



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

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

### **► To cite this version:**

Jérôme Vialard, Frédéric Vitart, Magdalena Balmaseda, Timothy N. Stockdale, David L. T. Anderson. An ensemble generation method for seasonal forecasting with an ocean-atmosphere coupled model. *Monthly Weather Review*, 2005, 133 (2), pp.441-453. <10.1175/MWR-2863.1>. <hal-00155077>

**HAL Id: hal-00155077**

**<https://hal.science/hal-00155077v1>**

Submitted on 17 Mar 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

## An Ensemble Generation Method for Seasonal Forecasting with an Ocean–Atmosphere Coupled Model

JÉRÔME VIALARD

*Seasonal Forecasting Group, ECMWF, Reading, United Kingdom, and LODYC, Paris, France*

FRÉDÉRIC VITART, MAGDALENA A. BALMASEDA, TIMOTHY N. STOCKDALE, AND DAVID L. T. ANDERSON

*Seasonal Forecasting Group, ECMWF, Reading, United Kingdom*

(Manuscript received 19 September 2003, in final form 26 March 2004)

### ABSTRACT

Seasonal forecasts are subject to various types of errors: amplification of errors in oceanic initial conditions, errors due to the unpredictable nature of the synoptic atmospheric variability, and coupled model error. Ensemble forecasting is usually used in an attempt to sample some or all of these various sources of error. How to build an ensemble forecasting system in the seasonal range remains a largely unexplored area. In this paper, various ensemble generation methodologies for the European Centre for Medium-Range Weather Forecasts (ECMWF) seasonal forecasting system are compared. A series of experiments using wind perturbations (applied when generating the oceanic initial conditions), sea surface temperature (SST) perturbations to those initial conditions, and random perturbation to the atmosphere during the forecast, individually and collectively, is presented and compared with the more usual lagged-average approach. SST perturbations are important during the first 2 months of the forecast to ensure a spread at least equal to the uncertainty level on the SST measure. From month 3 onward, all methods give a similar spread. This spread is significantly smaller than the rms error of the forecasts. There is also no clear link between the spread of the ensemble and the ensemble mean forecast error. These two facts suggest that factors not presently sampled in the ensemble, such as model error, act to limit the forecast skill. Methods that allow sampling of model error, such as multimodel ensembles, should be beneficial to seasonal forecasting.

### 1. Introduction

Reliable sea surface temperature (SST) forecasts are a prerequisite 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. However, any SST forecast is subject to various types of errors. First, it is widely accepted that the details of atmospheric variability cannot be forecasted deterministically beyond a few days. Since synoptic atmospheric variability can affect the SST field, through the paradigm of integrated noise in midlatitudes (Frankignoul and Hasselmann 1977) and through westerly wind bursts—for example, in the case of El Niño (e.g., McPhaden 1999)—there is a source of uncertainty in the forecast stemming from the chaotic nature of the atmosphere. Second, various stability analyses of the coupled ocean–atmosphere system (e.g.,

Moore and Kleeman 1996) suggest that errors in oceanic initial conditions can be amplified by the large-scale ocean–atmosphere dynamics during the course of the forecast. Several studies have indeed shown that improving the ocean initial state using data assimilation did improve the skill of seasonal forecasts (Ji and Leetma 1997; Rosati et al. 1997; Alves et al. 2004). However, even in the presence of data assimilation, the ocean initial state is not perfectly known. So this is another source of error in seasonal forecasts. Finally, coupled ocean–atmosphere models used to perform the seasonal forecasts are subject to model error. This includes systematic bias (which can be relatively easily subtracted for seasonal forecasts, see, e.g., Stockdale 1997), but also error in the variability at various time scales, which can affect the quality of the forecast.

There are thus primarily three factors that limit the skill of SST seasonal forecasts: coupled model error, error in the estimate of the ocean initial state, and the unpredictable nature of atmospheric synoptic variability. It is necessary to sample the effect of these uncertainties in a seasonal forecast system. This can be done through ensemble forecasting, whereby not just one, but many forecasts are made, perturbing the forecasts

---

*Corresponding author address:* Dr. David Anderson, European Centre for Medium-Range Forecasting, Shinfield Park, Reading RG2 9AX, United Kingdom.  
E-mail: d.anderson@ecmwf.int

in agreement with the known statistics of error sources (Palmer 2000). In general, the perturbations are applied to initial conditions. The spread of the ensemble should then provide some measure of the level of uncertainty attached to the forecast.

This approach was followed to some extent in the first seasonal forecasting system at European Centre for Medium-Range Weather Forecasts (ECMWF), denoted System 1 (S1). A new ocean analysis was generated each day by the ocean data assimilation system and used as initial conditions for a forecast. The atmospheric initial conditions came from the operational numerical weather prediction analysis system. Starting from these initial conditions, the coupled model was integrated forward for 200 days. 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). The above approach, known as lagged average, samples some of the error sources mentioned above. It is quite widely used in climate modeling as a method for generating ensembles because of its simplicity. Since it uses several ocean initial conditions, it measures to some extent error linked to oceanic initial conditions. Furthermore, different realizations of the synoptic atmospheric variability are realized in the various forecasts, allowing some sampling of the effects of chaotic atmospheric variability. However, this lagged-average approach has three main drawbacks. First, since daily integrations are grouped into 1-month ensembles, this introduces a delay of at least 15 days in the forecast delivery date. Second, the initial states of the ocean in S1 span 1 month but the typical change of the ocean in this time is not necessarily representative of the actual error in oceanic initial conditions. Third, members of the ensemble start from oceanic initial conditions up to 30 days apart, which is not statistically indistinguishable given the variation of model forecast statistics with lead time.

In this paper, we investigate ensemble generation methods for seasonal forecasting which attempt to overcome these drawbacks. We start all the integrations in one burst at the start of the month to allow a more timely delivery of the forecasts. Three types of perturbations to generate the ensemble are investigated in this paper. In the first approach, the forecasts are started from a (small) ensemble of ocean analyses to represent to some degree uncertainty in ocean initial conditions arising from error in wind forcing, one of the main sources of error in deriving ocean initial conditions. Uncertainty in SST is also important and needs to be taken into account. This is done in the second approach by perturbing the upper-layer temperature of

the ocean analyses with SST patterns representative of typical errors in the SST estimate (how we do this will be discussed later). Finally, some random perturbations are applied to the atmosphere throughout the coupled integrations. In section 2, we describe the forecasting system and the details of those perturbation strategies. In section 3, we compare these ensemble generation methodologies (applied separately or collectively) to the more traditional lagged-average approach. In section 4, we focus on the strategy where the three types of perturbations are applied, which is the one that has been used in designing the ECMWF seasonal forecasting System 2 (S2). We will in particular investigate if the spread of the ensemble is sufficient to cover the error in the hindcasts. In section 5, a summary of the results of the paper is given. The limits of the present approach are then discussed and we make some suggestions for improving ensemble strategies for seasonal forecasting.

## 2. The seasonal forecasting system

### *a. The ocean data assimilation scheme*

The ocean model used in this study is HOPE (Wolff et al. 1997), with a  $2^\circ$  zonal resolution and a meridional resolution of  $0.5^\circ$  near the equator. There are 20 vertical levels, with a resolution of 20 m near the surface. The model is integrated forward over periods of 10 days, forced by the wind stress, heat, and freshwater fluxes from the atmospheric analysis system. Every 10 days, this estimate of the ocean state is combined with all the available oceanic temperature data (gathered over a  $-5$  days to  $+5$  days period), to generate an oceanic analysis, using an optimal interpolation scheme (Smith et al. 1991; Alves et al. 2004). 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). No ocean analysis of SST is performed. Rather, a strong relaxation to the (Reynolds et al. 2002) SST OIv2 product is applied to ensure that the coupled model starts from SSTs close to observations.

### *b. An ensemble of ocean analyses*

To sample the uncertainty in the forecasts arising from uncertainty in ocean conditions, we create an ensemble of ocean analyses. 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, errors in the winds have the potential to be a major source of error in deriving oceanic initial conditions. Ocean model errors are quite difficult to sample other than by methods of multimodel en-

sembles. So, in this paper only perturbations in wind stress will be considered in the analysis ensemble as a first step toward generating uncertainties in the oceanic initial conditions. We present below a method designed to construct perturbation patterns in wind stress, representative of the typical random part of the error in these fields.

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 1980–97. These differences between two state-of-the-art estimations of wind stress using observations must be representative of the typical uncertainties on the knowledge of the wind stress field. The wind stress perturbations are stratified by calendar month. By linearly interpolating two randomly picked wind stress patterns representative of consecutive months (the full pattern being applied to the middle of each month), daily perturbations can be obtained. These are then used to randomly perturb the daily wind stress that forces the ocean model.

Wind stress perturbations are stronger in high latitudes as seen in Fig. 1a. 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 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 a spread is introduced. This is a relatively small ensemble, the size of which was constrained by numerical cost issues. The consequences of this relatively small ensemble will be discussed later in the paper.

The impact of the wind stress perturbations on the ocean is illustrated in Fig. 2. If there were sufficient ocean data, then wind perturbations would have little effect: the ocean data would be able to define the ocean state. However, it is highly unlikely that this situation is

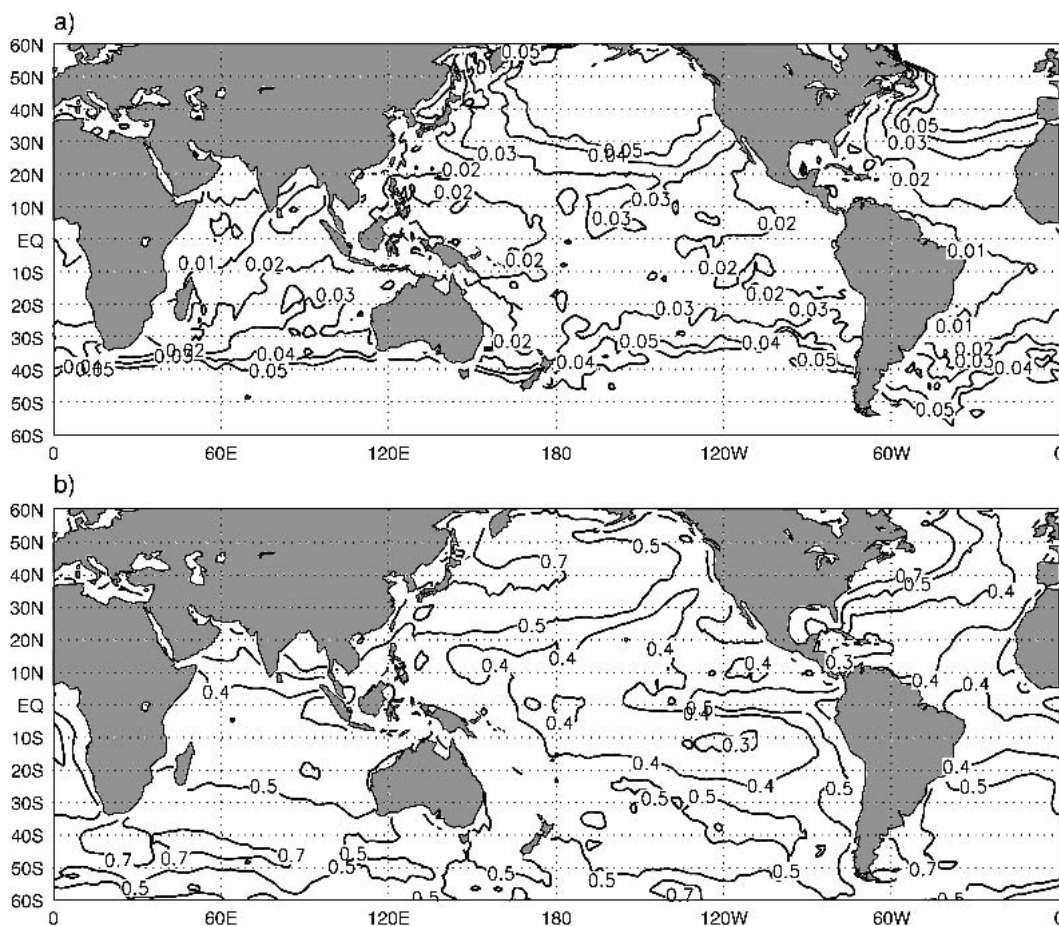


FIG. 1. (a) Rms of the monthly mean wind stress perturbation patterns used to generate the ensemble. Contour interval every  $0.01 \text{ N m}^{-2}$  up to  $0.05 \text{ N m}^{-2}$ . (b) Rms of the SST perturbation patterns used to generate the ensemble. Contour interval every  $0.1^\circ\text{C}$  up to  $0.5^\circ\text{C}$  and then  $0.2^\circ\text{C}$ .

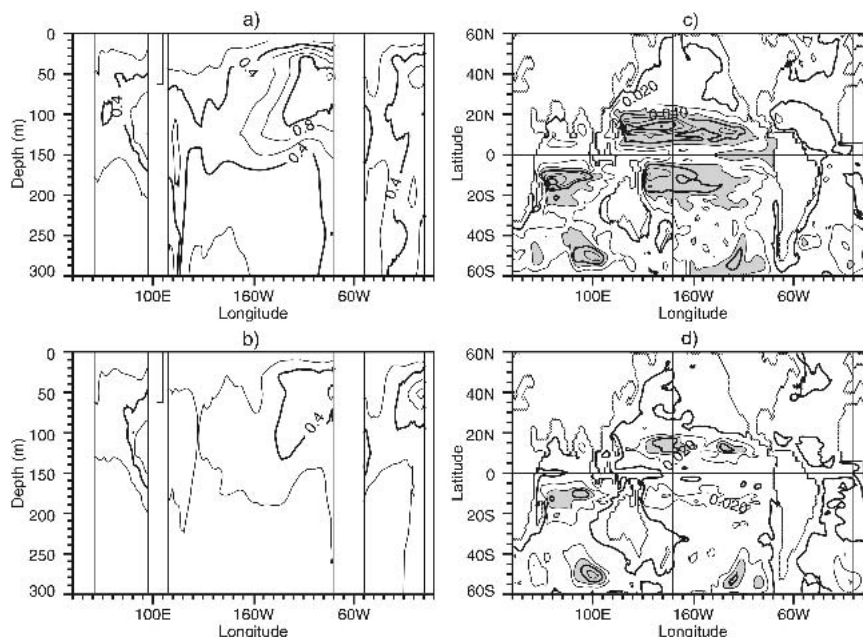


FIG. 2. (a) Vertical section along the equator of the rms spread of the five-member ensemble oceanic subsurface temperature analysis generated by applying wind stress perturbations in the absence of oceanic data assimilation. Contour interval every  $0.2^{\circ}\text{C}$  (b) As in (a), but for an experiment with oceanic data assimilation. (c) Map of the rms 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.

realized. We can get some idea of the importance of ocean data in limiting the effect of wind uncertainty by performing another set of ocean “analyses.” These are obtained by forcing the ocean model with the same wind and wind perturbations as described, but no subsurface ocean data are assimilated. In both cases, SST is relaxed to observations in the same way, however.

Figure 2a shows the spread in the analyses from such an experiment. Figure 2b shows the equivalent but from the ensemble of analyses in which subsurface ocean data have been assimilated. Figures 2a and 2b illustrate the effects of wind perturbations on the temperature along the equator. The largest impacts are located near the thermocline, which is not surprising since the wind stress induces vertical displacement of the thermocline via Ekman pumping. Fig. 2a shows how uncertainties in the wind forcing translate into large uncertainties in the oceanic state when data assimilation is not used, while Fig. 2b 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  $0.1^{\circ}\text{--}0.2^{\circ}\text{C}$ ) close to the surface because of the strong relaxation to observed sea surface temperature.

Figures 2c and 2d 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, (and 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. Figure 2c 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. Figure 2d shows how data assimilation acts to constrain the upper ocean heat content and collapses the spread. In the Tropical Atmosphere Ocean (TAO) region, the spread goes down from 3–4 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.

### c. Forecast system

Seasonal forecasts are made using a coupled model. The ocean model is the same as that used in the ocean analysis system described above. The atmospheric model is the ECMWF Integrated Forecasting System in its cycle 23r4 version. It is run at  $T_L95$  resolution with 40 levels. 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. We describe below the perturbation strategies that we have designed to account for errors linked to the uncertainties in SST and to the atmospheric stochastic forcing.

#### 1) ATMOSPHERIC PERTURBATIONS (STOCHASTIC PHYSICS)

Since a burst mode is used for ensemble generation rather than the lagged-average approach, all ensemble members start on the same day. To sample well the effect of different atmospheric forcing on the SST, we need to ensure that the different ensemble members follow a different sequence of synoptic variability after a few days. One way of doing this is to use the so-called stochastic physics (Buizza et al. 1999). The use of stochastic physics is also to represent uncertainties in the parameterization of subgrid-scale processes. These parameterizations are meant to represent the average effect of subgrid scale processes on the large-scale flow, but there is also a random component to this effect (e.g., for the same value of the average cloud cover in one model cell, there are many possible vertical and horizontal distributions of the clouds, and thus a range of radiative forcing of the flow). Stochastic physics is an attempt to take into account these uncertainties in the physical parameterizations by randomly perturbing the atmospheric parameterized physical tendencies at each time step of the model integration. This introduces a random component in the atmosphere that results in a divergence of synoptic systems in the early range of the forecast. This approach is used in the ECMWF medium range weather ensemble prediction system (Buizza et al. 1999). We could of course obtain divergence of the synoptic variability in many other ways. Even for the tiniest of initial perturbations, the level of synoptic variation between ensemble members is saturated after 20 days or so, and essentially independent of how the differences were triggered. Stochastic physics helps us reach this saturated level a little quicker than some other methods might, and thus ensures that an appropriate

level of atmospheric “noise” acts on the coupled system from an early stage.

#### 2) SST PERTURBATIONS

We saw in section 2b that the spread of the ocean analyses in SST was weak, due to the strong SST relaxation to the SST analysis denoted OIv2 by Reynolds et al. (2002). This is, however, a problem, since SST is not perfectly known and as SST is central to seasonal forecasting, we should take into account uncertainties in SST when generating the ensemble. We do this in a similar way to what we did with wind stress: we estimate SST patterns that should be representative of the typical errors in SST products. One set of perturbation patterns has been constructed by taking the difference between two different weekly mean SST analyses (Reynolds OI and Reynolds 2DVAR) from 1985 to 1999 (Reynolds et al. 2002). A second set of SST perturbations has been constructed by taking the difference between Reynolds 2DVAR SSTs and its 1-week persistence. The first set of SST perturbations samples the uncertainties in the SST analysis, whereas the second difference samples the uncertainties due to the fact that the SSTs from the National Centers for Environmental Prediction (NCEP) are a weekly mean product. For each starting date, two combinations from these two different sets of perturbations are randomly selected and are added to the SSTs produced by the operational ocean analyses with a + and – sign, creating four perturbed initial states. The perturbation has full value at the surface but is ramped down to zero at 40-m depth. The SST perturbations are not present during the analysis phase, but are added to the ocean initial conditions at the start of a forecast.

The standard deviation of SST perturbations (Fig. 1b) is particularly strong over the ice margin regions in the Northern Hemisphere and over the dynamically most active regions (Gulf Stream, Kuroshio, circumpolar current, etc.) where the perturbations can exceed  $1^{\circ}\text{C}$ , 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  $60^{\circ}\text{S}$  to  $60^{\circ}\text{N}$ .

### 3. Comparison of the various ensemble strategies

#### a. The ensemble experiments

In this section, we want to compare the perturbation strategies discussed in section 2 to the more traditional lagged-average approach. Five-member ensemble forecasts experiments were run for the 1991–98 period, starting from 1 January, 1 April, 1 July, and 1 October of each year. In three of these experiments, the ensemble was generated using stochastic physics (experiment SP), wind perturbations (WP), and temperature

perturbations (TP) individually. In the fourth experiment (SWT), all three perturbations (stochastic physics, wind perturbations and SST perturbations) were applied.

In experiment TP we use the member of the ocean analysis, which has no wind perturbation, and create an ensemble by perturbing the SST from this analysis. Two groups of two SST forecasts are made using symmetric SST perturbation patterns. Experiment SP also starts from the ocean analysis, which has no wind perturbation, and four different seeds are used for the random number generator used in perturbing the physical tendencies. In experiment SWT, we use the five-member ocean analyses but further perturb them with SST perturbations, and apply stochastic physics during the forecast with four different seeds.

The fifth experiment used perturbations as in SWT but no subsurface ocean data were assimilated in preparing the ocean initial conditions; it is denoted NDA.

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 five-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  $+15$ ). One difference is that the same atmospheric conditions are used for all five start dates. The different ocean dates are used as if they applied at the first of the month. In S1, the atmospheric initial conditions also contain information on soil wetness, sea-ice distribution, and snow cover, and so LA as in S1 would also include some perturbations of the land surface boundary conditions and of the ice edges. These perturbations are not included here to allow closer comparison with the other experiments that do not have such perturbations. A summary of the various experiments is given in Table 1.

### *b. Ensemble spread of the various experiments*

We now want to investigate if the different ensemble generation strategies lead to different properties of the ensemble spread. Figure 3 shows the ensemble spread

of the third-to-fifth month SST forecast for experiments SP in Fig. 3a and WP in Fig. 3b. 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, and LA may come from amplification of perturbations in the oceanic initial conditions or from forcing by atmospheric internal variability. The fact that the spread is very similar in all four experiments suggests that 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 synoptic atmospheric forcing. In fact, at midlatitudes, the ensemble spread in Fig. 3 is similar to the observed midlatitude interannual variability (in Fig. 4a). This means that the amplitude of the signal generated by stochastic forcing over a few months in the coupled model matches the amplitude of interannual variability in observations. This also seems to confirm the paradigm of integrated noise. Figure 4b has weaker than observed interannual variability because it is based on a five-member ensemble mean which smoothes the variability resulting from stochastic (synoptic) forcing.

In the equatorial Pacific the size of the interannual variability shown in Fig. 4 is much larger than the ensemble spread. The signal-to-noise ratio, often used as an indicator of potential predictability, is quite high in this model, which would indicate that a large part of the variability is predictable. However, it is difficult to assess how realistic the model signal-to-noise ratio is. While the model interannual variability can be compared with the observations (and Figs. 4a and 4b show that the model estimate is quite good), it is more difficult to estimate the level of noise in the real world. We will come back to this point later.

Figure 5 shows the spread for the different experiments as a function of the forecast lead time in the Niño-3 region. This area is of interest for seasonal forecasting as the tropical Pacific has a marked variability associated with El Niño and Niño-3 is a good marker of this variability. During the first month, experiments WP and SP display a different behavior 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. 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 that might correspond to the

TABLE 1. Summary of the various components of the experiments discussed, indicating whether they have wind perturbations, SST perturbations, stochastic physics, or data assimilation. The size of the ensemble ("Ens") is also shown.

Expt	Wind perturbations	SST perturbations	Stoch physics	Ens size	Data assimilation
WP	Y	N	N	5	Y
TP	N	Y	N	5	Y
SP	N	N	Y	5	Y
SWT	Y	Y	Y	5	Y
NDA	Y	Y	Y	5	N
LA	N	N	N	5	Y
S2	Y	Y	Y	40	Y

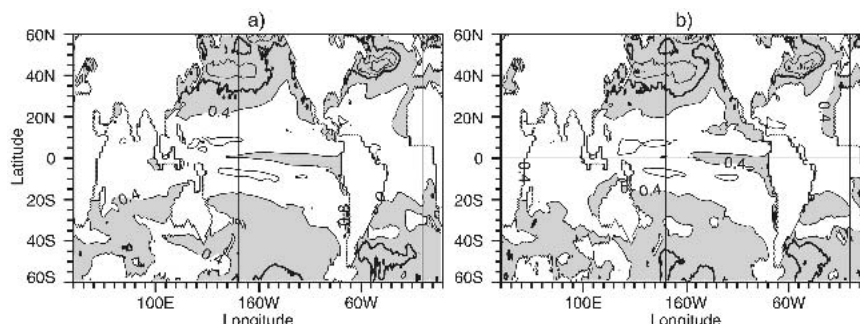


FIG. 3. Ensemble spread (std dev of the SST interannual anomaly forecasts) for months 3, 4, and 5 for the ensembles constructed using (a) stochastic physics and (b) wind stress perturbations. Contour interval  $0.2^{\circ}\text{C}$ , shading above  $0.4^{\circ}\text{C}$ .

noisy component of the SST perturbations being dissipated in the coupled model because it does not have a physical structure.

For all lead times beyond month 3, the spread in SST forecasts given by the TP, WP, and SP methods is very similar. What differences there are might be linked to sampling issues. 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 indistinguishable from the spread arising from purely internal atmospheric variability (experiment SP). In this system, uncertainty in initial conditions is not the main source of spread several months into the forecast.

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 Fig. 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, and suggesting that the low spread in experiment SWT is the result of a good control of the ocean initial conditions by observations. The difference in spread between SWT and NDA 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.

#### 4. Is the ensemble strategy sampling the sources of error in the system?

The purpose of this paper is to assess the performance of different methods of ensemble generation for

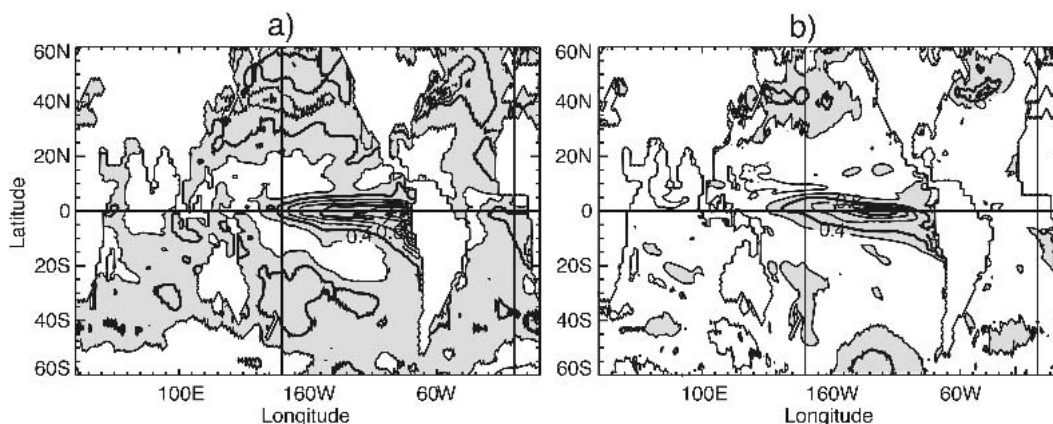


FIG. 4. Map of the root mean square of interannual SST anomalies for (a) observations and (b) months 3–5 ensemble-mean hindcasts for experiment SWT. Contour interval  $0.2^{\circ}\text{C}$ , shading above  $0.4^{\circ}\text{C}$ .



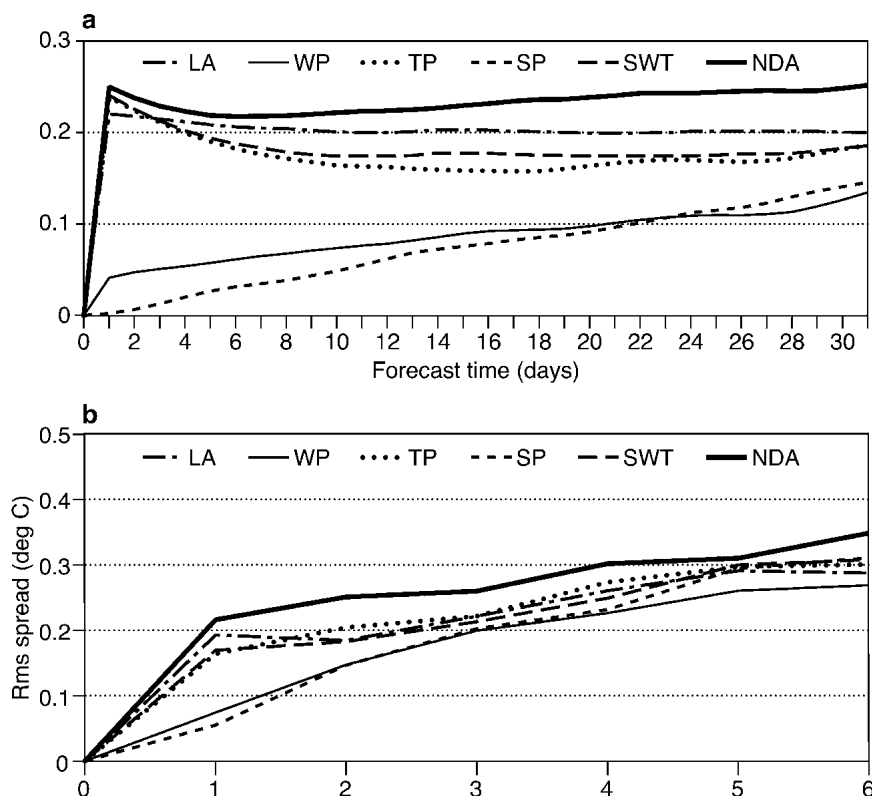


FIG. 5. Ensemble spread of SST forecasts as a function of lead time: spread in the (a), (b) Niño-3 region ( $5^{\circ}\text{N}$ – $5^{\circ}\text{S}$ ,  $150^{\circ}$ – $90^{\circ}\text{W}$ ) and (c) the Niño-3.4 region ( $5^{\circ}\text{N}$ – $5^{\circ}\text{S}$ ,  $170^{\circ}$ – $120^{\circ}\text{W}$ ). (a) The daily evolution of the spread during the first month; (b) and (c) the time evolution (in months) for the whole forecast period.

seasonal forecasting. 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, and SWT give a similar spread. Since TP and SWT allow a more timely delivery of the forecasts than LA, they 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. In this section, we will explore experiment SWT 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.

#### a. Reliability of the ensemble generation method

In a perfect ensemble forecasting system, the verifying SST should fall within the ensemble range most of the time. Specifically, the root-mean-square error of the ensemble-mean SST forecast should be equal to the standard deviation (spread) 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 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.

Figure 7a shows the ensemble-mean rms error (solid line) and the ensemble spread (dashed) as a function of the forecast lead time in region Niño-3 for experiment SWT. In this figure, as in all others, the ensemble spread is calculated as the square root of the mean value of the unbiased estimator of the population variance for each forecast date. Although the ensemble size of 5 means that the spread estimate for any particular date is fairly uncertain, the fact that we average over a large number of start dates results in a robust estimate of the (rms) ensemble spread. Tests with the 40-member ensemble of the ECMWF operational system confirm that the 5-member ensemble is quite adequate to estimate the ensemble spread. As a reference, the rms error of a forecast using persistence is also shown

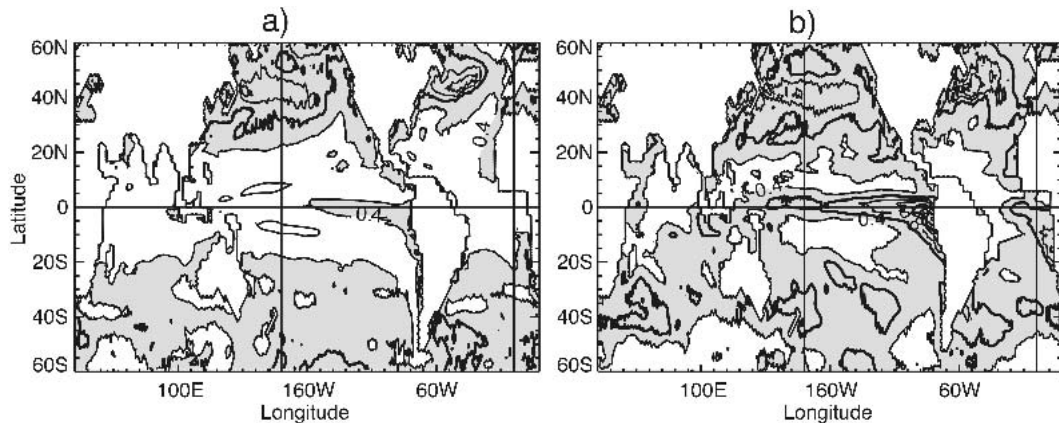


FIG. 6. (a) Ensemble spread (std dev of the SST interannual anomaly forecasts) for months 3, 4, and 5 for experiment SWT. (b) Rms error of the ensemble mean interannual SST anomaly for months 3, 4, and 5 for experiment SWT. Contour interval  $0.2^{\circ}\text{C}$ , shading above  $0.4^{\circ}\text{C}$ .

(dashed-dotted line). In the first month, the ensemble spread is close to the rms error. However, by month 3 the ensemble spread is already less than half the size of the error. This is an indication that significant sources

of forecast error were not taken into account in the ensemble generation method.

There are several possibilities to explain this gap. One is that we do not sample properly the error in

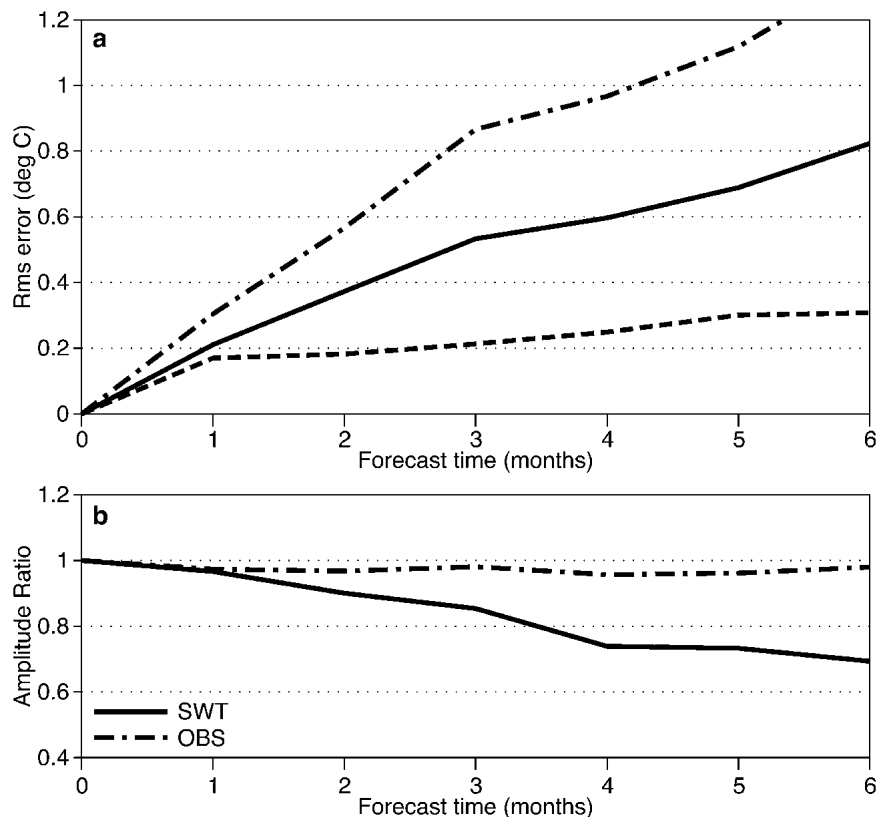


FIG. 7. (a) Evolution of the rms error of the ensemble mean (solid line) and the ensemble spread (dashed line) for the Niño-3 SST as a function of the hindcast lead time. The dot-dashed curve represents the error from using persistence as predictor. (b) The model/observed amplitude ratio for the Niño-3 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.

initial conditions, but it is unlikely that unsampled errors in oceanic initial conditions alone can bridge the gap between the spread of the SWT ensemble and its rms error. Experiment NDA has a much larger spread in oceanic initial conditions than SWT, but the spread of NDA forecasts is still significantly smaller than the rms error of either SWT or NDA. If we reasonably take the error in NDA to be indicative of an upper bound on the error amplitude in ocean initial conditions, then unsampled error in oceanic initial conditions is probably only a small component of the errors that are not taken into account by the ensemble. In other words, the ensemble spread of NDA in Fig. 5 is still far from the rms error curve in Fig. 7a.

A second possibility is that the effect of the unpredictable component of atmospheric forcing is underestimated leading to an underestimation of ensemble spread. Vitart et al. (2003) have shown that the ECMWF coupled model does not faithfully represent Madden-Julian oscillations (MJOs) that are thought to influence El Niño (e.g., McPhaden 1999; van Oldenborgh 2002). Vitart (2003) has shown that the poor representation of the MJO is linked to atmospheric model deficiencies.

Another problem with 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 Fig. 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 primarily to weak interannual surface wind variability in the atmospheric model (Anderson et al. 2003; Vitart et al. 2003). A consequence of the damping of the model interannual variability might be to reduce the ensemble spread that would otherwise exist in a perfect forecasting system with realistic initial errors.

Note that in both of these examples, model error did play an important role in leading to an underestimated ensemble spread. But a further straightforward explanation for the gap between the ensemble estimate of expected errors and the level of errors actually observed is that model errors are also 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. As a simple example from this paper note that in Fig. 4 the model and observations have 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 (see also Anderson et al. 2003 for a thorough discussion of some of the model errors implicated in attempts to hindcast the 1997 El Niño). If model errors are reduced and the forecasts improved, then the gap between ensemble spread and forecast error will be

reduced. Although we cannot quantify the total impact of model error on forecast skill, we believe that a forecasting system based on a near-perfect model would have a considerably reduced rms forecast error and an increased ensemble spread.

With a perfect model, predictability estimates depend on adequate sampling of the uncertainty in the initial conditions and in the directions of maximum error growth. The results presented in this section suggest that the seasonal forecasting system is far from this perfect model hypothesis: the mismatch between forecast error and ensemble spread is dominated by model errors affecting both the accuracy of the forecasts and the size of the ensemble spread. Therefore, methods to sample model error (multimodel or others) are a priority when designing future strategies for ensemble generation in seasonal forecasting.

### *b. 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 an ensemble forecast for a particular date should indicate that uncertainties are larger and that less confidence should be given to that forecast. One way to assess whether an ensemble forecasting system displays this desirable feature is to look for spread-skill relationships. Figure 8 shows a scatterplot of the ensemble-mean absolute error of the Niño-3 SST against the spread of the forecast for experiment SWT. In an ensemble with a spread-skill relationship, one would generally expect larger errors to occur preferentially in situations where the spread of the ensemble is larger. There is no indication of such a relationship in Fig. 8, neither at a 1 month nor at 4-to-6 months lead time. However, the problem with analysing experiment SWT is that the ensemble contains only five members, which makes the estimate of the spread for *individual* forecast dates not very reliable.

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. A 40-member set of hindcasts from S2 is available for all years from 1987 to 2002 but only for November and May starts. The estimate of the spread in this case is much better by dint of having a 40-member ensemble. However, no clear spread-skill relationship emerges.

Others have sought spread-skill relationships with very limited success. Moore and Kleeman (1998) did not find a spread-skill relationship when they used model (rms) error as a measure of skill, but did find something when they used correlation. Jewson et al. 2003, personal communication, available online at [http://arxiv.org/find/physics/1/au:+Jewson\\_s/0/1/0/all/0/](http://arxiv.org/find/physics/1/au:+Jewson_s/0/1/0/all/0/)

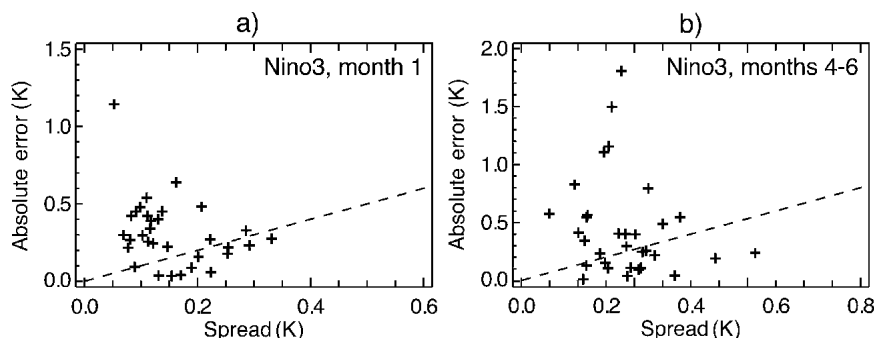


FIG. 8. Scatterplot of the SWT ensemble-mean Niño-3 SST forecast absolute error against the ensemble spread for (a) month-1 hindcasts and (b) months 4–6 hindcasts.

1) and Coelho et al. (2004) found no useful information in the spread.

This suggests that although there may be a small 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.

## 5. Summary and discussion

### a. 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 initial 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 3–6) all of the methods explored gave 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 thus required for more reliable seasonal forecasting systems.

Error-spread relationships were sought to see if the spread of the ensemble can confidently be used to define some uncertainties in the forecasts. No clear relationship was found in ensembles with 5 members nor in 40-member hindcasts made using the ECMWF S2 seasonal forecasting system. This 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 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 smaller than in experiments without data assimilation.

### b. 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, 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 level may be underrepresented in our present model), and that the still-substantial errors in our forecasts are largely due to errors in the coupled model. This seems to be confirmed by the fact that experiments with much larger errors in ocean initial conditions (experiment NDA) produce a spread that is still quite far from the actual level of error of the forecast, underlining the importance of errors coming from factors other than ocean initial conditions.

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, and 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 1-month time scale (by construction), while errors more persistent in time probably exist and might have a larger impact on the ocean analyses.

More work on defining proper perturbations for ocean initial conditions is needed. It has been shown by, for example, Moore et al. (2003), 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; Klee-man and Moore 1999) allow the computation of the forcing perturbation that maximises the spread of the forecast ensemble. Stochastic optimals have been calculated for several hybrid coupled models used by

Moore et al. 2003. 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 and by better estimating their uncertainties, 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 multimodel ensemble based on different oceanic and atmospheric models and perhaps different data-assimilation strategies. The use of multi-model ensembles is investigated in the framework of the European Union projects Demeter (Palmer et al. 2004) and Enact. Experience shows that it is relatively easy to improve the ensemble-mean forecast and widen the ensemble spread using a multimodel approach. How *well* the uncertainties due to model error can be represented by multimodel 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 (Fig. 6b) to those close to the limit of predictability (Fig. 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.

*Acknowledgments.* The authors would like to thank Simon Jose for providing some of the wind fields used in this study. Part of the work was carried out in support of EU project ENACT.

#### 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 1997/98 El Niño. ECMWF Tech. Memo. 404, 93 pp. [Also available online at [www.ecmwf.int](http://www.ecmwf.int).]
- Buizza, R., M. J. Miller, and T. N. Palmer, 1999: Stochastic simulation of uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 1935–1960.
- Burgers, G., M. Balmaseda, F. Vossepoel, G. J. van Oldenburgh, and P. J. van Leeuwen, 2002: Balanced ocean data assimilation near the equator. *J. Phys. Oceanogr.*, **32**, 2509–2519.
- Coelho, C. A. S., S. Pezzulli, M. Balmaseda, F. J. Doblas-Reyes, and D. B. Stephenson, 2004: Forecast calibration and combination: A simple Bayesian approach for ENSO. *J. Climate*, **17**, 1504–1516.
- 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. Part 2: Application to sea surface temperature anomalies and thermocline variability. *Tellus*, **29**, 289–305.
- 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. Oceanogr.*, **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. *Quart. J. Roy. Meteor. Soc.*, **122**, 1405–1446.
- , and —, 1998: Skill assessment for ENSO using ensemble prediction. *Quart. J. Roy. Meteor. Soc.*, **124**, 557–584.
- , J. Vialard, A. T. Weaver, D. L. T. Anderson, R. Kleeman, and J. R. Johnson, 2003: 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**, 14 261–14 290.
- Palmer, T. N., 2000: Predicting uncertainty in forecasts of weather and climate. *Rep. Prog. Phys.*, **63**, 71–116.
- , and D. L. T. Anderson, 1994: The prospects for seasonal forecasting—A review paper. *Quart. J. Roy. Meteor. Soc.*, **120**, 755–793.
- , and Coauthors, 2004: Development of a European multi-model ensemble system for seasonal to interannual prediction (DEMETER). *Bull. Amer. Meteor. Soc.*, **85**, 853–872.
- 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. Climate*, **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. *Progress in Oceanography*, Vol. 28, Pergamon, 219–256.
- Stockdale, T. N., 1997: Coupled ocean atmosphere forecasts in the presence of climate drift. *Mon. Wea. Rev.*, **125**, 809–818.
- , 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 Tech. Rep. TR/CMGC/95-46, 123 pp.
- 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.
- van Oldenborgh, G. J., 2000: What caused the onset of the 1997/98 El Niño. *Mon. Wea. Rev.*, **128**, 2601–2607.
- Vitart, F., 2003: Monthly forecasting system. ECMWF Tech. Memo. 424, 68 pp. [Also available online at [www.ecmwf.int](http://www.ecmwf.int).]
- , M. A. Balmaseda, L. Ferranti, and D. Anderson, 2003: Westerly wind events and the 1997/98 El Niño event in the ECMWF seasonal forecasting system. *J. Climate*, **16**, 3153–3170.
- Wolff, J. O., E. Maier-Raimer, and S. Legutke, 1997: The Hamburg Ocean Primitive Equation Model. Deutsches Klimarechenzentrum Tech. Rep. 13, 98 pp.