Project no. GOCE-CT-2003-505539

Project acronym: ENSEMBLES

Project title: ENSEMBLE-based Predictions of Climate Changes and their Impacts

Instrument: Integrated Project

Thematic Priority: Global Change and Ecosystems

D5.3: Scientific article/report and R software on optimal statistical methods for combining multi-model forecasts to make probabilistic forecasts of rare extreme event

Due date of deliverable: 28 February 2006

Actual submission date: 17 March 2006

Start date of project: 1 September 2004 Duration: 60 Months

Organisation name of lead contractor for this deliverable: ECMWF

Revision [draft, 1, 2, ..]

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	x
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the Consortium (including the Commission Services)	

ENSEMBLES D5.3

“Scientific article/report and R software on optimal statistical methods for combining multi-model forecasts to make probabilistic forecasts of rare extreme events”

F. J. Doblas-Reyes

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

C. Primo, D. B. Stephenson

Department of Meteorology, University of Reading, Reading, UK

Abstract

This report attempts to address the problem of formulating probabilistic predictions of extreme events at seasonal and interannual time scales using ensembles of dynamical simulations from multiple ocean-atmosphere coupled models. Although the prediction of extreme events in seasonal and interannual time scales can benefit from the large body of research carried out for the short-range prediction and climate change problems, its own unique requirements and difficulties have to be thoroughly addressed. In particular, the need of clear definitions of relevant extreme events for those time scales and the high importance of appropriate calibration and verification methods are discussed.

The report is intended as a discussion paper rather than a list of results and recommendations and so comments would be most appreciated.

1. Introduction

Recent natural disasters have highlighted the fact that human populations around the world remain at risk of certain weather and climatic events. Such events happen infrequently and tend to occur with little or no warning, sometimes with devastating impact. Despite modern technology, populations even in the most developed countries may be quite vulnerable to them, as hurricane Katrina in 2005 and the heat wave of Western Europe in July-August 2003 have shown. Such vulnerability should not be surprising given the evident capacity of extreme events to overwhelm normal protection mechanisms. Vulnerability is a very complex concept that takes into account many aspects of human and environmental systems, but it depends somehow on the risk of a certain event, in particular extreme events, to happen.

There are several scientific aspects of extreme events that have attracted attention, among which there is the detection, the analysis of the mechanisms involved, the understanding of the factors that modify the impact on human populations, and the prediction of the risk associated. These aspects have been addressed differently depending on the time scale involved. Weather and climate extremes can involve various time scales, from tornadoes and hail storms lasting only minutes or hours to droughts of years’ duration. Concerning prediction, an important effort has been carried out to predict extremes with a lead time of a few hours or days (Lalaurette, 2002) or to determine the risk of change in extreme event occurrence under different climate change scenarios (Palmer and Räisanen, 2002; Barnett et al., 2006; Benestad, 2006). However, not much effort has been carried out to predict extremes in intermediate time scales such as seasonal and interannual, even if such information would be of high value for the preparedness and planning of early warning systems, and the understanding of adaptation and resilience factors and system responses to reduce negative impacts. To bridge this gap, this document aims to discuss issues related to the formulation of forecasts of extreme events on the seasonal-to-decadal (s2d) time scale.

The note is organized as follows. Section 2 describes the type of events that can be considered extreme events in s2d time scales. Section 3 summarizes the characteristics of the dynamical forecast systems used in s2d forecasting. Section 4 addresses practical problems in the formulation of probabilistic forecasts, while Section 5 introduces the requirement of an adequate forecast quality assessment process. Section 6 summarizes the main issues and introduces the future challenges of the problem.

2. Extremes in seasonal-to-decadal time scales

There is no clear definition of what can be considered an extreme event in s2d time scales. For instance, while a large seasonally-averaged anomaly can be considered as a climatic extreme (with respect to a specific percentile of the climatological distribution) event (e.g. Weisheimer and Palmer, 2005), a succession of rare weather events, such as a long hot spell within a season, is also of relevance even if the seasonally-averaged temperature anomaly is not particularly large. Both examples refer to different characteristics of an extreme event. While the first example considers the interannual variability of the seasonal average, the other one looks at the extreme properties of the intraseasonal variability in a certain period. In addition, as the final value of s2d predictions is estimated by the users of the climate information (Thomson et al., 2006), extremes defined by users are of great importance. Users may need to define certain events that, even if they cannot be considered as extreme from the meteorological or climatological point of view, have anomalously high impact in their systems. For instance, short episodes of high temperature (~30°C) around the flowering time are expected to strongly affect seasonal yields of wheat or groundnut in India (Challinor et al., 2005).

Extremes can be classified in *simple* events, i.e. events based on a single variable, or *complex* events, which involve a critical combination of variables associated with a particular event (McGregor et al., 2005). Values of temperature or precipitation above a certain threshold are common examples of simple extremes, while a specific storm with moderately strong winds, associated with low temperatures and some form of strong precipitation could be considered as a complex extreme. End-user extremes can also be classified as simple, if only one specific end-user variable is used, or complex, if an end-user extreme event is analyzed in terms of the combination of the different meteorological or climatological variables that increase their vulnerability and exposure. Given that the analysis of extreme events in s2d time scales has been relatively scarce up to now, to easy the task we will concentrate on simple events in the rest of this paper.

The consideration of certain s2d situations as extreme events ought to take into account the specific time scale. For instance, while a temperature anomaly of 5 K in an extra-tropical region can be extreme for a seasonal average, it might not be so for a specific week or month because the frequency distribution of the same variable using weekly, monthly or seasonal averages is different. As another example, a continuous succession of five years of low precipitation could be considered an extreme event, at least from the socioeconomic point of view, even if the average precipitation over the period is similar to a sample with three very dry and two normal years. In a more general way, heat and cold waves, using either monthly or seasonal averages (Beniston, 2004; Schär and Jendritzky, 2004; Schär et al., 2004), clusters of tropical or extra-tropical storms (Kelman, 2001; Vitart and Stockdale, 2001), anomalous persistence of some very stable situations such as blocking highs (Cassou et al., 2005; Scherrer et al., 2005) or long-lasting (several years) droughts measured as a strong rainfall deficit (Woodhouse et al., 2002) can all be considered as extreme events at the s2d time scales. To clarify our definition of extreme events, we consider the simulation and prediction of two types of variables:

- Averages of a meteorological variable over a certain period, e.g., the mean seasonal temperature in a region.

- Clusters of anomalous, in the sense of either exceeding a certain pre-defined threshold or as meteorological/climatological entities, events. Examples of this type of variable are the exceedances of a 30 mm/day threshold over a calendar year or the number of tropical cyclones within a season.

Once the s2d variable is defined, extreme events are those for which it has a value either above or below a given percentile chosen preferably near the tail of its distribution. The distinction between extreme events in variables based on time averages and those based on exceedances over a certain threshold is due to the fact that most of the work in s2d forecasting has been carried out using seasonal or annual means of meteorological variables (Massacand, 2003; Cassou et al., 2005), while useful information about processes might be obtained by defining seasonal or annual variables defined as the exceedances of a certain threshold or the number of meteorological entities (e.g. cyclones). In fact, both types of variables are somehow related because large mean anomalies can also be interpreted as the temporal integration of changes in the occurrence of entities of a shorter time scale.

As stated above, many of the problems of how to formulate forecasts of extreme, high impact climatic events in s2d time scales and whether there is any skill in their prediction remain unanswered. Using the methodology adopted in other disciplines and time scales to detect extreme events, the assessment of their predictability requires a clear definition of the event, the characterization of the main features in homogeneous datasets and a detailed analysis of the physical mechanisms involved. These steps should be undertaken not only in the observational datasets, but also in the simulations made with the dynamical models used to formulate the forecasts. The ability of these models to adequately simulate if not the magnitude of the extreme events, at least the large-scale conditions associated to them, needs to be thoroughly assessed. Therefore, before discussing the predictability of extreme events, a brief description of the tools used to issue dynamical predictions of climate in s2d time scales is offered next.

3. Seasonal, interannual and decadal forecasts

Seasonal time scale dynamical climate predictions are now made routinely at a number of operational meteorological centres around the world, using comprehensive coupled models of the atmosphere, ocean, and land surface (e.g. Stockdale et al., 1998; Mason et al., 1999; Kanamitsu et al., 2002). Coupling the different components allows for complex feedbacks, a feature expected to be relevant in the simulation of most extreme climatic events. In contrast to seasonal forecasting, interannual and decadal forecasting are at their earliest stages (Boer, 2000). Preliminary assessments show that there are signs of ensemble-mean skill in near-surface temperature in multi-annual time scales (Smith et al., 2006), which is partly due to the impact of the increase of greenhouse gases in the atmosphere. Recent results point out that the effects of anthropogenic climate forcing need to be included in all these forecast systems (Doblas-Reyes et al., 2006).

In spite of the fact that predictable signals can arise from atmosphere-land-ocean interaction, the overlying atmosphere is intrinsically chaotic. This implies that predicted day-to-day evolution of weather is necessarily sensitive to initial conditions (Palmer, 1993). In practice, the impact of such sensitivity can be determined by integrating forward in time ensembles of forecasts of a model, the individual members of the ensemble differing by small perturbations to the starting conditions. However, if uncertainties in initial conditions are the only perturbations represented in a climate forecast ensemble, the resulting measures of predictability will not be reliable because the model equations are also uncertain. At present, there is no underlying theoretical formalism from which model uncertainty can be estimated (Palmer et al., 2005) and more pragmatic approaches must be sought. One such approach relies on the fact that global climate models have been developed somewhat independently at different climate institutes, using different numerical approaches to represent the climate dynamics and applying different parameterizations of physical processes. An ensemble comprising such quasi-independent models, referred to as a multi-model ensemble (Palmer et al., 2004), can be used to sample model uncertainty. Other ways to represent model uncertainty are based on the stochastic physics (Palmer, 2001) or the perturbed-parameter (Murphy et al., 2004) approaches. All these methods are being investigated in ENSEMBLES by performing co-ordinated experiments for which a large set of ensemble forecasts are carried out for a long period in the past. In all the forecast systems mentioned above, a forecast ensemble valid for a specific time is computed, while only one observation is available at a given time. This means that simulated samples are somewhat larger than observed samples, although the number of independent events is the same.

Dynamical predictions do not automatically provide a simulation of an observable due to model systematic error (Stockdale, 1997; Marzban et al., 2006). In particular, an ensemble may reflect errors both in statistical location and dispersion (Wilks, 2006). Biases in simulated variables (understood as the difference in statistical properties between model simulations and observations) are a consequence of model systematic error, which is due to the model phase space being different to the actual climate system phase space. The presence of systematic error also affects the tails of the distributions. In a multi-model context, the systematic error is, in addition, significantly different between the single models. It is therefore necessary 1) to estimate these errors in the simulations before formulating any forecasts to understand the limitations of the dynamical models and 2) to design a probabilistic model to convert the model output into probabilistic predictions of observables in a process known as calibration.

The calibration (Stephenson et al., 2005) and the assessment of the forecast quality (Hagedorn et al., 2005) of s2d forecast systems require a comprehensive set of forecasts over a substantial period of the past. This means that the dynamical models commonly used in climate forecasting have a fairly low horizontal resolution (~250 km), which might compromise the characteristics of the extreme events simulated by the model when compared to those observed in the actual climate system. This is more so because some s2d extreme events such as an abnormal number of convective events in a region might be of very small spatial scale. Besides, low resolution is also an undesirable feature because most of the end-user applications require high-resolution data to force their models. As a consequence, calibration methods to formulate probabilistic forecasts of extreme events should also deal with an increase in spatial resolution in a process known as downscaling (Coelho et al., 2006).

4. Prediction of extremes in seasonal, interannual and decadal time scales

While both deterministic and probabilistic predictions of extreme events can be issued using ensemble simulations with dynamical models, this note will focus only on the second type. The reason is that a probabilistic approach can take full advantage of the information contained in the ensemble, especially for events that happen rarely.

Predictability and forecast quality assessment of extreme events for s2d time scales is rarely found in the literature or in operational forecasting centres. Extreme events such as heat waves, understood as a number of consecutive days during which threshold temperatures (e.g. 32°C for London) are recorded, or high precipitation, defined as a series of days with precipitation above a certain threshold (e.g. 30 mm/day in the Netherlands), are expected to have a certain degree of predictability, although no forecasts are available yet. Some of the scarce operational examples of extreme event forecasts available consider situations that would not be regarded as extreme in other time scales. For instance, the Met Office issues monthly probability forecasts for the upper and lower quintile categories:

http://www.metoffice.com/monoutlook/index.html

while ECMWF issues seasonal predictions for the upper and lower 15th percentile category:

http://www.ecmwf.int/products/forecasts/d/charts/seasonal/forecast

One difficulty encountered when studying extremes is that the statistical analysis is very limited by a low number of events for finite and relatively short datasets. Although skilful predictions of a record event (i.e., one that happens only once in the available sample) would be very useful, it would be impossible to estimate its forecast quality. Every prediction requires an estimate of its quality to be integrated in an end-user system. As with any statistical analysis, a robust forecast quality assessment of a forecast system requires large enough samples. Therefore, the choice of the extreme event categories mentioned in the operational examples above is mainly determined by the need of a large enough number of events (25% of the sample size in the case of quartile categories). This sort of not-so-extreme categories has also been employed by end users. For instance, Thomson et al. (2006) show that seasonal predictions of quartile categories of both seasonal precipitation and annual malaria incidence based on multi-model dynamical forecasts are useful and more skilful than categories closer to the median.

As stated in previous sections, a previous analysis of the ability of the dynamical models to simulate extreme events in a similar way as in the real climate system is recommended. This includes not only a robust estimate of the changes in probability distributions of the meteorological variables with methods similar to those suggested in Ferro et al. (2005), but also an analysis of the physical causes of the extreme events, either in terms of the large-scale (Cassou et al., 2005) or surface

(Ferranti and Viterbo, 2006) conditions forcing the event. Once again, these two tasks have been seldom undertaken using s2d simulations. There is a pressing need to provide users of s2d forecasts with this sort of information.

In the examples of extreme event forecasts mentioned, forecast probabilities are computed as the proportion of ensemble members that are beyond the percentile that defines the category. In spite of being widely used, this method does not take into account important aspects of the model systematic error discussed in the previous section. The problem of creating sound probability forecasts is not simple. For instance, the size of the seasonal forecast samples available to date is typically 30 years, which proves to be too small to fit complex models for calibration and combination. In spite of that, some methodologies have been devised and successfully tested to formulate probability forecasts from dynamical models. The forecast assimilation method (Stephenson et al., 2005) has proved that calibrated multi-model seasonal forecasts are more reliable and skilful than probability forecasts from a simple multi-model. This method has been coded using the R language and made available from:

http://www.met.reading.ac.uk/cag/rclim/

to test its ability to formulate probabilistic seasonal forecasts of extreme events.

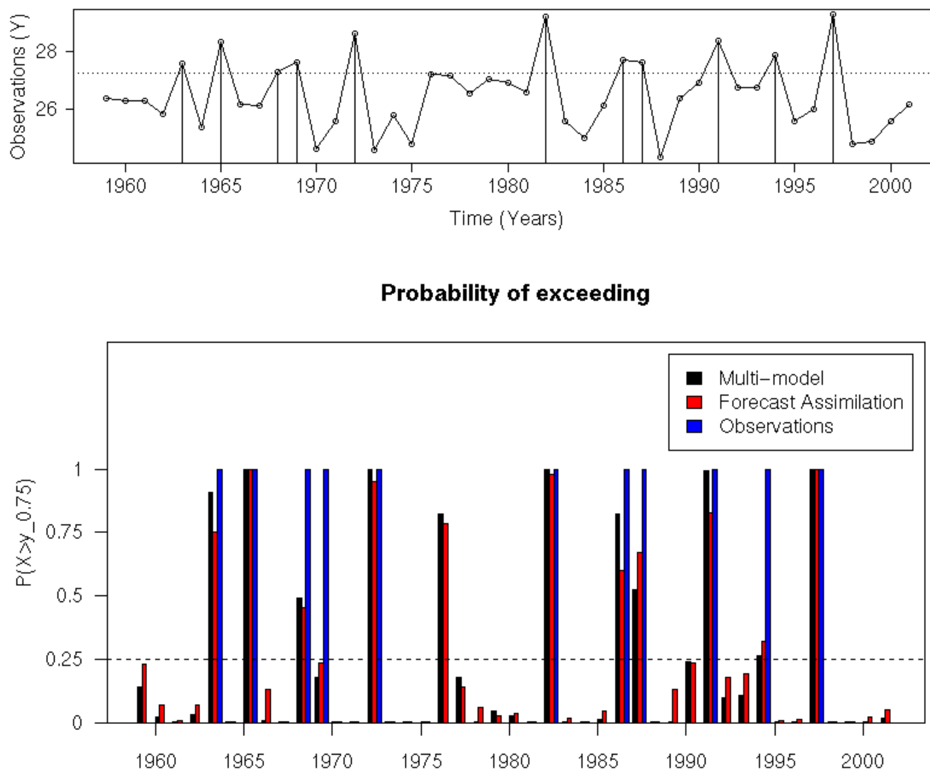

Figure 1: (Top panel) December sea surface temperature (SST) over Niño3.4 (5° N-5° S

and 170° W-120° W) for the period 1959-2001. The vertical lines correspond to the eleven dates when the SST was found to be above the upper quartile, which is represented with a horizontal dotted line. (Bottom panel) Observed (blue bars, taken the value of one when the event verifies and zero otherwise) and forecast probability of the SST exceeding the upper quartile. The forecasts have been formulated using a simple multi-model (black bars) and forecast assimilation (red bars) methods. DEMETER seasonal hindcasts with three coupled models and nine ensemble members initialized on the 1st of August of each year have been used. The predictions have a lead time of five months. The climatological 0.25 value is shown with a horizontal dotted line.

Figure 1 shows an example of probabilistic predictions of a variable exceeding the upper quartile. The variable is monthly mean December sea surface temperature (SST) averaged over the region in the equatorial Pacific known as Niño3.4 (5° N-5° S and 170° W-120° W). The predictions have been formulated using hindcasts from three models (ECMWF, Met Office and Météo-France) of the DEMETER experiment (Palmer et al., 2004). The hindcasts have been initialized on the 1st of August of each year, so that the predictions have a lead time of five months. The top panel displays the observed time series, with the vertical lines corresponding to the eleven dates when the SST was found to be above the upper quartile, represented by the horizontal dotted line. The lower panel shows the observed (one when the event verifies and zero otherwise) and forecast probability of the SST exceeding the upper quartile. The forecasts have been formulated using a simple multi-model (Hagedorn et al., 2005) and forecast assimilation methods. The simple multi-model predictions have been constructed by previously correcting the mean and variance biases of each individual model. Forecasts for which the forecast probability is above (below) the climatological 0.25 value (horizontal dotted line) when the event did (not) verify can be considered as correct. This happens in almost all cases with both sets of forecasts agreeing quite well with the observed probability, the number of correct rejections being outstanding. However, both formulations have a false alarm in 1976 and a miss in 1969. The reader should be reminded that a deterministic prediction would only provide yes/no forecasts, missing completely the richer information a probabilistic prediction provides.

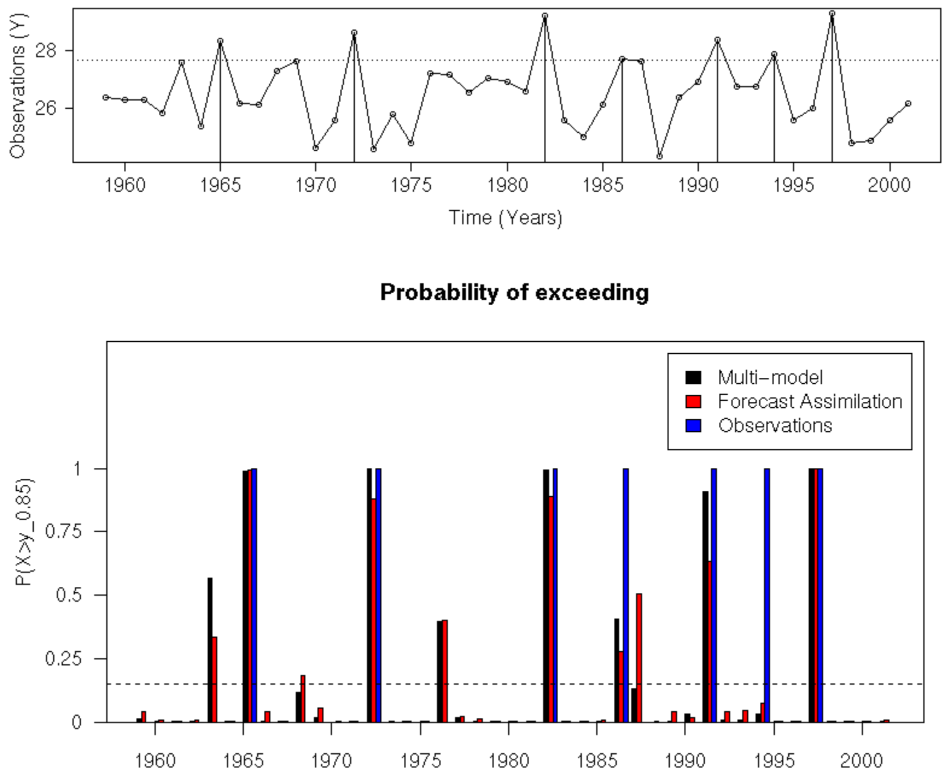

Figure 2: As Figure 1, but for the 85th percentile.

Figure 2 illustrates an example for the SSTs exceeding the 85th percentile. The event is observed in seven occasions (top panel), which clearly shows the reduction in sample size when the percentile defining the category is closer to the tail of the distribution. While the forecast probabilities still agree with the observed probability in most cases (lower panel), the number of false alarms has increased to two and four for the simple multi-model and the forecast assimilation, respectively. As a consequence of the relatively high number of false alarms, the user of the forecasts may feel inclined to hedge for more extreme events, which is a clearly undesirable option from a forecaster's perspective (Thornes and Stephenson, 2001).

Forecast assimilation has been considered here as a first step in the comparison, sample size permitting, with more adequate methods. A proper formulation of probability forecasts of extreme events requires a characterization of the tail of the joint distribution of the observations and forecasts (based e.g. on extreme value theory).

To conclude this section, the role of the ensemble size is discussed. As extreme events have low-probability of occurrence by definition, most simulations might fail to reproduce them even given the right large-scale forcing. In this context, ensemble size plays an important role in the detection of these events because a large ensemble size may allow for a better representation of the tails of the forecast probability distribution function. Optimal calibration and combination methods still need to be developed to assess the effect of ensemble size in the prediction of extreme events and to be able to extract the maximum information from the current small ensemble sizes.

5. Forecast quality assessment of extreme event predictions

Given the low-frequency of occurrence of extreme events, the assessment of the forecast quality requires both adequate methods that take into account their rarity and an analysis of the physical processes involved. As in the case of the formulation of the probability forecasts of extreme events, the forecast quality assessment process has to deal with small sample sizes. This means that, for some purposes, forecast quality assessment may be more meaningful for relatively-frequent, high-amplitude severe events (for instance events defined with the 75^{th} percentile). Alternatively, under certain assumptions, it is expected that moderate extremes, for which there are larger samples, may be used to provide information about rarer extremes (C. Ferro, personal communication).

Although there is a plethora of skill scores available, deliverable D5.4 promotes the extreme dependency score and the odds ratio for the verification of extreme events and mentions some of their desirable characteristics, such as their consistency for categorical forecasts based on extreme percentiles and their lack of sensitivity to hedging. Given that this is a subject discussed in detail in D5.4, the reader is referred to that text for more information on this issue.

6. Summary and challenges

This document discusses a series of issues that need to be addressed for the formulation of probabilistic predictions of extreme events in s2d time scales. Specifically, we raise the need of further work in the following aspects:

- Identification of relevant events, either from the climatological or the end-user perspective, and the use of this knowledge to make a classification of extreme events for these time scales.

- Analysis of the suitability of s2d forecast systems to simulate extreme events and assessment of the systematic error by comparing with the characteristics of the extreme events and their associated large-scale conditions in observational datasets.
- Development of sound probabilistic models for the calibration and combination of dynamical forecasts to formulate predictions of extreme events. These models should help to determine minimum ensemble sizes for a skilful detection of the extreme event signal.
- Development of forecast quality assessment methods specifically adapted to the rarity of the extreme events and to the small sample sizes available to determine whether there is any useful skill to predict these events and whether the skill is superior to that found for more frequent events. These tools should be used in the forecast quality assessment of extreme events with the stream 2 ENSEMBLES experiments.

Although the document deals with predictions of climate, a close link to the developments carried out in the weather prediction community, where samples are sensibly larger than at the s2d time scale should be established. In this respect, the largely artificial distinction between climate and weather prediction is intended to disappear in the near future in a unified weather-climate forecast approach, leading to a so-called *seamless suite* of forecast products (Rodwell and Doblas-Reyes, 2006) applicable on all relevant decision making time and space scales. From the perspective of the users, the distinction between weather and climate forecasts is somewhat abstract and arbitrary. Users are concerned about certain weather or climate events. Hence, they require forecasts with particular lead-time ranges that usually do not correspond to either weather or climate scales. How these forecasts are produced is mostly out of their concern. If the users need forecasts with multiple lead-time ranges, for ease of use, these forecasts should preferably be framed in the same, probabilistic forecast format. A seamless suite should bridge the gap between weather and climate forecasting, leading to better understanding, improved forecast techniques, and more skilful and useful forecasts.

Similarly, estimates of long-term change in climate extremes are available (e.g. Weisheimer and Palmer, 2005) and users of climate information are trying to design adaptation strategies to cope with the risks. Given that long-term decisions made by most end users are made at the interannual time scale, adaptation to ongoing climate change can be achieved by training the end-user systems with climate forecast information from s2d predictions, which have the desirable property of being verifiable. Once again, the ENSEMBLES project offers a unique framework to investigate these possibilities.

Acknowledgments

This work would not have been possible without the discussions with and insightful input of many colleagues: Caio Coelho, Chris Ferro, Clare Goodess, Ian Jolliffe, Geert Jan van Oldenborgh, Tim Palmer and Antje Weisheimer.

References

Barnett, D. N., S. J. Brown, J. M. Murphy, D. M. H. Sexton and M. J. Webb, 2006. Quantifying uncertainty in changes in extreme event frequency in response to doubled CO_2 using a large ensemble of GCM simulations. *Climate Dyn.*, **25**, 489-511.

Benestad, R. E., 2006. Can we expect more extreme precipitation on the monthly time scale? *J. Climate*, **19**, 630-637.

Beniston, M., 2004: The 2003 heat wave in Europe. A shape of things to come? *Geophys. Res. Lett.*, **31**, 2022-2026.

Boer, G. J., 2000. A study of atmosphere-ocean predictability on long time scales. *Climate Dyn.*, **16**, 469-472.

Cassou, C., L. Terray and A. S. Phillips. 2005. Tropical Atlantic influence on European heat waves. *J. Climate*, **18**, 2805-2811.

Challinor, A. J., J. M. Slingo, T. R. Wheeler and F. J. Doblas-Reyes, 2005. Probabilistic simulations of crop yield over western India using the DEMETER seasonal hindcast ensembles. *Tellus A*, **57**, 498-512.

Coelho, C. A. S., D. B. Stephenson, F. J. Doblas-Reyes, M. Balmaseda, A. Guetter and G. J. van Oldenborgh, 2006. A Bayesian approach for multi-model downscaling: Seasonal forecasting of regional rainfall and river flows in South America. *Meteorol. Appl.*, **13**, 73-82.

Doblas-Reyes, F. J., R. Hagedorn, T. N. Palmer and J.-J. Morcrette, 2006. Impact of increasing greenhouse gas concentrations in seasonal ensemble forecasts. *Geophys. Res. Let.*, in press.

Ferranti, L. and P. A. Viterbo, 2006. The European summer of 2003: sensitivity to soil water initial conditions. *J. Climate*, in press.

Ferro, C. A. T., A. Hannachi and D. B. Stephenson, 2005. Simple nonparametric techniques for exploring changing probability distributions of weather. J. Climate, 18, 4344-4354.

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

Kanamitsu, M., A. Kumar, H.-M Juang, J.-K. Schemm, W. Wang, F. Yang, S.-Y. Hong, P. Peng, W. Chen, S. Moorthi and M. Ji, 2002. NCEP dynamical seasonal forecast system 2000. *Bull. Amer. Meteor. Soc.*, **83**, 1019-1037.

Kelman, I., 2001. The autumn 2000 floods in England and flood management. *Weather*, **56**, 353-360.

Lalaurette, F., 2002. Early detection of abnormal weather using a probabilistic Extreme Forecast Index. ECMWF Technical Memorandum, 373, 27 pp.

Marzban, C., S. Sandgathe and E. Kalnay, 2006. MOS, perfect prog and reanalysis. *Mon. Weather Rev.*, **134**, 657-.663

Mason, S. J., L. Goddard, N. E. Graham, E. Yulaeva, L. Sun and P. A. Arkin, 1999. The IRI seasonal climate prediction system and the 1997/98 El Niño event. *Bull. Amer. Meteor. Soc.*, **80**, 1853-1873.

Massacan, A., 2003. Forecasting of extreme seasonal precipitation: Insight into the ECMWF potential. ECMWF Technical Memorandum, 415, 16 pp.

McGregor, G. R., C. A. T. Ferro and D. B. Stephenson, 2005. Projected changes in extreme weather and climate events in Europe. In *Extreme Weather Events and Public Health Responses*, W. Kirch, B. Menne and R. Bertollini eds, Springer, Heidelberg, 13-23.

Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, D. A. Stainforth, 2004. Quantifying uncertainties in climate change from a large ensemble of general circulation model predictions. *Nature*, **430**, 768-772.

Palmer, T. N., 1993. Extended-range atmospheric prediction and the Lorenz model. *Bull. Amer. Meteor. Soc.*, **74**, 49-65.

Palmer, T. N., 2001. A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Quart. J. Roy. Meteor. Soc.*, **127**, 279-304.

Palmer, T. N. and J. Räisanen, 2002. Quantifying the risk of extreme seasonal precipitation events in a changing climate. *Nature*, **415**, 512-514.

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

Palmer, T. N., G. J. Shutts, R. Hagedorn, F. J. Doblas-Reyes, T. Jung and M. Leutbecher, 2005. Representing model uncertainty in weather and climate prediction. *Annual Review of Earth and Planetary Sciences*, **33**, 163-193.

Rodwell, M. and F. J. Doblas-Reyes, 2006. Predictability and prediction of European monthly to seasonal climate anomalies. *J. Climate*, in press.

Schär, C. and G. Jendritzky, 2004. Hot news from summer 2003. *Nature*, **432**, 559-560.

Schär, C., P. L. Vidale, D. Lüthi, C. Frei, C. Häberli, M. A. Liniger and C. Appenzeller, 2004. The role of increasing temperature variability in European summer heatwaves. *Nature*, **427**, 332-336.

Scherrer, S. C., M. Croci-Maspoli, C. B. Schwierz and C. Appenzeller, 2005. Two dimensional indices of atmospheric blocking and their statistical relationship with winter climate patterns on the Euro-Atlantic region. *Int. J. Climatol.*, **26**, 233-249.

Smith, D. M., A. W. Colman, S. Cusack, C. K. Folland, S. Ineson and J. M. Murphy, 2006. Predicting surface temperature for the coming decade using a global climate model. *Nature*, submitted.

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

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

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

Thomson, M. C., F. J. Doblas-Reyes, S. J. Mason, R. Hagedorn, S. J. Connor, T. Phindela, A. P. Morse and T. N. Palmer, 2006. Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. *Nature*,

439, 576-579.

Thornes, J. E. and D. B. Stephenson, 2001. How to judge the quality and value of weather forecast products. *Meteorol. Appl.*, **8**, 307-314.

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

Weisheimer, A. and T. N. Palmer, 2005. Changing frequency of occurrence of extreme seasonal temperatures under global warming. *Geophys. Res. Lett.*, **32**, L20721, doi:10.1029/2005GL023365.

Wilks, D., 2006. Comparison of ensemble-MOS methods in the Lorenz '96 setting. *Meteorol. Appl.*, submitted.

Woodhouse, C. A., J. L. Lukas and P. M Brown, 2002. Drought in the Western Great Plains, 1845-56. Impacts and implications. *Bull. Amer. Meteorol. Soc.*, **83**, 1485-1493.