TECHNICAL MEMORANDUM

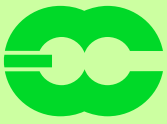
417

An ensemble generation method for seasonal forecasting with an ocean-atmosphere coupled model.

Jérôme Vialard, Frederic Vitart, Magdalena A. Balmaseda, Timothy N. Stockdale, David L. T. Anderson.

Research Department

18/09/2003



European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen terme

For additional copies please contact

The Library
ECMWF
Shinfield Park
Reading
RG2 9AX
library@ecmwf.int

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications/

Abstract

The methodology for ensemble generation for the ECMWF seasonal forecasting system is presented. It samples forecast errors due to unpredictable atmospheric synoptic variability, errors in the oceanic initial conditions due to uncertainties in the wind stress and errors in sea surface temperature (SST) analyses. A series of experiments using wind perturbations, SST perturbations, stochastic physics individually and all three together is presented and compared with the more usual lagged-average approach. During the first two months of the forecasts, experiments with only stochastic physics or only wind perturbations give a smaller spread of the ensemble than those with SST perturbations. From month three onward, all perturbation methods give a similar spread. Even with all perturbations applied the spread is still significantly smaller than the rms-error of the forecasts, suggesting that factors not presently sampled in the ensemble, such as model error and suboptimal ocean assimilation procedures act to limit the forecast skill. This is confirmed by analyses that do not show a clear link between the spread of the ensemble and the ensemble mean forecast error, suggesting that model error is a bigger source of erroneous forecasts than the uncertainties in initial conditions sampled by our method. These analyses thus suggest that methods that allow sampling of model error, such as multi-model ensembles, should be beneficial to seasonal forecasting.

1 Introduction

Reliable sea surface temperature (SST) forecasts are a pre-requisite to forecast the climate one to two seasons ahead (Palmer and Anderson 1994). One key-region is the tropical Pacific ocean, where the SST variability associated with *El Niño* affects the weather on a near-global scale. Besides model error, there are primarily two factors that limit the skill of SST seasonal forecasts. Firstly, the estimate of the ocean initial state through data assimilation is not perfect, and El Niño forecasts are sensitive to the accuracy of oceanic initial conditions (Ji and Leetma 1997, Rosati *et al.* 1997, Alves *et al.* 2002). Secondly, synoptic atmospheric variability which is to a large extent unpredictable beyond a few days, can also significantly impact the SST evolution.

It is thus necessary to sample the effect of the uncertainties in oceanic initial state and atmospheric synoptic variability in a seasonal forecast system. This can be done through ensemble forecasting, whereby not just one, but many integrations are run, starting from slightly different initial conditions, the distribution of which should be representative of the actual errors in the analyses (Palmer 2000). The spread of the ensemble then provides some measure of the level of uncertainty attached to the forecast.

To some extent, this approach was followed in the first seasonal forecasting system at ECMWF, denoted System1 (S1). Each day, a new ocean analysis was generated by the ocean data assimilation system. The atmospheric analysis was from the operational numerical weather prediction system. Starting from these initial conditions, the coupled model was then integrated forward for 6 months. Over a one month period, between 28 and 31 forecasts were made which were then grouped into an ensemble which was corrected to account for the climate drift of the coupled model (Stockdale 1997). Statistical treatment was applied to deliver ensemble-mean forecasts and probability distributions for SST, precipitation and 2-m temperature seasonal forecasts (Stockdale *et al.* 1998).

This approach has three main drawbacks. Firstly, since daily integrations are grouped into one-month ensembles, this introduces a delay of at least 15 days in the forecast delivery date. Secondly, members of the ensemble start from oceanic initial conditions up to 30-days apart, which are not statistically indistinguishable. Thirdly, the ensemble should sample the effects of unpredictable atmospheric internal variability and the error in the oceanic initial conditions. Because of the chaotic behaviour of the atmospheric synoptic variability, the choice of atmospheric initial states is relatively unimportant: one month into the integration, it is likely that all the members will have diverged to produce completely different realizations of the synoptic atmospheric flow. With 30 different realizations, the effect of atmospheric internal variability on the evolution of the SST field is

well sampled in S1. The situation is different for the ocean: some patterns of error in the oceanic initial conditions can grow much more quickly than others (e.g. Moore and Kleeman 1996). The spread of the forecast will thus depend strongly on the choice of the oceanic initial states of the ensemble. The initial states of the ocean in S1 span one month but the typical change of the ocean in this time is not necessarily representative of the actual error in oceanic initial conditions.

In this paper, we present an ensemble generation method for seasonal forecasting which attempts to overcome those three drawbacks. We will start all the integrations in one burst at the start of the month to allow a more timely delivery of the forecasts. We also design a perturbation technique to generate initial states of the ocean which are statistically indistinguishable. Finally, the perturbation technique is designed to create realistic estimates of uncertainties in the oceanic analysis originating from the main sources of error in the boundary conditions (namely wind forcing and SST). This will hopefully allow a better representation of the uncertainty in the oceanic initial conditions. The methodology presented in this paper was used in designing the ECMWF seasonal forecasting system 2 (S2), and some results from S2 will be presented at the end of section 3.2.

2 The seasonal forecasting system

2.1 The coupled model and ocean data assimilation scheme

The forecasting system that is used in the present paper is very similar to S2, except for the ocean resolution, which is lower. However, this should not be a major factor in the results of this paper, where we only want to evaluate the spread in the ensemble caused by various ensemble generation methods. We will show some results from S2 at the end of this paper that confirm that both forecasting systems perform very similarly in terms of ensemble spread.

The ocean model used in this study is HOPE (Wolff *et al.* 1997), with a 2 degree zonal resolution (against 1 degree in S2), and a meridional resolution of 0.5 degree (against 0.33 degree in S2) near the equator. There are 20 vertical levels, with a resolution of 20m near the surface (versus 29 levels and 10m in S2). During the analysis phase, the ocean model is forced by the wind stress, and heat and freshwater fluxes from the ECMWF operational numerical weather prediction (NWP) system. A strong relaxation to the Reynolds and Smith *et al.* 2002 SST product is applied to ensure that the coupled model starts from SSTs close to observations. Every 10 days, all the available oceanic temperature data are gathered and an analysis is generated, using an optimal interpolation scheme (Smith *et al.* 1991, Alves *et al.* 2002). The ocean model salinity is corrected using in situ temperature and the model temperature-salinity relation, according to the method of (Troccoli *et al.* 2002). Velocity corrections are also applied using a geostrophic balance relation, following (Burgers *et al.* 2002).

The atmospheric model is the ECMWF Integrated Forecasting System in its cycle 23r4 version. It is run at T_L95 resolution with 40 levels, just as in S2. The initial conditions for the atmosphere are provided from the ECMWF operational analysis.

In forecast mode, the oceanic and atmospheric models exchange fluxes every day through the Oasis software (Terray *et al.* 1995). No flux correction is applied to the exchanged fluxes. The climate drift is corrected by removing the mean coupled model drift, computed over the whole forecast. In this study, 5-member ensemble forecasts are run for the 1991-98 period, starting from the 1st of January, April, July and October of each year. The methods used for generating the various ensembles are discussed in the next section.

2.2 Perturbation technique to generate the ensemble

To sample the uncertainty in the forecasts, three kinds of perturbation techniques have been introduced. One is applied to the atmosphere during the course of the coupled model integration (stochastic physics). The two others (wind stress and SST perturbations) are applied to the ocean, when generating the ocean initial conditions.

The atmosphere is perturbed throughout the coupled integration by using stochastic physics (Buizza *et al.* 1999). This method mimics the effect of the uncertainties in the physical parametrization on the ensemble spread, and is used in the ECMWF medium range weather ensemble prediction system. It consists in randomly perturbing the atmospheric parameterised physical tendencies at each time step of the model integration. This introduces a random component in the atmosphere which results in a divergence of synoptic systems in the early range of the forecast, and thus contributes to spread in the SST forecast.

Errors in oceanic initial conditions can originate from errors in the forcing, in the analysis method, in the ocean model and in the observations. Since the equatorial circulation is largely driven by wind stress, and SSTs have an important role in the forecast, errors in the winds and SSTs are likely to represent important sources of error in the oceanic initial conditions. Ocean model errors are quite difficult to sample other than by methods of multi-model ensembles. So, in this paper only perturbations in SSTs and wind stress will be considered in a first step toward generating uncertainties in the oceanic initial conditions. We present below a method designed to construct perturbation patterns in SST and wind stress, representative of the typical random part of the error in these fields.

The SSTs used in S2 come from an SST analysis performed at NCEP and denoted OIv2. Errors in SST analyses can arise through errors in time interpolation, error in the analysis process and error in the observations of SST. The first is evaluated by comparing the SST analysis with an analysis one week later. The second is evaluated by forming the difference between interannual weekly anomalies of two different SST products produced at NCEP viz the 2DVAR and OIv2 analyses (Reynolds *et al.* 2002). Typical patterns of uncertainties for the corresponding calendar week are then constructed by combining randomly these two types of patterns. These perturbations are added to the ECMWF ocean temperature analysis to create the initial conditions for a forecast. The perturbation has full value at the surface but is ramped down to zero at 40m depth. The standard deviation of SST perturbations (figure 1a) is particularly strong over the ice margin regions in the Northern Hemisphere and over the dynamically most active regions (Gulf Stream, Kuroshio, circumpolar current...) where the perturbations can exceed 1^o *C*, though differences in these regions are probably not very important in influencing equatorial SSTs. Over the tropical Pacific, the perturbations have an amplitude of about half a degree, but can exceed this over the eastern tropical Pacific. SST perturbations are only applied from 60S to 60N.

Patterns of wind stress perturbations are constructed from differences between interannual monthly anomalies of the ERA-15 reanalysis and Southampton Oceanography Centre (SOC) monthly mean wind stresses (Josey *et al.* 2002), for the period 1980-1997. Wind stress perturbations are stronger in high latitudes as seen in figure 1b. Although the amplitude may appear to be weak in the tropics, in fact the perturbation can reach up to 30% of the amplitude of the mean wind stress in that region, and is therefore not small. The wind stress perturbations are stratified by calendar month, and are then used to randomly perturb the daily wind stress that forces the ocean model. The daily perturbations are obtained by linearly interpolating two randomly picked wind stress patterns representative of consecutive months (the full pattern being applied to the middle of each month).

The ocean data assimilation is performed in the following way. An ocean Optimal Interpolation (OI) analysis is performed in which all subsurface ocean data 5 days before to 5 days after the nominal analysis time are

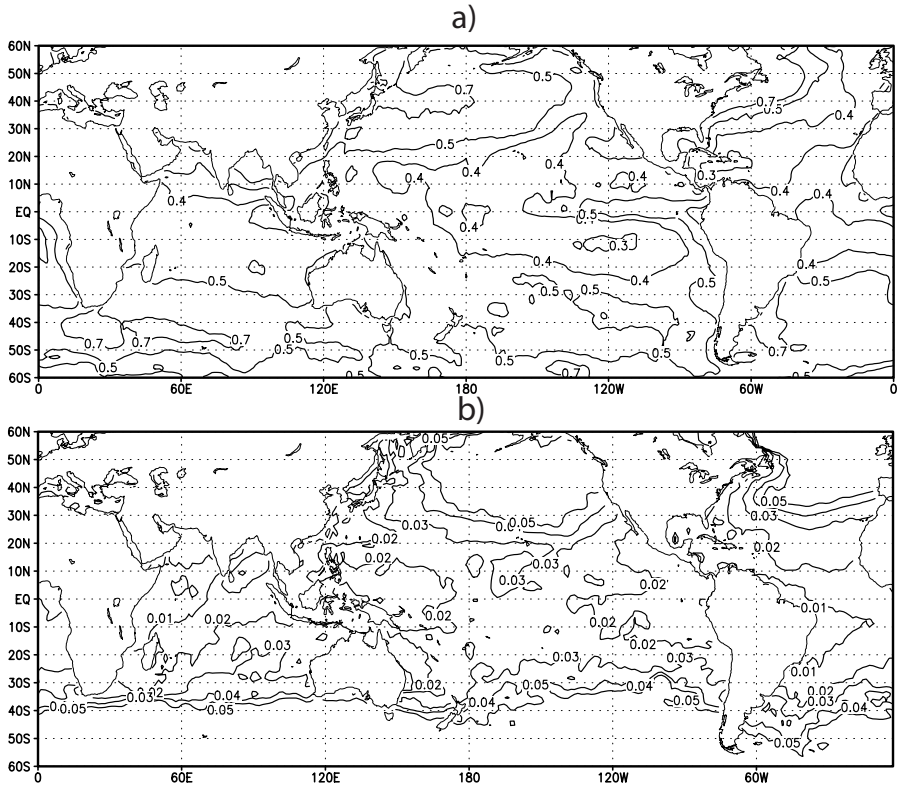


Figure 1: a) Root mean square of the SST perturbation patterns used to generate the ensemble. Contour interval every 0.1^{o}C up to 0.5^{o}C and then 0.2^{o}C b) Root mean square of the wind stress perturbation patterns used to generate the ensemble. Contour interval every 0.01 N/m^2 up to 0.05 N/m^2.

merged with a first guess. The ocean model is then integrated forward 10 days forced by the wind fields and heat and fresh water fluxes from the atmospheric analysis system; this provides the first guess for the next ocean analysis. No OI analysis of model SST is performed: rather, during the running of the model, the model surface layer is relaxed strongly to the observed SST. In order to represent some of the uncertainty in the ocean state arising from uncertainty in the wind field, the ocean analysis system consists of an ensemble of five independent ocean analyses, making use of the wind perturbations described above. Member 0 has no wind perturbations applied, members 1 and 2 have the same patterns but of opposite sign, and likewise for members 3 and 4. This method of ensemble generation means that the ensemble-mean winds are not biased relative to the unperturbed member: only the spread is increased to represent uncertainty in the winds. Consequently, the ensemble-mean ocean analysis is unbiased, (or nearly so); only the spread is increased. Wind perturbations are applied only during the analysis phase: they are not applied during the coupled integrations. The SST perturbations are also chosen to sum to zero and thus not to bias the ensemble mean. The SST perturbations are not present during the analysis phase, but are added to the initial conditions at the start of a forecast.

3 Results

3.1 Ensemble spread of the different experiments

The results of several experiments will be discussed in this section. In three of these experiments, the ensemble has been generated using stochastic physics (experiment SP), wind perturbations (WP), temperature perturbations (TP) individually. In the fourth experiment (SWT) all three perturbations (stochastic physics, wind perturbations and SST perturbations) are applied. The fifth experiment uses perturbations as in SWT but has no data assimilation; it is denoted NDA. A 5-member ensemble has been constructed for each forecast date. In each case, there is an unperturbed forecast and four other members constructed as follows. For the stochastic physics, four different seeds are used for the random number generator used in perturbing the physical tendencies. For both the wind and temperature perturbations, two groups of two experiments with symmetric perturbation patterns are used.

The sixth experiment (LA) is set to mimic the lagged-average approach used in S1. To be able to compare the spread in the LA ensemble with the other experiments, we created a 5-member ensemble of forecasts starting from initial dates at -12, -6, 0, +6, +12 days relative to the initial date of the other experiments. The range of start dates (-12 to +12) is similar to that used in S1 (-15 to +15). Since the atmospheric initial conditions also contain information on soil wetness, sea-ice distribution and snow cover, the LA experiment also includes some perturbations of the land surface boundary conditions and of the ice edges, which are not present in the other experiments.

The impact of the wind stress perturbations on the ocean is illustrated in figure 2. Panels a) and b) illustrate the effects of wind perturbations on the temperature along the equator. The largest errors are located near the thermocline, which is not surprising since the wind stress induces vertical displacement of the thermocline via Ekman pumping. Panel a) shows how uncertainties in the wind forcing translate into large uncertainties in the oceanic state when data assimilation is not used. Panel b) shows how data assimilation reduces uncertainties over much of the three oceans, though there is a limited region in the thermocline in the eastern Atlantic where the run with data assimilation has a larger spread than the run without. This is a region where not many data are available and where quality control might accept some observations in some members of the ensemble, and reject them in others, thus slightly increasing the spread. The spread of both experiments diminishes to weak values (of the order of 0.1 to 0.2^{o}C) close to the surface because of the strong relaxation to observed sea surface temperature. However, during the early stages of the forecast, temperature differences in the thermocline will

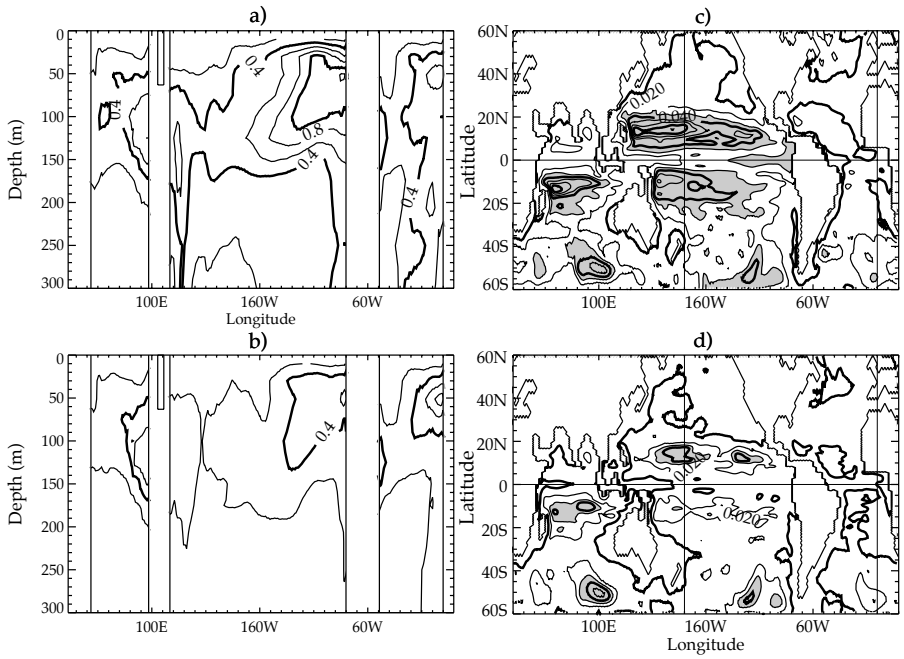


Figure 2: a) Vertical section along the equator of the root mean square spread of the five-member ensemble oceanic subsurface temperature analysis generated by applying wind stress perturbations in the absence of oceanic data assimilation. Contour interval every 0.2^{o}C b) As in a, but for an experiment with oceanic data assimilation. c) Map of the root mean square spread of the five-member ensemble oceanic sea-level analysis generated by applying wind stress perturbations in the absence of oceanic data assimilation. Contour interval every cm. d) As in c, but for an experiment with oceanic data assimilation.

start to influence the SST through modification of vertical advection and mixing, and will result in larger SST differences.

Panels c) and d) show the corresponding spread in the oceanic analyses of sea level, which is representative to some extent of the perturbations of the upper ocean heat content, thus highlighting regions where SST anomalies are likely to develop during the early months of the forecast. It is interesting to note that regions of largest spread in the wind do not immediately translate into regions of large spread of the sea level, but that the imprint of oceanic structures can be seen. Panel c) shows that the regions of largest spread are the equatorward part of the subtropical gyres in the Pacific and Indian oceans. These are where the horizontal gradients of upper ocean heat content are large and where the thermocline is closer to the surface and thus more sensitive to wind changes. Panel d) shows how data assimilation acts to constrain the upper ocean heat content and collapses the spread. In the TAO region, the spread goes down from 3-4 cm to 1-2 cm when data assimilation is applied. The remaining areas of high spread are representative of regions with a poor data coverage, such as the southern Ocean, the southern subtropical Indian Ocean and some areas in the north Pacific subtropical gyre. The overall picture from this analysis is that the uncertainties in the ocean state originating from uncertainties in the wind are considerably reduced when data assimilation is applied, in particular in the three equatorial oceans.

Figure 3 shows the ensemble spread of the third-to-fifth month SST forecast for experiments SP in panel a) and WP in panel b). As the spread for the WP, TP and LA experiments is very similar, only that from WP is plotted. This spread is in turn very similar to that of SP. From an ensemble generation perspective, this suggests that all the perturbations applied here are equivalent in terms on their impact on the spread of the third-to-fifth month forecasts. The spread of experiment SP is generated by atmospheric internal variability only whereas the spread in WP, TP, LA comes both from amplifications of perturbations in the oceanic initial conditions by unstable modes of the coupled system and stochastic forcing by atmospheric internal variability. The fact that the spread is very similar in all four experiments suggests that the atmospheric internal variability is the main source of spread in the coupled experiments. At mid latitudes, this is consistent with Frankignoul and Hasselmann (1977) who proposed that midlatitude SST variability is mainly driven by stochastic atmospheric forcing. In fact, at mid latitudes, the ensemble spread in figure 3 is similar to the observed mid-latitude interannual variability (in figure 4a).

Figure 4 shows the spatial structure of the interannual variability in observations and in the forecasts ensemble mean. Comparing with figure 3, we can see that the main difference between mid-latitudes and the equator is the signal to noise ratio. In mid-latitudes the ensemble spread in figure 3 is similar to the observed interannual variability (figure 4a), which confirms the paradigm of integrated noise. Figure 4b has weaker than observed interannual variability because it is based on a 5-member ensemble mean. The actual interannual variance of the model would be equal to the variance of the ensemble spread plus the interannual variance of the ensemble mean.

In the Equatorial Pacific the interannual variability is much larger than the ensemble spread. The signal-to-noise ratio is often used as an indicator of potential predictability. In the model, this ratio is quite high, which would be consistent with a large part of the variability being due to a predictable signal. However, it is difficult to assess how realistic is the model signal-to-noise ratio.

A region of interest for seasonal forecasting is the tropical Pacific which has a marked variability associated with El Niño. Figure 5 shows the spread for the different experiments as a function of the forecast lead time in two key regions usually associated with ENSO: Niño3 and Niño3.4, representing the Eastern and Central Pacific respectively. In the Eastern Pacific, during the first month, experiments WP and SP display a different behaviour to the other experiments. Neither experiment has perturbations of the initial SST analyses. Despite the wind perturbations applied in experiment WP, the strong relaxation to observed SST during the analysis prevents any significant spread in SST at initial time. The time that it takes for subsurface perturbations generated by the wind

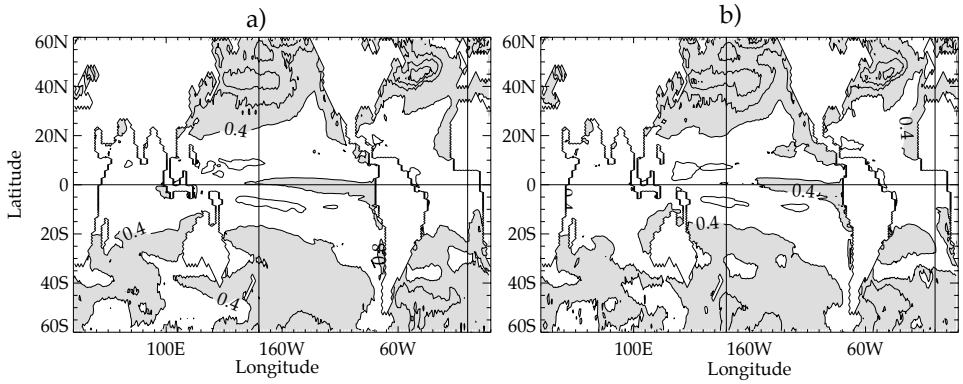


Figure 3: Ensemble spread (standard deviation of the SST interannual anomaly forecasts) for months 3, 4 and 5 for the ensembles constructed using: a) stochastic physics; b) wind stress perturbations. Contour interval 0.2°C, shading above 0.4°C

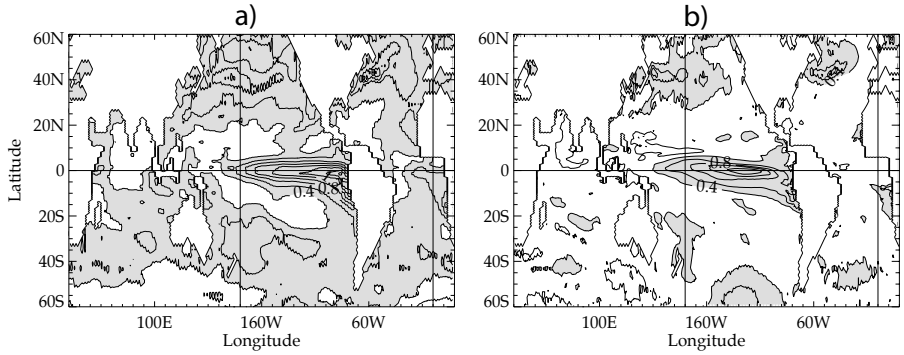


Figure 4: Map of the root mean square of interannual SST anomalies, for a) observations and b) months 3-5 ensemble-mean hindcasts for experiment SWT. Contour interval 0.2°C shading above 0.4°C

to influence the surface and to create a spread in SST may depend on the region, on the model physics and on the resolution. The higher resolution version of the same model shows larger sensitivity to wind perturbations during the first month, probably because the higher vertical resolution allows the subsurface differences to surface faster. In region Niño3.4 (figure 5c) the spread produced by experiment WP is not so different to the spread produced by TP, although the spread induced by TP is still the largest at month 1. Since the SST is not perfectly known, experiments TP, SWT and LA probably provide a better estimate of the uncertainties during the early range of the forecast than WP and SP. In TP and SWT, there is a slight decrease of the spread during the first week of the forecast. This might correspond to the noisy component of the SST perturbations that is dissipated in the coupled model because it does not have a physical structure.

For all the lead times beyond month 3, the spread in SST forecasts given by the TP, WP, and SP methods are very similar. What differences there are might be linked to sampling issues. More spread might have been expected in WP than in TP since the tropical eastern Pacific is known to sustain interannual variability (the El Niño phenomenon) that arises from unstable air-sea interactions (see Neelin *et al.* (1998) for a review). Since El Niño predictability is generally believed to stem from the knowledge of oceanic initial conditions, one would normally expect uncertainties in oceanic initial conditions such as those generated in the WP experiment, to give rise to corresponding uncertainties in El Niño forecasts. However, the spread in experiment WP (with uncertainties in initial conditions) is equivalent (or if anything, smaller) than the spread arising from purely internal atmospheric variability (experiment SP). The fact that all the ensemble methods give similar results suggests that all of the perturbation methods project efficiently onto the unstable mode or modes, and that the specific SST and wind perturbations used are not much better than atmospheric noise in this regard. Alternatively, there may be no unstable mode in the system, and the spread in equatorial SST may be the result of integrating noise, as in mid latitudes.

It was shown earlier that the oceanic initial state is well constrained in the tropical Pacific by observations. Because of this, the wind perturbations generate only a small spread in oceanic initial conditions (see figure 2). Figure 5 also shows the spread of an ensemble with stochastic physics, wind and temperature perturbations but without data assimilation (experiment NDA). This spread is larger than any of the other experiments at any lead time, thus underlining the potential of errors in oceanic initial conditions to lead to larger uncertainties in the forecast than the effect of atmospheric internal variability. The difference in spread for months 3 to 5 is not especially large, though, and the consequences of this will be discussed in more detail in the discussion section.

3.2 Reliability of the ensemble generation method

We have shown in the previous section that methods WP and SP tend to underestimate the spread during the early range of the forecast. Methods LA, TP, SWT give a similar spread over most of the period. Since they allow a more timely delivery of the forecasts, SWT should be preferred to LA in an operational system. Since, in principle, SWT samples more components of the forecast error sources, it was the method chosen to be used in the ECMWF seasonal forecasting System 2 (S2). We will thus now explore this method further to see if the ensemble generated is representative of the actual errors in the forecasts. Diagnostics similar to those presented below were applied to the other experiments, and gave results consistent with those presented here.

In a perfect ensemble forecasting system, the verifying SST should fall within the ensemble range most of the time. A precise requirement is that the root mean square error of the ensemble-mean SST forecast should be equal to the spread (standard deviation) of the ensemble, within sampling error. Figure 6 shows the spread of experiment SWT for months 3-5, and the associated rms error of the ensemble-mean forecast. Outside the tropics, the ensemble spread and rms error of the forecast have a similar amplitude and spatial pattern, suggesting that the ensemble system accounts reasonably well for forecast uncertainties. The area where most of the predictable signal is expected is the tropical region, and especially the tropical Pacific ocean. In this

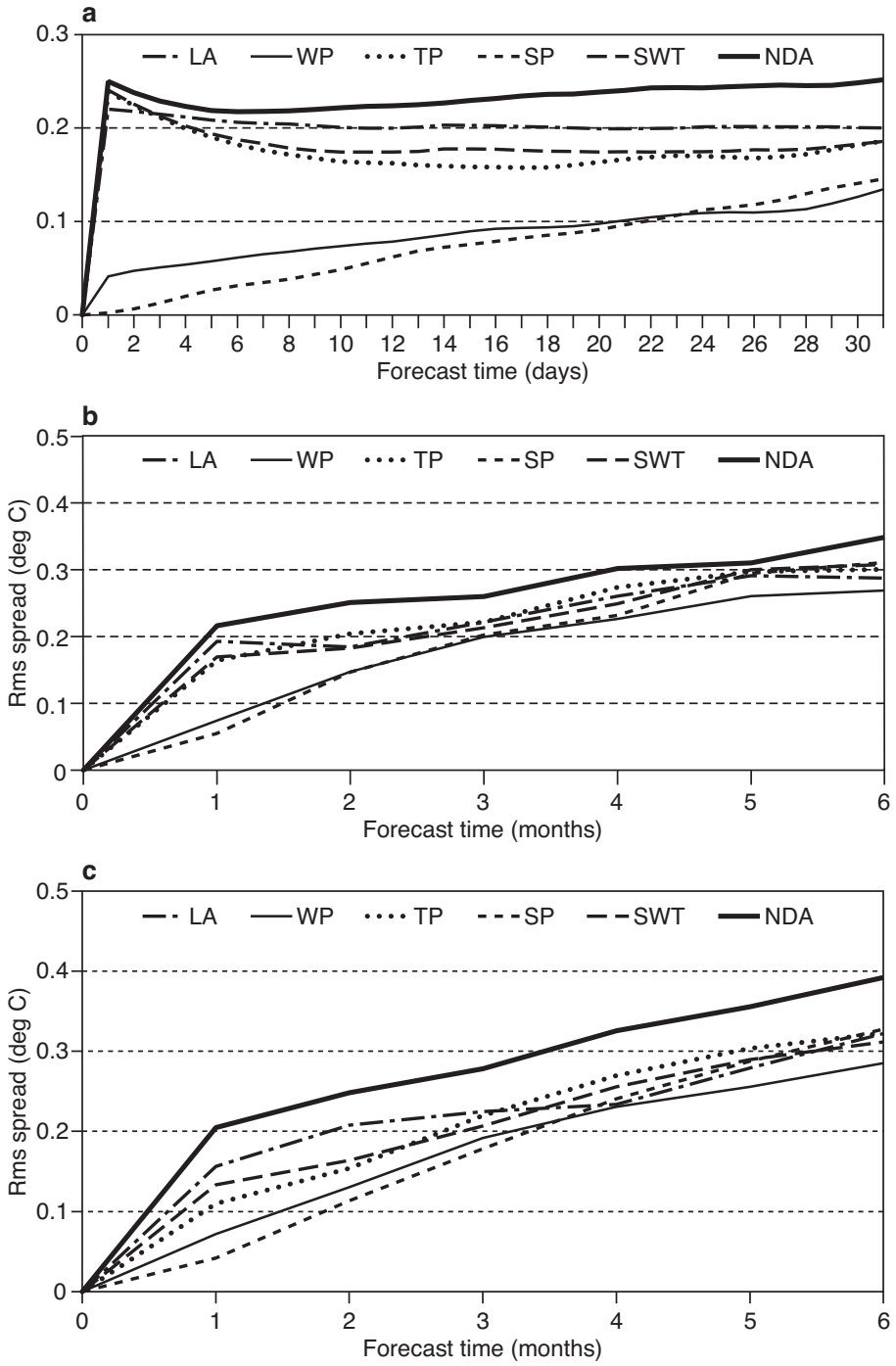


Figure 5: Ensemble spread of SST forecasts as a function of lead-time. Panels a) and b) show the spread in the Niño3 region (150°W-90°W, 5°N-5°S) and panel c) in the Niño3.4 region (170°W-120°W, 5°N-5°S). Panel a) shows the daily evolution of the spread during the first month, while panels b) and c) show the time evolution (in months) for the whole forecast period.

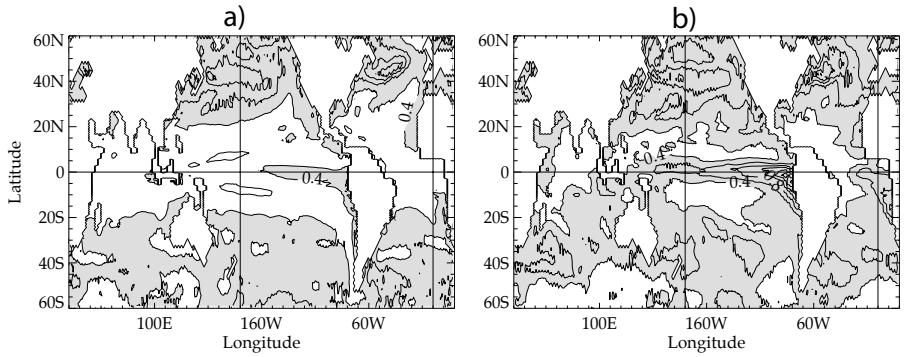


Figure 6: a) Ensemble spread (standard deviation of the SST interannual anomaly forecasts) for months 3, 4 and 5 for experiment SWT. b) Root mean square error of the ensemble mean interannual SST anomaly for months 3,4 and 5 for experiment SWT. Contour interval 0.2^{o}C, shading above 0.4^{o}C

region, although the ensemble technique did capture successfully the patterns of actual error, the amplitude of the error is clearly much larger than the spread of the ensemble. This is an indication that significant sources of forecast error were not taken into account in the ensemble generation method.

Figure 7a shows the ensemble mean rms error (solid line) and the ensemble spread (dashed) as a fuction of the forecast lead time in region Niño 3 for experiment SWT. As a reference, the rms error of a forecast using persistence is also shown (dashed-dotted line). In the first month, the ensemble spread is close to the rms error. However, by month 3 the ensemble spread size is already less than half the size of the error. The most straightforward explanation for the gap between the ensemble estimate of expected errors and the level of errors actually observed, is that model errors are degrading the forecasts. Indeed, we know that there are many sources of model errors in the coupled model, of types and magnitudes that might be expected to cause significant forecast errors. For instance, figure 4 showed that model and observations had different patterns of interannual variability at the Equator, with the maximum of equatorial variability in the model clearly shifted away from the coast and towards the central Pacific.

Another known error in our forecasts is the damping of the amplitude of interannual variability (Anderson *et al.* 2003). That is, as the coupled model forecasts run forward in time, there is an overall tendency for the amplitude of SST anomalies to reduce. This can be seen in figure 7b, which shows the rms amplitude of Niño 3 as a function of forecast lead time. The model amplitude is based on the individual model integrations, not the ensemble mean. This damping is believed to be due to errors in both components of the coupled model, with a particular problem in the surface wind variability in the atmospheric model.

An important point is that the damping of the model variability will also have an impact on the ensemble spread, reducing its size. That is, we have good reason to believe that our model ensemble underestimates the spread that would exist in a perfect forecasting system with realistic initial errors.

The mismatch between forecast error and ensemble spread is largest in the eastern Pacific. If the reduced amplitude of the model anomalies were the main source of error, then by scaling the ensemble such that the overall amplitude of model variability matches that of the observations, one might hope both to get a larger, more realistic value for the ensemble spread and to improve the ensemble-mean forecast. By such a scaling, we found that it was indeed possible to increase the spread and to narrow, though not close, the gap between model spread and model error, but the rms error was not reduced. In fact a reduction in rms error would not generally be the case as the scaling that would minimise the rms error is not the same as that needed to correct

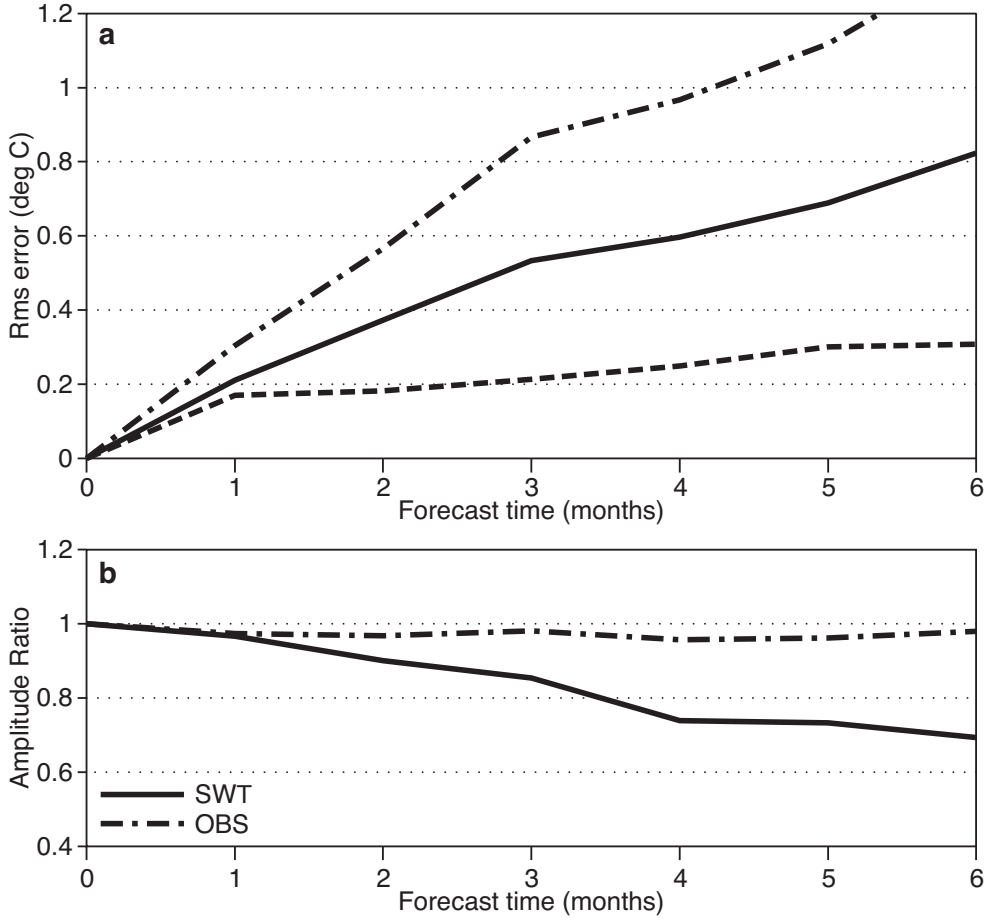


Figure 7: a) Evolution of the rms error of the ensemble mean (solid line) and the ensemble spread (dashed line) for the Niño 3 SST as a function of the hindcast lead-time. The dot-dashed curve represents the error from using persistence as predictor. b) The model/observed amplitude ratio for the Niño 3 SST interannual anomalies. Dot-dash is for persistence of observed SST. The solid curve is the average amplitude ratio of the individual ensemble members from experiment SWT.

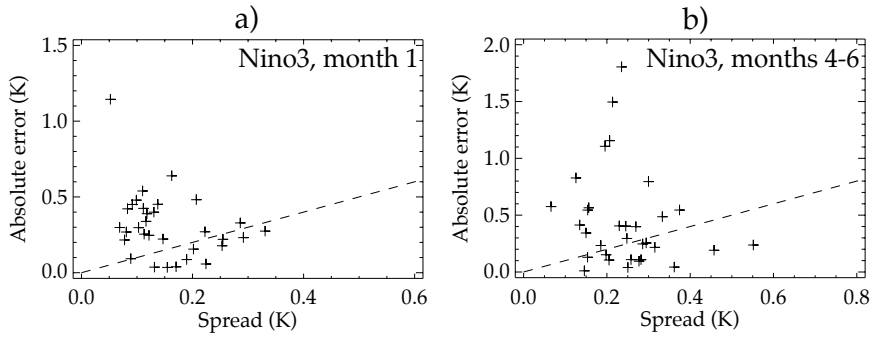


Figure 8: Scatterplot of the SWT ensemble-mean Niño3 SST forecast absolute error against the ensemble spread, a) for month 1 hindcasts and b) for months 4-6 hindcasts.

the model amplitude. There is evidently a broad spectrum of model error that cannot easily be removed by *a posteriori* processing of the data.

This calculation demonstrates that correcting the ensemble spread for one known type of error still leaves a large gap between the inferred predictability limit and the actual forecast performance. Other types of error may exist in the coupled GCM that also act to reduce the ensemble spread, and we cannot be certain of exactly where the predictability limit lies. It would be surprising, though, if the cumulative effect of model errors had a large impact on predictability, but did not degrade the skill of the ensemble mean forecasts. A large part of the mis-match between ensemble spread and actual forecast error is likely to be due to model error causing errors in the forecast.

In a perfect model scenario, predictability estimates depend a great deal on adequate sampling of the uncertainty in the initial conditions and the directions of maximum error growth. This is usually the assumption in the medium range weather forecasting. The results presented in this section suggest that the seasonal forecasting system is far from this perfect model hypothesis: errors in the ensemble mean and in the ensemble spread are dominated by model errors. Therefore, methods to sample model error (multi-model or others) are a priority when designing future strategies for ensemble generation in seasonal forecasting.

3.3 Spread/skill relationships

Another desirable feature of an ensemble generation method is to provide an estimate of the uncertainties of each particular forecast. Larger spread for one particular ensemble forecast should indicate that uncertainties are larger and that less confidence should be given to the forecast. One way to assess whether an ensemble forecasting system displays this desirable feature is to look for spread-skill relationships, as was done for example by Moore and Kleeman (1998) and Kleeman and Moore (1999). Figure 8 shows a scatterplot of the ensemble mean absolute error of the Niño3 SST against the spread of the forecast, for experiment SWT. In an ensemble with a spread-skill relation, one would generally expect larger errors to occur in situations where the spread of the ensemble is larger. There is no indication of such a relationship in figure 8, neither at a 1 month nor at 4-to-6 months lead time, due in particular to some outliers with large forecast error for weak or average spread values. However, the problem with experiment SWT is that the ensemble contains only 5 members, which makes the estimate of the spread not very reliable. Further, Moore and Kleeman (1998) did not find a spread skill relationship when they used model (rms) error as a measure of skill, but only when they used correlation.

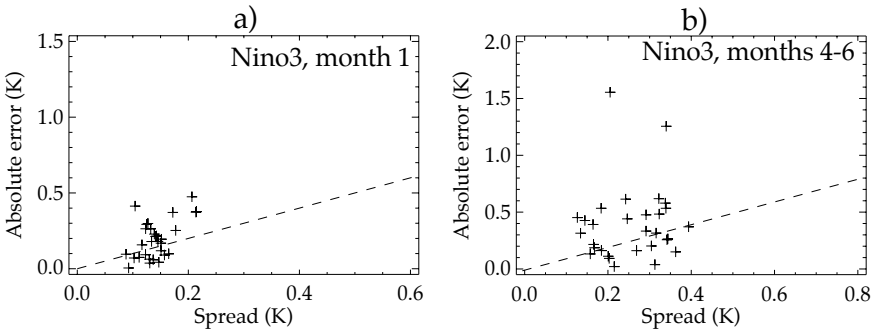


Figure 9: As in figure 8, but for the ECMWF seasonal forecasting system2 (S2) 40 members hindcasts starting in November and May, for the period 1987-2002.

We thus tried to investigate if the spread-skill relationship would hold better with an ensemble generation method similar to SWT, but with many more members. We based our analysis on a set of hindcasts performed with the ECMWF operational seasonal forecasting system S2. The main difference in the models and ocean data assimilation system between SWT and S2 is that the latter has higher oceanic resolution. The S2 hindcasts are for the 1987 to 2002 period, and were run for November and May starts only. Each ensemble hindcast has 40 members. The ensembles are constructed in the following way: 5 ocean analyses are performed, with wind perturbations as in WP. Each of these analyses is perturbed with 8 random SST patterns, applied as in TP. One hindcast is run from each of those 40 initial oceanic states, with stochastic physics activated in the atmosphere. The resulting experiment is thus similar to SWT for the ensemble generation method, but for a different period, different starting dates, and larger ensemble. In this experiment, the 40 member ensemble results in a much better estimate of the spread. Figure 9 shows the absolute error-spread relationship for those S2 hindcasts. Except for two forecasts with large error for months 4-6, there is less scatter for the S2 40-member ensemble than for the SWT 5-member ensemble. Some (weak) spread-error relation appears for both month 1 and months 4-6 hindcasts. However, this relation is not yet very clear, and there are evidently outliers in the error-spread relation. This suggest that although there is one (weak) component of the hindcast errors which can be linked to initial SST errors and uncertainties in the wind stress, there is another large component of forecast errors not yet sampled by our forecasting system. We will explore this issue further in the discussion section.

3.4 Time variability of ensemble spread

We have seen above that the use of 40 members in S2 provides a more robust estimate of the spread. It is thus interesting to explore in more detail how this spread varies in time and how it is linked to the seasonal cycle and interannual variability. Figure 10 shows the spread at months 1 and 5, of the S2 hindcasts described above as a function of the start date. The amplitude of the interannual observed SST signal is also indicated.

There is a hint of lower spread in the later part of the record, as might be expected if the ocean state is better constrained for more recent periods than for the earlier periods where less data were available. This is also true looking at regions other than Niño 3, and a bigger spread is particularly evident in 1987. However, in overall terms the level of uncertainty in the forecasts has remained approximately constant, despite the improvements in the observing system. Note that the perturbations that we apply to wind do not account for the improvement in our knowledge of this field in recent years. Note also that the spread in the first month is dominated by SST perturbations (whose impact is independent of our data assimilation system), and in the later months is dominated by the random noise component. This implies that the potential impact of improvements in the subsurface ocean observing system on the forecast spread will be limited, although observing system improvements

should have a beneficial impact on forecast skill. Finally, note that the impact of data on our analyses is of course a function of our analysis system.

Month-one spread is larger than usual in November 1988 and November 1999, i.e. at the peak of two rather large La Niña events. This might be explained by the fact that perturbations in the ocean initial subsurface state are more easily communicated to the surface during La Niña, i.e. when the thermocline is closer to the surface in the eastern Pacific. In the long range (5 months lead time) hindcasts, the most obvious signal in the spread is the seasonal cycle.

There is generally more spread in the forecasts starting in May (for which month 5 is September) than for the forecasts starting in November. This does not entirely accord with the so-called ”spring barrier hypothesis”, some versions of which suggest that March and April SSTs are less predictable than September and October SSTs. Nonetheless, there is an obvious mechanism for the seasonal cycle in spread as seen in our experiments. September is the month for which the easterlies and the intensity of the upwelling in Niño 3 are large. Those are thus the months for which the subsurface and surface ocean are best connected, and where subsurface temperature perturbations, either coming from the oceanic initial conditions or generated during the first 4 months of the coupled model integration by atmospheric internal variability, can most influence the SST. The fact that the spread is unusually low in September 1997, where the thermocline was much deeper than usual due to the El Niño conditions, seems to confirm this. Finally, the only year that displays an unusually strong spread for the November start hindcast is 1999. This might also be seen as a consequence of interannual variability. The La Niña at this period is associated with an unusually shallow thermocline, and thus a better connection between potential subsurface temperature anomalies and the surface.

The fact that there is not a one-to-one relationship between the interannual variability and the spread can be explained as follows. A shallow thermocline creates favourable conditions for subsurface anomalies to influence the SST. The subsurface anomalies themselves will depend on many factors, including observation distribution, the amplitude of the randomly picked wind patterns used to generate the ensemble, the amount of atmospheric internal variability during the early months of the forecast, etc.. . During some years where the interannual variability would favour a strong communication between sub-surface and surface, those factors might give only a small spread in the oceanic subsurface temperature, and thus only a small spread in the forecasts. The use of 40 different wind perturbations would certainly give us better sampling in this regard.

4 Summary and discussion

4.1 Summary

In this paper, we have explored ensemble generation methods for a seasonal forecasting system with a coupled general circulation model. The ensemble generation method we have constructed samples errors in the oceanic conditions due to uncertainties in the wind stress and sea surface temperature products, and the effect of atmospheric internal variability. The effects of those uncertainties on the forecast were examined individually and collectively, and compared to the more usual lagged-average approach where forecasts starting from consecutive dates are grouped to define an ensemble.

Uncertainties in observed SSTs define some lower bound on the uncertainties of the forecasts. It was found that experiments sampling only uncertainties due to wind stress or atmospheric internal variability underestimate the spread during the early months of the forecast, because it takes time for subsurface temperature anomalies or for atmospheric internal variability to generate some spread in SST. It is thus advisable to sample initial SST errors in an ensemble generation system. At longer lead times (months 3 to 6) all of the methods explored gave

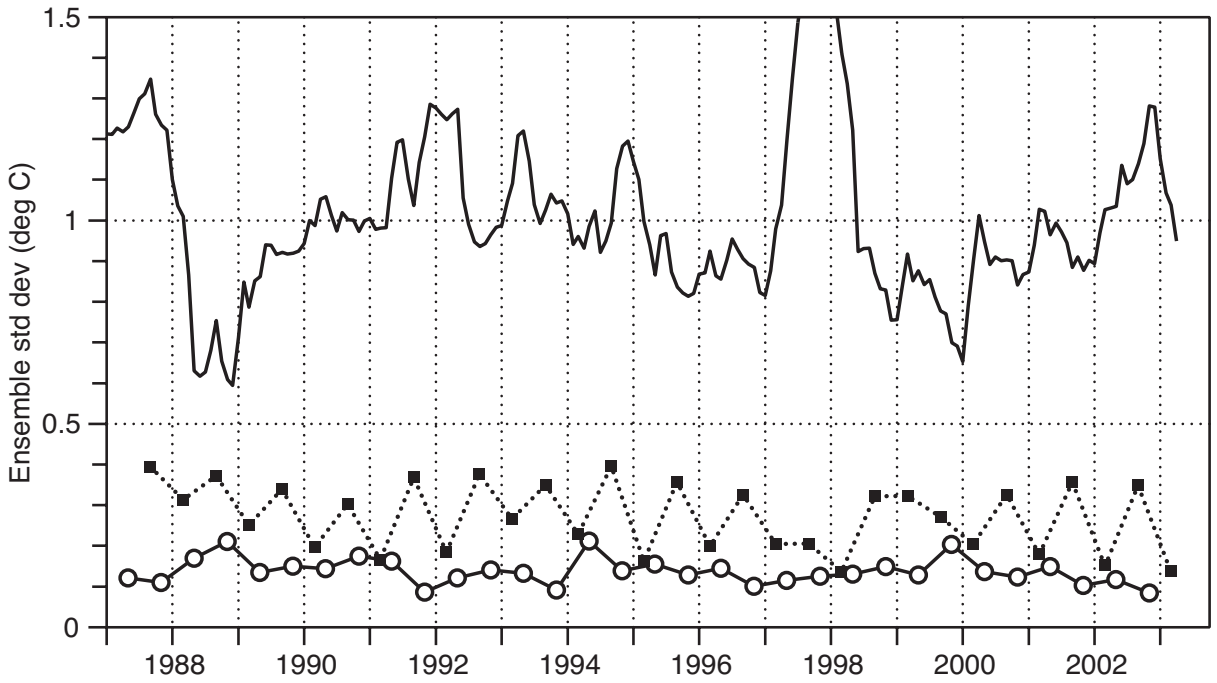


Figure 10: Spread of the ECMWF seasonal forecasting system2 (S2) 40 members Niño 3 SST hindcasts starting in November and May, for the period 1987-2002 as a function of the verification date. The circles are for month-1 hindcasts and the squares for month-5 hindcasts. The upper solid curve indicates the observed interannual SST anomaly in Niño 3, offset and at a reduced scale.

a similar spread. Because SWT allows a more timely delivery of the forecasts than the lagged-average method, and because it samples uncertainties in wind, SST and atmospheric variability, it is the method that was chosen to construct the ensemble forecasts for the ECMWF seasonal forecasting system 2.

The spread of the forecasts in all the methods is, however, significantly smaller than the rms-error of the forecasts. This indicates that uncertainties that are not sampled by the ensemble contribute significantly to forecast error. Results suggest that model error, and not error in the initial conditions, is the main source of forecast error, affecting not only the value of the ensemble mean but also the ensemble spread. Improved models and better sampling of model error are then required for more realiable seasonal forecasting systems.

Error-spread relationships were sought to see if the spread of the ensemble can confidently be used to define some uncertainties on the forecasts. No clear relation was found in ensembles with 5 members. Some error-spread relation begins to emerge in 40-member hindcasts made using the ECMWF S2 seasonal forecasting system, but this relationship is weak. This last point once again indicates that there are factors contributing to forecast error which are not accounted for by the present strategy.

The factors contributing to the spread in initial conditions and in the forecast were also examined in this paper. It was found that applying the wind perturbations during the analysis generation creates the largest spread in the thermocline region, but that the SST spread is weak due to the the strong SST relaxation. The spread is largely reduced over most of the tropics when data assimilation is applied, suggesting that the oceanic subsurface temperature structure is rather well constrained in those regions. As a result of this, the forecast spread in experiments with data assimilation and wind perturbations is much smaller than in experiments without data assimilation. Finally, the analysis of the S2 hindcasts suggests that the spread of the forecasts has a strong seasonal component and some interannual modulation. During years or seasons where the thermocline is closer to the surface, there is a bigger potential for subsurface anomalies to generate SST spread. However, the

multiplicity of factors affecting the subsurface temperature spread (observation coverage, amplitude of the wind patterns applied, atmospheric internal variability) prevents a one-to-one relation between interannual variability and forecast spread.

4.2 Discussion

We have seen above that after three months of coupled integration, experiments with perturbed oceanic initial conditions have a similar spread to those sampling only the effects of atmospheric internal variability. This is not surprising at middle latitudes where atmospheric stochastic forcing is expected to contribute significantly to interannual variability (Frankignoul and Hasselmann, 1977). In the tropical Pacific ocean, where the ocean provides the memory of the coupled El Niño phenomenon, one would expect uncertainties in oceanic initial conditions to lead to a larger spread. This is not the case, apparently because the assimilation of numerous ocean observations reduces substantially the effect of wind and SST uncertainties. There are several possible reasons for the weak impact of the uncertainty of the ocean initial conditions.

Firstly, one might believe that the result is essentially true. That is, our knowledge of the equatorial Pacific is good enough to initialize our forecast models, that the uncertainty in our forecasts is dominated by atmospheric noise (even if the spread is slightly under-represented in our present model), and that the still-substantial errors in our forecasts are largely due to errors in the coupled model. This is certainly plausible, given the evidence we have of inadequacies in both the ocean model and the response of the atmospheric model to SST (Anderson *et al.* 2003, Vitart *et al.* 2003).

Alternatively, one might worry about the extent to which our ocean analysis system draws to the observed data. The spread in our ocean analyses is relatively small, once the available data have been assimilated into our system. But it is conceivable that the error in the large scale state of the ocean is substantially larger than the spread (small scale errors might be assumed of limited importance for forecasts beyond 3 months or so). The large scale temperature field is moderately constrained by TAO, but the way in which we insert the temperature field into our model, and the way we handle the large scale salinity field could affect our forecasts. We may be underestimating our ocean initial condition error.

Developing this idea further, one has to be aware that errors in the wind stress only constitute a small part of the errors in the oceanic initial conditions. There might be other errors associated with other forcing fields (heat and freshwater fluxes), ocean model errors, suboptimal data assimilation procedures. Furthermore, the method we used to sample wind stress error is far from perfect. For example it only samples error in the wind with a one-month time scale (by construction), while errors more persistent in time probably exist and might have a larger impact on the ocean analyses. More importantly, the 5 member-ensemble we chose here (no perturbation, and two symmetric wind patterns applied) only samples a very small subspace of all the likely initial errors.

It has been shown by numerous studies (e.g. Moore *et al.* 2001) that some perturbation patterns to the ocean atmosphere system in the tropical Pacific grow much more than others, and that the spectrum of the growth rates of those patterns is very steep. Ordinarily, a large number of wind patterns is required to get a reasonable estimate of the uncertainty of a particular forecast linked to uncertainties in oceanic initial conditions. However, methods such as ”stochastic optimals” (Farrell and Ioannou 1993, Kleeman and Moore 1997) allow the computation of the forcing perturbation that maximises the spread of the forecast ensemble. These methods have been applied to compute the stochastic optimals in a hybrid coupled model like that used by Moore *et al.* 2001. Although not designed for this purpose, the wind patterns of the stochastic optimals could be applied while computing the oceanic analyses as an alternative method of generating an ensemble of ocean initial conditions. Only a limited number of wind perturbation patterns would be needed. In addition the initial oceanic perturbations would then contain information on both the dynamics of error growth in the system through the

stochastic optimals, and on how the data assimilation method and observation coverage constrain the initial state through applying these perturbations during the analysis. This will be explored in a future study.

If we consider the implications of this study for the ECMWF ensemble seasonal forecasting system, two strands of thought emerge. One is that more work is needed to improve the representation of uncertainty in the forecasts. In part this can be addressed by improving the preparation of the ocean initial conditions, perhaps following some of the ideas outlined above. More fundamentally, we need to address the issue of allowing for model error. A direct method of sampling model error (both its direct and indirect effects) is to use a multi-model ensemble based on different oceanic and atmospheric models and perhaps different data-assimilation strategies. The use of multi-model ensembles is currently under investigation in the framework of the European Union projects Demeter (Palmer *et al.* 2003) and Enact. Experience shows that it is relatively easy to improve the ensemble-mean forecast and widen the ensemble spread using a multi-model approach. How *well* the uncertainties due to model error can be represented by multi-model techniques remains to be seen.

The second strand of thought is an encouraging one. Consider that the spread in the SWT experiment might be approximately realistic in estimating the spread in a hypothetical near-perfect forecasting system in which the errors in the forecast come from unpredictable atmospheric synoptic variability and limited errors in the oceanic initial conditions. In this case, there is a very clear scope for improving the levels of forecast error in the equatorial ocean from their present values (figure 6b) to those close to the limit of predictability (figure 6a). The possibility of such an improvement is exciting, but underlines the importance of continued work to improve both models and ocean initialization methods for the purpose of seasonal forecasting.

REFERENCES

Alves O., M. Balmaseda, D Anderson, T Stockdale, 2002: Sensitivity of dynamical seasonal forecasts to ocean initial conditions. *Quart. J. Roy. Met. Soc.*, submitted. See also ECMWF Technical Memorandum 369, available at www.ecmwf.int

Anderson, D., T. Stockdale, M. Balmaseda, L. Ferranti, F. Vitart, P. Doblas-Reyes, R. Hagedorn, T. Jung, A. Vidard, A. Troccoli, T. Palmer, 2002: Comparison of the ECMWF seasonal forecast Systems 1 and 2, including the relative performance for the 1997/8 El Niño. *ECMWF Technical Memorandum 404*, available at www.ecmwf.int.

Blanke, B., J.D. Neelin and D. Gutzler, 1997: Estimating the effects of stochastic wind stress forcing on ENSO irregularity, *J. Clim.*, *10*, 1473-1485.

Buizza, R., M.J. Miller and T.N. Palmer, 1999, Stochastic simulation of uncertainties in the ECMWF ensemble prediction system, *Q. J. Roy. Meteorol. Soc.*, *125*, 1935-1960.

Burgers G., M. Balmaseda, F. Vossepoel, G.J van Oldenburgh, P.J. van Leeuwen, 2002: Balanced ocean data assimilation near the equator. *J. Phys. Ocean.*, 32, 2509-2519.

Eckert, C. and M. Latif, 1997: Predictability of a stochastically forced hybrid coupled model of El Niño *J. Clim.*, *10*, 1488-1504.

Farrell, B.F. and P.J. Ioannou, 1993: Stochastic forcing of the linearized Navier-Stokes equations. *Phys. Fluids A*, **5**, 2600-2609.

Frankignoul, C. and K. Hasselmann, 1977: Stochastic climate models, Pt. 2, Application to sea surface temperature anomalies and thermocline variability. *Tellus*, 29, 289-305.

Goswami B.N. and J. Shukla, 1991: Predictability of a coupled ocean-atmosphere model, *J. Climate, 4*, 3-22.

Ji, M. and A. Leetma, 1997: Impact of data assimilation on ocean initialization and El Niño prediction. *Mon. Wea. Rev., 125*, 742-753.

Josey, S.A., E.C. Kent and P.K. Taylor, 2002: Wind Stress Forcing Of The Ocean In The SOC climatology: Comparisons with the NCEP-NCAR, ECMWF, UWM/COADS and Hellerman and Rosenstein Datasets. *J. Phys. Oceanog.*, 32, 1993-2019.

Kleeman, R. and A.M. Moore, 1999: A new method for determining the reliability of dynamical ENSO predictions. *Mon. Wea. Rev.*, **127**, 694-705.

McPhaden, M.J., 1999: Genesis and evolution of the 1997-98 El Niño, *Science*, *283*, 950-954.

Moore, A.M. and R. Kleeman, 1996: The dynamics of error growth and predictability in a coupled model of ENSO, *Q. J. Roy. Meteorol. Soc.*, *122*, 1405-1446.

Moore, A.M. and R. Kleeman, 1998: Skill assessment for ENSO using ensemble prediction. *Q. J. R. Meteorol. Soc.*, **124**, 557-584.

Moore, A.M., J. Vialard, A.T. Weaver, D.L.T. Anderson, R. Kleeman and J.R. Johnson, 2001: The role of air-sea interaction in controlling the optimal perturbations of low-frequency tropical coupled ocean-atmosphere modes. *J. Climate*, **16**, 951-968.

Neelin, D., D.S. Battisti, A.C. Hirst, F-F. Jin, Y. Wakata, T. Yamagata, and S.E. Zebiak, 1998: ENSO Theory, *J. Geophys. Res. 103*, 14261-14290.

Palmer, T.N. and D.L.T. Anderson, 1994: The prospects for seasonal forecasting - a review paper, *Q. J. R.*

Meteorol. Soc., 120, 755-793.

Palmer, T.N., 2000: Predicting uncertainty in forecasts of weather and climate, *Rep. Prog. Phys., 63,* 71-116.

Palmer, T. N., A. Alessandri, U. Andersen, P. Cantelaube, M. Davey, P. Dlcluse, M. Dqu, E. Dez, F. J. Doblas-Reyes, H. Feddersen, R. Graham, S. Gualdi, J.-F. Gurmy, R. Hagedorn, M. Hoshen, N. Keenlyside, M. Latif, A. Lazar, E. Maisonnave, V. Marletto, A. P. Morse, B. Orfila, P. Rogel, J.-M. Terres, M. C. Thomson, 2003: Development of a European multi-model ensemble system for seasonal to inter-annual prediction (DEMETER), submitted to Bull. Amer. Meteor. Soc.

Reynolds, R.W. and T.M. Smith, 1995: A high resolution global sea surface temperature climatology, *J. Clim.,* 8, 1571-1583.

Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes and W. Wang, 2002: An improved in situ and satellite SST analysis for climate, *J. Clim.,* 15, 1609-1625.

Rosati, A., K. Miyakoda and R. Gudgel, 1997: The impact of ocean initial conditions on ENSO forecasting with a coupled model, *Mon. Wea. Rev., 125,* 754-772.

Smith N.R., J.E. Blomley and G. Meyers, 1991: A univariate statistical interpolation scheme for subsurface thermal analyses in the tropical oceans, *Prog. Ocean.,* 28, 219-256.

Smith N.R., 1995: An improved system for tropical ocean sub-surface temperature analyses. *J. Atmos. Ocean. Technol.* 12, 850-870.

Stockdale, T.N., 1997: Coupled ocean atmosphere forecasts in the presence of climate drift, *Mon. Wea. Rev., 125,* 809-818.

Stockdale, T.N., D.L.T. Anderson, J.O.S. Alves and M.A. Balmaseda, 1998: Global seasonal rainfall forecasts using a coupled ocean-atmosphere model, *Nature, 392,* 370-373.

Terray L., E. Sevault, E. Guilyardi and O. Thual, 1995: The OASIS Coupler User Guide Version 2.0. *CERFACS technical report TR/CMGC/95-46*

Troccoli A., M. Balmaseda, J. Segschneider, J. Vialard, D. Anderson, K. Haines, T. Stockdale and F. Vitart, 2002: Salinity adjustments in the presence of temperature data assimilation. *Mon. Wea. Rev.*, 130, 89-102. Also available as ECMWF Tech Memo 305, at www.ecmwf.int.

Vitart F., M. Balmaseda, L. Ferranti and D. Anderson, 2003: Westerly wind events and the 1997 El Niño event in the ECMWF seasonal forecasting system: a case study. *J. Clim.* in press. See also ECMWF Technical Memorandum 370, available at www.ecmwf.int

Wolff, J.O., E. Maier-Raimer and S. Legutke, 1997: The Hamburg Ocean Primitive Equation Model, *Technical report 13,* Deutsches Klimarechenzentrum, Hamburg, 98pp.