Tropical Atlantic SST Prediction with Coupled Ocean–Atmosphere GCMs

TIMOTHY N. STOCKDALE, MAGDALENA A. BALMASEDA, AND ARTHUR VIDARD

ECMWF, Shinfield Park, Reading, United Kingdom

(Manuscript received 11 March 2005, in final form 15 November 2005)

ABSTRACT

Variations in tropical Atlantic SST are an important factor in seasonal forecasts in the region and beyond. An analysis is given of the capabilities of the latest generation of coupled GCM seasonal forecast systems to predict tropical Atlantic SST anomalies. Skill above that of persistence is demonstrated in both the northern tropical and equatorial Atlantic, but not farther south. The inability of the coupled models to correctly represent the mean seasonal cycle is a major problem in attempts to forecast equatorial SST anomalies in the boreal summer. Even when forced with observed SST, atmosphere models have significant failings in this area. The quality of ocean initial conditions for coupled model forecasts is also a cause for concern, and the adequacy of the near-equatorial ocean observing system is in doubt. A multimodel approach improves forecast skill only modestly, and large errors remain in the southern tropical Atlantic. There is still much scope for improving forecasts of tropical Atlantic SST.

1. Introduction and motivation

On a global scale, it is the ENSO-related SST variability in the Pacific that is the biggest single driver of seasonally averaged climate anomalies. In comparison to ENSO, Atlantic SST variability is typically weaker, and is often given little attention in "global" seasonal forecast systems. Yet Atlantic SST variability is by no means negligible and can have a substantial impact on the atmosphere and on seasonal weather patterns. This impact is most visible on the regional scale (e.g., the influence of Atlantic SST on Brazilian Nordeste rainfall), but the influence of Atlantic SST anomalies can extend out of the Tropics, where it combines with forcing from the Pacific and Indian Oceans. In the case of the North Atlantic sector, the contribution from the Atlantic is often substantial (Mathieu et al. 2004).

In this paper we focus on our present capabilities to predict tropical Atlantic SST anomalies and the implications this has for seasonal prediction of climate anomalies in the Atlantic sector. Of course, an important part of what is predictable in Atlantic climate is driven from outside. Here we assume that these other factors are either already handled reasonably well (possibly true for some ENSO influences) or will be pursued elsewhere. There is a specific issue of teleconnections in coupled models, namely, how accurately signals propagate from a correctly represented tropical source outside the Atlantic, given the errors that exist in the mean state of coupled models. We will not examine this in detail, but the level of success (or otherwise) is implicitly included in our examination of SST prediction skill in ENSO-affected regions such as the north tropical and equatorial Atlantic.

The predictability of SST varies across the Atlantic, due to the variety of mechanisms that are operating. Equally, the physical importance of the SST in influencing climate anomalies varies, depending on the region of SST concerned. Our knowledge of exactly which regions of SST are most important in a coupled prediction system is still incomplete, but certain areas and seasons of established importance are given prominence in this paper. In particular, the interhemispheric SST gradient is important for March–May (MAM) precipitation in Northeast Brazil, while the equatorial SSTs are important for June–August (JJA) rainfall in West Africa. The paper by Kushnir et al. (2006) gives a good overview of the observed variability, which we are trying to predict, and the physical mechanisms known to be involved. Rather than provide a full discussion of the scientific background to our paper, we advise the reader to look through Kushnir et al. (2006).

In this paper we first give a brief review of coupled models used for predicting Atlantic SST. We then con-

Corresponding author address: Dr. T. Stockdale, ECMWF, Shinfield Park, Reading RG2 9AX, United Kingdom.
E-mail: T.Stockdale@ecmwf.int

© 2006 American Meteorological Society

sider the elements that are likely to be needed for a good numerical forecasting system, namely, a good model (section 3) and good initial conditions (section 4). In section 5, the actual skill and properties of coupled GCM SST forecasts are then examined and interpreted in the context of the quality of the models and initial conditions. Finally, in section 6 we summarize the state of the art of numerical SST forecasts in the Atlantic, and provide some tentative conclusions regarding the actual predictability that exists and the observing systems that may be needed to realize the potential predictability.

2. Overview of existing coupled prediction systems for the tropical Atlantic

Relatively few coupled models have been used to investigate predictability and prediction skill specifically in the Atlantic sector. Much seasonal prediction work with coupled models has used basinwide models of the Pacific or Indo–Pacific with Atlantic SSTs simply being specified. Indeed, many operational seasonal forecasting systems today are still using empirical methods to generate an SST forecast for the Atlantic basin. As will be clear by the end of this paper, such a strategy is not unreasonable, given the challenges involved in trying to get coupled model forecast systems to work.

Zebiak (1993) used a simplified coupled model together with data analysis to establish that coupled interaction in the equatorial Atlantic is a source of interannual variability. Compared to the Pacific the variability is much weaker (and subcritical with regard to self-sustaining oscillations, according to Zebiak's estimate), and SST variability is clearly driven by many other factors besides. As far as we are aware, this Atlantic model has not been used for further studies into predictability or prediction of the component of SST variability that it claims to identify. Chang et al. (2003) have developed a coupled model consisting of an atmospheric GCM and an ocean mixed-layer model. This ignores the role of ocean dynamics but allows study of the role of thermodynamic coupled processes on Atlantic SST predictability and prediction. The remote influence of ENSO and local thermodynamic processes both contribute toward predictive skill in their system, but they only considered forecasts of the north tropical Atlantic since equatorial forecasts showed no skill. Huang et al. (2002) used a coupled GCM to clarify the relative roles of ENSO and local coupling in driving Atlantic SST variability, showing that much of the variability in the equatorial and southern Atlantic is locally produced rather than ENSO forced; but they did not attempt prediction studies.

One interesting observation is the fair degree of correlation between interannual SST variability on the equator in the Atlantic and SST variability farther south. This correlation is seasonally varying and is most evident in the boreal summer. What this tells us about mechanisms and, thus, prediction strategies is unclear. Do seasonal/interannual cold or warm events in the south and equatorial Atlantic typically originate with subsurface anomalies in the western ocean propagating along the equator and then southwards along the coast, or are other mechanisms more important such as those that may derive predictability essentially from the SST field? Previous ocean modeling studies such as that by Carton et al. (1996) established that equatorial SST anomalies were dominated by equatorial wind variability, but for a coupled system this leaves much unresolved. Off-equatorial SST anomalies are normally ascribed to latent heat anomalies driven by wind variations, and on decadal time scales this is argued to be a possible mechanism for variability (Chang et al. 1997), but the exact roles of heat flux–SST feedbacks, cloud feedbacks, and teleconnections on the interannual variability remain to be established.

Several centers use coupled global seasonal forecasting systems for real-time forecasting, for example, the European Centre for Medium-Range Weather Forecasts (ECMWF), the Met Office (UKMO), and the National Aeronautics and Space Administration (NASA) Seasonal-to-Interannual Prediction Project (NSIPP), but to date their performance in the tropical Atlantic has not been documented. Another notable source of coupled GCM forecasts for the Atlantic sector are integrations from the European Union–funded Development of a European Multimodel Ensemble System for Seasonal to Internannual Prediction (DEMETER) project, which has run seasonal forecasts with a set of seven different global coupled models. This gives a rich dataset for investigating Atlantic predictability and assessing the model dependence of the results. In this paper we present results from the ECMWF operational forecasting system (Anderson et al. 2003) and from the DEMETER models (Palmer et al. 2004). In the case of DEMETER, we choose to look at only six of the seven models. This is because the quality of the Atlantic SST forecasts depends strongly on the wind field that was used to initialize the ocean component of the model. For the six models this wind field was provided by the 40-yr ECMWF Re-Analysis (ERA-40) Project (Uppala et al. 2005), which gives a reasonable specification of the winds. The seventh model instead used winds derived from an atmosphere GCM forced by observed SSTs. In the equatorial Atlantic unforced variability in the winds is important, and this does not give

an adequate basis for model initialization. The six DEMETER models have forecasts available on four start dates per year from 1982 to 2001. Each model has a nine member ensemble. Six of these ensemble members use ocean initial conditions prepared with the addition of wind perturbations in the surface forcing, simulating the uncertainty in the wind. Although these perturbed initial conditions represent an important part of the forecast uncertainty, they give less accurate forecasts on average, and in some of our analyses we take only the three "unperturbed" ensemble members per model, giving an 18-member multimodel ensemble. Some DEMETER integrations, for example, with the ECMWF model, covered much longer periods, and we use these where appropriate.

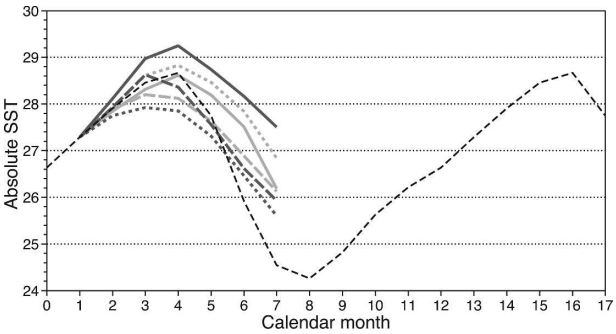

FIG. 1. The mean seasonal cycle of SST in the ATL3 region (3°N–3°S, 20°W–0°) from six coupled GCMs run in forecast mode from 1 February (thicker lines) compared to the observed mean seasonal cycle of SST (thin-dashed line). All models fail to reproduce properly the rapid cooling observed in June and July.

3. The ability of coupled GCMs to simulate the mean climate of the Atlantic sector

The first challenge for a coupled GCM prediction system is to produce a reasonable simulation of the mean state of the Atlantic sector. Moderate levels of error will not destroy all predictive skill—in the case of ENSO variability in the Pacific, for example, we know that coupled GCMs can give useful forecasts despite significant errors. The interplay between forecast errors, mean state error, and the model failings, which produce the mean state error, is typically complex and dependent on the physical mechanisms giving rise to predictability. Nonetheless, a simple a priori starting point is that the greater the errors in the model mean state, the more likely we are to have trouble with our forecasts, and that, if the mean state starts to look qualitatively different to reality, then we should expect to be in difficulty.

Past experience has shown that simulating the mean state in the tropical Atlantic is not easy. For example, the Study of Tropical Oceans in Coupled Models (STOIC) project (Davey et al. 2002) examined the tropical climate in a set of nonflux corrected coupled GCMs. When looking at the SST along the equator, the models exhibited a considerable range of absolute values of equatorial SST in the Pacific and typically had problems near the eastern boundary, but in all cases the gradient in midocean was reasonably represented. In the Atlantic, all of the models, bar one, had the mean zonal gradient of SST the wrong way round!

The results from STOIC were based on the climate of long runs from coupled GCMs, and in the shorter runs typically used in seasonal forecasting systems (e.g., around 6 months) the model climate does not behave quite so badly in terms of SST. Nonetheless, the DEMETER runs show that models still have difficulty in reproducing the observed rate of seasonal cooling in the eastern equatorial Atlantic in July and that, at least at this time of the year, the zonal SST gradients are poorly represented. Figure 1 shows the time evolution of SST bias in the eastern equatorial Atlantic for the various DEMETER models for forecasts starting on 1 February. Although the absolute SSTs vary between the models, the observed rapid cooling in June/July is not reproduced by any of them. From short lead times (e.g., forecasts from 1 May), a minority of models are able to cool quite rapidly, and this better representation of the mean state does seem to be associated with different behavior of the SST forecasts. Sensitivity experiments on this topic will be presented in section 6.

It is important to note that the errors in SST are associated with errors in the dynamical forcing of the ocean within the coupled system. The easterly surface zonal winds in the western equatorial Atlantic are too weak in the coupled models in the months preceding (and often during) the period of rapid observing cooling in June and July. This error in the surface winds acts to depress the equatorial thermocline in the east relative to the position it should take, reducing the cooling effect of the upwelling. This will result in substantial changes in the sensitivity of the model SST to subsurface anomalies, with a deeper thermocline giving less sensitivity. The fact that the SST errors in the equatorial Atlantic appear to be driven by dynamical errors (rather than, e.g., just problems with heat flux) makes it likely that the coupled system will respond relatively poorly to anomalies in the ocean initial conditions.

The errors in the zonal wind in the coupled integrations can be amplified by coupled interactions—a reduction in SST gradient will reduce the zonal winds that help maintain that gradient—but uncoupled integrations show that the atmosphere GCMs have significant wind errors even when run with observed SST. For

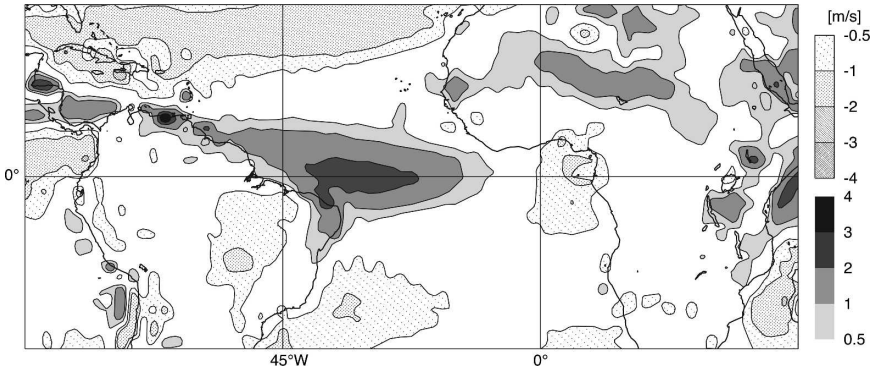

FIG. 2. The bias in 10-m zonal wind, averaged over April–June, for the ECMWF atmosphere model when run with observed SST. Based on model version Cy29r1, run as a five-member ensemble at T95 for 1 April starts from 1987 to 2001, compared to ERA-40 reanalysis. The westerly bias in the western equatorial ocean will depress the thermocline in the eastern Atlantic, reducing the potential for a rapid cooling in June/July.

example, Fig. 2 shows the April–June bias in 10-m zonal wind from the latest version of the ECMWF atmosphere model. The substantial westerly bias at the equator is clearly visible. The proper analysis of errors in atmospheric GCMs is a substantial task that we will not undertake here. Nonetheless, one common difficulty in GCMs is getting correct precipitation over tropical land areas, with insufficient precipitation in the Amazon being a particular problem. It may well be that the wind errors are related to this problem. Clearly in the Atlantic basin, a failure to position convection correctly with respect to land/ocean is serious and is likely to cause a range of problems in simulating both the mean state and interannual variability.

In summary, the mean climate simulated by coupled GCMs in the Atlantic sector still has significant room for improvement. The models are not so bad as to preclude the possibility of obtaining something useful in forecast mode, but the problems are serious enough for us to expect a significant degradation of skill.

4. The quality of ocean analyses in the tropical Atlantic

Knowledge of the state of the ocean is essential for initialization of seasonal forecasts. The ocean state can be estimated by forcing an ocean GCM with prescribed atmospheric fluxes. But uncertainty in the time evolution of the wind stress results in significant uncertainty in the interannual and decadal variability of the upper ocean, and model errors further contaminate the results. Combining observations with wind-forced ocean models through data assimilation techniques is in principle the optimal solution for estimating the initial state of the ocean.

The impact of data assimilation can be evaluated by comparison with independent data. Figure 3 (top right) shows the correlation during the period 1993 to mid-2003 of the sea level from altimeter data with the sea level from the ECMWF operational ocean analysis (System 2). This analysis assimilates only temperature data, although multivariate adjustments are made to salinity to conserve water mass properties (Troccoli et al. 2002) and geostrophic corrections are made to the velocity field (Burgers et al. 2002). For a detailed description, see Balmaseda (2003). Altimeter data are not used. For comparison with the assimilation-based analyses, Fig. 3 (top left) shows the correlation for an experiment without data assimilation (i.e., a forced run). In the tropical Pacific, data assimilation improves the representation of the ocean state: it increases the peak value of the correlation as well as the area with correlations above 0.8, which now covers most of the band of 10°N–10°S. TAO mooring data are the main source of the improvement (Vidard et al. 2005).

In the Atlantic, the correlation is much lower than in the Pacific with or without data assimilation. The forced run shows peak values above 0.7 in a small area around the east-central equatorial Atlantic. In the assimilation run, the extension of the 0.7 contour is slightly increased but remains low compared to the equatorial Pacific. The equatorial Atlantic is a problematic area. To begin with, forced ocean models tend to produce a very diffuse thermocline. Additionally, the strong salinity stratification contributes significantly to the vertical stability, while in these experiments only temperature data are assimilated since observations of salinity were scarce until the recent development of the Array for Real-Time Geostrophic Oceanography (ARGO) network. This makes the equatorial Atlantic a very demanding test for the multivariate tempera-

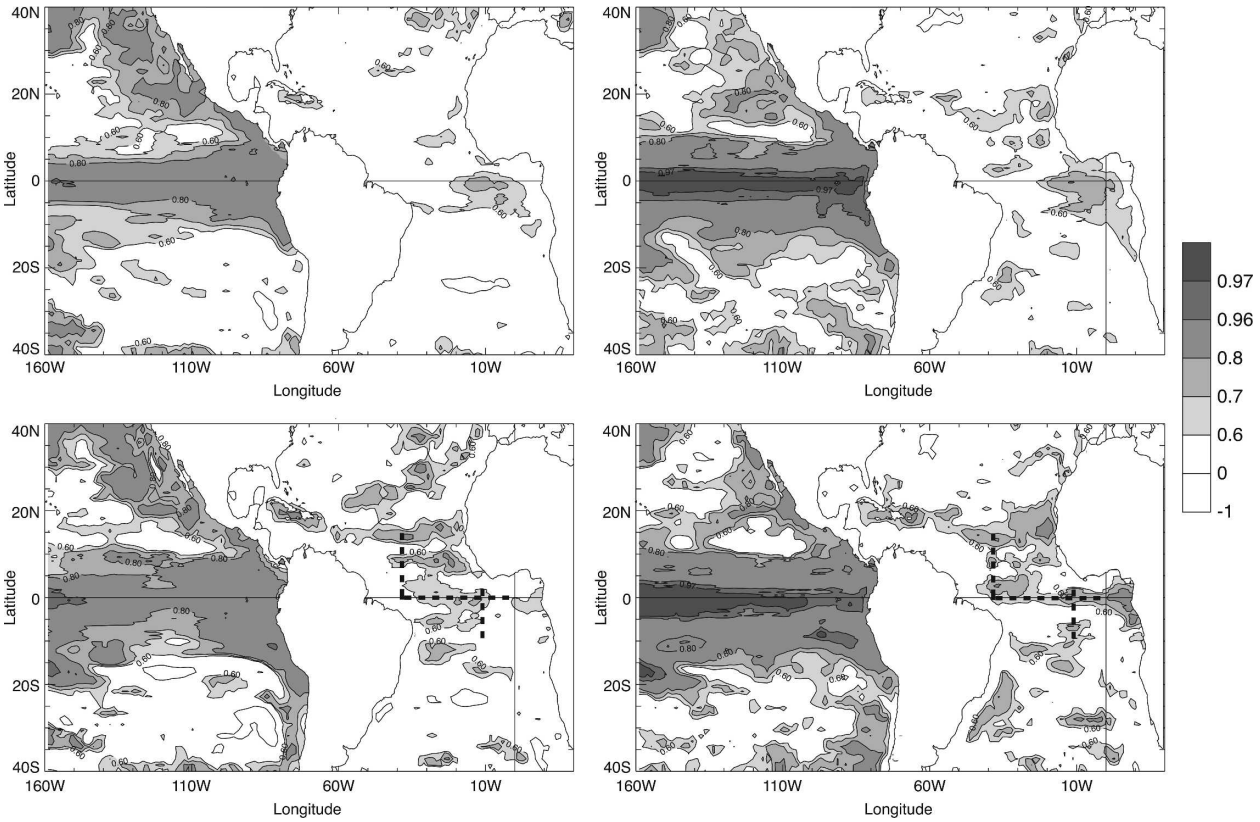

FIG. 3. Correlation of sea level from altimeter data with sea level from the ECMWF operational ocean analysis, for the periods (top) January 1993–June 2003 and (bottom) January 1998–January 2003, with (left) no data assimilation and (right) with in situ temperature assimilation.

ture–salinity relationship. Furthermore, the freshwater fluxes (river discharge, precipitation − evaporation) are poorly known, which causes errors in the representation of the water mass characteristics.

The presence of systematic error can also be damaging for the representation of interannual and decadal variability since it can lead to aliasing of variability with changes in the observing system. In fact, one factor that contributes to the degradation of sea level in the assimilation run is the change in the observing system, namely, the introduction of the Pilot Research Array in the Tropical Atlantic (PIRATA) buoys after 1998. The impact of PIRATA manifests itself as a systematic decrease in the sea level over the equatorial Atlantic (Segschneider et al. 2000; Vidard et al. 2005). This difference in the mean state leads to an artificial variability in sea level and therefore to an apparent degradation to the correlation with altimeter (Vidard et al. 2005). Moreover, several studies have shown that ocean data assimilation is typically correcting the systematic error, which can be caused by errors in the forcing, errors in the models, or errors in the assimilation methods (Alves et al. 2004; Vialard et al. 2003; Balmaseda 2003; Bell et al. 2004). The presence of systematic error can deteriorate the state estimation if the assimilation methods are not robust enough to cope with it. In fact, in several cases the assimilation method itself can be causing the error. These results suggest that appropriate methods to handle systematic error, such as those advocated in Dee and Da Silva (1998), Bell et al. (2004), or Vidard et al. (2004), are needed in order to obtain consistent climate reanalyses that are not contaminated by the developments of the observing system. This does not imply that the effect of PIRATA is in itself damaging. In fact, if the correlation with the altimeter is computed only for the post-PIRATA era, the results are more optimistic. Figure 3 (bottom) shows the correlation during the period from 1998 to mid-2003 for both forced (left) and assimilation (right) runs. The correlations for the forced run are higher for the recent period (apart from the central/east equatorial Atlantic), probably due to more accurate wind fields. With assimilation the correlation increases, with peak values of 0.8 located around the PIRATA array (black thick-dashed lines), which suggests that the data provided by the moorings are valuable.

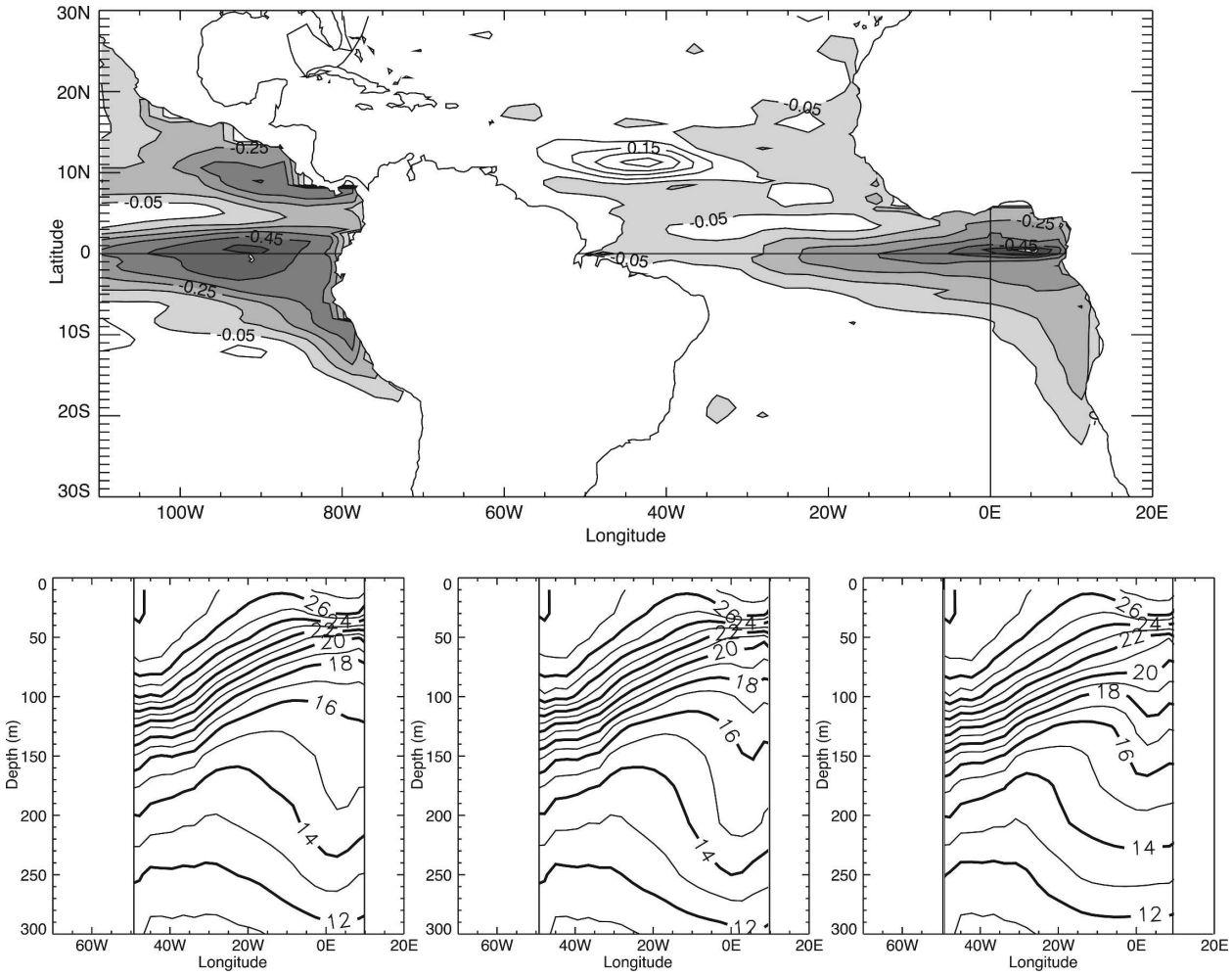

FIG. 4. (top) Mean differences in averaged temperature over the top 300 m for the years 2002–03 between one assimilation experiment using all XBT, mooring, and Argo temperature data and one using only XBT and Argo data. (bottom) Zonal sections of temperature at the equator for (from left to right) assimilation of all data, all data but mooring, and no assimiliation.

Figure 4 (top panel) shows the impact of PIRATA moorings on the top 300-m temperatures. This is obtained through an Observing System Experiment (OSE). In these experiments, permutations of combinations of the available observation systems are used in an analysis of the oceanic state, in which one system is excluded from the analysis, so providing an estimate of the impact of the omitted system [for more global consideration about ocean observing system, refer to Vidard et al. (2005)]. In this case, it shows that the main impact of PIRATA is a cooling of the equatorial Atlantic, significantly stronger in the east. The bottom panels show a cross section of the mean temperature in 2002–03 along the equatorial Atlantic for the forced run (right), the assimilation of Argo and XBTs (middle), and the assimilation of all data (left). The assimilation tends to tighten the thermocline and make it steeper, which is a desirable feature. Notice that both components of the observing system contribute to this enhancement.

Despite this improvement of the mean temperature fields, the inability of the system to produce high correlation with altimetry in the tropical Atlantic implies a very clear need to improve key components of data assimilation techniques such as the multivariate constraints, as well as development to make direct use of the salinity data that is now becoming available from ARGO. Such developments are now in an advanced phase at ECMWF, and early results are promising (Haines et al. 2006).

As well as improving the assimilation methods it is desirable to reduce the error of the ocean simulations, both by improving the ocean models (mixing physics, resolution) and by improving the surface fluxes. The

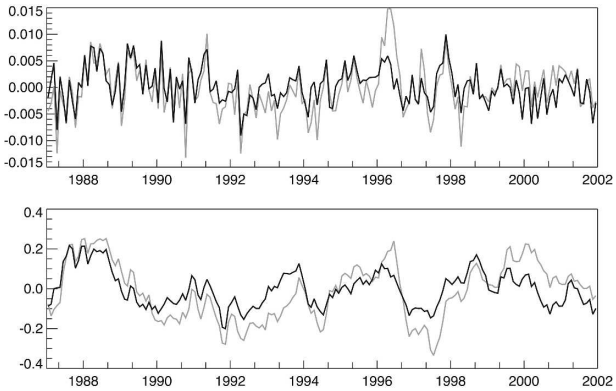

FIG. 5. (top) The time evolution of the wind stress anomalies averaged over the equatorial Atlantic from ERA-15/OPS (gray) and for ERA-40 (black). Note the large difference in 1996. (bottom) The evolution of the upper-ocean heat content (average temperature in the upper 300 m) when the different fluxes are used to drive the same ocean model.

importance of surface fluxes can be illustrated by looking at results produced by the same GCM forced by different wind stresses. Figure 5a shows the time evolution (averaged over the equatorial Atlantic 5°N–5°S) of two different wind stress products: one used by the operational ocean analysis (referred as ERA-15/OPS in what follows, represented by gray) and one provided by the ERA-40 atmospheric reanalysis (black). The largest differences occur in the mid-1990s, particularly in 1996. In general we believe the ERA-40 stresses to be more accurate, and wind-forced runs of the ocean correlate better with altimetry data when ERA-40 forcings are used. Nonetheless, the differences between the two wind stress products are thought to be a reasonable proxy for the true uncertainty. The impact on the ocean can be measured by the evolution of upper-ocean heat content (average T in the upper 300 m), shown in Fig. 5b. In the equatorial Atlantic, uncertainty in the stresses leads to uncertainties in the ocean state that are almost as large as the interannual variability. Data assimilation should reduce this uncertainty but, as we have shown, our ability to do this effectively in the equatorial Atlantic is still rather limited.

5. The skill of coupled GCM forecasts of tropical Atlantic SST

a. Results from ECMWF models

A set of ensemble forecasts with a single coupled GCM provides an estimate of both the predictability limit and the actual prediction skill. The predictability limit relates directly only to the model used; how good an estimate it is of the real-world predictability depends on the verisimilitude of the model. The model predictability limit for a given lead time is defined here as the unbiased estimator of the standard deviation of the forecast ensemble at that lead time. To form the average across different start dates, the mean variance is calculated and the square root then taken. This definition of model predictability is directly related to rms error: in a perfect model scenario, the verifying "truth" and the ensemble members are indistinguishable, and their rms distance from the ensemble mean over a large enough sample will be the same; that is, for a perfect ensemble forecasting system, the rms error will equal the predictability limit. In discussing the estimated skill of tropical Atlantic SST forecasts, an issue to note is that the statistics vary appreciably depending on the period being verified. There are plausible hypotheses that might explain this (low frequency variations of Atlantic climate; a diverse range of processes that drive SST variability, with different predictability and model skill; sampling from only a limited number of significant events), but rather than explore these we simply stress that caution is needed when trying to draw absolute conclusions about forecast skill based on results from a particular period.

Figures 6 and 7 show forecast statistics for the longest period available to us, the 43-yr period from 1959 to 2001. The forecasts come from the ECMWF model used in DEMETER and consist of a nine member ensemble made four times per year with February, May, August, and November start dates. The ocean model was initialized by driving it with fluxes from ERA-40, while additionally relaxing the model SST to observed values with a feedback of 400 W m^{-2} $°C^{-1}$. Statistics are shown for three indices of SST. The first is the ATL3 index (3°N–3°S, 20°W–0°) introduced by Zebiak (1993). Variability in this region is thought to be partly due to an ENSO-like mechanism (where ocean initial conditions should be important) and partly forced by teleconnections from ENSO. The other two indices are NTA (5°–28°N, 80°W–20°E) and EQSTA (20°S–5°N, 60°W–20°E), covering the north tropical Atlantic and the equatorial and south tropical Atlantic, respectively. The definition of these regions follows Servain (1991), who used these indices to form the much-discussed "dipole" of Atlantic SST variability. Here we simply look at the two regions separately. On interannual time scales there is a moderately strong correlation between the EQSTA and ATL3 indices. This is not simply due to EQSTA including the equatorial region, but involves true covariability of SST in equatorial and off-equatorial regions. Our primary references as to how well the forecast system works are to compare it to

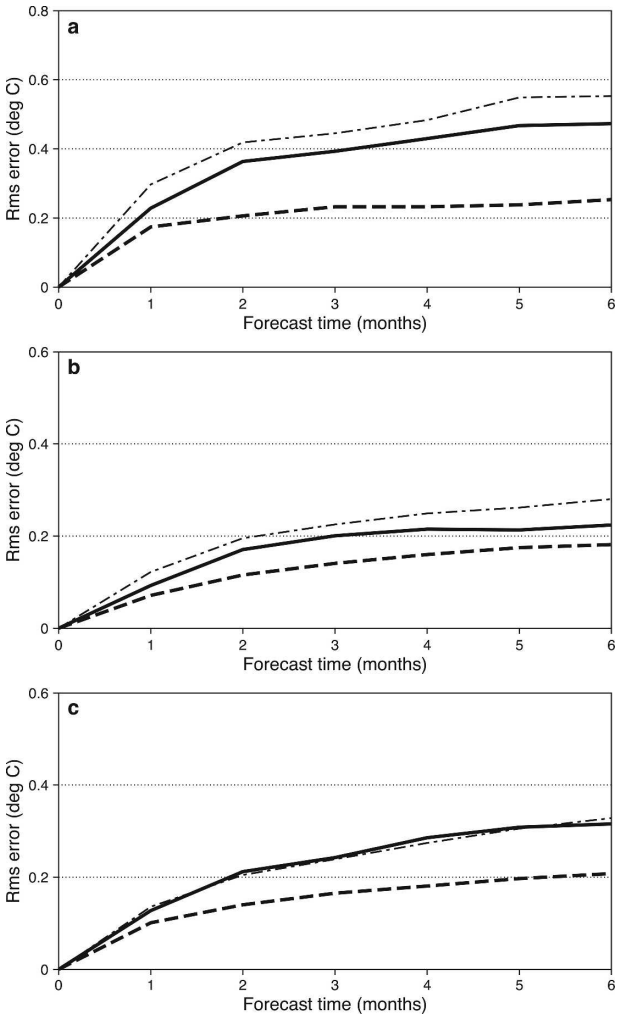

FIG. 6. Rms error of SST forecasts from 1959 to 2001 in (a) ATL3, (b) NTA, and (c) EQSTA regions. Solid line shows forecast skill, dashed line shows model estimate of predictability limit, and thin dot–dash line shows skill of persistence forecast.

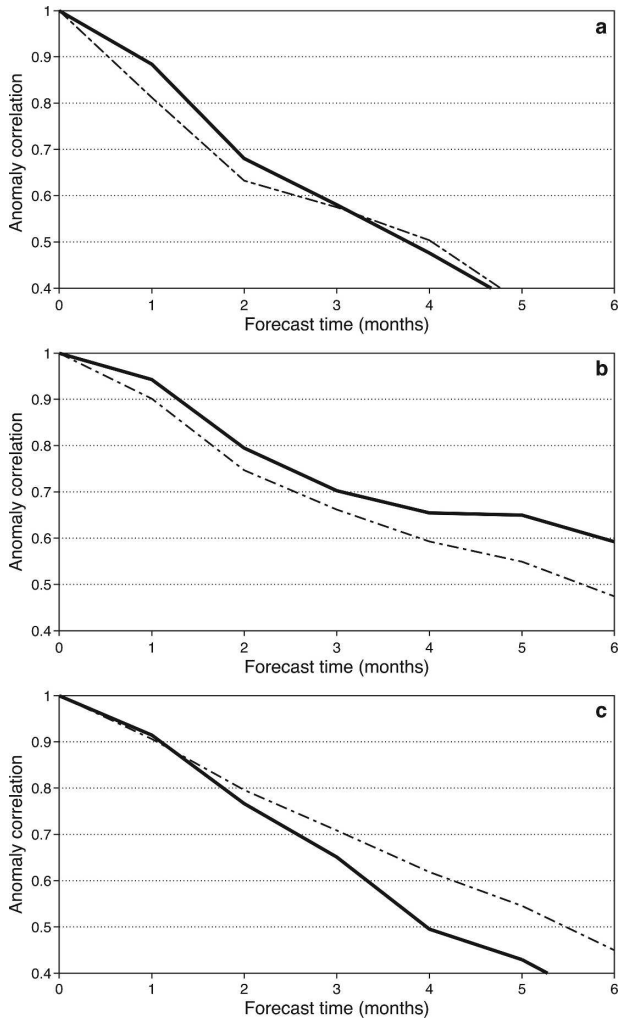

FIG. 7. Anomaly correlation of SST forecasts from 1959 to 2001 in (a) ATL3, (b) NTA, and (c) EQSTA regions. Solid line shows forecast skill and thin dot–dash line shows skill of persistence forecast.

anomaly persistence, on the one hand, and the model's own diagnosis of the predictability limit, on the other.

Figure 6a shows that in rms error terms, ATL3 forecasts are better than persistence at all lead times although much worse than the model predictability limit. In fact, the rms error is helped at longer lead times by an underprediction of the amplitude of observed variability, and Fig. 7a shows that, in anomaly correlation terms, the forecast has an appreciable benefit over persistence only for the first two months. Observed ATL3 SST variability shows relatively strong, short-lived anomalies, thus making persistence a poor predictor more than two or three months ahead. One might hope that models have the potential to do somewhat better, but it is doubtful that a high level of predictability exists beyond the first three months.

In the NTA (Figs. 6b and 7b), the actual rms errors are only modestly larger than the predictability estimate, and the forecasts beat persistence clearly in both rms and anomaly correlation terms. A time series of ensemble mean forecast values shows that the model is fairly active and often reproduces the growth of anomalies (e.g., positive in 1966, 1983, 1987, and 1998; negative in 1976 and 1984) as well as their decay. In general terms, it seems that the coupled model forecasts are doing a reasonable job of picking up a substantial portion of the "remotely driven" SST variability in this part of the ocean and, although there are clearly more errors than would be expected in a perfect forecasting system, the overall performance is not too bad. This conclusion is consistent with the work by Huang et al. (2002), who

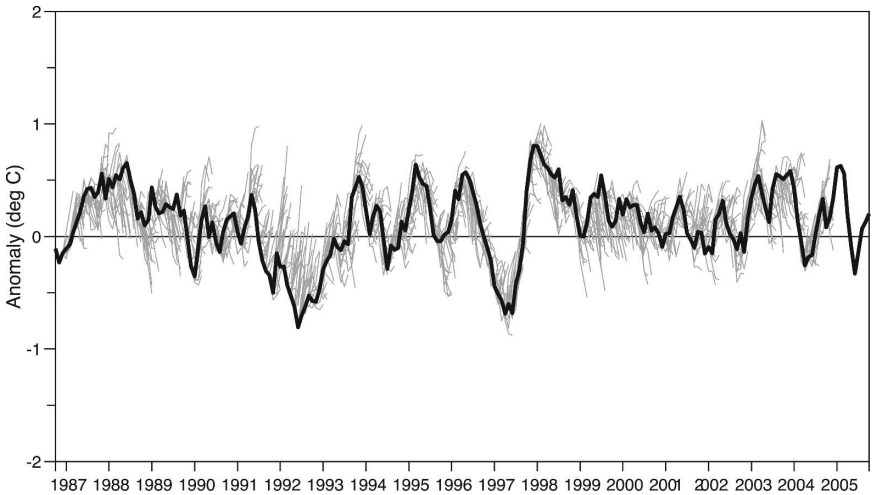

FIG. 8. SST forecasts for the EQSTA region (20°S–5°N, 60°W–20°E) from the operational ECMWF forecast system. Each forecast from a five-member ensemble is plotted starting every month, and only the first three months of each forecast are plotted.

showed that in the north subtropical Atlantic much of the SST variability is remotely driven.

The plots for EQSTA (Figs. 6c and 7c) show that forecast performance is not as good as for the north subtropical Atlantic: rms errors are no better than persistence and much worse than the model predictability limit, and anomaly correlation is substantially worse than persistence.

We can look more closely at forecast performance in EQSTA by using the ECMWF operational forecast system, for which retrospective forecasts exist only since 1987, but which is run at one-month intervals rather than the three-month intervals of the 1959–2001 runs. Figure 8 shows the first 3 months of every forecast in the period from 1987 to 2004, plotting each of five ensemble members separately to allow a reasonable assessment of whether the observations and forecast are consistent for each forecast. Some periods stand out as successful—notably 1997/98, but also the warmings of 1987 and 2003. Others stand out as being consistently poor—1991/92 and 1999/2000 both contain extended periods where every month, every ensemble member wanted to go in completely the wrong direction. This is clearly not a case of "lack of predictability," but tells us that something was seriously wrong with the initialization, the external forcings, or the physical evolution of the model. Which combination of these factors is hurting the forecasts is unknown.

When we look at the equator, for example ATL3, different issues are at play. A notable factor is that the forecasts are more successful at some times of year than others. A particularly difficult period is boreal summer (JJA): SST anomalies can develop at the start of this period, with the model completely failing to capture them. Unfortunately, JJA is a key period in which SST in the equatorial Atlantic and Gulf of Guinea has a significant impact on the West African monsoon, so we would like to understand why our forecasts often fail at this point. Figure 9 shows the rms errors for the ATL3 region as a function of the seasonal cycle. Forecasts verifying in July have high errors regardless of lead time. Persistence is also a poor predictor for July, even at short leads. January is well predicted by the model, however. The poor performance of forecasts for July is also evident in anomaly correlation scores (not shown): the high rms errors are not just due to the higher amplitude of SST anomalies at this point in the seasonal cycle. There is some modest seasonality in forecast behavior in the other Atlantic regions, but it is not sufficiently important to be discussed here.

b. Multimodel results from DEMETER

So far we have looked at the SST prediction results only from ECMWF models. What happens when we broaden our view to include the other DEMETER models? A multimodel ensemble is formed by taking the ensemble forecast for each model (corrected by the mean forecast bias specific to the model concerned) and putting them together to form a large ensemble. The ensemble mean of the multimodel forecast is simply the mean of the ensemble means of the individual models. The ensemble spread is the standard deviation of the whole ensemble and is contributed to both by the spread within the individual model ensembles and the difference between different model ensemble means.

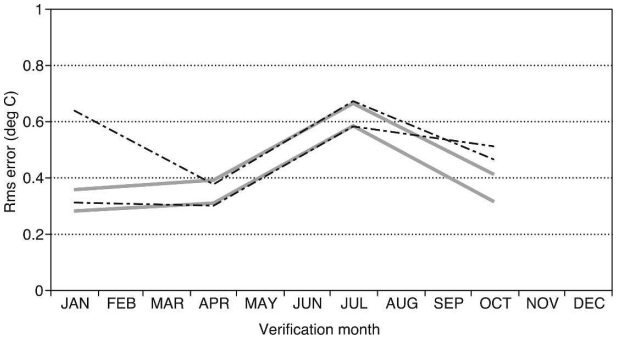

FIG. 9. SST prediction rms errors for the ATL3 region as a function of the seasonal cycle, based on ECMWF forecasts from 1959 to 2001. Thick gray line is from the model, thin dot–dash line is persistence. Curves are plotted for both 3-month and 6-month lead times. Forecasts verifying in July have high errors, regardless of lead time.

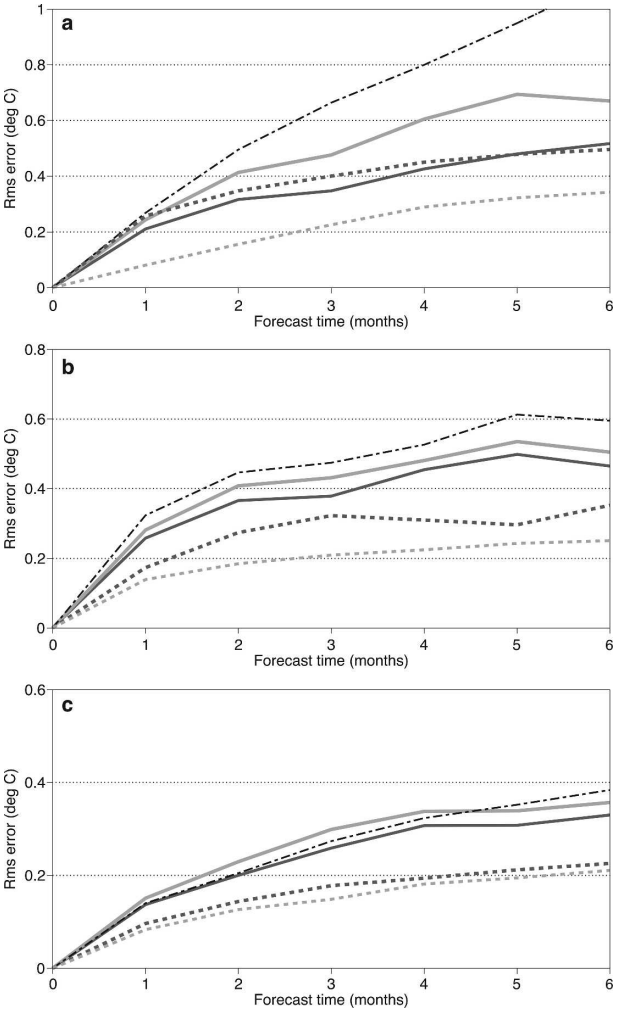

FIG. 10. The impact of multimodel averaging on SST errors and ensemble spread in (a) Niño-3.4, (b) ATL3, and (c) EQSTA. Solid lines are rms errors; dotted lines are ensemble spread. Gray is for the ECMWF model only; black is for the multimodel ensemble. The multimodel ensemble is bigger, which reduces its rms errors slightly just because of ensemble size. Ensemble spread estimates are not affected by ensemble size due to use of unbiased estimators.

On one view of multimodel ensembles, a "perfectly sampled" multimodel ensemble will have a variance associated with the model error of the ensemble mean, which is only a factor of $1/n$ of the variance associated with the model differences, where n is the number of independent models. So for a perfect sampling of model error, the overall multimodel ensemble spread will be somewhat larger than the rms forecast error of the multimodel ensemble mean. In practice, a collection of models will not properly span and perfectly sample the space of model errors, so the spread will be less than the theoretical maximum; but an enhanced ensemble spread is nonetheless a good sign that we are effective in sampling a significant amount of model error.

Figure 10 shows the impact of multimodel averaging on rms error in three regions where the multimodel ensemble consists of three forecasts from each of six models. In the Pacific, the error is substantially reduced and the spread is substantially enhanced such that the ensemble spread is, in fact, a little larger than the forecast error. In ATL3, the spread is increased somewhat, especially in the first two or three months, but the skill improvement is small, and a large gap remains between ensemble spread and forecast error. Further investigation shows that the extra spread in the first months is greatly contributed to by short range forecasts for June/July SSTs, where one or two of the DEMETER models produces a reasonable mean state cooling, facilitating extra spread both within and between different models. The longer range forecasts for June/July SST, by contrast, have very little spread to accompany their large errors—consistent with the fact that none of the DEMETER models produce the observed mean state cooling at this forecast range. The relationship between mean state and forecast error is discussed below. In EQSTA the situation is even more hopeless: the multimodel ensemble has only a very modest impact on both spread and skill, resulting in an ensemble that does not come close to sampling the forecast error. Seasonality is not evident, and specific mechanisms responsible for the failure do not suggest themselves.

Figure 11 shows a concrete example of a "failed" longer-lead multimodel forecast for ATL3 SST, with significant errors from the first month onward and the substantial warming in May–August being missed entirely. On this occasion the EQSTA forecast failed in a similar way. It is possible that in this particular case the

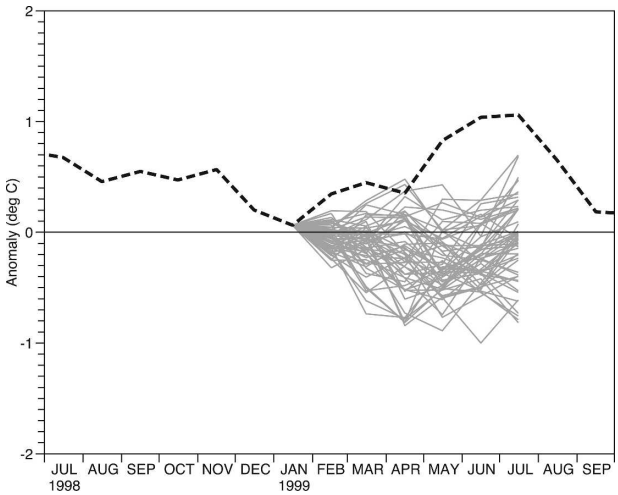

FIG. 11. A multimodel forecast of ATL3 SST, initialized on 1 Feb 1999, using nine ensemble members from each of the six models. The warming of May–August 1999 was not captured by any of the ensemble members.

relatively large wind perturbations still fail to capture the initial condition uncertainty. But the general increase in initial condition uncertainty over that already diagnosed would need to be implausibly large to explain the forecast error statistics. The hypothesis that the lack of mean upwelling in the models is a substantial contributor to the forecast error is plausible, however. The evolution of the observed SST anomaly in this (not atypical) case is suggestive of the upwelling of a preexisting subsurface temperature anomaly. If this is true, then it is possible to predict such events: all we need is knowledge of the subsurface initial conditions, and realistic models. An alternative hypothesis is that, while the mean upwelling is necessary to "sensitize" the coupled system, the actual development of the anomaly is spontaneously produced by unpredictable wind fluctuations, and that consequently SST predictability will remain low even with good models and good data. Where the truth lies between these two viewpoints remains to be established.

c. *Role of the subsurface in SST predictability*

We can use a set of experiments with the ECMWF model to gain some insight into the role of the subsurface in determining SST predictability. Figure 12 shows the estimated SST predictability (i.e., the unbiased estimator of the standard deviation of the ensemble about its mean) from two pairs of experiments for each of several regions, covering the years 1987–2001. The black solid line shows the result when using wind perturbations to create an ensemble of initial conditions when no data assimilation is used. The black dotted line shows the result when wind perturbations are not used so that the ocean subsurface is the same in the initial conditions of all ensemble members. In both cases, SST perturbations are applied in the surface layers at the start of the forecast. Comparison of the ensemble spread from these two experiments reveals what impact typical uncertainties in the unconstrained subsurface ocean have on SST predictability. Figure 12a shows that in the Niño-3.4 region of the Pacific, the wind perturbations create a very large spread in the SST forecasts, additional to that which depends only on surface perturbations. This is a nice demonstration that the subsurface state of the ocean is important for predicting SST. We now again consider forecasts with and without wind perturbations, but this time in a system that assimilates subsurface ocean data (light gray curves, solid and dotted as before). The data assimilation greatly reduces the spread produced by the wind perturbations (gray solid versus black solid), although the spread is still larger than in the case of no wind perturbations (gray solid versus gray dashed), so that the data assimilation has not completely removed the effect of the uncertainty in the wind. Thus for ENSO forecasting we can demonstrate (i) that the subsurface is important and (ii) that for the period 1987–present, we are able to use in situ data to constrain the system against any significant uncertainties in the wind field that may exist.

We now look at the tropical Atlantic. Figure 12b shows the same curves for the ATL3 region. Comparing the black solid and dashed curves shows that the influence of the wind perturbations is much less than in the Pacific. Indeed, assuming that the wind perturbations are reasonably efficient at perturbing the ocean subsurface, as we would expect in the equatorial ocean, this shows that in the model the subsurface is only modestly important in determining SST variability in ATL3 beyond the first month. Note that "in the model" is an important caveat. Comparing the gray solid and dashed curves shows that data assimilation does not reduce the admittedly modest impact of the wind perturbations on the forecast spread. This is consistent with the rather pessimistic assessment of the Atlantic Ocean analyses given earlier and suggests that for this period (1987–2001) the in situ data and our methods of using them are not sufficient to overcome errors in the wind forcing.

Results from EQSTA are shown in Fig. 12c. The spread grows more slowly than in the equatorial Atlantic in all cases, perhaps due to the larger area and slower physical processes being important for the SST evolution. In relative terms, the contribution of the

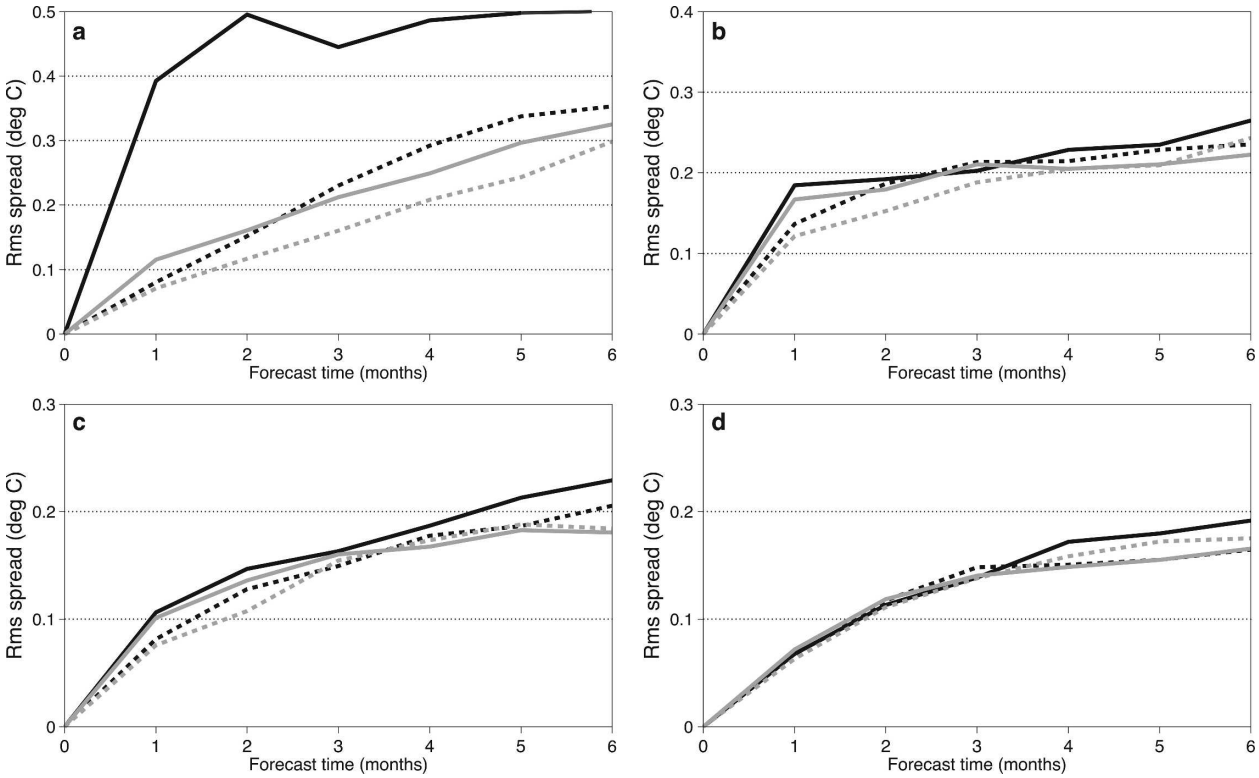

FIG. 12. SST ensemble spread as given by four experiments for (a) Niño-3.4, (b) ATL3, (c) EQSTA, and (d) NTA. Black lines are from experiments which have ocean initial conditions prepared without data assimilation; gray lines are from experiments that did assimilate subsurface ocean data. Solid lines are from experiments with wind perturbations applied, and dotted lines from experiments without. The vertical scale varies between the panels. See text for interpretation.

wind perturbations is similarly modest, however, and again the assimilation of in situ data does not usefully constrain the system. Finally, Fig. 12d shows NTA, where the spread in the early months is essentially unaffected by the subsurface ocean perturbations and the data assimilation becomes irrelevant. Analysis of a longer period (1958–2001) shows that there is a small component to the spread that comes from the wind perturbations, but it is the smallest of the regions considered here.

These results graphically illustrate one reason why coupled ocean–atmosphere SST forecasting systems for the Atlantic are little developed in comparison to the Pacific, at least given the present state of the models.

We have already noted that the models systematically fail to produce the observed strong seasonal cooling in ATL3 in June/July, and that this failure is plausibly related to a lack of spread in forecast SST and insensitivity to the ocean subsurface at this time. A direct way to address this question is to run a coupled GCM with a flux correction applied to the zonal equatorial wind stress, designed to ensure that the coupled model does upwell water in its mean seasonal cycle. Such an experiment has been made, in idealized form, using a recent version of the ECMWF coupled model (Cycle 28r3) and covering the years 1993–2002. The additional wind stress is applied in the equatorial band only and is constant in time and longitude around the globe with a strength of -0.02 N m^{-2}. This is equal and opposite to the peak seasonal error in the uncorrected model in May, averaged across the equatorial Atlantic, and compares to the observed climatological zonal stress that varies seasonally between -0.02 and -0.03 N m^{-2}. The evolution of mean SST in the ATL3 region shows a much stronger seasonal cooling with the additional wind stress term although, if anything, the cooling is still slightly underestimated despite the stress being on average a little too strong. The additional wind term puts the coupled model into a different regime that is closer to observations, but still not perfectly realistic.

Figure 13 shows the impact on the ATL3 SST forecasts. Only results from forecasts including July as a verification month are shown to focus attention on this

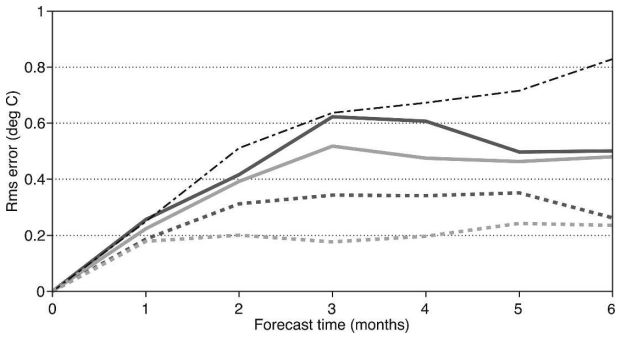

FIG. 13. ATL3 SST rms error (solid lines) and ensemble spread (dotted lines) from two experiments, verifying forecasts from 1 April and 1 July start dates only. Gray: a recent version of the ECMWF coupled model (28r3). Black: the same model, but with a constant additional wind stress term. The additional mean wind causes a big increase in ensemble spread.

part of the seasonal cycle. The ensemble spread of the forecasts is dramatically increased by the change in coupled model mean state, confirming that at this time of year the coupled models are very sensitive to the mean state. However, note that the rms error of the forecasts is also increased and, indeed, the anomaly correlation deteriorates slightly also. The better mean state, and the stronger communication of the subsurface to the surface, has not resulted in better forecasts. There are numerous other problems still affecting the forecasts, but the lack of skill improvement might again point to limitations in the quality of the ocean subsurface initial conditions.

6. Conclusions

Present day SST forecasts in the tropical Atlantic on the seasonal time scale leave much to be desired. This statement is based on comparing the quality of forecasts against the model-estimated predictability limit and also against the skill achieved by a very simple empirical scheme—persistence. The best region is the north tropical Atlantic (NTA), where model forecasts are acceptably good, beating persistence on all measures and not being so much worse than the estimated predictability limit. In the equatorial ocean, the margin over persistence is modest, and largely restricted to the first two months. In the southern Tropics, model skill is often worse than persistence. It is not just the ECMWF forecasts that perform poorly—all of the DEMETER forecasts do, as does the multimodel ensemble.

It is apparent that the coupled GCMs are typically unsatisfactory when measured against what is needed to ensure reasonable seasonal forecast performance. The most obvious manifestation of model imperfections is the evolution of the mean state in June/July, when the models do not produce the observed intensity of SST cooling. As a direct consequence of this, they substantially underestimate the spread in SST forecasts that should occur at this time of year and are unable to represent important processes that drive observed SST variability. But the failure of the models is more serious than this: misrepresentation of convection and winds in and around the Atlantic, which causes the mean state errors, is liable to produce many other erroneous responses when used in a seasonal forecasting system. Again, problems are not restricted to one or two models but appear to be endemic.

Next, it seems that ocean analyses in the tropical Atlantic are of limited quality, as is apparent when comparing to independent data such as altimetry. At least in the past, in situ data has been insufficient to overcome the uncertainties in the wind forcing. There are particular difficulties relating to salinity, which have been compounded by a lack of data. We believe there is real scope for improving the ocean analyses for past dates, but effort is required and the eventual level of success remains uncertain. The adequacy of today's observing system is still uncertain. Our analyses from the PIRATA era (1998 onward) give slightly better equatorial analyses, but these are still poor relative to what is desired. The era of ARGO float coverage is still very short. There are good physical arguments for the desirability of enhanced coverage by equatorial and near-equatorial moorings, but convincing experimental evidence for the potential impact on seasonal forecast scores may be hard to produce anytime soon.

Given the above problems, our ability at this time to produce reliable insights into the physics of the forecasting problem is limited. Our analysis of the role of subsurface variability on SST forecast spread demonstrated that there was in general only a small role for the subsurface in the Atlantic compared to the Pacific. In the case of the equatorial Atlantic in the June/July period, this conclusion is clearly untrustworthy, given the evident model problems. But in other regions and at other times the results may not be that far from reality.

An outstanding question from this paper is the true level of predictability for the ATL3 region in June–August. If we have poor initialization of the ocean subsurface, practical predictability may be limited to periods of strong remote forcing. If we can specify the ocean subsurface correctly, then it could be that predictability, at least for two months or so, is relatively high. However, a reliable assessment of this will need models with a good mean state and a good representation of atmospheric variability.

The models perform particularly badly in the

EQSTA region with extended periods of forecasts inconsistent with observations. This is a large region and, although we might expect a role for processes along the eastern boundary not fully resolved by the model (Benguela Niños; Florenchie et al. 2004), without some unknown feedbacks it is not clear that these will be sufficient to explain the amplitude of the discrepancies. Forecast models do capture some of the processes that drive SST variability in this region, as shown in Fig. 8, but the reasons behind the overall poor performance remain unclear.

So, how is our seasonal forecasting capability in the tropical Atlantic to be improved? As mentioned, we need better ocean analyses. High quality forcing fields are important, and the combination of scatterometer winds and NWP systems should give us reasonable winds in the future. The in situ ocean observing system is better than at any time in the past. Progress will be made on improving the assimilating systems themselves. Nonetheless, the signals in the equatorial Atlantic are relatively small in amplitude and spatial scale, certainly compared to the interannual variability of the Pacific. Geophysical noise from the eddy field is large. There are only two moorings in the western half of the Atlantic equatorial waveguide, the key source region for shorter range seasonal prediction of equatorial Atlantic SST, and it is doubtful that these are sufficient to constrain the equatorial initial conditions to the extent that a reliable forecast system would want. Further attention to the equatorial Atlantic observing system is warranted.

There is also a clear need for better coupled models. The usual list of *desiderata* apply: better convection, wind and clouds in the atmosphere models, more credible treatment of equatorial and eastern boundary upwelling, and mixing in the ocean models. Other factors may be particularly important in the Atlantic sector due to the relative proximity of land: variations in soil moisture, vegetation and albedo, and the role of aerosols. Significant improvements take time, but will come eventually. As models become more realistic, our understanding of the predictability of the Atlantic sector will mature.

Finally, for predicting equatorial SST in particular, it may be most valuable to pursue relatively short range forecasts with lead times of one or two months. These are the time scales where most predictability is likely to exist, and reliable and informative forecasts at these lead times are still of great potential value. Present day forecast systems have many weaknesses, but this implies that there is scope for substantial progress in the years to come.

REFERENCES

Alves, O., M. Balmaseda, D. Anderson, and T. Stockdale, 2004: Sensitivity of dynamical seasonal forecasts to ocean initial conditions. *Quart. J. Roy. Meteor. Soc.,* **130,** 647–668.

Anderson, D., and Coauthors, 2003: Comparison of the ECMWF seasonal forecast systems 1 and 2, including the relative performance for the 1997/8 El Nino. ECMWF Tech. Memo. 404, 95 pp. [Available online at www.ecmwf.int/publications.]

Balmaseda, M. A., 2003: Ocean data assimilation for seasonal forecasts. *Proc. ECMWF Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean,* Reading, United Kingdom, ECMWF, 301–326. [Available online at www.ecmwf.int/publications.]

Bell, M. J., M. J. Martin, and N. K. Nichols, 2004: Assimilation of data into an ocean model with systematic errors near the equator. *Quart. J. Roy. Meteor. Soc.,* **130,** 873–893.

Burgers, G., M. A. Balmaseda, F. C. Vossepoel, G. J. van Oldenborgh, and P. J. van Leeuwen, 2002: Balanced ocean data assimilation near the equator. *J. Phys. Oceanogr.,* **32,** 2509–2519.

Carton, J. A., X. Cao, B. S. Giese, and A. M. Da Silva, 1996: Decadal and interannual SST variability in the tropical Atlantic Ocean. *J. Phys. Oceanogr.,* **26,** 1165–1175.

Chang, P., L. Ji, and H. Li, 1997: A decadal climate variation in the tropical Atlantic Ocean from thermodynamic air–sea interactions. *Nature,* **385,** 516–518.

——, R. Saravanan, and L. Ji, 2003: Tropical Atlantic seasonal predictability: The roles of El Nino remote influence and thermodynamic air–sea feedback. *Geophys. Res. Lett.,* **30,** 1501, doi:10.1029/2002GL016119.

Davey, M. K., and Coauthors, 2002: STOIC: A study of coupled model climatology and variability in tropical ocean regions. *Climate Dyn.,* **18,** 403–420.

Dee, D. P., and A. M. Da Silva, 1998: Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc.,* **124,** 269–295.

Florenchie, P., C. J. C. Reason, J. R. E. Lutjeharms, M. Rouault, C. Roy, and S. Masson, 2004: Evolution of interannual warm and cold events in the southeast Atlantic Ocean. *J. Climate,* **17,** 2318–2334.

Haines, K., J. Blower, J.-P. Drecourt, A. Vidard, C. Liu, I. Astin, and X. Zhou, 2006: Salinity assimilation using $S(T)$ relationships. *Mon. Wea. Rev.,* **134,** 759–771.

Huang, B, P. S. Schopf, and Z. Pan, 2002: The ENSO effect on the tropical Atlantic variability: A regionally coupled model study. *Geophys. Res. Lett.,* **29,** 2039, doi:10.1029/2002GL014872.

Kushnir, Y., W. A. Robinson, P. Chang, and A. W. Robertson, 2006: The physical basis for predicting Atlantic sector seasonal-to-interannual climate variability. *J. Climate,* **19,** 5949–5970.

Mathieu, P.-P., R. T. Sutton, B. Dong, and M. Collins, 2004: Predictability of winter climate over the North Atlantic European region during ENSO events. *J. Climate,* **17,** 1953–1974.

Palmer, T. N., and Coauthors, 2004: Development of a European Multimodel Ensemble System for Seasonal-to-Interannual Prediction (DEMETER). *Bull. Amer. Meteor. Soc.,* **85,** 853–872.

Servain, J., 1991: Simple climatic indices for the tropical Atlantic Ocean and some applications. *J. Geophys. Res.,* **96,** 15 137–15 146.

Segschneider, J., M. Balmaseda, and D. L. T. Anderson, 2000:

Anomalous temperature and salinity variations in the tropical Atlantic: Possible causes and implications for the use of altimeter data. *Geophys. Res. Lett.,* **27,** 2281–2284.

Troccoli, A., and Coauthors, 2002: Salinity adjustments in the presence of temperature data assimilation. *Mon. Wea. Rev.,* **130,** 89–102.

Uppala, S. M., and Coauthors, 2005: The ERA-40 re-analysis. *Quart. J. Roy. Meteor. Soc.,* **131,** 2961–3012.

Vialard, J., A. T. Weaver, D. L. T. Anderson, and P. Delecluse, 2003: Three- and four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean. Part II: Physical validation. *Mon. Wea. Rev.,* **131,** 1379–1395.

Vidard, A., A. Piacentini, and F.-X. Le Dimet, 2004: Variational data analysis with control of the forecast bias. *Tellus,* **56A,** 177–188.

——, D. L. T. Anderson, and M. Balmaseda, 2005: Impact of ocean observation systems on seasonal forecasts. ECMWF Tech. Memo. 460, 32 pp. [Available online at www.ecmwf.int/publications.]

Zebiak, S. E., 1993: Air–sea interaction in the equatorial Atlantic region. *J. Climate,* **6,** 1567–1586.