



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

D1.13 Scientific report/paper on documenting the seasonal hindcast skill of the most recent version of the stochastic physics scheme developed at ECMWF

Due date of deliverable: 29 February 2008

Actual submission date: 20 February 2008

Start date of project: 1 September 2004

Duration: 60 Months

Organisation name of lead contractor for this deliverable: ECMWF

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)	

**Impact of a Quasi-Stochastic Cellular Automaton Backscatter
Scheme on the Systematic Error and Seasonal Prediction Skill of a
Global Climate Model**

J. Berner, F. J. Doblas-Reyes*, T. N. Palmer, G. Shutts and A. Weisheimer

* Corresponding author: p. +44 118 9499655, f. +44 118 9869450, f.doblas-reyes@ecmwf.int

Short title for page headings: “Impact of CA-stochastic backscatter scheme on model error”

Abstract

The impact of a nonlinear-dynamic cellular automaton model, as a representation of the partially-stochastic aspects of unresolved scales in global climate models, is studied in the European Centre for Medium-Range Weather Forecasts coupled ocean-atmosphere model. Two separate aspects are discussed: impact on the systematic error of the model, and impact on the skill of seasonal forecasts. Significant reductions of systematic error are found both in the tropics and in the extratropics. Such reductions can be understood in terms of the inherently nonlinear nature of climate, in particular how energy injected by the cellular automaton at the near-grid scale can backscatter nonlinearly to larger scales. In addition, significant improvements in the probabilistic skill of seasonal forecasts are found in terms of a number of different variables such as temperature, precipitation and sea-level pressure. Such increases in skill can be understood both in terms of the reduction of systematic error as mentioned above, and in terms of the impact on ensemble spread of the cellular automaton's representation of inherent model uncertainty.

Key words (3-6): model uncertainty, backscatter, cellular automaton, seasonal forecasting

1. Introduction

Ever since their introduction, numerical climate models have been formulated according to a rather precise prescription: represent the equations of motion as accurately as possible by projection onto a Galerkin basis down to some truncation scale, and represent the effect of unresolved scales on the resolved-scale motions through deterministic bulk-formula parametrisations. The tendencies associated with such parametrisations are determined by, and hence slave to, the resolved-scale flow, typically at the truncation scale.

Such bulk-formula parametrisations are motivated by concepts in statistical mechanics. For example, the formulation of the effect of molecular viscosity is given by considering an ensemble of randomly-moving molecules whose mean-free path is small compared with some macroscale of interest. Similarly, climate-model parametrisations (e.g., of convection or gravity-wave drag) are formulated by supposing that there exist an ensemble of subgrid processes (e.g., convective plumes or gravity waves) in quasi-equilibrium with the resolved-scale flow.

By using a simplified Cloud Resolving Model, Shutts and Palmer (2007) studied this statistical-mechanical assumption quantitatively, in the case of convection in the tropics. In regions of strong convection it was found that the bulk-formula assumptions were in error by an order of magnitude; rather than determining the sub-grid tendency, the truncation scale motions provided only a partial constraint on some underlying probability distribution of sub-grid tendencies.

Consistent with these findings, Palmer (2001) has suggested that sub-grid motions should be represented by simplified nonlinear stochastic-dynamic models, as an alternative to the deterministic bulk-formula approach. One consequence of such an approach is that a stochastic-dynamic model can allow sub-grid-scale energy to be backscattered onto the resolved-scale grid.

Stochastic kinetic energy backscatter has its origins in early three-dimensional turbulence closures that recognized the deficiencies of the eddy viscosity concept (Kraichnan, 1976). Leith (1990) and Mason and Thomson (1992) demonstrated the benefits of using a stochastic backscatter term in large eddy simulations of turbulence and Frederiksen and Davies (1997) extended the technique to large-scale planetary motion. The idea is based on the notion that the turbulent dissipation rate is the difference between upscale and downscale spectral transfer and that the upscale component is available to the resolved flow as a kinetic energy source. In

a nonlinear system, such backscattered energy can in principle reduce the large-scale systematic error of the model. In addition, through multiple independent random samplings of the underlying stochastic processes, an ensemble of integrations is readily generated by such models, thus providing a representation of underlying model uncertainties.

One simple type of stochastic-dynamic model considered by Palmer (2001) is based on the cellular automaton. This approach has been further developed by Shutts (2005) and applied to the medium-range forecast problem. In this paper, the cellular automaton scheme of Shutts is applied to climate timescales. Specifically, the impact of the scheme, both on the systematic error of the ECMWF coupled ocean-atmosphere model, and on the seasonal forecast skill of the same model, is studied.

A brief summary of the cellular automaton scheme is described in Section 2. The experimental design is given in Section 3, and results in terms of impact on systematic error and on seasonal forecast skill are given in Section 4. Conclusions are summarised in Section 5.

2. The cellular-automaton backscatter scheme (CASBS)

In this section we give a brief summary of Shutts' kinetic-energy cellular-automaton backscatter scheme (CASBS) and point out some minor differences from the original scheme. An extension of this scheme and its impact on medium-range ensemble forecasting and flow-dependent predictability and error-growth is discussed in Berner et al. (2008).

Shutts (2005) argues that in numerical weather prediction (NWP) systematic kinetic energy is lost in both numerical integration schemes and parametrisations. For instance, errors in semi-Lagrangian departure-point interpolation cause a net energy sink and kinetic energy released in deep convection does not find its way sufficiently into balanced flows and gravity wave generation. Consequently, the backscatter scheme aims at representing upscale error-growth on synoptic and sub-synoptic scales from convection, gravity and mountain/wave drag and numerical dissipation.

The backscatter scheme generates a flow-dependent stochastic kinetic energy source by introducing a streamfunction forcing

$$\psi(\varphi, \lambda, z, t) = \sqrt{D(\varphi, \lambda, z, t)}\phi(\varphi, \lambda, t). \quad (1)$$

where $D(\varphi, \lambda, z, t)$ is the instantaneous dissipation rate and $\phi(\varphi, \lambda, t)$ a realization of the cellular automaton. The resulting streamfunction forcing is then transformed to spectral space,

converted into a vorticity increment and added to the dynamical equations. True to its nature as a kinetic-energy backscatter scheme, this will affect the kinetic energy spectrum of the model. In ensemble forecasts, CASBS is initialised with different choices of the cellular automaton pattern. Since these patterns can be chosen randomly, CASBS is effectively a stochastic sub-grid parametrisation.

In the following paragraphs we give some more details on the pattern generator and the computation of the total dissipation rate $D(\varphi, \lambda, z, t)$. The cellular automaton (CA) is a variant of the cellular automaton known as Conway's Game of Life (e.g., Gardner, 1983). The CA-domain covers the entire globe in a fine grid of cells that are regular in latitude/longitude coordinates, here 720 x 360 cells (Figure 1a). Each cell can be alive or dead and living cells have up to 32 lives. Dead cells come to life if their nearest neighbours satisfy certain 'birth' conditions (e.g. when the number of them with 32 lives equals either 3, 4 or 5). When a dead cell comes to life it is given 32 lives and these are progressively lost in later steps when certain survival rules fail to be met. Only cells with the maximum number of lives (32) participate in the rule counts. In order to present a sufficiently smooth pattern to the forecast model the cell values are averaged into 2-degree 'macro' cells comprising 4x4 CA cells and the subsequent array is smoothed again using passes of a 1-2-1 filter first in x (longitude) then in y (latitude). The smoothed values are then normalized so that the domain average is unity (Figure 1b). The resulting pattern looks very "organic" and consists of clusters of densely populated cells with distinct fronts, reminiscent of mesoscale structures in the atmosphere. Further detail on the CASBS algorithm can be found in Shutts (2005).

The spatial structure of the smoothed coarse-grained cellular automaton pattern is analysed by calculating the spatial correlation as function of the distance r between the macro cells, in units of the width of a single macro cell. The spatial correlation decreases with distance and oscillates around zero for macro cells of 7 or more units apart (Figure 1c). The e-folding distance, i.e. the decorrelation scale, equals 4 units. Since the macro cell represents an area of 2 x 2 degrees the decorrelation scale is effectively around 8 degrees or 800 kilometres.

The total instantaneous dissipation rate $D(\varphi, \lambda, z, t)$ contains contributions from deep convection, numerical dissipation and gravity/mountain wave drag. We refer to Shutts (2005) for details on the computation of the numerical dissipation and contribution from gravity/mountain wave drag. Changes in the NWP model made it necessary to no longer base the dissipation from deep convection on convective updraught speeds but instead a mass-flux formulation was used:

$$D_{conv}(\varphi, \lambda, z, t) = \frac{\delta(\varphi, \lambda, z, t)}{\rho(\varphi, \lambda, z, t)^3} \frac{M(\varphi, \lambda, z, t)^2}{\beta^2},$$

where $M(\varphi, \lambda, z, t)$ is the updraft convective mass flux rate in $\text{kg}/(\text{m}^3\text{s})$ from the convective parametrisations, $\delta(\varphi, \lambda, z, t)$ the updraft detrainment rate in $\text{kg}/(\text{m}^3 \text{s})$, $\rho(\varphi, \lambda, z, t)$ the density and β an assumed detrainment cloud fraction for deep convection of $\beta=2.6 \cdot 10^{-2}$. To focus on the large-scale structure, the total dissipation rate $D(\varphi, \lambda, z, t)$ is subsequently smoothed by applying a spectral filter that completely retains wavenumbers $n < 20$ and gradually reduces to zero for wavenumbers $20 < n < 30$.

The total dissipation rate per unit area typical of the ECMWF atmospheric model for the months December through February and its contributions from deep convection, numerical dissipation and gravity/mountain wave drag are shown in Figure 2. For illustration only we show density-weighted vertical averages of the 3D-dissipation fields. The dominating contributor to the total dissipation rate with a global mean of 0.99 Wm^{-2} is deep convection. Its maxima are in the deep convective regions of the Tropics, especially over Indonesia, but also downstream of the Andes. The most active regions are seen in the summer hemisphere, and for the period June through August there are larger contributions from the Northern hemisphere (not shown). With a global average of 0.56 Wm^{-2} the second largest contribution comes from the numerical dissipation. It is largest in the storm track regions and downstream of high orography like Greenland, the Himalayas and the Andes. The global-mean dissipation from gravity and mountain wave drag is much smaller and occurs mainly in the lower tropospheric levels over orography. However locally, over the major mountain ranges, the dissipation rates may be very large indeed (e.g. over 100 Wm^{-2}).

3. Experimental setup

The experiments discussed in this paper were performed using an ensemble of dynamical seasonal re-forecasts. Two sets of experiments with and without the stochastic backscatter scheme were performed. In the following, the control model simulations without stochastic backscatter will be denoted as CTRL, whereas the experiments using the new stochastic physics scheme will be referred to as CASBS.

The model used to produce these two sets of re-forecasts was integrated over 7 months initialised twice a year on May 1st and November 1st at 00 GMT over the 11-year period 1991-2001. Ensembles of 9 members generated by different initial conditions were used for

each start date, so that a total set of 22 nine-member seasonal re-forecasts was available to analyse the results for both CTRL and CASBS.

As a forecast model we used the ECMWF coupled atmosphere-ocean general circulation model (Anderson et al. 2007). The atmospheric component is the Integrated Forecasting System (IFS) of ECMWF as used in operational medium-range weather forecasting in its version CY29R2. The model was run with a horizontal truncation of T_L95 and 40 vertical levels. The HOPE (Hamburg Ocean Primitive Equation) ocean model has a horizontal resolution of 1° , with an equatorial refinement of 0.3° , and 29 levels in the vertical. The coupler OASIS (version 2) is used to interpolate once per day between the oceanic and atmospheric grids.

The atmospheric initial conditions, including land surface conditions, come from ERA-40 (Uppala et al. 2005); the ocean initial conditions come from an ensemble of ocean analyses (Balmaseda et al. 2007). The nine initial-condition ensemble members were generated by adding small perturbations to the atmospheric and oceanic fields. For the initialisation of the atmosphere, perturbations based on singular vectors were applied in a similar way as in the operational medium-range ensemble forecasts (Buizza & Palmer 1995; Rodwell & Doblas-Reyes 2006). For the initialisation of the ocean, three different ocean analyses were created by adding daily wind stress perturbations on the basis of the differences between two quasi-independent datasets. In addition, four sea-surface temperature (SST) perturbations, again based on differences between two SST datasets, were added to or subtracted from the initial fields.

4. Results

4.1 *Effect on kinetic energy spectrum*

Figure 3 shows the rotational and divergent part of the kinetic energy spectrum at 500 hPa for ERA40-data, a control integration without stochastic backscatter (CTRL) and an integration with CASBS. The kinetic energy in CTRL is less than in ERA40, suggesting that the model is under-active for all wavenumbers n , but especially so for $n > 10$. The stochastic backscatter scheme injects energy, so that the kinetic energy increases for all wavenumbers. This leads to an improvement in the rotational part of the kinetic energy spectrum in the synoptic band (wavenumbers 2-12) but to a slight overestimation for wavenumbers $n > 12$. By introducing a streamfunction forcing, only the rotational part of the spectrum is directly forced.

Nevertheless, a substantial impact is also seen in the divergent part of the kinetic energy

spectrum for wavenumbers larger than $n=5$. The divergent part of the spectrum is now much closer to that of ERA40 (Figure 3), although there is now slightly too much energy for wavenumbers between 20 and 40. From this we conclude that the activity in the model has been changed in such a way that the model with CASBS is better at producing and maintaining divergent modes, which is especially important for tropical variability.

4.2 Systematic model error

In this section the impact of the stochastic backscatter scheme on the systematic errors that develop during the course of the coupled seasonal integrations is described. We focus on the atmosphere and discuss results for the boreal winter (December-February, DJF) re-forecasts starting in November and for the summer (June-August, JJA) re-forecasts starting in May.

Figure 4 shows the systematic errors in simulating DJF mean total precipitation. The difference between CTRL and the GPCP verification data set (Adler et al. 2003) is displayed in the upper panel. The control model generates excessive precipitation over the tropics. The Inter-Tropical Convergence Zone (ITCZ) over the tropical Pacific is particularly wet. The control re-forecasts also generate too much precipitation in the Indonesian warm pool area. A dry bias can be seen over large parts of South America and Northern Australia. The errors in mid-latitudes are in general smaller than in the tropics.

The systematic error of the CASBS runs is shown in the middle panel of Figure 4, and the lower panel of that figure displays the difference CASBS minus CTRL. The impact of the stochastic backscatter scheme on precipitation errors can be summarized as reducing the wet bias over the tropical ITCZ, in particular over the Pacific north of the equator. This is due to both a southward shift of the convergence zone and a reduction of the overall rainfall amounts. A decrease of the excessive precipitation over the tropical Atlantic and the Indian Ocean can also be detected. The systematic errors in precipitation over land are mainly unaffected by the stochastic parametrisation scheme, although a positive impact can be found over Northern Australia.

Jung et al. (2005) showed that previous versions of the stochastic physics scheme induced significant changes in the extratropical tropospheric circulation in an uncoupled model. As a diagnostic of this impact, the Tibaldi and Molteni (1990) blocking detection index has been used to estimate the frequency of persistent blocking anticyclones in the simulations. The index computes the 500hPa geopotential height meridional-gradient for specific mid-latitude bands (of latitude between 40°N - 60°N and 60°N - 80°N) and classifies a longitude for a given time step as blocked, if in the northernmost latitude band a reverse of the westerly flow is

detected. An event is classified as blocked if at least 10 degrees of longitude are blocked over more than five consecutive days. Figure 5 shows the mean blocking frequency for DJF averaged over the years 1991-2001 and the nine ensemble members. 95% confidence intervals have been computed using a bootstrap method, where the original data were resampled with replacement 500 times. The grey dots (solid for CTRL and open for CASBS) in the top of the panel show the longitudes where the model climatology is not significantly different from the ERA40 results, using a two-sample test based on the bootstrap estimates (Nicholls 2001; Lanzante 2005). The black squares correspond to those longitudes where both experiments are not significantly different. Although both experiments significantly underestimate the frequency of blocking events, the CASBS simulation reduces this error over the North Pacific to the point that the simulations are significantly different. The improvement of the mean state over the North Pacific is a robust feature of CASBS in seasonal integrations (Jung et al. 2005). In the experiments presented here, CASBS has no significant impact on Atlantic blocking.

In Figure 6 we show the growth of systematic errors in sea surface temperature (SST) over the Niño3.4 region (5°S - 5°N , 170° - 120°W) in the eastern-to-central tropical Pacific, a key region for the development of the El Niño-Southern Oscillation phenomenon, over the 7-month lead time of the re-forecasts. The dashed curves show the SST bias for the CTRL simulation for the two start dates in May and November; the solid ones for the CASBS simulations. For both start dates and the two sets of experiments, the modelled SSTs are warmer than the climatological temperatures. However, for the May start dates, this warm bias is systematically reduced in the CASBS runs. For example, whereas in the CTRL experiment the SST errors after 3 and 7 months lead time come close to 1.5 K and 2.5 K respectively, they are reduced by nearly 30%, that is about 0.5 K after 3 months and almost 1 K after 7 months, in the CASBS experiments. For the November start dates no difference between the two were found.

4.3 Forecast quality assessment

Various measures of forecast quality have been used to assess the relative merit of CASBS when compared to CTRL. The scores include the anomaly correlation of the ensemble mean (ACC) and, for dichotomous probability forecasts, the Brier skill score (BSS) with respect to climatology and the ROC skill score (ROCSS) (Jolliffe and Stephenson 2003). The BSS has also been decomposed into the sum of two components (Murphy 1986): the reliability (RELSS) term that measures the relative bias of conditional means, and the resolution

(RESSS) term that measures the relative variance of the conditional means. The BSS decomposition used here includes two additional terms in the resolution component that account for the within-bin variance and covariance of the probability forecasts, as described in Stephenson et al. (2007). As the BSS depends strongly on the ensemble size for small ensembles, Ferro (2007) has developed an analytical expression to estimate the Brier score for different ensemble sizes using the Brier score value obtained from the sample. This expression, which depends on the variance (also known as sharpness) of the probability forecasts, allows the computation of a theoretical BSS for infinite ensemble sizes (BSSI henceforth).

The seasonal forecast system performance has been traditionally tested on tropical Pacific SSTs. This is because the main source of seasonal predictability is assumed to come from the interannual variability related to ENSO. Figure 7 shows the root mean square error (RMSE) and the anomaly correlation (ACC) of ensemble-mean surface temperature anomalies over the Niño3.4 region as a function of lead time, for the May and November start dates over the period 1991-2001. For comparison, the RMSE and ACC of a simple statistical model based on persisting the anomaly of the month previous to the start date are also shown. The accuracy of the re-forecasts generally decreases with lead time, although both experiments have higher skill than this simple persistence. The smaller RMSE and higher ACC for CASBS are evident after the third month of the forecast. This improvement of CASBS over CTRL is mainly associated with the forecasts started in May; for these start-dates CASBS also leads to a large reduction in the systematic error. Panel (a) also shows the spread of both experiments measured by the standard deviation of the ensemble members around the ensemble mean. For a reliable ensemble system, state-dependent uncertainty is expected to be modelled by the spread. In a well-calibrated system, spread and RMSE should have a similar magnitude, which is not the case for CTRL. However, both experiments underestimate the forecast uncertainty, though CASBS has an increased spread with respect to CTRL from the first month of the integration.

Figure 8 illustrates the different ensemble forecasts for each experiment in the Niño3.4 region. In general, CASBS has a larger spread than CTRL, and its ensemble members encompass the CTRL ensemble in most cases. Furthermore, there are cases where CASBS seems to yield a better forecast. For example, the onset of the extremely warm ENSO event in 1997/98 was better predicted in CASBS than in CTRL.

The full set of forecast quality measures has been computed for several variables (500hPa geopotential height, 850hPa temperature, precipitation, 2-metre temperature and mean sea

level pressure) over a large number of regions (Table 1) and for three different events in the case of the probability forecasts: anomalies above the upper tercile, above the median, and below the lower tercile. The terciles and the median have been computed separately for each experiment and for the reference dataset to take into account the different systematic errors in the simulated distributions. Figure 9 shows the scatter plots of BSSI and ROCSS of CTRL versus CASBS for both start dates, the first four regions in Table 1, the three events mentioned above and several forecast ranges (first month and forecast periods 1-3, 2-4, 3-5, and 4-6 months), making a total of 600 cases. For both skill scores there is a large range of values, from unskilful predictions (lower than zero) to values close to 0.5 and 1 for BSSI and ROCSS, respectively. Frequently CASBS performs better than CTRL, as the larger number of points above the diagonal suggests. The difference of the skill scores for a given variable, region, lead time, start date and event, between CASBS and CTRL, has been tested for significance using a two-sample test and a bootstrap method with 1,000 samples. While CTRL is significantly better (with 95% confidence) than CASBS in 3 cases for BSSI and in 6 cases for ROCSS, CASBS is significantly better than CTRL in 251 and 118 cases, respectively.

This overwhelming superiority of CASBS over CTRL in terms of forecast quality is also evidenced in Table 2, which summarizes the results for several skill scores and all the regions in Table 1. The large proportion of cases where RELSS is better for CASBS than for CTRL suggests that the increase in BSS and BSI can be largely attributed to an improvement in reliability. This might be linked to the alleviated underdispersion of the ensemble when using CASBS. A less intuitive result is the large number of cases where the ROCSS in CASBS is significantly higher than in CTRL. This improvement in terms of forecast resolution (of which ROCSS is a measure) is especially important because the resolution of a forecast system can only be enhanced using additional sources of forecast information, while the reliability can be improved a posteriori using climatological information. Improved resolution could arise because of the smaller systematic error in CASBS.

The improvements due to CASBS can be highly relevant in specific cases. Figure 9 shows that most of the cases with negative skill scores for the European region for CTRL have positive skill for CASBS, especially for the May start date (note that the negative skill cases over Europe are displayed with blue dots in Figure 9). This improvement can be appreciated in more detail in Figure 10, where the ROCSS for predictions of 500 hPa geopotential height and 850 hPa temperature are shown as a function of the start date and the forecast range. While the forecast quality decreases with lead time (from left to right, with the November

start date results starting after the first five sets of bars), all the sample values (i.e., those obtained with the re-forecasts without resampling, which are displayed with black dots) for CASBS are positive, which is not always the case for CTRL. Confidence intervals (95%) obtained with the bootstrap method described above are displayed around the skill score sample value. They suggest that it is more likely to obtain a ROCSS significantly different from zero (colour bar clear of the zero-skill horizontal line) with CASBS than with the control.

Attribute diagrams (Hsu and Murphy 1986) offer a comprehensive illustration of the benefits of CASBS in terms of forecast quality. These diagrams allow the visualization of the reliability, resolution and sharpness of a system for a specific event. Figure 11 shows the attribute diagrams for first-month re-forecasts of precipitation anomalies over the tropics started on the 1st of May. The diagrams illustrate the conditional relative frequency of occurrence of the events as a function of their forecast probability, based on a discrete binning of many forecast probabilities taken over a region. Each forecast probability bin in the diagrams is represented by a solid circle whose area is proportional to the bin sample size. The vertical line represents the average forecast probability, while the horizontal line is for the climatological frequency of the event. In the idealised case of infinite sample and ensemble sizes, the diagonal line represents perfect probabilistic reliability: if from a set of cases where an event is forecast with probability p , the event actually occurs on a fraction p of occasions. If, on the occasions where an ensemble forecast predicts some event with probability p , the event occurs in reality on a fraction q of times, then if p is sufficiently different from q , the ensemble forecast probabilities are not reliable. This case will appear in the diagram as a point away from the diagonal. If the corresponding curve is shallower than the diagonal, the forecast system is said to be over-confident, while if it is steeper the system will be under-confident. The sum of the horizontal squared distance of all the points to the diagonal (weighted by the sample size of each bin) is a measure of the lack of reliability of the system as measured by the Brier score. In the same way, the sum of the vertical distance of the points to the horizontal line corresponding to the climatological frequency of the event measures the forecast resolution, i.e., the ability of the system to produce reliable forecasts that differ from the naive probability. This means that if a reliability curve were to be horizontal, the frequency of occurrence would not depend on the forecast probabilities and the system would have zero resolution. The black dashed line in the diagram separates skilful from unskilful regions in the diagram: points with forecast probabilities smaller (larger) than the climatological frequency that fall below (above) this line, contribute to positive BSS;

otherwise they contribute negatively to the BSS. Finally, the sharpness is the variance of the forecast probabilities and is shown in the diagram by the proportion of probability forecasts away from the average probability.

The diagrams in Figure 11 show that the unreliable curves obtained for CTRL are much steeper for CASBS, bringing the curve into the skilful area determined by the black dashed line (note the significant improvement of the corresponding BSSI). This happens for both events considered. The increased steepness improves both the reliability (by decreasing the distance to the diagonal) and the resolution (by increasing the distance to the horizontal line corresponding to the climatological frequency). This occurs at the expense of a reduction in the variance of the forecast probability, or sharpness, illustrated by a reduction in the size of the dots for the extreme forecast probabilities in CASBS with respect to CTRL. Although a sharp set of probabilities, i.e. forecast probabilities with larger variance, is a desirable feature of a forecast system, large sharpness may be harmful in the case of overconfident reliability diagrams, such as the ones for CTRL.

5. Summary and conclusions

In this paper, we have studied the impact on the climate of the ECMWF coupled model of a cellular automaton scheme, as first proposed by Palmer (2001), with specific implementation by Shutts (2005), as a partial representation of stochastic sub-grid processes. Two aspects of this impact have been studied: on the systematic error of the model, and on seasonal forecast skill. For the former, some remarkable reductions in the error in tropical seasonal-mean rainfall, extratropical blocking frequency, and temperature drift in the tropical Pacific were found. The skill of the model, measured both in terms of ensemble-mean accuracy in predicting tropical Pacific SST variability, and in terms of probabilistic seasonal predictions of temperature, precipitation and mean sea level pressure, were notably improved by the addition of the cellular automaton scheme.

The reduction of systematic error can be understood by the fact that climate is a profoundly nonlinear dynamical system, and that the proposed cellular automaton acts as a source of sub-synoptic scale and mesoscale energy that can be backscattered to the planetary-scale components of the flow. A simple analogy is a double-well potential. By adding stochastic noise to the system, the sub-dominant well may be visited more frequently, thus altering the mean state of the system. The increase in forecast skill arises from a combination of two effects. The first is the reduction in systematic error as discussed above. The second arises from the fact that the model equations are a key source of forecast uncertainty for climate

prediction and, through different random initialisations, the cellular automaton provides a representation of model uncertainty in ensemble predictions. Thus the cellular automaton increases the reliability of probabilistic seasonal forecasts thereby impacting on standard skill scores.

In this paper, a relatively simple cellular automaton has been implemented in which the accurate representation of the dynamics of atmospheric sub-grid motions is somewhat minimal. As such, one line of development in the cellular automaton model is the inclusion of more explicit atmospheric sub-grid dynamics; for example, a “Mexican wave” representation of equatorially-trapped Kelvin wave dynamics can be coded very simply in a cellular automaton model. Hence, for example, one can envisage representing through a cellular automaton, the sort of convectively-driven wave modes in the tropical atmosphere which are not well resolved in a standard global climate model. This is work for the future.

This work was supported by the ENSEMBLES project (GOCE-CT-2003-505539). The authors would like to acknowledge Dr. Thomas Jung for his contribution to the analysis of the experiments.

References

- Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P. & Nelkin, E. 2003 The version 2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present). *J. Hydrometeor.* **4**, 1147-1167.
- Anderson, D. L. T., Stockdale, T., Balmaseda, M. A., Ferranti, L., Vitart, F., Molteni, F., Doblas-Reyes, F. J., Mogensen, K. & Vidard, A. 2007 *Development of the ECMWF seasonal forecast System 3*. ECMWF Technical Memorandum 503 [Available from <http://www.ecmwf.int/publications/library/do/references/show?id=87744>].
- Balmaseda, M., Vidard, A. & Anderson, D. 2007 *The ECMWF System 3 ocean analysis system*. ECMWF Technical Memorandum 508 [Available from <http://www.ecmwf.int/publications/library/do/references/show?id=87667>].
- Berner, J., Shutts, G. J., Leutbecher, M., & Palmer, T. N. 2008 A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *J. Atmos. Sci.*, submitted
- Buizza, R. & Palmer, T. N. 1995 The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.* **52**, 1434-1456.
- Ferro, C. A. T. 2007 Comparing probabilistic forecasting systems with the Brier score. *Wea. Forecasting* **22**, 1076-1088.
- Frederiksen, J. S. & Davies, A. G. 1997 Eddy viscosity and stochastic backscatter parametrizations on the sphere for atmospheric circulation models. *J. Atmos. Sci.* **54**, 2475-2492.
- Gardner, M. 1983 The Game of Life, Parts I-III. In *Wheels, Life, and other Mathematical Amusements*. New York: W. H. Freeman.
- Hsu, W.-R. & Murphy, A. H. 1986 The attributes diagram: a geometrical framework for assessing the quality of probability forecasts. *Int. J. Forecasting* **2**, 285-293.
- Jolliffe, I. T. & Stephenson, D. B. 2003 *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. London, UK: Wiley and Sons.
- Jung, T., Palmer, T. N. & Shutts, G. J. 2005 Influence of stochastic parameterization on the frequency of occurrence of North Pacific weather regimes in the ECMWF model. *Geophys. Res. Lett.* **32**, L23811. (doi:10.1029/2005GL024248)
- Kraichnan, R. H. 1976 Eddy viscosity in two and three dimensions. *J. Atmos. Sci.* **33**, 1521-1536.
- Lanzante, J. R. 2005 A cautionary note on the use of error bars. *J. Clim.* **18**, 3699-3703.
- Leith, C. E. 1990 Stochastic backscatter in a sub-grid-scale model: Plane shear mixing layer. *Phys. Fluids A* **2**, 297-299.
- Mason, P. J. & Thomson, D. J. 1992 Stochastic backscatter in large-eddy simulations of boundary layers. *J. Fluid Mech.* **242**, 51-78.
- Murphy, A. H. 1986 A new decomposition of the Brier score: Formulation and interpretation. *Mon. Wea. Rev.* **114**, 2671-2673.
- Nicholls, N. 2001 The insignificance of significance testing. *Bull. Amer. Meteor. Soc.* **82**, 981-986.
- Palmer, T. N. 2001 A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parameterization in weather and climate prediction. *Quart. J. Roy. Meteor. Soc.* **127**, 279-304.
- Rodwell, M. & Doblas-Reyes, F. J. 2006 Predictability and prediction of European monthly to seasonal climate anomalies. *J. Clim.* **19**, 6025-6046.

- Shutts, G. J. 2005 Kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.* **131**, 079-3102.
- Shutts, G. J. & Palmer, T. N. 2007 Convective Forcing fluctuations in a cloud-resolving model: Relevance to the stochastic parameterization. *J. Clim.* **20**, 187-202.
- Stephenson, D. B., Coelho, C. A. S. & Jolliffe, I. T. 2007 Two extra components of the Brier score decomposition. *Wea. Forecasting*, in press.
- Tibaldi, S. & Molteni, F. 1990 On the operational predictability of blocking. *Tellus A* **42**, 343-365.
- Uppala, S. M. & 45 others 2005 The ERA-40 reanalysis. *Quart. J. Roy. Meteor. Soc.* **131**, 2961-3012.

Tables

Table 1: Regions used in the computation of the forecast quality measures. The first four regions include land and ocean grid points, while only land points have been considered in the rest of the regions.

	Latitude (south,north)	Longitude (west, east)
Europe	35° to 75°	-12.5° to 42.5°
North America	30° to 70°	-130° to -60°
Northern Hemisphere	30° to 87.5°	0° to 360°
Tropics	-20° to 20°	0° to 360°
Mediterranean	30° to 47.5°	-10° to 40°
Australia	-45° to -11°	110° to 155°
Amazon	-20° to 12°	-82.5° to -35°
Southern South America	-55° to -20°	-75° to -35°
Western North America	30° to 60°	-130° to -82,5°
Eastern North America	25° to 50°	-85° to -60°
Northern Europe	47.5° to 75°	-10° to 40°
West Africa	-12.5° to 17.5°	-20° to 22.5°
East Africa	-12.5° to 17.5°	22.5° to 52.5°
Southern Africa	-35° to -12.5°	-10° to 52.5°
Southeast Asia	-10° to 20°	95° to 155°
East Asia	20° to 50°	100° to 145°
Southern Asia	5° to 30°	65° to 100°
Central Asia	30° to 50°	40° to 75°
North Asia	50° to 70°	40° to 180°

Table 2: Number of cases in which one of the experiments is significantly better (with 95% confidence) than the other for different forecast quality measures. The skill scores have been computed for both start dates, all the regions in Table 1, the variables 500 hPa geopotential height, 850 hPa temperature, precipitation, 2-metre temperature and mean sea level pressure, and several forecast times (first month and seasonal forecast periods 1-3, 2-4, 3-5, and 4-6 months). This gives 380 cases for the ACC and 1,140 for the rest of the scores. The events anomalies above the upper tercile and the median and below the lower tercile have been considered for the probability forecasts.

	ACC	BSS	BSSI	RELSS	RESSS	ROCSS
CTRL	23	62	58	23	58	66
CASBS	114	644	729	735	149	303

Figures

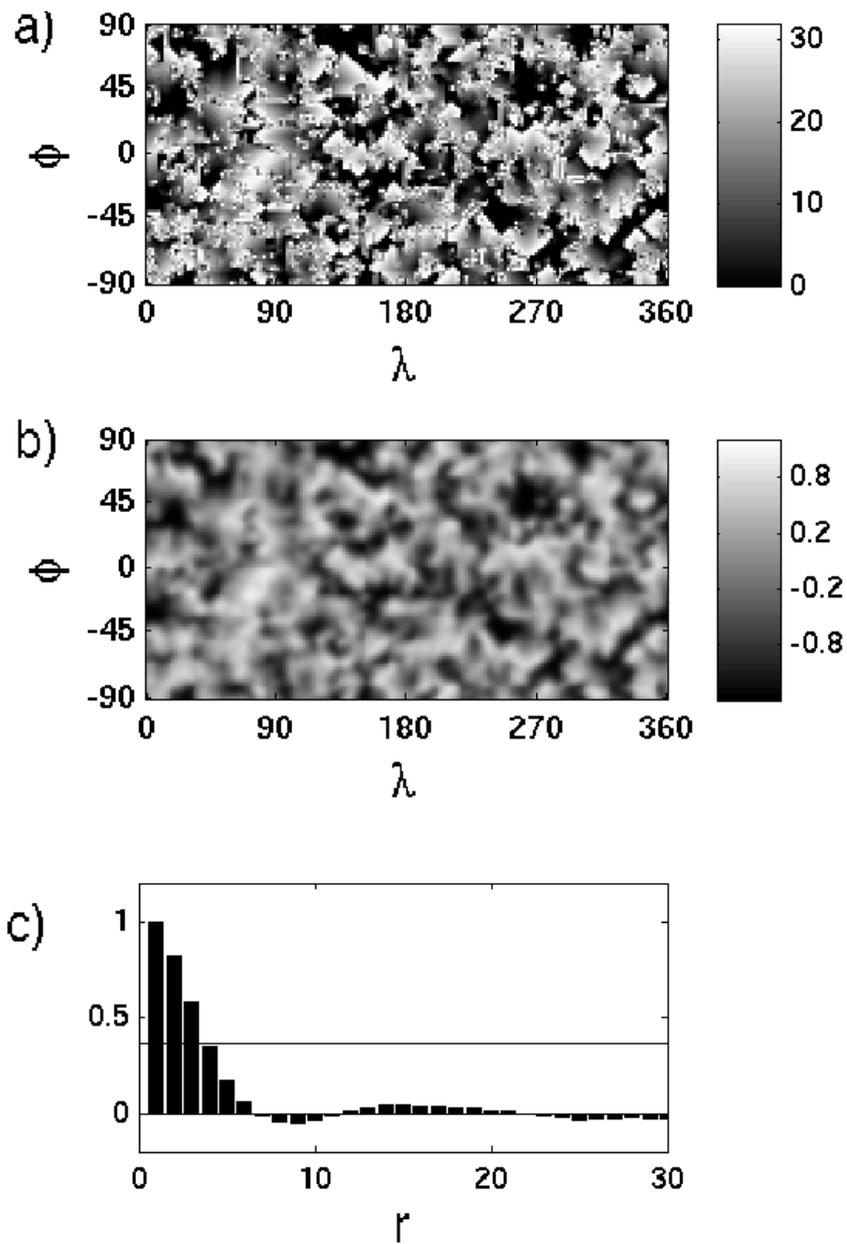


Figure 1: a) Raw pattern of 720 x 360 cells generated by the fine-scale cellular automaton used in the Cellular-Automaton Stochastic Backscatter Scheme CASBS. The shading depicts the number of lives. b) Pattern after coarse-graining onto 180 x 90 macro cells, smoothing and normalization (see text). The shading shows the pattern values after normalisation. c) Spatial correlation as function of the distance r between macro cells, in units of the width of a single macro cell. The horizontal line signifies the e-folding distance.

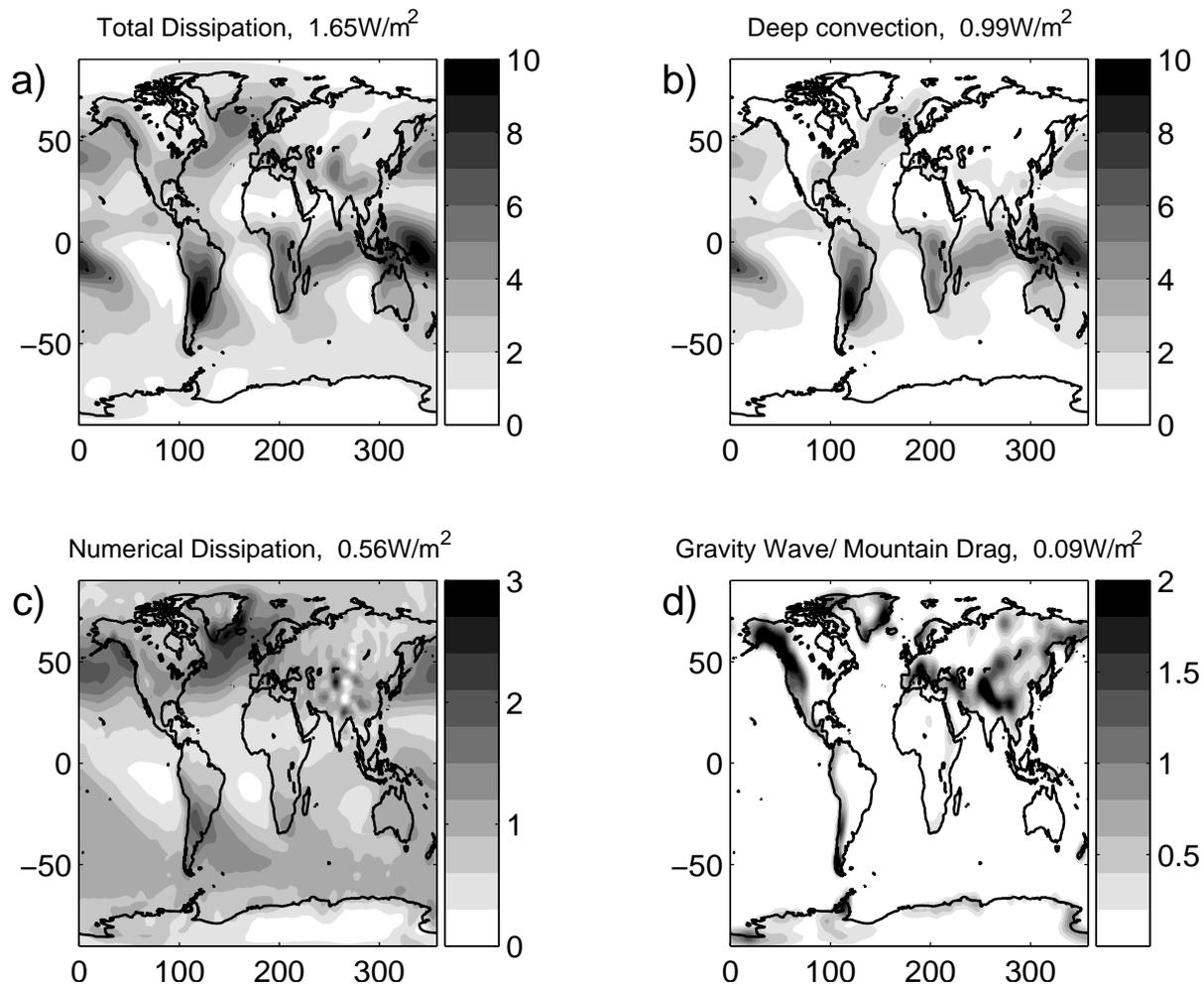


Figure 2: Vertically averaged annual mean total dissipation rate per unit area for the period December to February (a) with contributions from deep convection (b), numerical dissipation (c), gravity/mountain wave drag (d). Units are Wm^{-2} . The global average is indicated at the top of each panel.

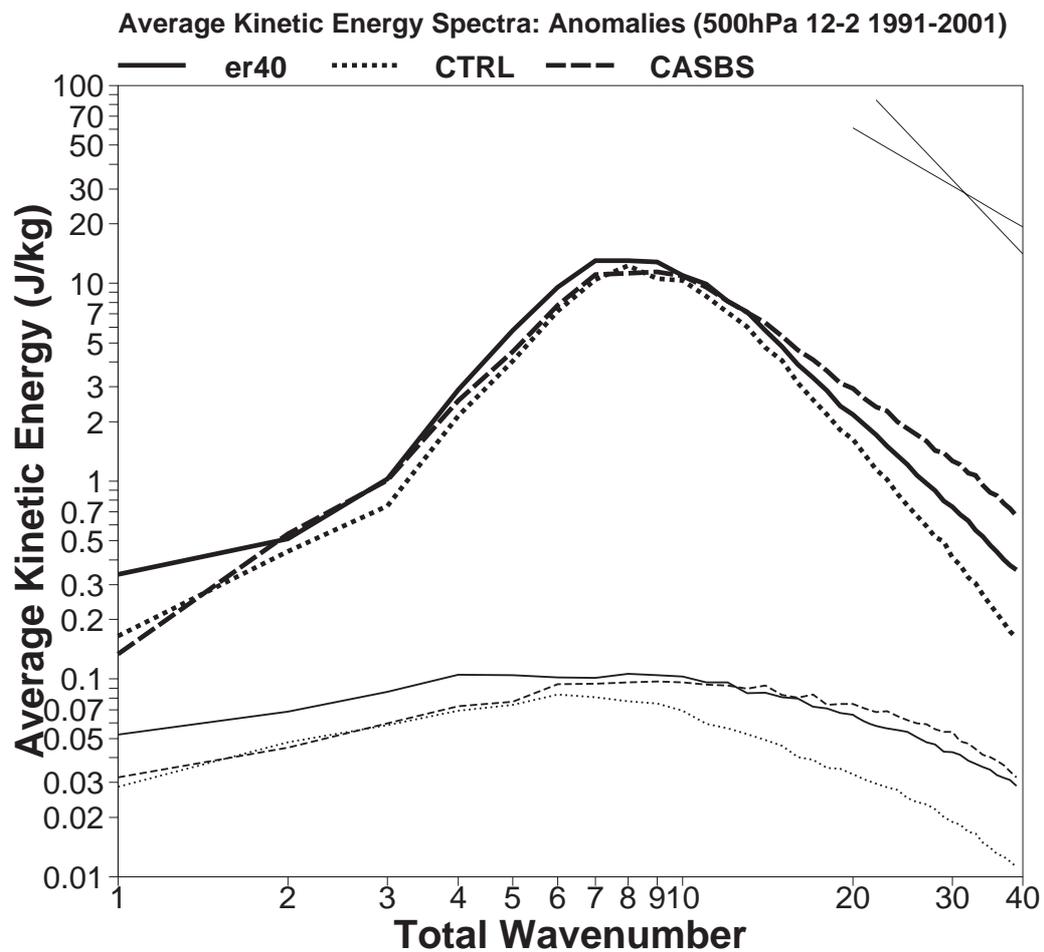


Figure 3: Mean kinetic energy spectra at 500hPa of rotational (upper) and divergent (lower) component for simulations starting on 1st November over the period 1991-2001. The anomalies were computed as departures from the respective long-term mean. Solid: ERA-40; dotted: CTRL; dashed: CASBS. Lines in upper right corner denote power-law behaviour with slopes of -3 and -5/3.

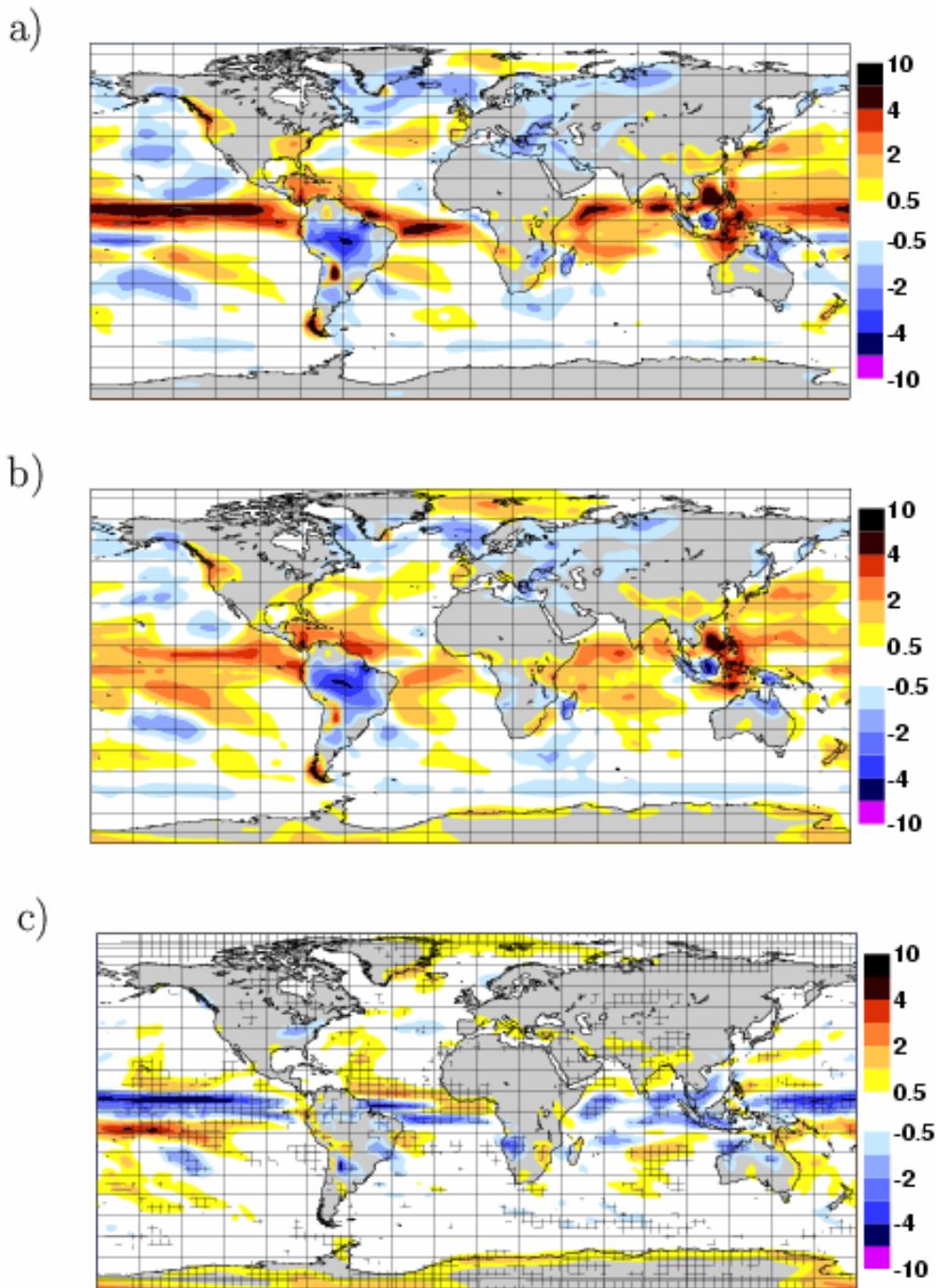


Figure 4: Systematic error of DJF total precipitation with respect to GPCP for simulations starting on 1st November over the period 1991-2001 for a) CTRL and b) CASBS. (c) Difference between CASBS and CTRL experiments; hatching indicates areas where the experiments are statistically significantly different with 95% confidence. Units are mm/day.

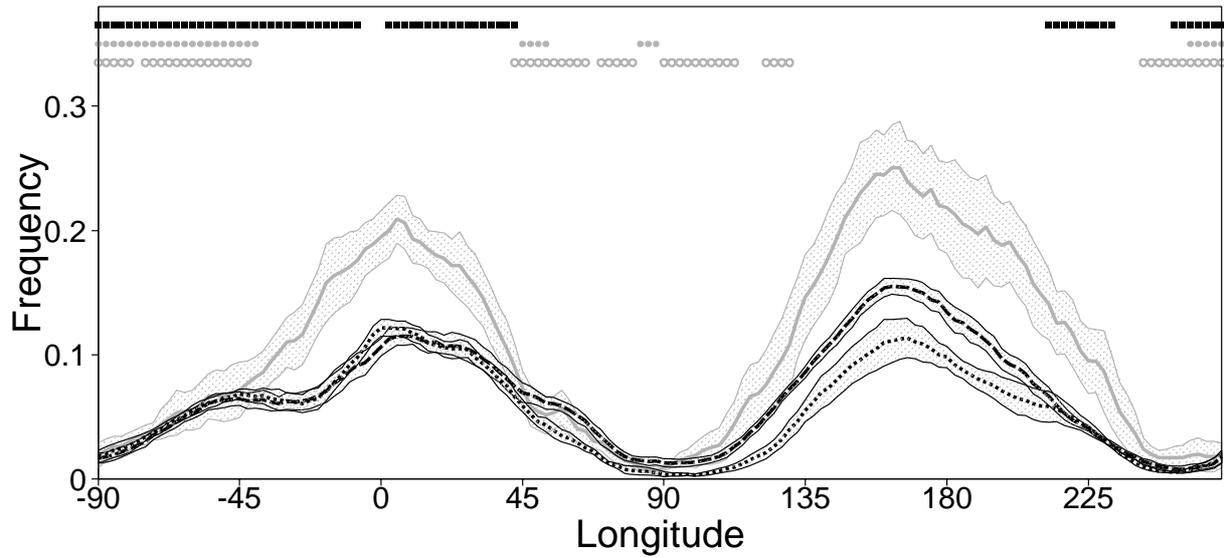


Figure 5: Frequency of blocking days in DJF for simulations starting on 1st November over the period 1991-2001. Gray solid: ERA-40; dotted: CTRL; dashed: CASBS. The grey symbols in the top of the panel show the longitudes where the experiment climatology is not significantly different from the ERA40 results, using a two-sample test based on the bootstrap estimates (solid dots for CTRL and open dots for CASBS). The black squares correspond to those longitudes where both experiments are not significantly different.

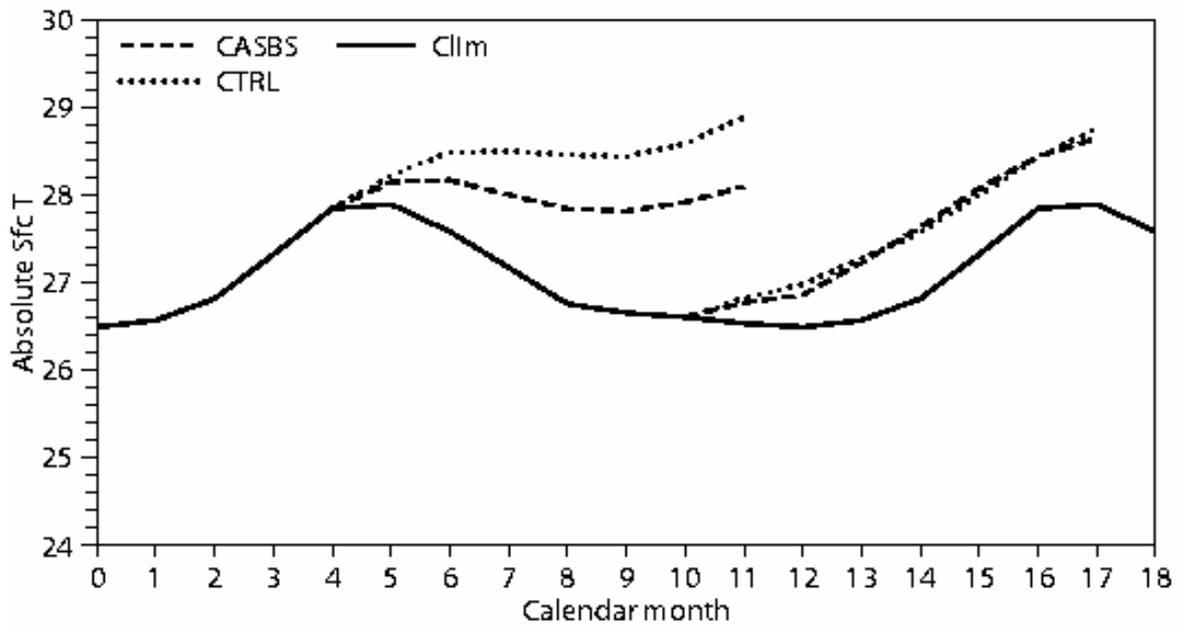


Figure 6: Sea surface temperature ($^{\circ}\text{C}$) over seven months in the Niño3.4 region (5°S - 5°N , 170° - 120°W) for seasonal re-forecasts starting on 1st May and 1st November over the period 1991-2001. Solid line: climatological annual cycle; dotted: CTRL; dashed: CASBS.

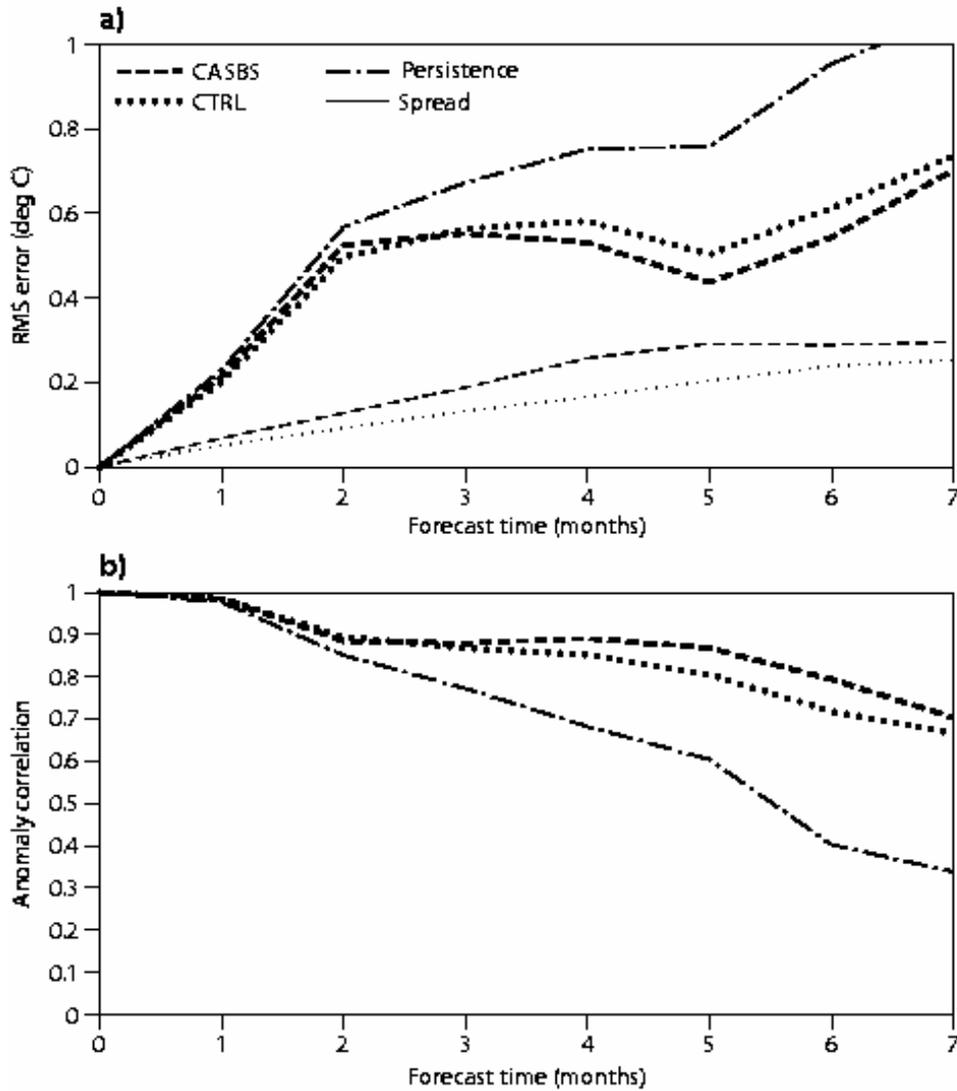


Figure 7: (a) Root mean square error of the ensemble mean (bold lines) versus lead time for anomalies of surface temperature over the Niño3.4 region (5°S - 5°N , 170° - 120°W) of the CTRL (dotted) and CASBS (dashed) experiments for the May and November start dates in the period 1991-2001. The dot-dashed black line corresponds to a simple statistical model based on persisting the anomaly of the month previous to the start date. The thin lines show the spread estimated as the standard deviation of the ensemble members around the ensemble mean. (b) Anomaly correlation coefficient of the ensemble mean for CTRL (dotted), CASBS (dashed) and persistence (dot-dashed line).

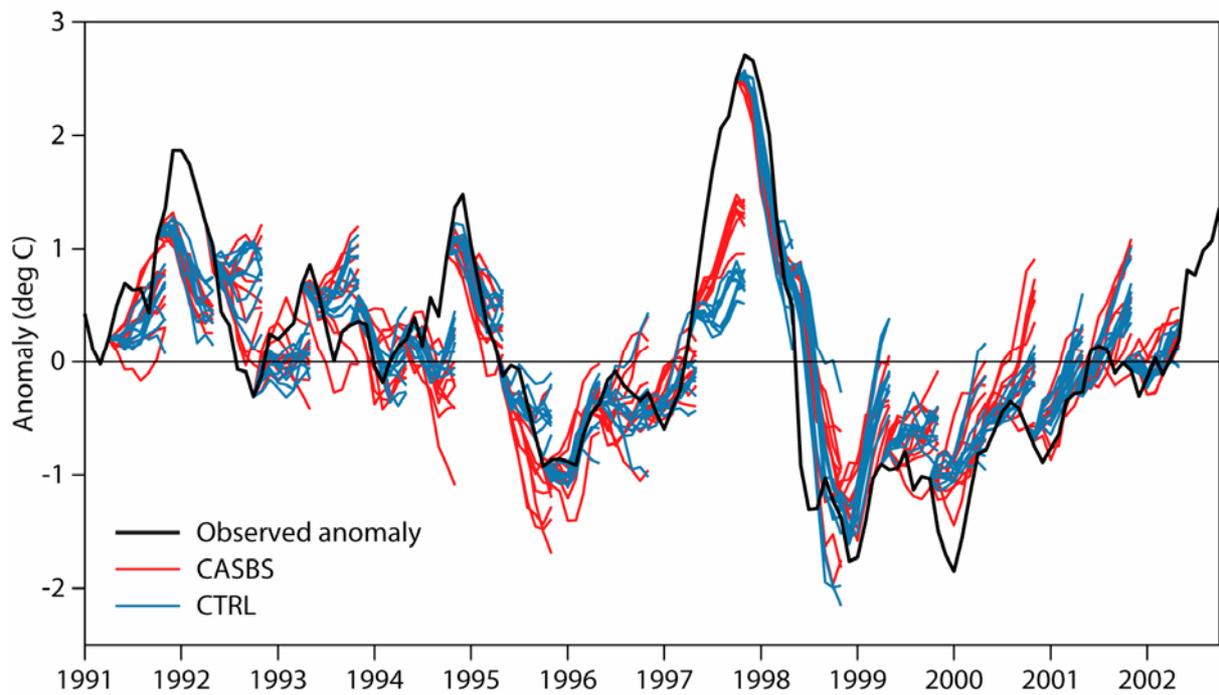


Figure 8: Ensemble forecasts of sea surface temperature anomalies over the Niño3.4 region (5°S-5°N, 170°-120°W) of the CTRL (blue) and CASBS (red) experiments for the May and November start dates in the period 1991-2001. The solid black line represents the observed monthly mean anomalies with respect to the mean seasonal cycle of the 1971-2000 period.

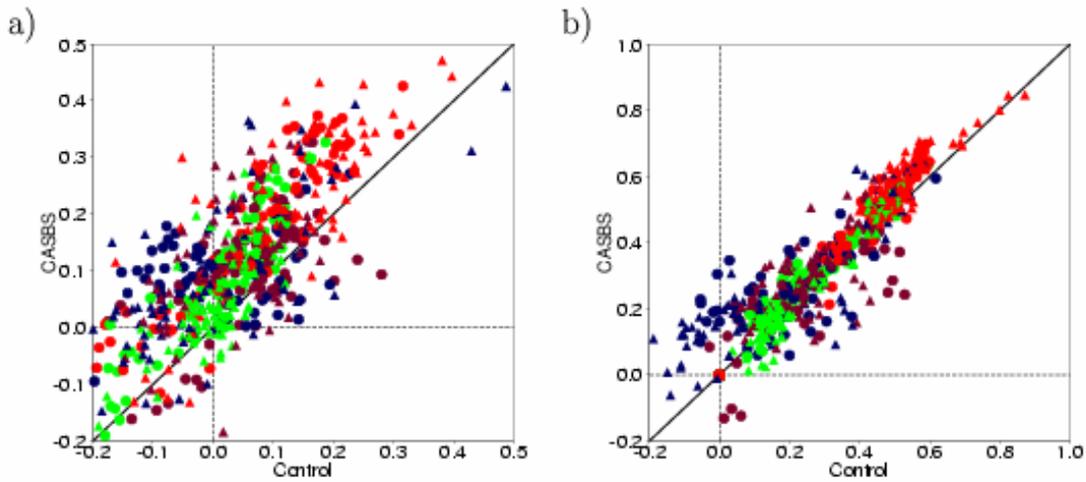


Figure 9: Brier skill score for infinite ensemble size (a) and ROC skill score (b) for CASBS (ordinate) versus CTRL (abscissa). The scatter plots collect results for both start dates, the first four regions in Table 1 (tropics in red, Northern Hemisphere in green, Europe in dark blue and North America in purple), the variables 500 hPa geopotential height, 850 hPa temperature, precipitation, 2-metre temperature and mean sea level pressure, several forecast times (first month and forecast periods 1-3, 2-4, 3-5, and 4-6 months) and the events anomalies above the upper tercile and the median and below the lower tercile. Circles (triangles) have been used for the results of the May (November) start date. Symbols below and to the left of the no-skill dotted lines have less skill than a climatological forecast.

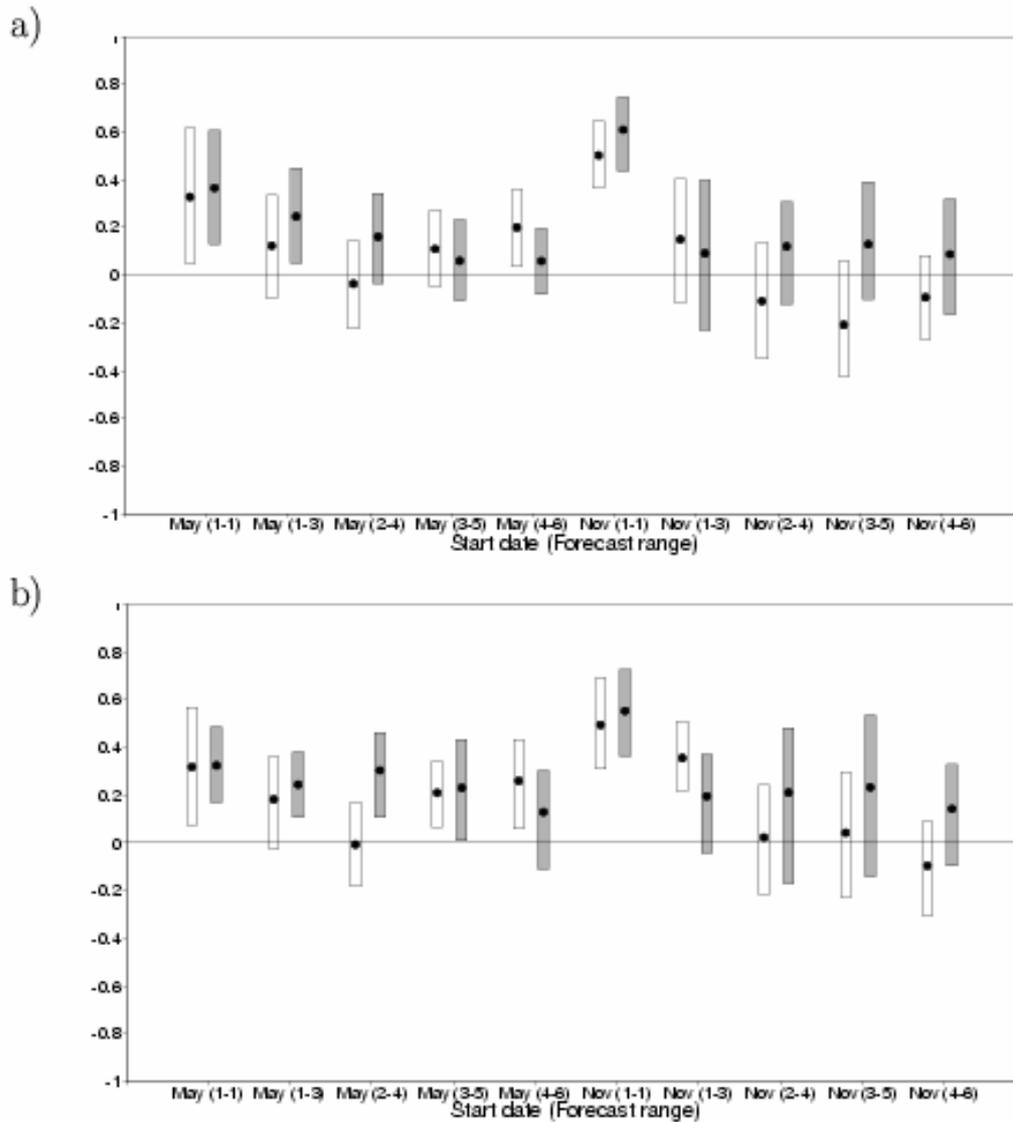


Figure 10: ROC skill score for the event “anomalies above the upper tercile” of 500 hPa geopotential height (a) and 850 hPa temperature (b) for the CTRL (unfilled bars) and CASBS (filled bars) experiments as a function of the start date and the forecast range (in brackets). The scores have been computed over Europe. The sample values are represented with a black circle, while the bars correspond to the 95% confidence intervals obtained with a bootstrap method (see text for details).

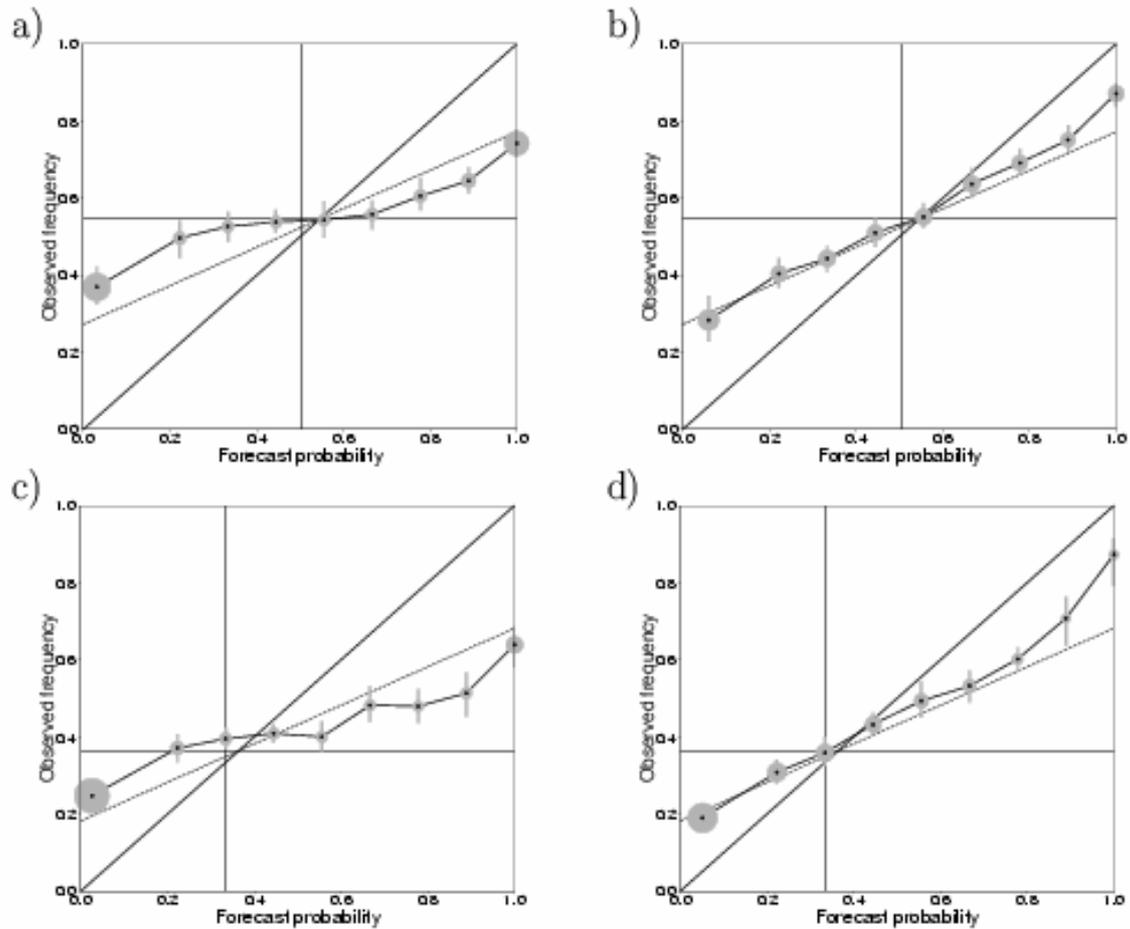


Figure 11: Attribute diagrams for first-month forecasts of precipitation over the tropics started on the 1st of May during the period 1991-2001 for the CTRL (a,c) and CASBS (b,d) experiments. Panels (a,b) show the diagrams for the event “anomalies above the median”, while panels (c,d) are for “anomalies above the upper tercile”. Each forecast probability bin in the diagrams is represented by a solid circle whose area is proportional to the bin sample size. The diagrams have been generated using nine probability bins. The grey horizontal and vertical lines indicate the climatological frequency of the event in the observations and forecasts, respectively. The diagonal line represents a perfectly reliable system where the forecast probability matches the mean observed frequency. The black dashed line separates skilful from unskilful regions in the diagram: points with forecast probabilities smaller (larger) than the climatological frequency which fall below (above) this line, contribute to positive BSS; otherwise they contribute negatively to the BSS. Grey vertical bars over the dots indicate the 95% confidence intervals of the estimated observed frequency based on a 1,000 bootstrap re-sampling procedure. The Brier skill scores for infinite ensemble size and corresponding 95% confidence intervals are (a) -0.091 (-0.149,-0.034), (b) 0.153 (0.120,0.187), (c) -0.085 (-0.146,-0.025), (d) 0.160 (0.119, 0.203).