Q. J. R. Meteorol. Soc. (2006), **132**, pp. 647–666 doi: 10.1256/qj.05.65

Seasonal forecasting of tropical storm frequency using a multi-model ensemble

By FRÉDÉRIC VITART*
European Centre for Medium-Range Weather Forecasts, Reading, UK

(Received 13 April 2005; revised 11 October 2005)

SUMMARY

The skill of seven coupled ocean–atmosphere models to predict the frequency of tropical storms from 1987 to 2001 has been assessed using a procedure for tracking model tropical storms. The tropical storm tracker takes account of the difference of atmospheric horizontal resolution between the different models. Results indicate that the models display some skill in predicting the interannual variability of tropical storms over the Atlantic, the eastern North Pacific, the western North Pacific, the Australian basin and the South Pacific. A simple multi-model forecast has been built by adding all the seven ensemble forecasts together after calibration. The skill of the simple multi-model system is overall better than the skill of any individual model. Over a specific basin, combining several models leads to better forecasts than the best individual model. This indicates that the multi-model approach could benefit the dynamical seasonal forecast of tropical storms.

This conclusion is also valid for a longer time period (1959–2001). However, the individual models and the simple multi-model display less skill in predicting the interannual variability of tropical storms in the earlier decades.

KEYWORDS: Interannual variability Interdecadal variability Tropical storms

1. INTRODUCTION

Manabe *et al.* (1970) noticed for the first time that low-resolution atmospheric general circulation models (AGCMs) are able to create tropical depressions reminiscent of observed tropical storms. Several studies (Bengtsson *et al.* 1982, 1995; Haarsma *et al.* 1993) have shown that the simulated tropical storms have climatologies and physical characteristics close to those of observed tropical storms. Vitart *et al.* (1997, 1999) reported on simulated tropical storms with a seasonal evolution and interannual variability consistent with observations over the western North Atlantic, eastern North Pacific and western North Pacific. The frequency of simulated tropical storms is strongly correlated with the interannual variability of the simulated large-scale circulation as in observations (Vitart *et al.* 1999). In particular, the interannual variability of simulated tropical storms over the North Atlantic is significantly correlated to ENSO as in observations. This suggests that GCMs can be valuable tools to study the variability of tropical storm frequency.

A dynamical seasonal forecasting system based on ensembles of coupled ocean–atmosphere integrations has been set up at ECMWF (Stockdale *et al.* 1998). The coupled ocean–atmosphere model is integrated for about six months. The skill of this coupled system to predict a few months in advance the frequency of tropical storms has been discussed in Vitart and Stockdale (2001). Results suggest that the model has some skill in predicting the interannual variability of tropical storm frequency over basins like the North Atlantic, the western North Pacific and the South Pacific. Following this study, operational seasonal forecasts of tropical storm frequency are produced at ECMWF every month.

The quality of the seasonal prediction of tropical storms may be improved by using combined ensemble forecasts produced by different models (multi-model ensemble forecasts). This method is efficient in filtering out model errors present in the individual

* Corresponding address: European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, Berkshire RG2 9AX, UK. e-mail: frederic.vitart@ecmwf.int

ensemble forecasts. Krishnamurti *et al.* (2000) demonstrated that a multi-model ensemble outperforms all the individual models for hurricane track and intensity forecasts. Results from the PROVOST (PRediction Of climate Variations On Seasonal to interannual Time-scales) project demonstrate that the reliability of a seasonal forecast can be strongly enhanced by the use of multi-model ensembles (Palmer and Shukla 2000). PROVOST was based on atmosphere-only integrations forced by observed SSTs. The DEMETER (DEvelopment of a Multi-model Ensemble system for seasonal to inTERannual prediction) project (Palmer *et al.* 2004) has been set up in a way similar to PROVOST but this time with fully coupled ocean–atmosphere integrations. The present paper evaluates the benefit of multi-model ensemble forecasting for the seasonal prediction of tropical storms using the DEMETER model integrations.

The DEMETER project is described in section 2. The tracking of model tropical storms is explained in section 3, followed by an evaluation of the statistics of model tropical storm frequency (section 4). Section 5 explores the frequency of tropical storms over a longer time-range (1959–2001). Finally, section 6 discusses the main results of this paper.

2. Description of the multi-model system

The DEMETER project has been funded under the European Union Vth Framework Environment Programme to assess the skill and potential economic value of multi-model ensemble seasonal forecasts. The principal aims of DEMETER have been to advance the concept of multi-model ensemble prediction by installing a number of state-of-the-art global coupled ocean–atmosphere models on a single supercomputer, to produce a series of multi-model ensemble hindcasts with common archiving and common diagnostic software, and to assess the utility of multi-model hindcasts in specific quantitative applications, notably health and agriculture.

The DEMETER multi-model prediction system comprises the global coupled ocean–atmosphere models of the following institutions: CERFACS (European Centre for Research and Advanced Training in Scientific Computation, France), ECMWF (European Centre for Medium-range Weather Forecasts), INGV (Istituto Nazionale de Geofisica e Vulcanologia, Italy), LODYC (Laboratoire d'Océanographie Dynamique et de Climatologie, France), CNRM (Centre National de Recherches Météorologiques, Météo-France, France), UKMO (Met Office, UK) and MPI (Max-Planck Institut für Meteorologie, Germany). In order to assess seasonal dependence of forecast skill, the DEMETER hindcasts have been started four times a year from 1 February, 1 May, 1 August, and 1 November at 00 GMT. The atmospheric and land-surface initial conditions are taken from the ECMWF ReAnalysis (ERA-40) dataset (Uppala *et al.* 2005). The ocean initial conditions are obtained from ocean-only runs forced by ERA-40 fluxes, except in the case of MPI that used a coupled initialization method. Each hindcast has been integrated for 6 months and comprises an ensemble of 9 members. The hindcast period common to all the models is 1980–2001, although some models produced hindcasts over the extended period 1958–2001.

3. Tropical storm detection

The objective procedure for detecting model tropical storms tracks low vortices with a warm core between 500 and 200 hPa (see Vitart *et al.* 2003 for more details). A major problem is that the models, which are part of the multi-model ensemble, have different characteristics and in particular have different atmospheric horizontal

TABLE 1. PERCENTAGE OF TROPICAL STORMS DETECTED BY THE AUTOMATIC PROCEDURE

Basin	ATL	ENP	WNP	NIO	SIO	AUS	SPAC	TOTAL
T159	73(91)	61(73)	80(83)	73(100)	80(90)	72(81)	60(100)	78(90)
T95	68(86)	52(73)	80(83)	69(100)	78(90)	61(81)	58(100)	72(85)
T42	62(85)	61(73)	80(83)	65(90)	71(90)	61(71)	50(90)	70(80)

Percentage of tropical storms detected by the automatic procedure when applied to the ECMWF reanalysis over 15 years (1987 to 2001) as a function of the horizontal resolution. The numbers in parenthesis represent the percentage of tropical cyclones detected with hurricane intensity (maximum wind speed exceeding 32 m s^{-1}). The tropical storms have been detected over seven basins: Atlantic (ATL), Eastern North Pacific (ENP), Western North Pacific (WNP), North Indian Ocean (NIO), South Indian Ocean (SIO), Australian Basin (AUS) and South Pacific (SPAC). The last column represents the percentages over all the basins.

resolutions. For instance, the ECMWF atmospheric model has a spectral resolution of T_L95 corresponding to a grid resolution of 1.875° (T_L95 indicates spectral triangular truncation at wave-number 95 with linear grid), whereas the UKMO atmospheric model is a grid-point model with a resolution of 2.5° × 3.75°. Therefore, the algorithm for tracking model tropical storms had to be adapted to the different models. In Camargo and Zebiak (2002), the thresholds used for the detection criteria are basin dependent as well as model dependent; they take account of the model biases and deficiencies. The approach chosen in the present paper is slightly different. As in Camargo and Zebiak (2002), the criteria are basin dependent, but instead of being model dependent they are resolution dependent. The parameters have been tuned so that the objective procedure detects as many observed tropical storms as possible and as few non-tropical storms as possible when applied to the ECMWF reanalysis (ERA-40) *interpolated to the same grid as the model*. In this approach, model biases and deficiencies will affect the statistics of the tropical storms detected, but this method ensures that the tropical storms detected share exactly the same characteristics as the tropical storms in the ECMWF reanalysis (ERA-40).

In the present paper, ‘observed tropical storms’ refers to the storms that have been officially classified as tropical storms by agencies that are monitoring tropical storm activity like the National Hurricane Center for the Atlantic and eastern North Pacific. All the historical records of observed tropical cyclones have been obtained from Neumann *et al.* (1993), http://weather.unisys.com/hurricane/index.html and the Joint Typhoon Warning Center Products website (http://199.10.200.33/jtwc.html). Table 1 shows the percentage of observed tropical storms detected over each ocean basin by applying the tropical storm detection to the ECMWF Reanalysis interpolated at different resolutions from 1987 to 2001. According to Table 1, more than 70% of the observed tropical storms that are listed in the historical records over all the ocean basins during the period from 1 January 1987 to 31 December 2001 have been detected by the procedure for tracking model tropical storms, for all the different resolutions. This number increases to 80% for the tropical storms with hurricane intensity. There is a slight decrease in the percentage of detected tropical storms when the resolution gets coarser. On the other hand, the percentage of non-observed tropical storms (tropical storms that are detected but are not listed as tropical storms in the historical records) increases quite significantly when the resolution gets coarser. This illustrates the fact that when the resolution gets coarser, tropical cyclones get less intense, and it is more difficult to distinguish them from tropical depressions. The majority of non-observed tropical storms are in fact observed tropical depressions (intensity less than 17 m s^{-1}) that are too intense in the reanalysis.

TABLE 2. CLIMATOLOGICAL NUMBER OF TROPICAL STORMS PER YEAR

Basin	ATL	ENP	WNP	NIO	SIO	AUS	SPAC	RMS error
LODYC	4.6	5.2	6.3	0.8	7.6	10.7	6.4	10.5
ECMWF	3.1	4.8	11.8	2.4	5.3	6.3	3.7	9.3
CNRM	8.8	11.7	54	12.3	19.6	22.3	14.4	12.9
CRFC	9.2	9.4	51	12.6	18.4	24.2	14.2	12.4
UKMO	4	11	25	6	12	10	6.7	4.8
MPI	2.6	3.8	5.5	4.8	11.1	6.3	4.2	10.7
SCNR	10.3	11.6	38	12	23	17	14	8.6
MULTI	6.1	8.22	27.4	7.2	14	14	9	5.6
OBS	11	17	27.7	6.4	12.7	9	6.1	–

Climatological number of tropical storms per year for the period 1987 to 2001, for each model (LODYC, ECMWF, CNRM, CRFC, UKMO, MPI, SCNR), for the multi-model ensemble (MULTI) and for observations (OBS). The last column indicates the root-mean-square error relative to the observed tropical storm climatology.

4. RESULTS FROM THE PERIOD 1987–2001

The objective procedure for tracking model tropical storms has been applied to all the integrations of DEMETER for the period 1987 to 2001 (15 years). The number of detected tropical storms has been counted, and the present section discusses the statistics of the model tropical storm frequency.

(*a*) *Tropical storm climatology*

Table 2 displays the mean number of tropical storms per year over each ocean basin, along with the observed annual frequency during the same period. The frequency of tropical storms varies significantly from one model to another. This could be due for instance to differences in the cumulus parametrization (Vitart and Stockdale 2001). Other differences in the model physics (boundary layer, radiative scheme) can also have a strong impact on the climatology of tropical storms. Both atmospheric and oceanic components seem to have an impact on the model climatology. LODYC and ECMWF models have the same atmospheric component, with the same resolution, but they have different oceanic circulation models. The ECMWF model simulates about twice as many tropical cyclones as the LODYC model over the western North Pacific. This may be explained by the fact that climatological sea surface temperatures (SSTs) in the ECMWF model are significantly warmer than the climatological SSTs in the LODYC model over the western North Pacific in summer and autumn. However, the differences are generally smaller than between the coupled models with different atmospheric components. LODYC and CERFACS which use the same oceanic model coupled with different atmospheric models display different climatologies of model tropical storms.

The multi-model obtained by simply averaging all the models (hereafter referred to as the simple multi-model) produces a climatology that is overall closer to observations than all the individual models, except the UKMO model. The tropical storm frequency in the multi-model is indeed very close to observations over the western North Pacific, the North Indian Ocean and the South Indian Ocean.

In the present section, the number of tropical storms was not calibrated. In the next sections, the number of model tropical storms has then been multiplied by a calibration factor, so that the total annual number of model tropical storms from 1987 to 2001, with the exclusion of the year of the forecast (cross-validation), is equal to the observed frequency for each region. Each model has been calibrated separately. A simple multi-model forecast is built by putting all the ensemble members of all the different models together, after the different models have been calibrated.

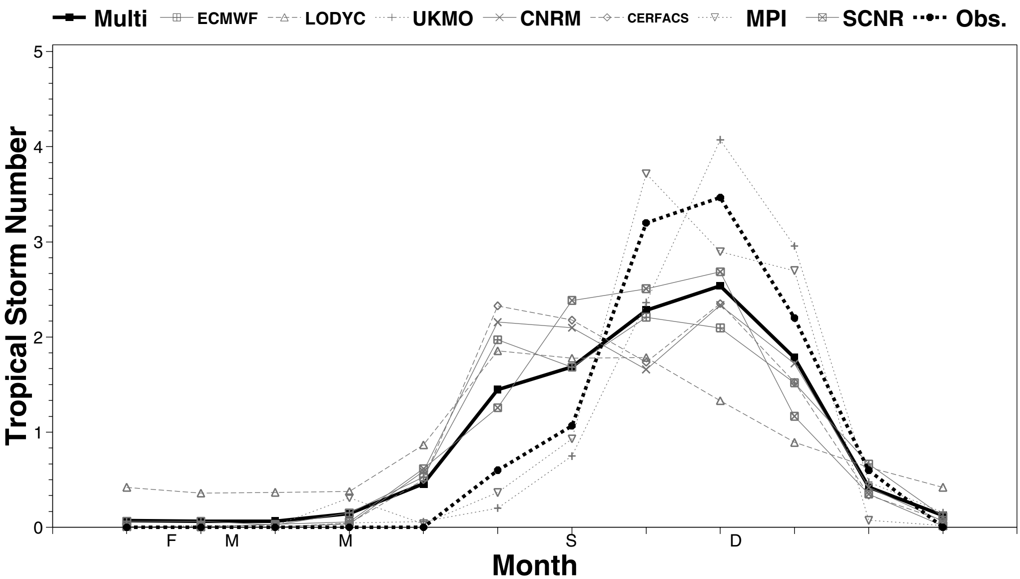

Figure 1. Seasonal cycle of tropical storm frequency in each individual model (grey lines) and in the multi-model ensemble (solid black line) over the North Atlantic for the period 1988–2001. Each model has been calibrated so that the annual mean number of tropical storms is equal to the observed annual mean tropical storm frequency. The calibration has been applied to each model individually. The dotted black line represents the observed seasonal cycle.

(*b*) *Seasonal variability*

For each individual model, the number of model tropical storms has been calculated for each month. Over the North Atlantic, all models, except the UKMO and MPI models, display too many tropical storms at the beginning of the tropical storm season, and too few after July (Figure 1). The combination of all 7 models (the simple multi-model) displays a peak activity in September, as observed, but with too many tropical storms before July and too few after July. Over the eastern North Pacific, most models present a peak activity in September and October, whereas the observed peak activity is from July to September. Over the western North Pacific, the models tend to predict a peak activity one month later than observed. The UKMO model seems to simulate the most realistic seasonal variability in this region. Over the North Indian Ocean, there are two tropical cyclone seasons: from April to June and from September to December. Most models fail to simulate this seasonal evolution. Only two models, MPI and UKMO predict a suppressed activity during the northern hemisphere summer. Over the southern hemisphere, all models successfully simulate the seasonal cycle of tropical storm frequency.

The simple multi-model ensemble displays a seasonal variability generally more realistic than for the majority of individual models. The UKMO model, which displays the best tropical storm climatology, seems to produce also the best simulation of tropical storm seasonality. ECMWF and LODYC coupled models display a very similar monthly variability. Both models share the same atmospheric component, but use different ocean models. This is also the case for the CNRM and CERFACS models, both producing also a very similar seasonal evolution of tropical storm frequency. This suggests that the atmospheric component determines most of the tropical storm seasonality.

TABLE 3. LINEAR CORRELATION

Basin	ATL	ENP	WNP	NIO	SIO	AUS	SPAC
ECMWF	0.37	**0.49**	**0.63**	0.29	−0.07	**0.45**	**0.63**
LODYC	−0.15	**0.58**	**0.67**	0.37	−0.05	0.41	**0.61**
UKMO	0.43	−0.28	0.44	0.14	−0.1	0.09	**0.52**
CNRM	**0.63**	0.25	0.39	−0.09	−0.1	0.33	**0.54**
CERFACS	0.28	**0.62**	**0.59**	−0.2	0.05	0.4	**0.59**
MPI	**0.54**	−0.2	0.39	0.01	−0.23	−0.32	0.21
SCNR	0.36	0.29	0.25	0.4	−0.04	0.31	0.14
MULTIMODEL	**0.61**	**0.56**	**0.72**	0.32	0.12	0.34	**0.62**

Linear correlation between the interannual variability of tropical storm frequency predicted by each model and observations from 1987 to 2001. For each model, we consider the ensemble mean. The simple multi-model ensemble has been built by averaging all the models after calibration. The starting date for the North Atlantic basin, eastern North Pacific and western North Pacific is the 1 May and the tropical storm frequency has been calculated over the period June to October. For the North Indian Ocean, the starting date is 1 August and the tropical storm period extends from 1 September until 31 January. For the southern hemisphere, the starting date is 1 November and the tropical storm period extends from 1 December until 30 April. Numbers in bold indicate a level of confidence larger than 90%.

The deficiencies of the models to reproduce the observed seasonal cycle of tropical storm frequency are consistent with their deficiencies in simulating a correct seasonal cycle of SSTs. For example, the LODYC or ECMWF models produce tropical Atlantic SSTs in September and October that are much colder than in observations, and in particular below 26 °C over a large portion of the main tropical storm development region. Since tropical storms tend to form when SSTs exceed 26 °C (see for instance Gray 1979), this cold bias of SSTs may explain why the ECMWF and LODYC models display so few tropical storms in September and October. Over the western North Pacific, most models display a seasonal cycle of SSTs which peaks one month later than in observations, which may explain why those models display a peak tropical storm activity also one month later than observed. Therefore it is not surprising to notice that the UKMO model, which displays the most realistic seasonal cycle of tropical storm frequency, is the model which displays the most realistic seasonal cycle of SSTs.

(*c*) *Interannual variability*

The number of tropical storms has been counted for each individual ensemble member, for each year and for each model. As underlined above, the models have been calibrated separately with cross-validation. Each ocean basin has a specific tropical cyclone season. Therefore for the Atlantic, eastern North Pacific and western North Pacific, we focus on the forecasts starting on 1 May and covering the period June to October (the first month has been discarded, since the atmospheric initial conditions may impact the tropical storm frequency during that month). For the North Indian Ocean, the forecasts starting on 1 August and covering the period September to January are considered. This period includes the second and biggest peak in the tropical cyclone activity over the North Indian Ocean (see section 4(b)). Finally, over the southern hemisphere where the tropical cyclone season extends from November to May, forecasts starting on 1 November and covering the period from December to April are evaluated. The skill of the models is measured by the linear correlation between the interannual variability predicted by the ensemble mean of the models and the observed interannual variability (Table 3).

According to Table 3, some of the models display skill (positive correlation with a level of confidence larger than 90%) in predicting the interannual variability of tropical

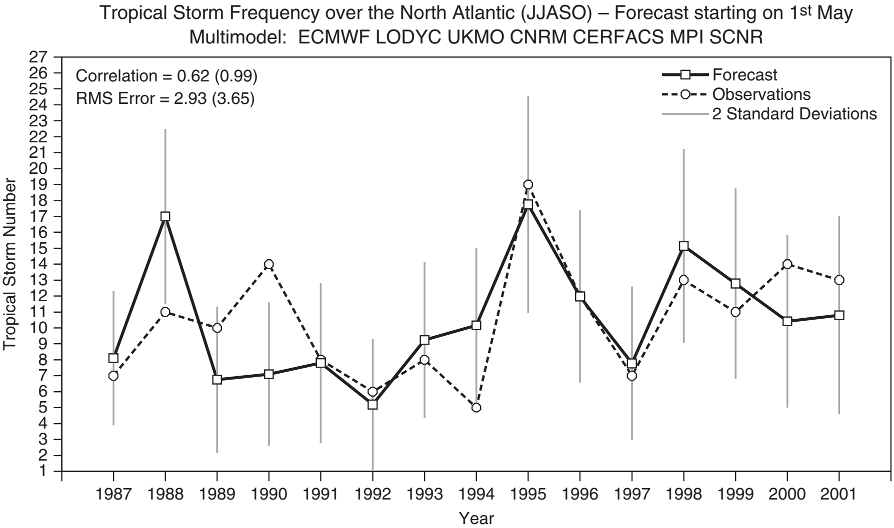

Figure 2. Interannual variability of tropical storm frequency in the multi-model ensemble (solid line) and in observations (dashed line) over the North Atlantic for the period 1987–2001. The simple multi-model is the average of the ensemble means of the 7 individual models. The vertical lines represent 2 standard deviations.

storm frequency over the North Atlantic, the eastern North Pacific, the western North Pacific and South Pacific. Over these four basins, the multi-model ensemble displays a linear correlation with the observed frequency larger than 0.5. Over the South and North Indian Oceans and over the Australian basin, the skill is much lower.

Figure 2 displays the interannual variability of tropical storm frequency for the simple multi-model over the North Atlantic along with the observed frequency (obtained from historical data). The correlation between the multi-model ensemble and the observed frequency is 0.62. Over the Atlantic there was a record number of tropical storms in 1995, which was likely associated with warm Atlantic SSTs (Saunders and Harris 1997). The multi-model forecast captures well this maximum in 1995. In fact, 5 out of 7 models predict more than 15 tropical storms in 1995. This suggests that the strong tropical storm activity in 1995 was predictable a few months before the start of the tropical storm season. The impact of the 1997 El Niño and 1998–1999 La Niña events are well predicted by most models. However, all the models predict fewer tropical storms in 1990 than observed and more tropical storms than observed in 1988 and 1994.

The simple multi-model seems to perform better *overall* (by averaging the scores over all the ocean basins) than any individual model and has only positive correlations (Table 3). This is consistent with the main conclusion of the DEMETER project (Hagedorn *et al.* 2005) and this suggests that the multi-model approach can be valuable for the prediction of tropical storms over different basins. Over a specific basin there is generally a model that performs better than the simple multi-model, like the CNRM model over the North Atlantic. However, the simple multi-model does not represent the optimal way of combining the different models. Giving different weights to the different models based on their past performances for each individual basin is likely to give better results. For example, if we reject the two models with the lowest performance over each individual basin, which is equivalent to giving them a weight of 0, then the multi-model performs significantly better than the best model over each individual basin. For instance, the linear correlation obtained with the combination of the CNRM, MPI, ECMWF, UKMO and SCNR model over the North Atlantic is 0.7, which is significantly higher than the linear correlation obtained with the CNRM model alone (0.61). This may be explained by the fact that the CNRM model has a tendency to overestimate the impact of ENSO on the Atlantic tropical storms, while some other models, such as ECMWF or SCNR, underestimate its impact. By combining the different models, some model errors are filtered out. This is consistent with Yoo and Kang (2005) who showed that the highest skill is obtained by selecting several skilful models which are less dependent on each other rather than adding poor models which degrade the multi-model composite prediction. On the other hand, the multi-model approach does not improve the forecasts if all the models display the same errors or have no skill as is the case over the South Indian Ocean, or if one of the models is so outstandingly better than all the others, that any combination with another model would reduce the skill. This latter situation did not happen with the seven models of the DEMETER project.

The performance of the individual models varies significantly from one basin to another. For instance the LODYC model performs very poorly over the Atlantic (correlation −0.15), but performs very well (compared to the other models) over the eastern North Pacific, the western North Pacific and the South Pacific. Interestingly, the models which perform the best over the Atlantic (CNRM, MPI, UKMO) perform very poorly over the eastern North Pacific, and vice versa. On the other hand, the multi-model ensemble performs equally well over both basins. Over the western North Pacific, it performs even better than any individual model. The fact that the performances are poorer over the North and South Indian Ocean and the Australian basin has already been noticed in Vitart and Stockdale (2001), and attributed to the lower impact of ENSO on tropical storm frequency over these basins.

The performance of the individual models over each ocean basin can be related to their skill in predicting the interannual variability of the large-scale circulation, and most especially to their skill in predicting ENSO teleconnections. For example, the poor performance of the ECMWF and LODYC model over the North Atlantic may be explained by the very weak correlation between ENSO and the vertical wind shear over the Atlantic in those models. The UKMO, CNRM and MPI models, which have the best performance over the North Atlantic, are the models which display the strongest impact of ENSO on the Atlantic vertical wind shear, probably because they display the strongest variance in NINO3 SSTs (5°N–5°S, 170°W–120°W), whereas ECMWF and LODYC display a variance in NINO3 SSTs well below the observed variance. However, the UKMO, CNRM and MPI models tend to overestimate the impact of ENSO on the vertical wind shear over the eastern North Pacific. As a consequence, the frequency of tropical storms in the eastern North Pacific predicted by those three models is too strongly correlated to the vertical wind shear, and therefore negatively correlated to ENSO, which is not the case in observations, where local SSTs (positively correlated to ENSO) play a stronger role.

The performance of the CERFACS model is closer to the performance of the LODYC model (same oceanic component) than to the performance of the CNRM model (same atmospheric component) over the North Atlantic, the eastern North Pacific and the western North Pacific. This suggests that the oceanic component of the coupled GCM may have a significant impact on the interannual variability of tropical storms, although it does not seem to have such an impact on their seasonality.

A similar study has been done using ranked correlation instead of linear correlation. Results obtained with the ranked correlation are consistent with those obtained with linear correlation (not shown).

According to the previous section, the UKMO model displays the most realistic climatology. However, in terms of interannual variability, this model does not perform particularly well compared to the ECMWF or CNRM models, which have a much poorer climatology and seasonality. This suggests that tuning a model to get the best tropical storm climatology may not necessarily improve the interannual variability. This also suggests that the fact that models have very large biases in the frequency of tropical storms does not disqualify them from having skill in predicting the interannual variability of tropical storms.

(*d*) *Some sensitivity studies*

(i) *Impact of SST forecasts.* A comparison of the scores obtained with the coupled and atmosphere-only experiments forced with observed SSTs should give an indication of the impact of the SST forecasts on the prediction of tropical storm frequency. For this purpose, a 15-member ensemble using the atmospheric component of the ECMWF operational seasonal forecasting system forced by observed SSTs and starting on 1 May and 1 November from 1987 to 2001 has been created.

The ECMWF operational seasonal forecasting system is close to the system used in DEMETER. The main differences include the use of ERA-15 and operational analysis instead of ERA-40 for the generation of the atmospheric and oceanic initial conditions. Figure 3 suggests that the scores obtained with the atmospheric component forced with observed SSTs are generally slightly better than those obtained with the coupled GCM. The most significant difference is over the North Atlantic, where the atmospheric component forced with observed SSTs performs significantly better than the coupled model. This could be due to the lack of ENSO variability in the ECMWF coupled model (see discussion in section 4(c)) or to the low skill of the model in predicting Atlantic SSTs. The linear correlation between the observed SSTs and the SSTs predicted by the ECMWF model are below 0.5 for the forecasts starting on 1 May and averaged over the period August to October. This low correlation is consistent with the known problem that coupled models have with predicting SSTs (see for example Davey *et al.* 2002).

(ii) *Impact of oceanic data assimilation.* The ECMWF model used in DEMETER takes its oceanic initial conditions from the ECMWF ocean analysis. This analysis was built using oceanic data assimilation with an OI (ocean initialization) scheme (Alves *et al.* 2004). In addition to surface wind fluxes from the ECMWF analysis/reanalysis, subsurface observational data are used to construct the oceanic analysis. In order to evaluate the impact of the oceanic data assimilation, a second set of ECMWF integrations has been set up, where the oceanic initial conditions have been produced by running the oceanic model for 44 years forced by analysed surface fluxes. No *in situ* data have been used.

Figure 4 displays the scores obtained with and without ocean data assimilation over the period 1987 to 2001. The results are better with data assimilation than without over five ocean basins. This is particularly true over the Atlantic with a correlation of 0.37 with data assimilation, in contrast with zero correlation without data assimilation. Over the western North Pacific, the run without data assimilation has a correlation slightly larger than that with data assimilation (0.69 instead of 0.63). Over the South Indian Ocean, the correlation is very low in both cases, and not significant.

(iii) *Impact of the ensemble size.* In order to investigate the impact of the size of the ensemble, the ECMWF model has been integrated with 48 members without oceanic data assimilation. Figure 5 displays the anomaly correlation obtained using different ensemble sizes over the North Atlantic, the eastern Pacific and the western Pacific. For each ensemble size, all possible combinations of ensemble members were considered.

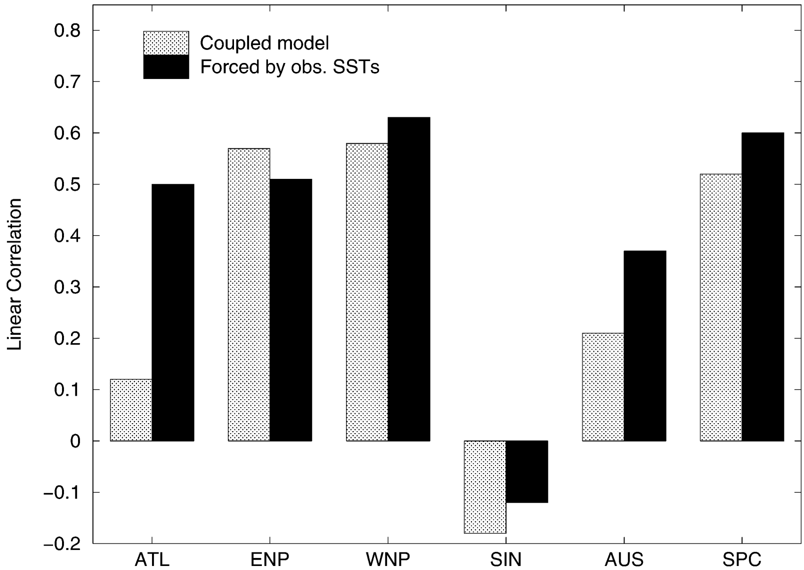

Figure 3. Linear correlation between the ensemble mean interannual variability of tropical storms and observations over the North Atlantic, the eastern North Pacific, the western North Pacific, the South Indian Ocean, the South Pacific, the Australian Basin and the South Pacific with the ECMWF operational coupled model and the ECMWF atmospheric model forced by observed SSTs. Solid bars correspond to forcing the atmospheric model with observed SSTs. Hatching corresponds to the coupled ocean–atmosphere model.

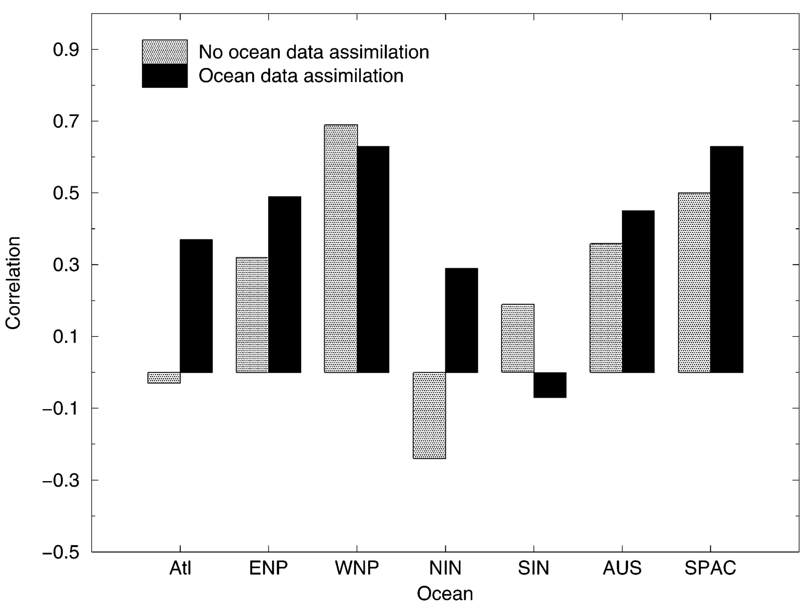

Figure 4. Linear correlation between the ensemble mean interannual variability of tropical storms and observations over the North Atlantic, the eastern North Pacific, the western North Pacific, the North Indian Ocean, the South Indian Ocean, the Australian Basin and the South Pacific with (solid bars) and without (hatching) oceanic data assimilation. Unlike in Figure 3, the ensemble integrations are from DEMETER. This explains why the solid column is not the same as the hatched column in Figure 3.

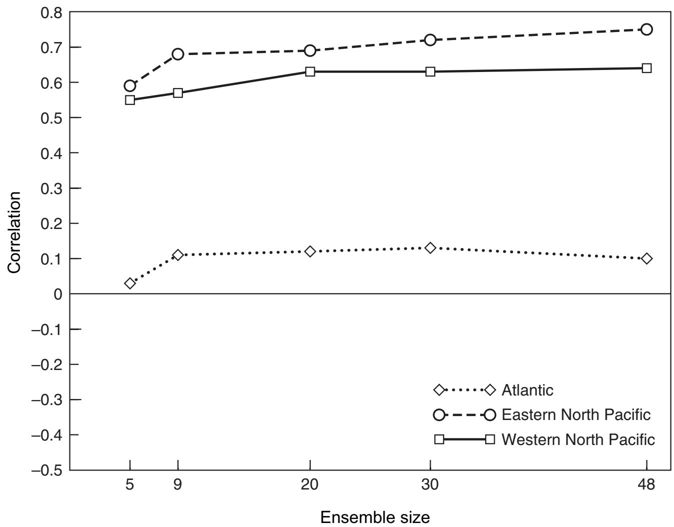

Figure 5. Linear correlation between the ensemble mean interannual variability of tropical storms and observations as a function of ensemble size over the North Atlantic, the eastern North Pacific and the western North Pacific.

TABLE 4. LINEAR CORRELATION

Basin	ATL	ENP	WNP	NIO	SIO	AUS	SPAC
ECMWF	0.03	**0.27**	0.08	0.0	0.17	0.20	**0.52**
CNRM	**0.30**	0.14	**0.27**	0.13	0.13	**0.26**	0.23
UKMO	0.22	**0.38**	0.14	0.04	0.05	**0.54**	**0.50**
MULTIMODEL	**0.24**	**0.48**	0.21	0.07	0.20	**0.41**	**0.53**

Same as Table 3 but for the period 1959–2001.

Results suggest that the scores obtained with the 5-member ensemble are not that far from those obtained with a 48-member ensemble. Increasing the ensemble size from 5 to 9 slightly improves the scores. Increasing the ensemble size from 9 to 48 does not seem to have a significant impact over the western North Pacific and the Atlantic, but seems to slightly increase the scores over the eastern North Pacific. Therefore, the fact that the multi-model ensemble generally outperforms the individual models is unlikely to be due to its larger ensemble size.

5. Results from the extended period 1959–2001

(*a*) *Interannual variability*

Three of the models' datasets (CNRM, ECMWF and UKMO) have been extended to cover the period 1959 to 2001. Over the whole period 1959–2001, the linear correlations are particularly low over the North Atlantic, western North Pacific, the North Indian and the South Indian Ocean, but still remain relatively high over the eastern North Pacific, the Australian Basin and the South Pacific (Table 4). The simple multi-model displays linear correlations (Table 4) or ranked correlations (not shown) that are overall higher than those obtained with the individual models.

The 1959–2001 period has been divided into three different periods of about the same size (15 years): 1959–1973, 1973–1987 and 1987–2001, in order to assess the variability of the scores from one decade to another. Figure 6 shows the score of the

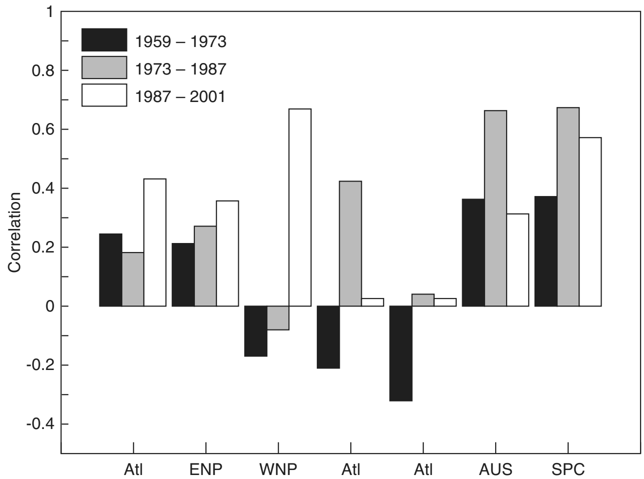

Figure 6. Linear correlation between the simple multi-model (based on CNRM, ECMWF and UKMO) ensemble mean interannual variability of tropical storm frequency and observations over the North Atlantic, the eastern North Pacific, the western North Pacific, the North Indian Ocean, the South Indian Ocean, the South Pacific, the Australian Basin and the South Pacific over three time periods: 1959–1973, 1973–1987 and 1987–2001.

simple multi-model ensemble over each basin for each individual period. The scores are higher during the most recent period (1987–2001) over the western North Pacific, the North Atlantic and the eastern North Pacific. This could be due in part to the fact that ENSO displayed more variance during the period 1987–2001 than during the two previous periods (the variance of NINO3 SSTs averaged over the period June to October was 0.9 degC during the period 1987–2001 instead of 0.7 degC during the periods 1959–1973 and 1973–1987). As a consequence, the frequency of tropical storms, particularly over the North Atlantic, may have been more predictable during the period 1987–2001. Over the western North Pacific, the simple multi-model has almost no skill at all during the previous decades in each individual model. Better ocean initial conditions over the western Pacific in the more recent period, thanks in particular to the TOGA observing system since 1991 may have contributed to the improvement in the seasonal prediction of tropical storm frequency over the western North Pacific during the last period. Over the North and South Indian Oceans, the scores are very low in all three periods, except for the period 1973–1987 over the North Indian Ocean. Over the southern hemisphere, the model displays some skill in all three periods.

The anomaly correlations are based on ensemble means, and do not take account of the individual realizations within each ensemble. Probabilistic scores such as the ranked probability skill score (RPSS) (Epstein 1969) are more suitable for ensemble forecasts since the ensemble mean is not always a good representation of the forecast, as for example in the case of a bimodal ensemble distribution. The period 1959–2001 (42 years) is probably sufficiently long to allow some comparison of the RPSS scores obtained with each individual ensemble and the score obtained with the simple multi-ensemble.

Table 5 displays the RPSS scores of the probability that the tropical storm frequency is in a given tercile. The RPSS scores of the simple multi-model ensemble are generally largely positive, suggesting that the multi-model has some skill in predicting the frequency of tropical storms. In addition, the simple multi-model ensemble has higher RPSS score than any individual model over each ocean basin, except over the

TABLE 5. RPSS SCORE

Basin	ATL	ENP	WNP	NIO	SIO	AUS	SPAC
ECMWF	0.12	0.24	0.17	0.11	0.39	0.28	0.29
CNRM	0.41	0.30	0.22	0.15	0.29	0.30	0.27
UKMO	0.27	0.31	0.22	0.15	0.37	0.35	0.33
MULTIMODEL	0.36	0.38	0.28	0.19	0.44	0.37	0.35

Same as Table 4 but for the RPSS score of the probability that the tropical storm frequency is in a given tercile.

North Atlantic, where the CNRM model displays a better score. However, the simple combination of the UKMO and CNRM model displays an RPSS score of 0.45, which is higher than the RPSS score obtained with the CNRM model alone. This suggests that combining different models improves the forecast of tropical storm frequency. The improvement is more visible in the ensemble distribution than in the ensemble mean.

(*b*) *Interdecadal variability*

In the previous section, the 1959–2001 period was divided in three 15-year-long periods. In the present section, the decadal variability of the frequency of model tropical storms is investigated over the whole period 1959–2001. Several observational studies have discussed the decadal variability of observed tropical storms over the Atlantic (Landsea and Gray 1992) or the South Pacific (Nguyen and Walsh 2001). The present section will investigate if the three models, which have been integrated from 1959 to 2001, reproduce such a signal. A 10-year running mean has been applied to the time series of the number of tropical storms predicted by each model, including the simple multi-model, and to the observed time series (from 1963 to 1997), in order to filter out the interannual signal.

The linear correlations between the observed and the multi-model ensemble mean time series of 10-year running mean of tropical storm frequency from 1963 to 1997 are particularly high over the eastern North Pacific (0.63), the western North Pacific (0.75), the Australian Basin (0.55) and the South Pacific (0.75) (Fig. 7). Such high correlations are likely to be explained by the fact that the models maintain the low frequency of SSTs during the 6-month period of model integrations, and 'translate' successfully this interdecadal variability of SSTs into an interdecadal variability of tropical storms generally consistent with observations. This result does not suggest that the numerical models can predict the interdecadal variability of SSTs responsible for the interdecadal variability of tropical storms; the 6-month integrations are too short to answer that question. Over the eastern North Pacific (Fig. 8(b)), the simple multi-model simulates an increase in tropical storm frequency in the 1980s, but fails to simulate a reduction of tropical storm activity in the 1990s. Over the western North Pacific the simple multi-model simulates a reduction of tropical storm activity in the 1970s and 1980s, and more tropical cyclone activity in the 1960s and 1980s (Fig. 8(c)). Over the South Pacific (Fig. 8(d)), the model tropical storms display an almost steady increase in the frequency since the 1960s, which is consistent with observations.

The North Atlantic basin (Fig. 8(a)) represents an interesting case. Observations suggest that the North Atlantic displays a significant decadal variability in the frequency of tropical storms, in phase with the decadal variability of rainfall over the Sahel (Landsea and Gray 1992). Several papers (Saunders and Harris 1997; Goldenberg and Landsea 1997; Landsea *et al.* 1999) have pointed out that the decadal variability of SSTs over the tropical Atlantic makes a key contribution to the observed decadal variability

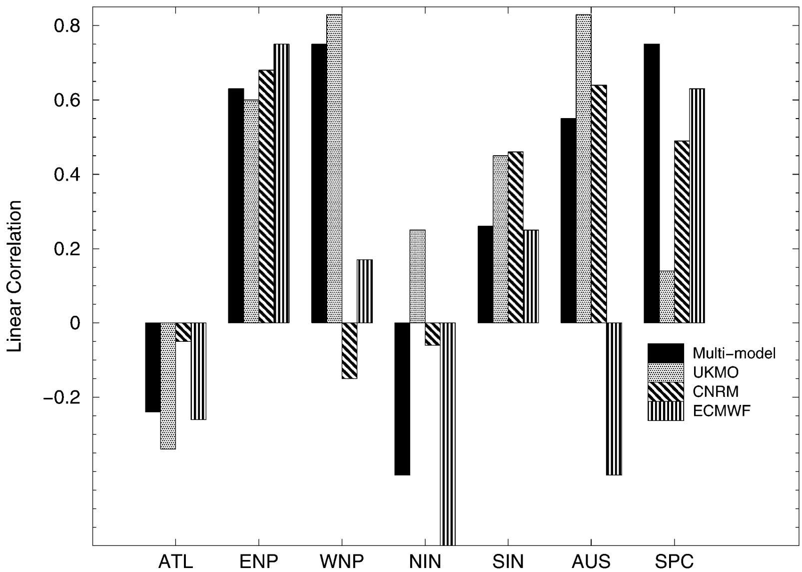

Figure 7. Linear correlation between the UKMO, ECMWF, CNRM and the simple multi-model ensemble mean 10-year running mean of tropical storm frequency and observations over the North Atlantic, the eastern North Pacific, the western North Pacific, the North Indian Ocean, the South Indian Ocean, the South Pacific, the Australian Basin and the South Pacific.

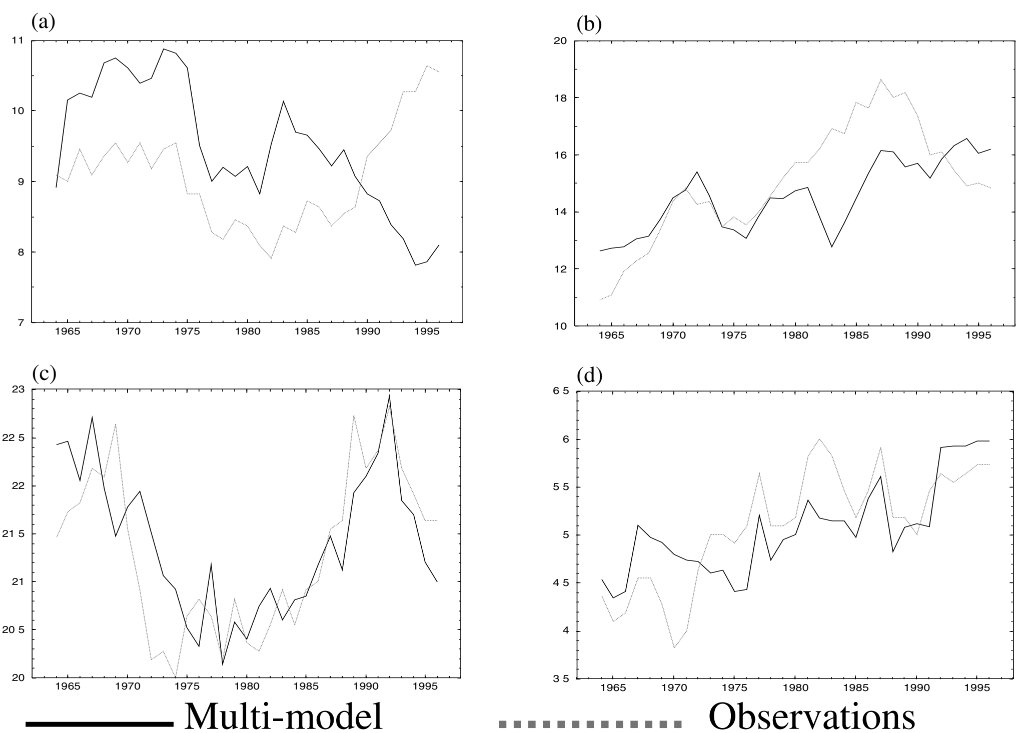

Figure 8. Time evolution of the 10-year running mean of multi-model tropical storm frequency and observations over (a) the North Atlantic, (b) the eastern North Pacific, (c) the western North Pacific and (d) the South Pacific.

of Atlantic tropical storm frequency. Anomalously warm tropical Atlantic SSTs in the 1950s are likely to explain the large number of Atlantic tropical storms during that decade, and anomalously cold tropical Atlantic SSTs may have caused the observed reduction of Atlantic tropical storms in the 1970s and 1980s. Vitart and Anderson (2001) have shown that the GFDL model forced by prescribed SSTs was able to simulate the impact of tropical Atlantic SSTs on the frequency of Atlantic tropical storms. In the present study, the models are coupled. Figure 8(a) shows that the simple multi-model ensemble displays some interdecadal variability over the Atlantic, with more tropical cyclone activity in the 1960s than in the 1970s, as in observations. However, the simple multi-model fails to simulate the increase of tropical cyclone activity in the 1990s. This difference between the observed and predicted frequency of Atlantic tropical storms is not due to one single year.

The inconsistency between the decadal variability of Atlantic tropical storms in the 6-month integrations from 1959 to 2001 and observations can be partially explained by the poor performance of the models to simulate the decadal variability of Atlantic SSTs. A linear correlation between the interdecadal variability of tropical storms and of SSTs reveals that for each model there is strong correlation over the Atlantic basin, as in observations (not shown). However, the simulated SSTs are not always consistent with the observed SSTs. Figure 9 shows the linear correlation between the interdecadal variability of predicted SSTs with each individual model for month 2 and month 6 and the corresponding observed SSTs. This figure shows that over most of the region where Atlantic tropical storms develop, the correlation is below 0.4 and is even negative over the Gulf of Mexico, and the eastern US coast by month 6. This is the case for the three models, and is an example where the simple multi-model approach does not improve the forecast. In addition, the three models display very low correlation (below 0.4) between the interdecadal variability of predicted Atlantic SSTs in month 1 and in month 6 (not shown). The correlations are indeed much lower than in observations and much lower than in the other basins. This indicates that the coupled models have problems in maintaining the interdecadal signal during the 6 months of the integrations over the tropical North Atlantic basin. Therefore, the low correlation between the interdecadal variability of Atlantic tropical storms in the simple multi-model and in observations is likely to be due to the poor performance of the models to predict Atlantic SSTs rather than to a poor performance in translating SST decadal variability into tropical storm decadal variability. This is consistent with Davey *et al.* (2002) who have shown that coupled models tend to perform poorly in this region.

6. CONCLUSION

A method for tracking tropical storms has been applied to the model integrations produced for the DEMETER project. The main result of this paper is that the multi-model ensemble technique applied to seasonal forecasts of tropical storm frequency produces forecasts that are overall better than any individual model. In addition, over each ocean basin, it is possible to find a combination of models which performs better than the best individual model. This is true for deterministic scores, such as anomaly correlations, as well as for probabilistic scores, such as RPSS. In principle, the multi-model approach may not always be the best choice if one model is so much better than all the others, that any combination with another model would reduce the skill. This situation did not happen with the seven models of the DEMETER project, because all the models displayed performances that were overall comparable, and most importantly because they presented different types of errors. Models that were skilful over the North

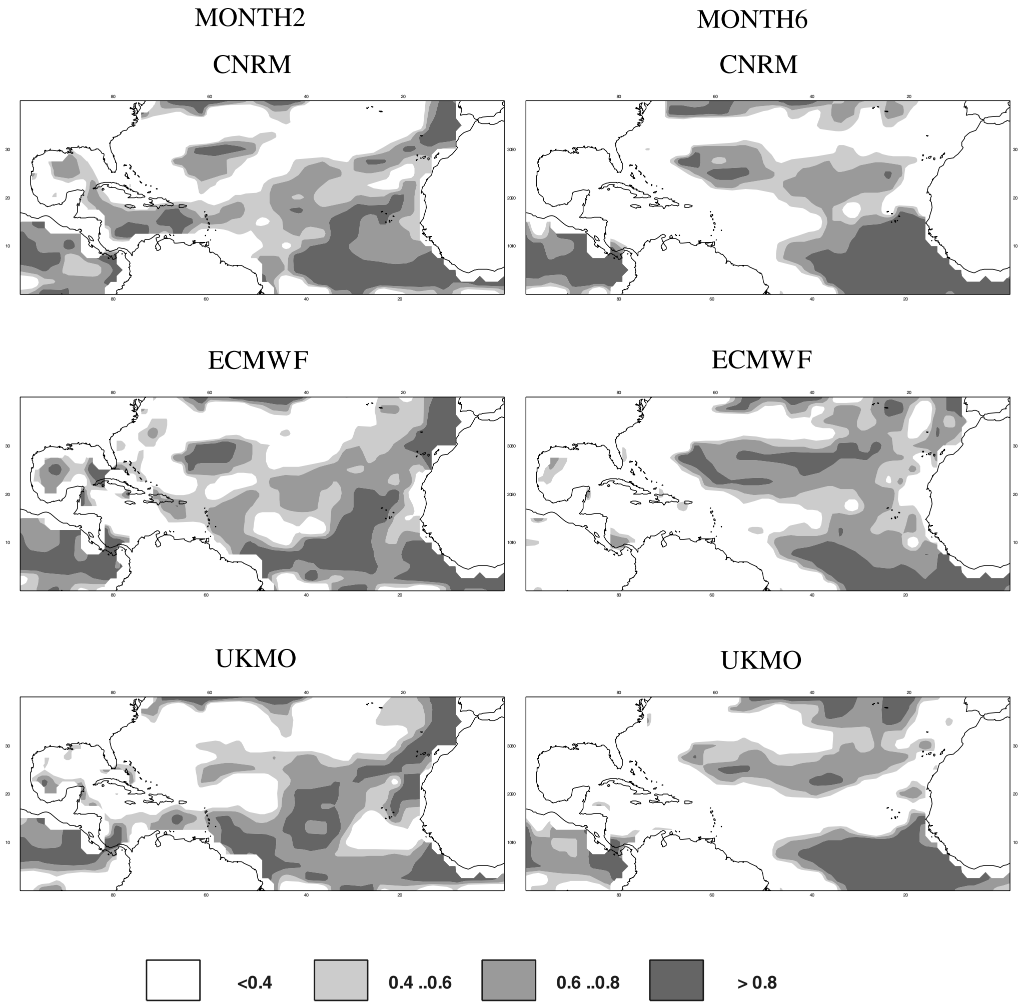

Figure 9. Linear correlation between the interdecadal variability of SSTs predicted by each individual model and observations for month 2 (left panels) and month 6 (right panel) over the North Atlantic.

Atlantic performed poorly over the eastern North Pacific, and vice versa. Even over a specific basin, the models displayed different model errors; some displayed a frequency of tropical storms strongly correlated to ENSO and were skilful during ENSO years, but underestimated the impact of local SSTs. Other models underestimated the impact of ENSO over the North Atlantic but were successful in predicting the record frequency of tropical storms in 1995 due to warmer local SSTs. Combining those models helped to filter out model errors.

The simple multi-model displays some skill over all the basins except the North and South Indian Oceans during the period 1987–2001. Over the Indian basins, the intraseasonal variability such as the Madden–Julian Oscillation (MJO) may play a strong role in the modulation of tropical storm activity, and GCMs are generally not very skilful in predicting the MJO (Slingo *et al.* 1996). The multi-model has a particularly high skill over the western North Pacific and the South Pacific for the period 1987–2001. All the models display high skill over these basins, whereas some models display poor skill over the Atlantic and eastern North Pacific. Therefore the multi-model approach seems to be beneficial for the seasonal prediction of tropical storms. A multi-model forecast of tropical storm frequency will become operational at ECMWF, based on three models: ECMWF, CNRM and UKMO.

In this study, models with finer resolution do not seem to perform significantly better than models with a lower resolution. In addition, models with the best climatology, like the UKMO model, do not display the best interannual variability of tropical storm frequency. The atmospheric component of the GCMs seems to have a more important impact on the scores than the oceanic component. For instance, ECMWF and the LODYC models, which share the same atmospheric component, but have different oceanic components, perform relatively similarly. Over the Atlantic, this could be due to the fact that all the ocean models share the same deficiencies in predicting Atlantic SSTs. When forced by observed SSTs, the atmospheric component of the ECMWF model produces an interannual variability of Atlantic tropical storms much more consistent with observations than the coupled model. Therefore, improvements in the prediction of Atlantic SSTs can significantly improve the prediction of Atlantic tropical storms. Over other basins, like the eastern North Pacific and the western North Pacific, the ECMWF atmospheric model forced by observed SSTs and the ECWMF coupled model have very similar performances, suggesting that further improvements in ocean modelling and ocean data assimilation are not likely to improve significantly the scores over those basins.

A study over a longer period (1959–2001) shows that the performance of the simple multi-model ensemble is much more modest over this 43-year period than during the past 15 years. Over this long time-scale, the simple multi-model has skill only over a few basins, such as the South Pacific and eastern North Pacific. Surprisingly, the simple multi-model has very low skill over the western North Pacific. Improvements in the observation data over the last 15 years may explain this discrepancy in the skill from one decade to another. Over most basins, the simple multi-model captures well the interdecadal variability. However, this is not the case over the North Atlantic, where all the models fail to simulate more tropical storms in the 1990s than in the other decades. This seems to be due at least partially to the difficulty of the models to maintain the interdecadal variability of Atlantic SSTs during the 6 months of integrations.

Overall the multi-model approach appears promising, and seems to help to filter out model error for the prediction of tropical storm frequency. Other statistics like tropical storm location and risk of tropical storm landfall could also benefit from the multi-model approach. However, the dynamical seasonal prediction of the risk of tropical storm landfall needs seasonal forecasting systems with a higher horizontal resolution than is presently available (Vitart *et al.* 2003). Future research will focus on those statistics. In addition, the present paper discussed just a simple way of combining the different models. Equal weight was given to all the models. The calibration of each individual model was also very simple. More sophisticated methods of combining the different models (Krishnamurti *et al.* 2000; Doblas-Reyes *et al.* 2005; Stephenson *et al.* 2005) and calibrating the forecasts (see e.g. Hamill *et al.* 2004) are likely to improve the scores of the multi-model ensemble. This will be explored in further studies.

Acknowledgements

The author wish to thanks David Anderson and Francisco Doblas-Reyes for their useful and stimulating remarks, and two anonymous reviewers whose comments proved invaluable in improving the presentation of the material.

REFERENCES

Alves, O., Balmaseda, M. A., Anderson, D. L. T. and Stockdale, T. 2004 Sensitivity of dynamical seasonal forecasts to ocean initial conditions. *Q. J. R. Meteorol. Soc.*, **130,** 647–667

Bengtsson, L., Böttger, H. and Kanamitsu, M. 1982 Simulation of hurricane-type vortices in a general circulation model. *Tellus*, **34,** 440–457

Bengtsson, L., Botzet, M. and Esch, M. 1995 Hurricane-type vortices in a general circulation model. *Tellus*, **47A,** 175–196

Camargo, S. J. and Zebiak, S. E. 2002 Improving the detection and tracking of tropical cyclones in atmospheric general circulation models. *Weather and Forecasting*, **17,** 1152–1162

Davey, M. K., Huddleston, M., Sperber, K. R., Braconnot, P., Bryan, F., Chen, D., Colman, R. A., Cooper, C., Cubasch, U., Délécluse, P., DeWitt, D., Fairhead, L., Flato, G., Gordon, C., Hogan, T., Ji, M., Kimoto, M., Kitoh, A., Knutson, T. R., Latif, M., Le Treut, H., Li, T., Manabe, S., Mechoso, C. R., Meehl, G. A., Power, S. B., Roeckner, E., Terray, L., Vintzileos, A., Voss, R., Wang, B., Washington, W. M., Yoshikawa, I., Yu, J.-Y., Yukimoto, S. and Zebiak, S. E. 2002 STOIC: A study of coupled model climatology and variability in tropical ocean regions. *Clim. Dyn.*, **18,** 403–420

Doblas-Reyes, F. J., Hagedorn, R. and Palmer, T. N. 2005 The rationale behind the success of multi-model ensembles in seasonal forecasting. Part II: Calibration and combination. *Tellus*, **57A,** 234–252

Epstein, E. S. 1969 A scoring system for probability forecasts of ranked categories. *J. Appl. Meteorol.*, **8,** 985–987

Goldenberg, S. G. and Landsea, C. W. 1997 'Are we seeing a long-term upturn in Atlantic basin major hurricane activity related to decadal-scale SST fluctuations?' Pp. 305–310 in Preprints, Seventh Conf. on Climate Variations, Long Beach, California. Amer. Meteorol. Soc.

Gray, W. M. 1979 Hurricanes: Their formations, structure and likely role in the tropical circulation. Pp. 155–218 in *Meteorology over the tropical oceans*. Ed. D. B. Shaw. R. Meteorol. Soc.

Haarsma, R. J., Mitchell, J. F. B. and Senior, C. A. 1993 Tropical disturbances in a GCM. *Climate Dyn.*, **8,** 247–257

Hagedorn, R., Doblas-Reyes, F. J. and Palmer, T. N. 2005 The rationale behind the success of multi-model ensembles in seasonal forecasting. Part I: Basic concept. *Tellus*, **57A,** 219–233

Hamill, T. M., Whitaker, J. S. and Wei, X. 2004 Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts. *Mon. Weather Rev.*, **132,** 1434–1447

Krishnamurti, T. N., Kishtawal, C. M., Zhang, Z., LaRow, T., Bachiochi, D., Williford, E., Gadgil, S. and Surendran, S. 2000 Multimodel ensemble forecasts for weather and seasonal climate. *J. Climate*, **13,** 4196–4216

Landsea, C. W. and Gray, W. M. 1992 The strong association between western Sahelian monsoon rainfall and intense Atlantic hurricanes. *J. Climate*, **5,** 435–453

Landsea, C. W., Pielke, R. A. Jr., Mestas-Nuñez, A. M. and Knaff, J. A. 1999 Atlantic basin hurricanes: Indices of climatic changes. *Clim. Change*, **42,** 89–129

Manabe, S., Holloway, J. L. Jr. and Stone, H. M. 1970 Tropical circulation in a time-integration of a global model of the atmosphere. *J. Atmos. Sci.*, **27,** 580–613

Neumann, C. J., Jarvinen, B. R., McAdie, C. J. and Elms, J. D. 1993 *Tropical cyclones of the North Atlantic Ocean, 1871–1992*. Historical Climatology Series, Vol. 6-2, NOAA

Nguyen, K. C. and Walsh, K. J. E. 2001 Interannual, decadal, and transient greenhouse simulation of tropical cyclone-like vortices in a regional climate model of the South Pacific. *J. Climate*, **14,** 3043–3054

Palmer, T. N. and Shukla, J. 2000 Editorial of the special issue DSP/PROVOST. *Q. J. R. Meteorol. Soc.*, **126,** 1989–1990

Palmer, T. N., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M., Délécluse, P., Déqué, M., Díez, E., Doblas-Reyes, F. J., Feddersen, H., Graham, R., Gualdi, S., Guérémy, J.-F., Hagedorn, R., Hoshen, M., Keenlyside, N., Latif, M., Lazar, A., Maisonnave, E., Marletto, V., Morse, A. P., Orfila, B., Rogel, P., Terres, J.-M. and Thomson, M. C. 2004 Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). *Bull. Am. Meteorol. Soc.*, **85,** 853–872

Saunders, M. A. and Harris, A. R. 1997 Statistical evidence links exceptional 1995 Atlantic hurricane season to record sea warming. *Geophys. Res. Lett.*, **24,** 1255–1258

Slingo, J. M., Sperber, K. R., Boyle, J. S., Ceron, J.-P., Dix, M., Dugas, B., Ebisuzaki, W., Fyfe, J., Gregory, D., Guérémy, J.-F., Hack, J., Harzallah, A., Inness, P., Kitoh, A., Lau, W. K.-M., McAvaney, B., Madden, R, Matthews, A., Palmer, T. N., Park, C.-K., Randall, D. and Renno, N. 1996 Atmospheric Model Intercomparison Project (AMIP): Intraseasonal oscillations in 15 atmospheric general circulation models (results from an AMIP diagnostic subproject). *Clim. Dyn.*, **12,** 325–357

Stephenson, D. B., Coelho, C. A., Doblas-Reyes, F. J. and Balmaseda, M. 2005 Forecast assimilation: A unified framework for the combination of multi-model weather and climate predictions. *Tellus*, **57A,** 253–264

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

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., da Costa Bechtold, V., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., McNally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P. and Woollen, J. 2005 The ERA-40 Reanalysis. *Q. J. R. Meteorol. Soc.*, **131,** 2961–3012

Vitart, F. and Anderson, J. L. 2001 Sensitivity of Atlantic tropical storm frequency to ENSO and interdecadal variability of SSTs in an ensemble of AGCM integrations. *J. Climate*, **14,** 533–545

Vitart, F. and Stockdale, T. N. 2001 Seasonal forecasting of tropical storms using coupled GCM integrations. *Mon. Weather Rev.*, **129,** 2521–2537

Vitart, F., Anderson, J. L. and Stern, W. F. 1997 Simulation of the interannual variability of tropical storm frequency in an ensemble of GCM integrations. *J. Climate*, **10,** 745–760

1999 Impact of large-scale circulation on tropical storm frequency, intensity and location simulated by an ensemble of GCM integrations. *J. Climate*, **12,** 3237–3254

Vitart, F., Anderson, D. and Stockdale, T. 2003 Seasonal forecasting of tropical cyclone landfall over Mozambique. *J. Climate*, **16,** 3932–3945

Yoo, J. H. and Kang, I.-S. 2005 Theoretical examination of a multi-model composite for seasonal prediction. *Geophys. Res. Lett.*, **32,** L18707, doi: 10.1029/2005GL023513