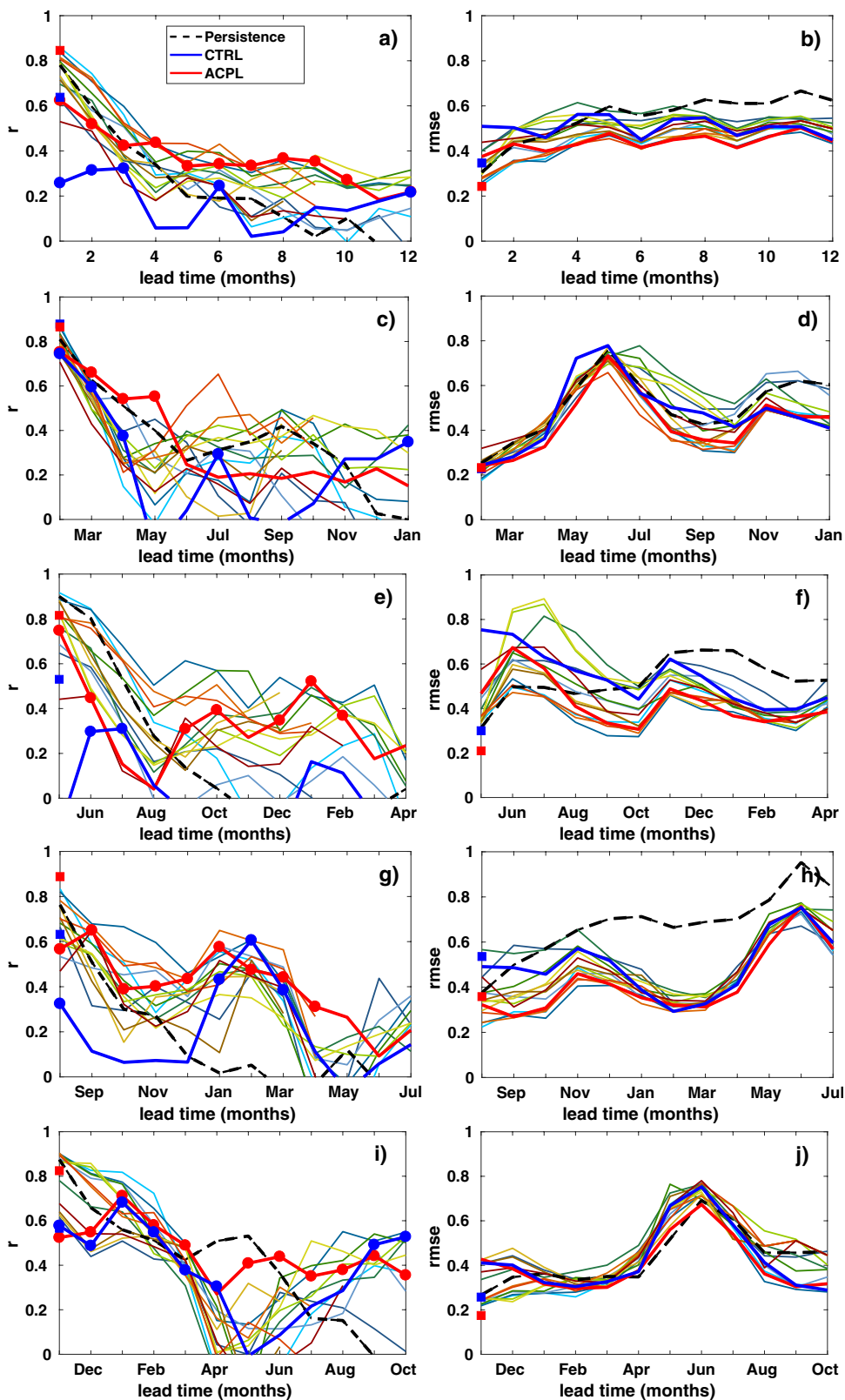
in equatorial SST variability (Fig. 2) and there is a rapid drop of skill in June for all start dates, which resemble the spring predictability barrier in May in the tropical Pacific.

Accordingly, for May starts, there is initially some large improvements for ACPL compared to CTRL prediction system, but the performance drops again very quickly and is

poorer than in most NMME systems and persistence. This is disappointing because June corresponds to the time when Atlantic equatorial variability and its impact are at their maximum. Also, one would expect to achieve skilful predictions when starting from May (as few NMME systems do). It should be acknowledged that the NMME systems achieving skill for May start (CMC1-CanCM3, CMC1-CanCM4, and NASA-GMAO) assimilate data in the atmospheric component. In NorCPM, we only initialise the ocean of these hindcasts on the 15th of April—i.e. 15 days prior to the effective start date (May 1st). We notice that ensemble members diverge rapidly during that time and that the variability of the ensemble mean flattens (not shown). This suggests that further initialisation of the atmospheric component and moving the assimilation closer to the effective start date can be important for improving prediction skill at this time of large internal atmospheric variability (Richter and Doi 2019). It is also possible that updating the ocean without the atmosphere component can introduce some imbalances (Penny and Hamill 2017), enhance the stochasticity of the atmosphere and accelerate the ensemble dispersion.

For the August starts, ACPL shows good improvement compared to CTRL for hindcasts. ACPL hindcasts sustain correlation skill between 0.4–0.6 up to 7 months lead time (February). Similarly, the RMSE of ATL3 in ACPL is reduced compared to CTRL. The skill relates to good subsurface initialisation that remerge to the surface when the second shoaling of the thermocline occurs, and it is tied to the seasonal cycle. The November–December upwelling is also associated with a shoaling thermocline and local weak intensification of the trade winds that drive upwelling. The thermocline anomalies partly originate from reflected Kelvin waves, and so from off-equatorial heat content anomalies (e.g. Fig. S5). Thus, the predictability comes from the propagation of anomalous heat content anomalies, which represent a modulation of the seasonal cycle and that become expressed as SST anomalies in November–February, when there is a local upwelling. In contrast, CTRL hindcasts have rapid drop in correlation skill during the first 2 months and this is followed by a recovery of skill to a level similar to ACPL at lead month 7. In CTRL, the upwelling seasons is very strong and delayed, this could explain why the SST anomalies are erroneous in the period before November. Many of the NMME models, ACPL and CTRL show relatively high skill in January and February, which indicates that this mechanism is highly predictable. The skill in February is larger in CTRL than in ACPL. In NorCPM, there is an overly strong ENSO teleconnection with equatorial Atlantic in February. In observations the correlation between in Nino3.4 and ATL3 is around 0.2 in February while in CTRL it is 0.4 and in ACPL it is 0.5. As a result, in NorCPM there is a spurious ENSO signal superposed on the observed low-frequency signal, degrading the ACPL prediction for

ATL3 more so than CTRL. It is interesting to notice that beyond June month, RMSE is reducing with lead time. This occurs for all start dates, and can be seen as a consequence of interannual variability getting weaker in these seasons.

For November starts, the initial SST condition is improved in ACPL compared to CTRL, but the performance of the two systems quickly converge and are similar to that of the NMME, until April of the following year—suggesting that the system is dynamically relatively stable prior to May. The highest skill of all systems is found in January–February as a reemergence of subsurface anomaly. In May and June, ACPL system achieves better performance than that of NMME systems and CTRL, and its skill approaches that achieved from May start.

It is encouraging that ACPL system shows some skill in predicting the second equatorial SST variability maximum occurring in November–December from August starts. This variability is related to the Atlantic Niño phenomenon, which is centred on the ATL2 region (within 15°W–5°W; 3°S–3°N, Okumura and Xie 2006). We further assess the skill of August started hindcasts in predicting such events. ACPL hindcasts capture the variability of the ATL2 SST index in November–December quite well with a correlation of around 0.5. We do not show the skill in predicting the ATL2 index as a function of lead time, as ATL2 and ATL3 are very similar and prediction skill is nearly identical to the one shown in Fig. 4g. For comparison with (Okumura and Xie 2006), we do show predictions of the ATL2 index in November–December (left panel of Fig. 5). ACPL is substantially more skilful than CTRL that has correlation skill around 0. The difference in correlation skill between ACPL and CTRL for November–December average calculated against NOAA OISST shows that ACPL improves the representation of the variability in the equatorial band (right panel of Fig. 5).

In Fig. 6, we show the evolution of the reliability (meaning satisfying Eq. 1) of the ensemble prediction systems as function of lead month for the ATL3 SST index. The standard deviation of observation error is about 0.1 °C. In both systems the estimated error is slightly lower than the RMSE. This is typical with the ensemble DA method, because DA assumes that models are unbiased (Dee 2005). This assumption is not satisfied even with anomaly assimilation, because it only correct the climatological bias while assimilation assumes that any error type is random, which can also influence the higher moments (e.g. variance, skewness, kurtosis). Breaking this assumption results in a spurious reduction of the ensemble spread (Anderson 2001; Raanes et al. 2019). The dispersion is much improved in ACPL compared to CTRL, in particular at analysis time when discrepancies between RMSE and estimated error is marginal. It is reassuring that assimilation with a prediction system that uses a model with
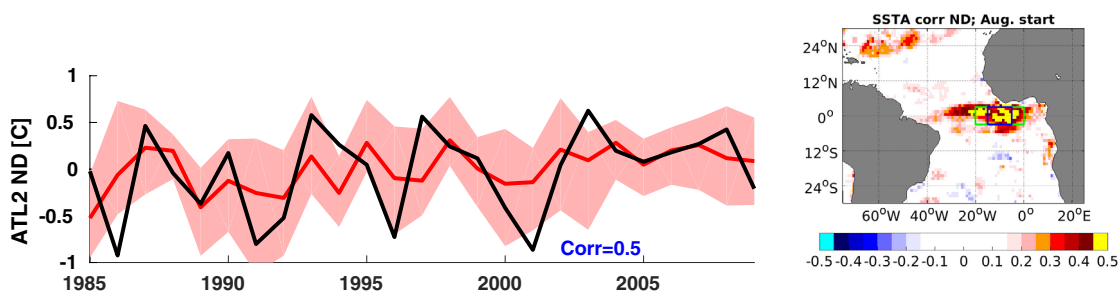
Fig. 5 Left is the time series of the ATL2 in November–December from NOAA OISST (in black) and from ACPL hindcasts started in August (red line, ensemble mean); the shading represents the ensemble envelope (min and max). Right is the pointwise correlation differ-ence between ACPL and CTRL for November–December predicted SST. The green box shows the ATL3 region and the blue box shows the ATL2 region
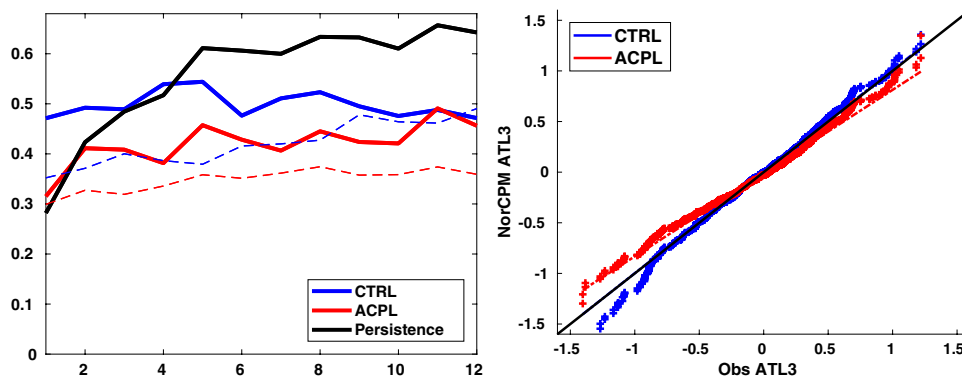

Fig. 6 In the left panel we assess the reliability of the ensemble prediction systems for the ATL3 index using HadISST2. The RMSE of persistence, ACPL and CTRL hindcasts are plotted with plain lines at different lead months. The dashed red and blue lines represent the estimated error (i.e. $\sqrt{\sigma_{ens}^2 + \sigma_{obs}^2}$, where σ_{ens} is the ensemble spread and σ_{obs} is the observation error). In the right-hand panel, we show the quantile–quantile graph of ATL3 predictions (all lead times) against observations

reduced bias improves this property. However, the ensemble spread does not grow at the same pace as the RMSE suggesting that the climatological correction term causes an artificial damping of the variability. By contrast, the RMSE in CTRL at the first lead month is not lower than in the rest lead month while the spread is reduced by assimilation. This suggests that assimilation in CTRL does not reduce substantially the error compared to climatological level but mostly reduces the ensemble spread. At 12 month lead time, the reliability is recovered in CTRL.

In Fig. 6, we assess how the probability density function of the ATL3 in ACPL and CTRL are aligned with the observations using a quantile–quantile plot of the index. In CTRL prediction system, the regression line is very well aligned although it tends to slightly overestimate the low values of ATL3. In ACPL, the prediction system underestimates the low values and the high value causing a tilt in the regression lines. We find very comparable results when breaking down the analysis for the different seasonal hindcasts or for hindcasts at different lead time (not shown).

6 Conclusions

Coupled Earth system models have very large biases and the impact of these on the prediction skill is yet to be determined. Here we investigate this issue for the tropical Atlantic, because this is a region where current models exhibit large biases compared to the amplitude of inter-annual variability and have demonstrated low prediction skill. Furthermore, several previous studies have reported an influence of model biases on the simulated variability and prediction skill in this region (Ding et al. 2015a; Dippe et al. 2018, 2019).

To investigate this question, we have used the anomaly coupled method of Toniazzo and Koseki (2018) in the NorESM model. The method has been shown to effectively reduce the systematic biases in the simulated tropical climatology of SST, wind stress and precipitation (Toniazzo and Koseki 2018). We compare performance of reanalyses and hindcasts with two versions of the NorCPM, one

using the standard version of NorESM (CTRL) and one using the anomaly coupled version of NorESM (ACPL). With both configurations, we assimilate SST and ocean temperature and salinity profile data with the EnKF, and we perform 12 month long seasonal predictions, with 9 ensemble members, and initiated once per season. We focus on the period 1985–2010, and we compare of prediction skill against that of the North American Multi-model Ensemble.

We show that reducing the climatological biases in the tropical Atlantic improves the ocean reanalysis. Using a model with reduced bias allows the data assimilation to find a better dynamical reconstruction that satisfies the observations and their likelihood. The benefits are largest from June to November when simulated equatorial variability is most improved by anomaly coupling.

We also show that the performance of the prediction system for the Atlantic Niño index (ATL3) is greatly enhanced in ACPL compared to CTRL. Beyond 3 months lead time, ACPL hindcasts reaches skill comparable to the better NMME systems and beats persistence. We show that the skill is most improved when the simulated variability and the initial condition are improved, i.e. during the second half of the calendar year. For certain lead times and validity dates, notable in May from February initialisation, the ACPL system shows better skill than the entire NMME system. The skill however rapidly drops in June, which corresponds to the start date with largest potential impact but also to the most challenging one. We speculate that poor skill at that times is related to the poor accuracy of the atmospheric state in our system, which is not initialised. We plan to move the assimilation close to the effective start date of the hindcast and further initialise the atmospheric state.

The reliability of the ensemble prediction system is well improved initially with ACPL, as the perfect model hypothesis in the data assimilation better holds. However, the reliability degrades with integration time as ACPL causes an artificial damping of the variability. In fact, the distribution of Atlantic Niño with ACPL is found to underestimate the low and high values.

References

Adler RF, Huffman GJ, Chang A, Ferraro R, Xie PP, Janowiak J, Rudolf B, Schneider U, Curtis S, Bolvin D et al (2003) The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present). J Hydrometeorol 4(6):1147–1167

Anderson J (2001) An ensemble adjustment Kalman filter for data assimilation. Mon Weather Rev 129(12):2884–2903

Barreiro M, Chang P, Ji L, Saravanan R, Giannini A (2005) Dynamical elements of predicting boreal spring tropical Atlantic sea-surface temperatures. Dyn Atmos Oceans 39(1–2):61–85

Bentsen M, Bethke I, Debernard JB, Iversen T, Kirkevåg A, Seland Ø, Drange H, Roelandt C, Seierstad IA, Hoose C, Kristjánsson JE (2012) The Norwegian Earth system model, NorESM1-M—part 1: description and basic evaluation. Geosci Model Dev Discuss 5:2843–2931. https://doi.org/10.5194/gmdd-5-2843-2012

Bleck R, Rooth C, Hu D, Smith LT (1992) Salinity-driven thermocline transients in a wind- and thermohaline-forced isopycnic coordinate model of the North Atlantic. J Phys Oceanogr 22:1486–1505

Brandt P, Funk A, Hormann V, Dengler M, Greatbatch RJ, Toole JM (2011) Interannual atmospheric variability forced by the deep equatorial Atlantic Ocean. Nature 473(7348):497–500

Corti S, Weisheimer A, Palmer T, Doblas-Reyes F, Magnusson L (2012) Reliability of decadal predictions. Geophys Res Lett. https://doi.org/10.1029/2012GL053354

Counillon F, Bethke I, Keenlyside N, Bentsen M, Bertino L, Zheng F (2014) Seasonal-to-decadal predictions with the ensemble Kalman filter and the Norwegian Earth System Model: a twin experiment. Tellus A 66:1. https://doi.org/10.3402/tellusa.v66.21074

Counillon F, Keenlyside N, Bethke I, Wang Y, Billeau S, Shen ML, Bentsen M (2016) Flow-dependent assimilation of sea surface temperature in isopycnal coordinates with the Norwegian climate prediction model. Tellus A: Dyn Meteorol Oceanogr 68(1):32437

Davey M, Huddleston M, Sperber K, Braconnot P, Bryan F, Chen D, Colman R, Cooper C, Cubasch U, Delecluse P et al (2002) STOIC: a study of coupled model climatology and variability in tropical ocean regions. Clim Dyn 18(5):403–420

Dee DP (2005) Bias and data assimilation. Q J R Meteorol Soc 131(613):3323–3343

de la Vara A, Cabos W, Sein DV, Sidorenko D, Koldunov NV, Koseki S, Soares PM, Danilov S (2020) On the impact of atmospheric vs oceanic resolutions on the representation of the sea surface temperature in the south eastern tropical Atlantic. Clim Dyn 54:4733–4757. https://doi.org/10.1007/s00382-020-05256-9

Delworth TL, Broccoli AJ, Rosati A, Stouffer RJ, Balaji V, Beesley JA, Cooke WF, Dixon KW, Dunne J, Dunne K et al (2006) GFDL's CM2 global coupled climate models. Part I: formulation and simulation characteristics. J Clim 19(5):643–674

Deppenmeier AL, Haarsma RJ, Hazeleger W (2016) The Bjerknes feedback in the tropical Atlantic in CMIP5 models. Clim Dyn 47(7–8):2691–2707

DeWitt DG (2005) Retrospective forecasts of interannual sea surface temperature anomalies from 1982 to present using a directly coupled atmosphere-ocean general circulation model. Mon Weather Rev 133(10):2972–2995

Ding H, Keenlyside NS, Latif M (2010) Equatorial atlantic interannual variability: role of heat content. J Geophys Res Oceans. https://doi.org/10.1029/2010JC006304

Ding H, Greatbatch RJ, Latif M, Park W (2015a) The impact of sea surface temperature bias on equatorial Atlantic interannual variability in partially coupled model experiments. Geophys Res Lett 42:5540–5546

Ding H, Keenlyside N, Latif M, Park W, Wahl S (2015b) The impact of mean state errors on equatorial Atlantic interannual variability in a climate model. J Geophys Res: Oceans 120(2):1133–1151

Dippe T, Greatbatch RJ, Ding H (2018) On the relationship between Atlantic niño variability and ocean dynamics. Clim Dyn 51(1–2):597–612

Dippe T, Greatbatch RJ, Ding H (2019) Seasonal prediction of equatorial Atlantic sea surface temperature using simple initialization and bias correction techniques. Atmos Sci Lett 20(5):e898

Evensen G (2003) The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dyn 53(4):343–367

Foltz GR, Brandt P, Richter I, Rodriguez-fonseca B, Hernandez F, Dengler M, Rodrigues RR, Schmidt JO, Yu L, Lefevre N et al (2019) The tropical Atlantic observing system. Front Mar Sci 6:206

Gent P, Danabasoglu G, Donner L, Holland M, Hunke E, Jayne S, Lawrence D, Neale R, Rasch P, Vertenstein M, Worley PH, Yang Z, Zhang M (2011) The community climate system model version 4. J Clim 24(19):4973–4991

Good SA, Martin MJ, Rayner NA (2013) EN4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. J Geophys Res: Oceans 118(12):6704–6716

Gouretski V, Reseghetti F (2010) On depth and temperature biases in bathythermograph data: development of a new correction scheme based on analysis of a global ocean database. Deep Sea Res Part I: Oceanogr Res Pap 57(6):812–833

Harlaß J, Latif M, Park W (2018) Alleviating tropical Atlantic sector biases in the Kiel climate model by enhancing horizontal and vertical atmosphere model resolution: climatology and interannual variability. Clim Dyn 50(7–8):2605–2635

Jansen MF, Dommenget D, Keenlyside N (2009) Tropical atmosphere–ocean interactions in a conceptual framework. J Clim 22(3):550–567

Jouanno J, Hernandez O, Sanchez-Gomez E (2017) Equatorial Atlantic interannual variability and its relation to dynamic and thermodynamic processes. Earth Syst Dyn 8(4):1061–1069

Karspeck AR (2016) An ensemble approach for the estimation of observational error illustrated for a nominal 1 global ocean model. Mon Weather Rev 144(5):1713–1728

Keenlyside NS, Latif M (2007) Understanding equatorial Atlantic interannual variability. J Clim 20(1):131–142

Kirkevåg A, Iversen T, Seland Ø, Hoose C, Kristjánsson J, Struthers H, Ekman AM, Ghan S, Griesfeller J, Nilsson ED et al (2013) Aerosol–climate interactions in the Norwegian Earth System Model—NorESM1-M. Geosci Model Dev 6(1):207–244

Kirtman BP, Min D (2009) Multimodel ensemble ENSO prediction with CCSM and CFS. Mon Weather Rev 137(9):2908–2930

Kirtman BP, Shukla J, Huang B, Zhu Z, Schneider EK (1997) Multiseasonal predictions with a coupled tropical ocean-global atmosphere system. Mon Weather Rev 125(5):789–808

Kirtman BP, Min D, Infanti JM, Kinter JL III, Paolino DA, Zhang Q, Van Den Dool H, Saha S, Mendez MP, Becker E et al (2014) The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. Bull Am Meteorol Soc 95(4):585–601

Koseki S, Keenlyside N, Demissie T, Toniazzo T, Counillon F, Bethke I, Ilicak M, Shen ML (2018) Causes of the large warm bias in the Angola–Benguela Frontal Zone in the Norwegian Earth system model. Clim Dyn 50(11–12):4651–4670

Losada T, Rodríguez-Fonseca B, Polo I, Janicot S, Gervois S, Chauvin F, Ruti P (2010) Tropical response to the Atlantic equatorial mode: AGCM multimodel approach. Clim Dyn 35(1):45–52

Lübbecke JF, McPhaden MJ (2013) A comparative stability analysis of Atlantic and Pacific Niño modes. J Clim 26(16):5965–5980

Lübbecke JF, McPhaden MJ (2017) Symmetry of the Atlantic Niño mode. Geophys Res Lett 44(2):965–973

Lübbecke JF, Rodríguez-Fonseca B, Richter I, Martín-Rey M, Losada T, Polo I, Keenlyside NS (2018) Equatorial Atlantic variability—modes, mechanisms, and global teleconnections. Wiley Interdiscip Rev: Clim Change 9(4):e527

Merryfield WJ, Lee WS, Boer GJ, Kharin VV, Scinocca JF, Flato GM, Ajayamohan R, Fyfe JC, Tang Y, Polavarapu S (2013) The Canadian seasonal to interannual prediction system. Part I: models and initialization. Mon Weather Rev 141(8):2910–2945

Milinski S, Bader J, Haak H, Siongco AC, Jungclaus JH (2016) High atmospheric horizontal resolution eliminates the wind-driven coastal warm bias in the southeastern tropical Atlantic. Geophys Res Lett 43:10455–10462

Murphy AH (1993) What is a good forecast? An essay on the nature of goodness in weather forecasting. Weather Forecast 8(2):281–293

Neale RB, Chen CC, Gettelman A, Lauritzen PH, Park S, Williamson DL, Conley AJ, Garcia R, Kinnison D, Lamarque JF, et al (2010) Description of the NCAR community atmosphere model (CAM 4.0). NCAR Tech Note NCAR/TN-486+ STR

Neelin JD, Dijkstra HA (1995) Ocean-atmosphere interaction and the tropical climatology. Part I: the dangers of flux correction. J Clim 8(5):1325–1342

Nnamchi HC, Li J, Kucharski F, Kang IS, Keenlyside NS, Chang P, Farneti R (2015) Thermodynamic controls of the Atlantic Niño. Nat Commun 6(1):1–10

Okumura Y, Xie SP (2006) Some overlooked features of tropical Atlantic climate leading to a new Niño-like phenomenon. J Clim 19(22):5859–5874

Penny SG, Hamill TM (2017) Coupled data assimilation for integrated earth system analysis and prediction. Bull Am Meteorol Soc 98(7):ES169–ES172

Raanes PN, Bocquet M, Carrassi A (2019) Adaptive covariance inflation in the ensemble Kalman filter by Gaussian scale mixtures. Q J R Meteorol Soc 145(718):53–75

Rayner N, Parker DE, Horton E, Folland CK, Alexander LV, Rowell D, Kent E, Kaplan A (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. J Geophys Res: Atmos 108(D14):4407. https://doi.org/10.1029/2002JD002670

Rebert JP, Donguy JR, Eldin G, Wyrtki K (1985) Relations between sea level, thermocline depth, heat content, and dynamic height in the tropical Pacific Ocean. J Geophys Res: Oceans 90(C6):11719–11725

Reynolds RW, Rayner NA, Smith TM, Stokes DC, Wang W (2002) An improved in situ and satellite SST analysis for climate. J Clim 15(13):1609–1625

Richter I, Doi T (2019) Estimating the role of SST in atmospheric surface wind variability over the tropical Atlantic and Pacific. J Clim 32(13):3899–3915

Richter I, Xie SP (2008) On the origin of equatorial Atlantic biases in coupled general circulation models. Clim Dyn 31(5):587–598

Richter I, Xie SP, Wittenberg AT, Masumoto Y (2012) Tropical Atlantic biases and their relation to surface wind stress and terrestrial precipitation. Clim Dyn 38(5–6):985–1001

Richter I, Behera SK, Masumoto Y, Taguchi B, Sasaki H, Yamagata T (2013) Multiple causes of interannual sea surface temperature variability in the equatorial Atlantic ocean. Nat Geosci 6(1):43–47

Richter I, Behera SK, Doi T, Taguchi B, Masumoto Y, Xie SP (2014a) What controls equatorial Atlantic winds in boreal spring? Clim Dyn 43(11):3091–3104

Richter I, Xie SP, Behera SK, Doi T, Masumoto Y (2014b) Equatorial Atlantic variability and its relation to mean state biases in CMIP5. Clim Dyn 42(1–2):171–188

Richter I, Doi T, Behera SK, Keenlyside N (2018) On the link between mean state biases and prediction skill in the tropics: an atmospheric perspective. Clim Dyn 50(9–10):3355–3374

Rodwell M, Lang S, Ingleby N, Bormann N, Hólm E, Rabier F, Richardson D, Yamaguchi M (2016) Reliability in ensemble data assimilation. Q J R Meteorol Soc 142(694):443–454

Saha S, Moorthi S, Wu X, Wang J, Nadiga S, Tripp P, Behringer D, Hou YT, Chuang Hy, Iredell M et al (2014) The NCEP climate forecast system version 2. J Clim 27(6):2185–2208

Sakov P, Counillon F, Bertino L, Lisæter K, Oke P, Korablev A (2012) TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic. Ocean Sci 8:633–656

Sausen R, Barthel K, Hasselmann K (1988) Coupled ocean–atmosphere models with flux correction. Clim Dyn 2(3):145–163

Small RJ, Curchitser E, Hedstrom K, Kauffman B, Large WG (2015) The Benguela upwelling system: quantifying the sensitivity to resolution and coastal wind representation in a global climate model. J Clim 28:9409–9432

Stockdale TN, Balmaseda MA, Vidard A (2006) Tropical Atlantic SST prediction with coupled ocean–atmosphere GCMs. J Clim 19(23):6047–6061

Stockdale TN, Anderson DL, Balmaseda MA, Doblas-Reyes F, Ferranti L, Mogensen K, Palmer TN, Molteni F, Vitart F (2011) ECMWF seasonal forecast system 3 and its prediction of sea surface temperature. Clim Dyn 37(3–4):455–471

Storto A, Alvera-Azcárate A, Balmaseda MA, Barth A, Chevallier M, Counillon F, Domingues CM, Drevillon M, Drillet Y, Forget G et al (2019) Ocean reanalyses: recent advances and unsolved challenges. Front Mar Sci 6:418

Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. Bull Am Meteorol Soc 93(4):485

Tjiputra J, Roelandt C, Bentsen M, Lawrence D, Lorentzen T, Schwinger J, Seland Ø, Heinze C (2013) Evaluation of the carbon cycle components in the Norwegian Earth System Model (NorESM). Geosci Model Dev 6(2):301–325

Toniazzo T, Koseki S (2018) A methodology for anomaly coupling in climate simulation. J Adv Model Earth Syst 10(8):2061–2079

Toniazzo T, Woolnough S (2014) Development of warm SST errors in the southern tropical Atlantic in CMIP5 decadal hindcasts. Clim Dyn 43(11):2889–2913

Vecchi GA, Delworth T, Gudgel R, Kapnick S, Rosati A, Wittenberg AT, Zeng F, Anderson W, Balaji V, Dixon K et al (2014) On the seasonal forecasting of regional tropical cyclone activity. J Clim 27(21):7994–8016

Vernieres G, Rienecker MM, Kovach R, Keppenne CL (2012) The GEOS-iODAS: description and evaluation. NASA technical report, NASA/TM-2012-104606. http://gmao.gsfc.nasa.gov/pubs/docs/Vernieres589.pdf. Accessed 9 Jan 2020

Vertenstein M, Craig T, Middleton A, Feddema D, Fischer C (2012) CESM1.0.3 user guide. http://www.cesm.ucar.edu/models/cesm1.0/cesm/cesmdoc104/ug.pdf. Accessed 9 Jan 2020

Voldoire A, Exarchou E, Sanchez-Gomez E, Demissie T, Deppenmeier AL, Frauen C, Goubanova K, Hazeleger W, Keenlyside N, Koseki S et al (2019) Role of wind stress in driving SST biases in the tropical Atlantic. Clim Dyn 53(5–6):3481–3504

Von Storch H, Zwiers FW (1999) Estimation, pp. 79–93. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. https://doi.org/10.1017/CBO9780511612336,1999

Wahl S, Latif M, Park W, Keenlyside N (2011) On the tropical Atlantic SST warm bias in the Kiel climate model. Clim Dyn 36:891–906

Wang Y, Counillon F, Bertino L (2016) Alleviating the bias induced by the linear analysis update with an isopycnal ocean model. Q J R Meteorol Soc 142(695):1064–1074

Wang Y, Counillon F, Bethke I, Keenlyside N, Bocquet M, Shen Ml (2017) Optimising assimilation of hydrographic profiles into isopycnal ocean models with ensemble data assimilation. Ocean Model 114:33–44

Wang Y, Counillon F, Keenlyside N, Svendsen L, Gleixner S, Kimmritz M, Dai P, Gao Y (2019) Seasonal predictions initialised by assimilating sea surface temperature observations with the EnKF. Clim Dyn 53(9–10):5777–5797

Wyrtki K, Kendall R (1967) Transports of the Pacific equatorial countercurrent. J Geophys Res 72(8):2073–2076

Xu Z, Chang P, Richter I, Kim W, Tang G (2014a) Diagnosing southeast tropical Atlantic SST and ocean circulation biases in the CMIP5 ensemble. Clim Dyn 43:3123–3145

Xu Z, Li M, Patricola CM, Chang P (2014b) Oceanic origin of southeast tropical Atlantic biases. Clim Dyn 43(11):2915–2930

Zebiak SE (1993) Air–sea interaction in the equatorial Atlantic region. J Clim 6(8):1567–1586

Zhang S, Harrison M, Rosati A, Wittenberg A (2007) System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. Mon Weather Rev 135(10):3541–3564

Zuidema P, Chang P, Medeiros B, Kirtman BP, Mechoso R, Schneider EK, Toniazzo T, Richter I, Small RJ, Bellomo K et al (2016) Challenges and prospects for reducing coupled climate model SST biases in the eastern tropical Atlantic and Pacific oceans: the US CLIVAR Eastern Tropical Oceans Synthesis Working Group. Bull Am Meteorol Soc 97(12):2305–2328