

Dynamically-based seasonal forecasts of Atlantic tropical storm activity issued in June by EUROSIP

F. Vitart,¹ M. R. Huddleston,² M. Déqué,³ D. Peake,² T. N. Palmer,¹ T. N. Stockdale,¹ M. K. Davey,^{2,4} S. Ineson,² and A. Weisheimer¹

Received 20 May 2007; revised 21 June 2007; accepted 19 July 2007; published 24 August 2007.

[1] Most seasonal forecasts of Atlantic tropical storm numbers are produced using statistical-empirical models. However, forecasts can also be made using numerical models which encode the laws of physics, here referred to as “dynamical models”. Based on 12 years of re-forecasts and 2 years of real-time forecasts, we show that the so-called EUROSIP (EUROpean Seasonal to Inter-annual Prediction) multi-model ensemble of coupled ocean atmosphere models has substantial skill in probabilistic prediction of the number of Atlantic tropical storms. The EUROSIP real-time forecasts correctly distinguished between the exceptional year of 2005 and the average hurricane year of 2006. These results have implications for the reliability of climate change predictions of tropical cyclone activity using similar dynamically-based coupled ocean-atmosphere models. **Citation:** Vitart, F., M. R. Huddleston, M. Déqué, D. Peake, T. N. Palmer, T. N. Stockdale, M. K. Davey, S. Ineson, and A. Weisheimer (2007), Dynamically-based seasonal forecasts of Atlantic tropical storm activity issued in June by EUROSIP, *Geophys. Res. Lett.*, *34*, L16815, doi:10.1029/2007GL030740.

1. Introduction

[2] Publicly-available forecasts of seasonal Atlantic tropical cyclone activity include the Colorado State University forecasts from P. J. Klotzbach and W. M. Gray (referred to as CSU forecasts; see Extended range forecast of Atlantic seasonal hurricane activity and U.S. landfall strike probability for 2006, available at <http://tropical.atmos.colostate.edu/forecasts/2006/june2006>), Tropical Storm Risk [Lea and Saunders, 2006] (referred to as TSR forecasts; see TSR forecasts, <http://tropicalstormrisk.com>) and NOAA (NOAA outlook, <http://www.cpc.ncep.noaa.gov/products/outlooks/hurricane.shtml>). Each of these is based on statistical models which encode empirical correlations between Atlantic tropical cyclone activity and precursor climatic predictors. The CSU forecasts also include an analog method and a final adjustment. In spring 2006, there was a consensus among these statistical empirical forecasts that the 2006 Atlantic season would be active, though less than the record-breaking 2005 season. Such forecasts proved inaccurate, as the 2006 tropical storm season with 10 tropical storms was slightly below the long term climatology of 10.4 from 1958 to 2001.

As highlighted in the media [Stevenson, 2006; Johnston, 2006], a number of insurance companies who took action based on these forecasts lost billions of dollars. Saunders [2006] attributes the low tropical storm activity in 2006 to El-Niño conditions and the presence of African dry air and Saharan dust.

[3] Dynamical coupled ocean-atmosphere climate models can also be used to predict seasonal tropical cyclone activity. Such global climate forecasts are made from sets of initial conditions preceding the onset of the hurricane season. These initial states are determined by atmosphere and ocean observations worldwide, and assimilated into the dynamical model. State dependent uncertainty in the resulting forecasts is estimated by running ensembles of integrations in which uncertainty in both initial conditions and model equation sets are represented explicitly [Palmer and Hagedorn, 2006]. There exist a number of different methods to represent model uncertainty [Palmer et al., 2005, Collins, 2007]. Here we focus on the use of the so-called multi-model technique, pioneered on the seasonal timescale by the DEMETER project [Palmer et al., 2004].

[4] Although the horizontal resolution of global operational dynamical seasonal forecasting models is generally insufficient to simulate the intensity of hurricanes, simulated tropical cyclonic systems are nevertheless realistic in other respects. For example, they develop a warm temperature anomaly above the centre of the vortex, which is a characteristic of observed tropical storms. Hindcast experiments have shown that dynamical models forced by observed SSTs [Vitart et al., 1997; Camargo et al., 2005] and fully coupled ocean-atmosphere models [Vitart, 2006] can simulate the inter-annual variability in Atlantic tropical storm activity.

2. The EUROSIP System

[5] The success of the DEMETER project led directly to the development of the operational EUROSIP multi-model ensemble. EUROSIP presently consists of 3 seasonal forecasting systems from ECMWF, Met Office and Météo-France and global probability forecasts are produced routinely for several variables. Each forecasting system is run in real-time with initial conditions from the 1st of each month and the forecasts are issued the 15th of the month (the delay allows acquisition of SST fields from the previous month, time to run the forecasts, and a margin to ensure a reliable operational schedule). In real-time mode, the ensemble size for each model is either 40 or 41. In addition, each coupled ocean-atmosphere model was used to produce a set of re-forecasts initialized with the ocean and atmosphere analyses of the 1st of the month for

¹European Centre for Medium-Range Weather Forecasts, Reading, UK.

²Met Office, Exeter, UK.

³Météo-France, Toulouse, France.

⁴Department of Mathematics, University College London, London, UK.

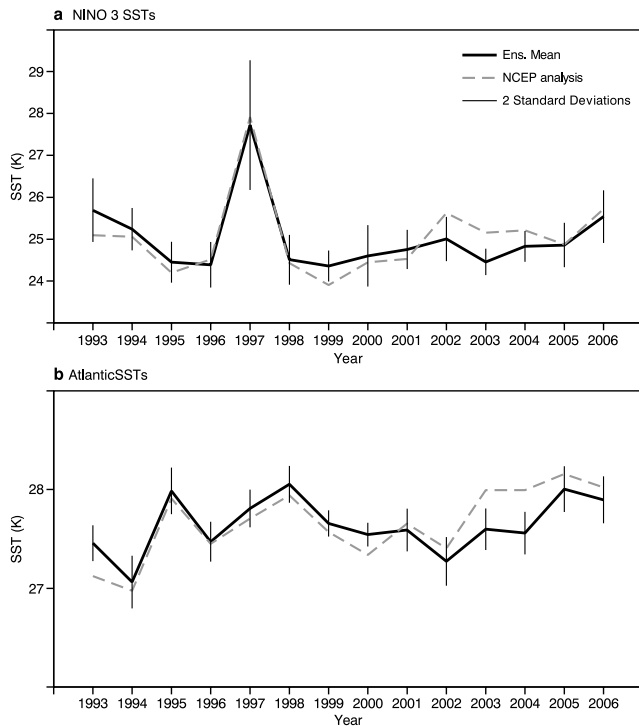


Figure 1. SST averaged over the period August to October predicted by EUROSIP after calibration (thick black line) starting on 1st June for (top) the NINO3 region (150W–90W, 5N–5S) and (bottom) the Atlantic main hurricane development region (20W–60W, 10N–20N). The dashed grey line represents the SST observations (analysis from NCEP OIv2 [Reynolds *et al.*, 2002]), the thick black line represents the ensemble mean and the vertical lines represent 2 standard deviations within the multi-model ensemble distribution.

each year (the SSTs are predicted by the coupled system). The number of re-forecasts varies between models. The re-forecasts share a common period of 1993–2004, with an ensemble size smaller than in real-time, but of at least 5 members.

[6] The tropical storms produced by each model component of EUROSIP are tracked using the method described by Vitart and Stockdale [2001]. Since the component models have biases which can vary from one model to another, the number of model tropical storms is calibrated a posteriori using a set of past integrations (see Appendix A).

[7] A large portion of the seasonal variability of Atlantic tropical storms is associated with sea surface temperature (SST) variations over the tropical Pacific [Gray, 1984; Shapiro, 1987; Goldenberg and Shapiro, 1997] and Atlantic [Saunders and Harris, 1997; Goldenberg and Shapiro,

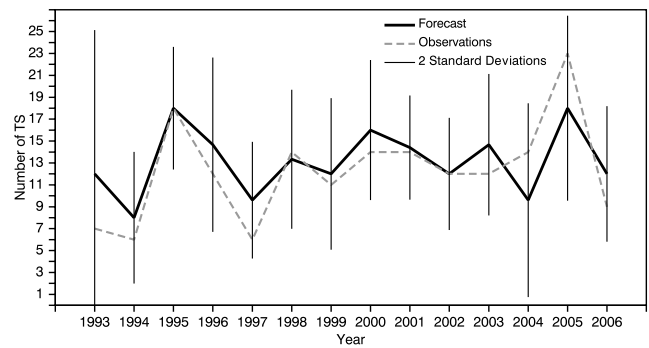


Figure 2. Number of tropical storms from July to November predicted by EUROSIP (median) starting on 1st June (thick black line) for the period 1993–2006. Re-forecasts were used for the period 1993–2004, and real-time forecast ensembles (calibrated using the median, as described in methods) were used for the period 2005–2006. The dashed grey line represents observations from July to November and the vertical lines represent 2 standard deviations within the multi-model ensemble distribution.

1997; Landsea *et al.*, 1999]. Therefore it is crucial for a coupled dynamical system to predict correctly the seasonal variability of SSTs in these two regions. Figure 1 displays the SSTs averaged over the peak August–October period of Atlantic tropical storm activity over the NINO3 region (top panel), often used as an index for ENSO activity, and over the hurricane main development region over the Atlantic (bottom panel), as predicted by the ensemble mean of EUROSIP after calibration for predictions starting from 1st June. The ensemble 2 standard deviation range is also indicated. EUROSIP has clear skill in predicting the evolution of SSTs in both regions, with a correlation between the EUROSIP ensemble mean and verification of 0.92 (p-value of 0.00001) over the NINO3 region and 0.81 (p-value of 0.001) over the Atlantic. Those correlations are larger than the correlations obtained by persisting SST anomalies from the previous month (respectively 0.47 and 0.73 which have p-values of 0.07 and 0.003 respectively).

[8] Figure 2 shows the EUROSIP re-forecasts of Atlantic tropical storm frequency (produced a posteriori) for the common re-forecast period 1993–2004 and the forecasts for 2005 and 2006 along with the observed frequency of Atlantic tropical storms for the period July to November (the forecasts are issued mid-June). The multi-model displays some skill in reproducing the observed inter-annual variability of Atlantic tropical storms, with the model successfully predicting intense tropical storm activity in 1995 and 2005 as in observations. Table 1 shows the linear correlation and RMS error between the inter-annual variability of the multi-model ensemble median and the ob-

Table 1. Linear Correlation and Root Mean Square Error Between the Number of Atlantic Tropical Storms Predicted by Each Individual Dynamical Model, EUROSIP, TSR and CSU and the Observed Number of Tropical Storms During the Whole Atlantic Hurricane Season Over the Period 1993–2006^a

	Model 1	Model 2	Model 3	EUROSIP	TSR	CSU
Correlation	0.67 (0.01)	0.78 (0.002)	0.59 (0.027)	0.81 (0.00036)	0.65 (0.004)	0.39 (0.16)
RMS error	3.35	3.65	4.4	3.26	4.70	4.76

^aObserved number of tropical storms is taken from <http://www.nhc.noaa.gov>. Numbers in parentheses correspond to the p-values of the correlation.

Table 2. Same as Table 1 but Over the Period 1987–2006^a

	ECMWF+UKMO	TSR	CSU
Correlation	0.80 (0.0002)	0.70 (0.0005)	0.54 (0.014)
RMS error	3.25	4.18	4.17

^aOnly the Met Office and ECMWF seasonal forecasting systems have re-forecasts covering the period 1987–2006. Numbers in parentheses correspond to the p-values of the correlation.

served frequency of Atlantic tropical storms over the full season, along with the same statistics for CSU and TSR forecasts over the same period. For the CSU forecasts, we use the modified final forecasts which were issued in real time. Re-forecasts from TSR were used to cover the full period 1993–2006 and are produced the same way as the real-time forecasts. Table 1 suggests that over the period 1993–2006, EUROSIP had higher skill than CSU and TSR forecasts. In particular EUROSIP displays an RMS error substantially smaller than the statistical models despite the fact that it displays more variance (2.91 instead of 2.64 for TSR and 2.15 for CSU) during this period. If we remove the year 2006, the TSR and CSU linear correlations increase significantly (0.77 and 0.67 respectively), but still remain lower than EUROSIP (0.815), and the RMS error of both statistical models is still much higher than with EUROSIP. Table 1 indicates also that EUROSIP multi-model produces better scores, particularly for RMS errors, than the individual model components, in agreement with Vitart [2006]. The dynamical models perform also better than the statistical methods over a longer period of time (20 years instead of 14 years) (Table 2). A 10,000 bootstrap re-sampling procedure indicates that the difference in skill between EUROSIP and CSU and EUROSIP and TSR is significant at the 1% level for both time periods. However, the comparison may not be entirely fair for the CSU forecasts, since the statistical model used by CSU before the final adjustment has changed several times in the past. In addition, the EUROSIP forecasts are issued slightly later than the TSR and CSU forecasts.

[9] EUROSIP ensemble forecasts can be issued as probabilities (for instance the probability of the number of

tropical storms to be above normal), but a large number of cases (larger than the current size of the model re-forecast common period 1993–2004) is needed to validate the reliability of the probability forecasts of tropical cyclone numbers. However, a version close to EUROSIP was integrated over a period of 49 years for 6-months starting on 1st May for the DEMETER project [Palmer *et al.*, 2004].

[10] The reliability of DEMETER Atlantic activity forecasts is measured using so-called attributes diagrams [Hsu and Murphy, 1986; Wilks, 2005] which show the conditional relative frequency of occurrence of an event as a function of its forecast probability (see Figure 3) with a perfectly reliable system having data close to the diagonal. For purposes of assessment, forecasts of above and below normal activity were grouped into three categories according to the forecast probability, 0–33.3%; 33.3% to 66.6% and 66.6 to 100%. Figure 3 suggests that the multi-model ensemble forecasts has more skill than the trivial climatological information in predicting the probability of a more intense or less intense Atlantic tropical storm season.

3. The 2005 and 2006 Seasons

[11] Global forecasts from EUROSIP have been produced in real-time each month since early 2005. This means that two real-time seasonal forecasts of the number of Atlantic tropical storms have been produced so far: the 2005 and the 2006 seasons. Table 3 shows the forecast of Atlantic tropical storms for each individual model and for EUROSIP for 2005 and 2006.

[12] 2005 was a record season for Atlantic tropical storm activity, with 27 tropical storms, which is well above the 1965–2004 climatology (10.6 Atlantic tropical storms per year). The multi-model forecast (16.2 Atlantic tropical storms) made in June 2005 was more accurate than the statistical predictions from CSU (15 tropical storms) and TSR (about 14 tropical storms). Although the ensemble mean and median of EUROSIP are well below the number of observed tropical storms in 2005, EUROSIP predicted probabilities of an extreme season in 2005 (37% and 13% probability for the number of Atlantic tropical storms to be above 20 and 27 respectively) higher than the probabilities

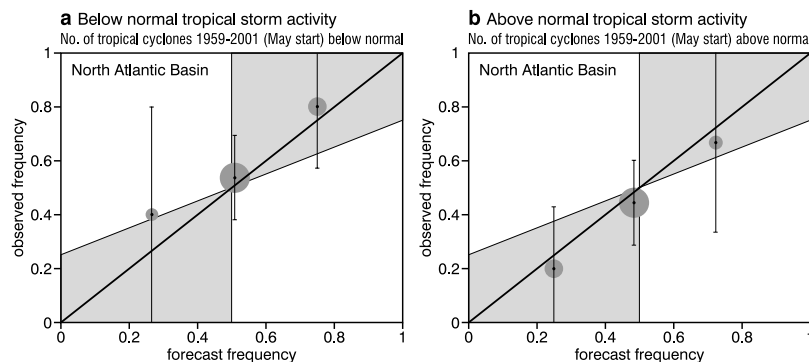


Figure 3. Attributes diagram for the number of tropical cyclones below and above normal in the North Atlantic Basin (1959–2001) from multi-model re-forecasts. The data are derived from the DEMETER ensemble re-forecasts and are based on 3 coupled circulation models with 9 ensemble members each. The size of the dots is proportional to the fraction of probability forecasts within the three probability categories. Grey areas indicate regions where the data would contribute to a positive Brier Skill Score [Wilks, 2005; Mason, 2004] if compared to a climatological reference forecasts. The error bars (90% level) were computed from a 10,000 bootstrap re-sampling procedure.

Table 3. Forecasts of the Number of Atlantic Tropical Storms Issued in June 2005 and 2006 for Each Component Model of EUROSIP and for the Multi-Model Ensemble, Along With Forecasts From CSU and TSR Issued in June and Forecasts From NOAA Issued in May^a

	Model 1	Model 2	Model 3	EUROSIP	NOAA	TSR	CSU	OBS
2005	14.3	15	19.4	16.2	12–15	13.8	15	27
2006	10.5	10.7	15.3	12.1	13–16	13.9	17	10

^aIf we had used the median as in Figure 2 instead of the mean, the EURO-SIP values would have been 18 and 12 for 2005 and 2006 respectively.

produced by EUROSIP for those thresholds during the re-forecast period 1993–2004 (18% and 4.2% probabilities respectively). In June 2006, EUROSIP predicted 12.1 tropical storms. This forecast issued in June was in contrast to the forecasts from statistical models, which predicted a very active hurricane season. For instance, CSU and TSR predicted 17 and about 14 tropical storms, respectively. Although EUROSIP predicted higher tropical storm activity than observed (12.1 instead of 10), the multi-model forecast proved more accurate than the statistical models in this case.

4. Further Implications

[13] Because of the coarse resolution of the dynamical models, the simulated tropical storm tracks tend to be unrealistically short (see for example Figure 4a) and the strength of the tropical storm is weaker than observed values. Because of this, the EUROSIP forecasts are at present limited to the frequency of tropical storms, unlike CSU and TSR forecasts which are more detailed and include the prediction of the risk of US landfall and the frequency of intense hurricanes. However, planned increases in the resolution of EUROSIP seasonal forecasting models (from about 200 km to about 100 km resolution) might make it possible to produce direct forecasts of tropical storm landfall, although the models still produces significantly less intense hurricanes than observed. Alternatively, statistical techniques could be applied to infer landfall probabilities from model outputs. An illustration of the expected impact of increased resolution is given in Figure 4b.

[14] In conclusion, this paper describes an inherently different approach from publicly available statistical methods to predict the frequency of tropical storms. Here we use a state-of-the-art dynamical seasonal forecasting system based directly on the laws of physics, which for tropical cyclone activity benefits from the skill of dynamical models to predict ENSO events [van Oldenborgh *et al.*, 2005] and Atlantic SSTs in particular. Results suggest that this method performs better than current publicly available statistical methods. In particular, EUROSIP predicted in June a 2006 season that would not be much different from climatology, unlike the statistical methods, which predicted an active season, causing some insurance companies to lose billions of dollars. The dynamical forecasts discussed in this paper represent a viable alternative to the statistical methods with tropical storm forecasts available over all ocean basins, since the dynamical models are global. Indeed, tropical storm forecasts are just one of the many outputs of dynamical seasonal forecasting systems.

[15] Another important conclusion of the present paper is that the dynamical multi-model ensemble technique produ-

ces probabilistic forecasts of the number of Atlantic tropical storms which have some reliability. Since we use the same class of dynamical models as in the Intergovernmental Panel on Climate Change (IPCC), the investigation of the reliability of seasonal forecasts of hurricane frequency provides an opportunity to assess confidence in predictions of changing frequency of Atlantic tropical storms with climate change.

[16] Dynamical climate forecasting of tropical storms is primarily serendipitous and an outcome of ENSO-focused research that has led to dynamical coupled ocean-atmosphere forecast model development. As such, future improvements in skill cannot be guaranteed until more targeted research is undertaken to understand the full-range of processes within the dynamical models.

Appendix A: Model Calibration

[17] Dynamical models tend to drift towards a climate that is somewhat different from the observed climate. The effect of the drift on the model calculations is estimated from previous integrations of the model in previous years (the re-forecast). The drift is then removed from the model solution a posteriori (the calibration). For most model variables, including SSTs the drift is treated as a bias and removed additively. We calibrate the number of tropical

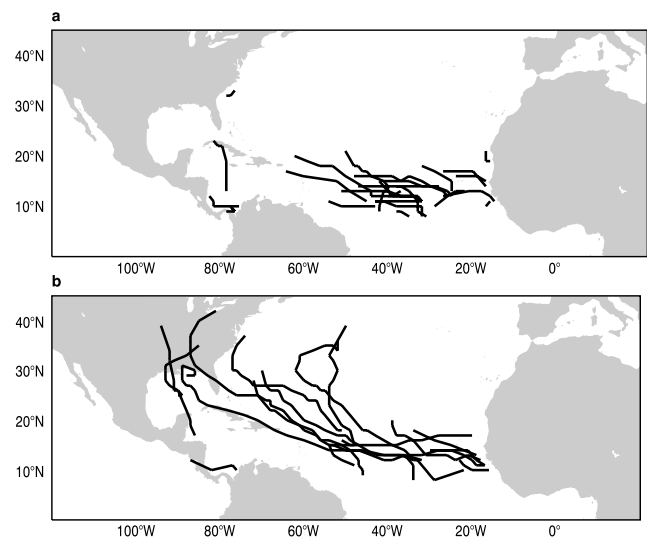


Figure 4. An example of an ensemble of trajectories (a) in the current seasonal forecasting system and (b) in a prototype version of the next ECMWF seasonal forecasting system. The tracks are produced by forecasts starting on 1st June 2000. For reason of clarity just 3 ensemble members are shown.

storms in a given year by multiplying the number of model storms by a factor such that the median of the model climate equals the median of the observed climate. The calibration of the re-forecast is performed using cross-validation and independently for each model.

[18] The real-time forecasts of 2005 and 2006 were produced using the mean. The median is however a more stable measure of the centre of the distribution, since the model distributions of tropical storm frequency tend to be strongly skewed towards very large values. The observed distribution of tropical storms is also far from being a normal distribution and the 2005 extremely active tropical storm season poses a problem of stability when performing cross-validation with the ensemble mean. Most results are calculated using the median, although we also give the actual forecasts issued in real-time (Table 3). We plan to issue future real-time forecasts using the median instead of the mean.

[19] **Acknowledgments.** The authors wish to thank the two anonymous reviewers whose comments proved invaluable in improving the presentation of the material.

References

- Camargo, S. J., A. G. Barnston, and S. E. Zebiak (2005), A statistical assessment of tropical cyclone activity in atmospheric general circulation models, *Tellus, Ser. A*, *57*, 589–604.
- Collins, M. (2007), Ensembles and probabilities: A new era in the prediction of climate change, *Philos. Trans. R. Soc.*, in press.
- Goldenberg, S. G., and L. J. Shapiro (1997), Physical mechanism for the association of El-Niño and West African rainfall with Atlantic major hurricane activity, *J. Clim.*, *9*, 1169–1187.
- Gray, W. M. (1984), Atlantic seasonal hurricane frequency. part I: El-Niño and 30 mb quasi-biennial oscillation influences, *Mon. Weather Rev.*, *112*, 1649–1668.
- Hsu, W.-R., and A. H. Murphy (1986), The attributes diagram a geometrical framework for assessing the quality of probability forecasts, *Int. J. Forecasting*, *2*, 185–193.
- Landsea, C. W., R. A. Pielke Jr., A. M. Mestas-Nunez, and J. A. Knaff (1999), Atlantic basin hurricanes: Indices of climate changes, *Clim. Change*, *42*, 89–129.
- Lea, A. S., and M. A. Saunders (2006), How well forecast were the 2004 and 2005 Atlantic and US hurricane seasons?, *Weather*, *61*(9), 245–249.
- Mason, S. J. (2004), On using “climatology” as a reference strategy in the Brier and ranked probability skill scores, *Mon. Weather Rev.*, *132*, 1891–1895.
- Palmer, T. N., et al. (2004), Development of a European multimodel ensemble system for seasonal-to-inter-annual prediction (DEMETER), *Bull. Am. Meteorol. Soc.*, *85*, 853–872.
- Palmer, T. N., and R. Hagedorn (Eds.) (2006), *Predictability of Weather and Climate*, 702 pp., Cambridge Univ. Press, New York.
- Palmer, T. N., G. J. Shutts, R. Hagedorn, F. J. Doblas-Reyes, T. Jung, and M. Leutbecher (2005), Representing uncertainty in weather and climate prediction, *Annu. Rev. Earth Planet. Sci.*, *33*, 163–195.
- Johnston, H. (2006), Faulty hurricane forecasts prove costly, *Physics World*, *19*, 12.
- Reynolds, R. W., N. A. Rayner, T. M. Smith, M. Thomas, D. C. Stokes, and W. Wang (2002), An improved in-situ satellite SST analysis for climate, *J. Clim.*, *15*, 73–87.
- Saunders, M. A. (2006), Winds of change, *Post Mag. Risk Rep.*, 28–29, 9 Nov.
- Saunders, M. A., and A. R. Harris (1997), Sea warming as a dominant factor behind near-record number of Atlantic hurricanes, *Geophys. Res. Lett.*, *24*, 1255–2614.
- Shapiro, L. J. (1987), Month-to-month variability of the Atlantic tropical circulation and its relationship to tropical storm formation, *Mon. Weather Rev.*, *115*, 2598–2614.
- Stevenson, T. (2006), The hurricanes that didn’t blow, *Daily Telegr.*, 7 Oct. (Available at <http://www.telegraph.co.uk/money/main.jhtml?xml=/money/2006/10/07/cnwind07.xml>.)
- Van Oldenborgh, G. J., M. A. Balmaseda, L. Ferranti, T. N. Stockdale, and D. L. T. Anderson (2005), Did the ECMWF seasonal forecast model outperform statistical ENSO forecast models over the last 15 years?, *J. Clim.*, *18*, 3240–3249.
- Vitart, F. (2006), Seasonal forecasting of tropical storm frequency using a multi-model ensemble, *Q. J. R. Meteorol. Soc.*, *132*, 647–666.
- Vitart, F., and T. N. Stockdale (2001), Seasonal forecasting of tropical storms using coupled GCM integrations, *Mon. Weather Rev.*, *129*, 2521–2537.
- Vitart, F., J. L. Anderson, and W. F. Stern (1997), Simulation of the inter-annual variability of tropical storm frequency in an ensemble of GCM integrations, *J. Clim.*, *10*, 745–760.
- Wilks, D. S. (2005), *Statistical Methods in the Atmospheric Sciences*, 2nd ed., 611 pp., Academic, San Diego, Calif.

M. K. Davey, M. R. Huddleston, S. Ineson, and D. Peake, Met Office, Exeter EX1 3PB, UK.

M. Déqué, Météo-France, F-31057 Toulouse Cedex, France.

T. N. Palmer, T. N. Stockdale, F. Vitart, and A. Weisheimer, European Centre for Medium-Range Weather Forecasts, Reading RG2 9AX, UK. (nec@ecmwf.int)