TECHNICAL MEMORANDUM

476

Impact of Increasing Greenhouse Gas Concentrations in Seasonal Ensemble Forecasts

F. J. Doblas-Reyes, R. Hagedorn,
T. N. Palmer, J.-J. Morcrette

Research Department

October 2005

Submitted to Geophysical Research Letters

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications.html

Contact: library@ecmwf.int

Abstract

A set of 44-year 6-month long seasonal ensemble coupled model forecasts performed with annually updated greenhouse gas concentrations is compared to a standard seasonal ensemble forecast experiment performed with fixed concentrations. The former has more realistic temperature variability and clearly better forecast quality. The improvement in model variability is due to a better simulation of climate trends. Differences indicate that realistic initial conditions are not enough to reproduce this long-term variability and provide new evidence of the climatic effects of anthropogenic changes in atmospheric composition. The improvement in probabilistic forecast quality is mostly due to an increase in the ability to reliably discriminate the occurrence of events and non-events. The relevance of these results for the improvement of operational seasonal predictions is discussed.

1. Introduction

Studies of historical temperature records provide evidence for global atmospheric warming since the early 20th century, especially for the last two decades (IPCC, 2001). The warming may be a response to anthropogenic and natural (changing solar or volcanic activity, cloud amounts or albedo) forcings, the climate system's intrinsic natural variability, or a combination of both. Long-term simulations of the Earth's climate reveal the important role of increasing greenhouse gas (GHG) concentrations in explaining this warming (e.g., Meehl et al., 2004). However, it has long been assumed that the effect of long-term variations in GHG concentrations may be rather small in shorter time scale simulations such as those performed in medium-range and seasonal forecasting. For instance, Cai and Kalnay (2005) showed that an atmospheric data assimilation system is able to capture the full strength of a climate trend caused by an external forcing even if this forcing is absent from the model, as long as the observations are frequently available for twice-daily assimilation. Zhao and Dirmeyer (2004) found little impact when GHG concentrations are annually updated in global atmospheric seasonal integrations forced by observed sea surface temperatures (SSTs).

Dynamical seasonal forecasts have been considered as a problem mainly determined by ocean initial conditions and, therefore, been produced with models using constant GHG concentrations. However, the potential effect of increasing GHG concentrations in seasonal integrations via the interaction between the atmosphere and the ocean has not been documented. The goal of this paper is to assess the size of this impact in the context of an operational ensemble seasonal forecast system with a global coupled model. To address this issue, we examine the temperature signal and changes in forecast quality associated with the long-term variations in GHG concentrations.

2. Model, data and experimental design

The ECMWF coupled general circulation model has been used in this experiment. The model uses IFS cycle 23R4 and HOPE-E as atmospheric and ocean components (Palmer et al., 2004), respectively. The data consisted of sets of 6-month long seasonal integrations, each set comprising 88 simulations started from realistic ocean, land and atmosphere initial conditions on the first day of May and November of each year over the period 1958-2001. Each integration consists of an ensemble of nine members at T95 horizontal resolution in the atmosphere. The experiment mimicks the setup used to produce ECMWF's operational seasonal forecasts. The initialization of the integrations followed the method described in Palmer et al. (2004), with ERA40 (Uppala et al., 2005) data for the atmosphere and soil initial conditions and a set of ocean runs forced with ERA40 fluxes with wind stress and SST perturbations to generate the ensemble.

Operational seasonal forecast systems around the world usually apply a fixed GHG concentration during the simulations. This is not only the case for the forecasts themselves, but also for the set of 20-30 years of past predictions required for calibration (Doblas-Reyes et al., 2005; Stephenson et al., 2005) and to estimate the forecast quality of the system (Jolliffe and Stephenson, 2003; Hagedorn et al., 2005). The need of long series of past integrations and the noticeable increase in GHG concentrations in the last 50 years imply that this external forcing might play an important role in seasonal forecasting. A control experiment (CONS henceforth) with greenhouse concentrations of the gases CO_2, N_2O, CH_4, CFC11 and CFC12 fixed to their 1990 concentrations (353 ppmv, 1.72 ppmv, 310 ppbv, 280 pptv and 484 pptv, respectively) was performed. A similar experiment was carried out with concentrations updated every year (VARI). The yearly values of GHG concentrations were taken from IPCC until 2000 and then completed beyond that year with the scenario A1B. A linear regression was applied to the mean annual values, assumed to be a representation of mid-year conditions. No seasonal cycle was used.

ERA40 (Uppala et al., 2005) atmospheric fields were used as reference dataset. In spite of the limitations of reanalysis data being used as surrogates for observations, they ensure global coverage of the fields. In addition, Agudelo and Curry (2004) showed the good agreement between ERA40 temperature trends and those obtained from a radiosonde network.

3. Results

As in other state-of-the-art seasonal forecast dynamical models, there are systematic errors in the simulated variables. For instance, the ocean bias at the end of the six months remains close to 1K. Seasonal-average model biases have been estimated as the differences between the model and the ERA40 climates for the same period. A thorough assessment of model biases in this particular version of the coupled model is available from http://www.ecmwf.int/research/demeter/d/charts/verification/bias/. The biases are almost identical for both experiments (not shown). This is not only the case when the complete period (1958-2001) is considered, but also for the earliest (1958-1969) and latest (1990-2001) sections of the experiments (not shown), when differences in GHG concentrations are the largest.

In the following, only anomalies will be considered. The anomalies have been computed as the difference between the model value and an estimate of the model climate. A 1-year out cross-validation strategy (Michaelsen, 1987) was employed wherein, for each year, an estimate of the model climate is computed with all the simulations available except the target one. A similar procedure has been used to compute the reference anomalies.

The anomaly fields have been used to assess how well the experiments simulate the observed variability for each season. Figure 1 shows the 44 ensemble hindcasts and ERA40 global-average 2-metre temperature. These are 3-month lead time August-to-October (ASO) seasonal average results for the ensembles started on the first of May of each year. The values are constructed as an area-weighted average of all the grid boxes from 87.5°N to 87.5°S. The ERA40 data (red dots) display a distinctive warming tendency, with an abrupt shift in the late 1970s (Graham, 1994). Before that date, most of the ERA40 anomalies are below the zero line, and vice versa for the period after 1980. A similar behaviour with a smaller warming rate has been found for winter. The box-and-whisker symbols display the range of ensemble values, with each whisker representing a third of the members. The CONS experiment clearly underestimates the observed variability, the red dots falling most of the time outside the range of the ensemble. In addition, the model anomalies do not follow the ERA40 behaviour before and after 1980. A slightly better match between model and ERA40 2-metre temperature time series can be found for the VARI experiment. In this case, the model anomalies are more frequently below the zero line for the first part of the period and above for the second. This better match is reflected in some measures of forecast quality such as the correlation between the ensemble mean (blue dots) and the reference and the ranked probability skill score (RPSS; Jolliffe and Stephenson, 2003) for tercile categories, as shown in Table 1. The VARI experiment outperforms the CONS one, having more statistically significant values. This not only happens for the late part of the integration, but as early as the first month, especially in winter, and can be attributed to a better simulation of the rising trend, as the interannual variability is similar in both experiments. Table 1 also shows that the improvement is due to a better simulation over both land and oceans and at different time scales. This is a rather surprising result given that, relative to the atmosphere, the oceans are expected to keep the memory of the realistic initial conditions for several months to avoid sensitivity to changes in GHG concentrations.

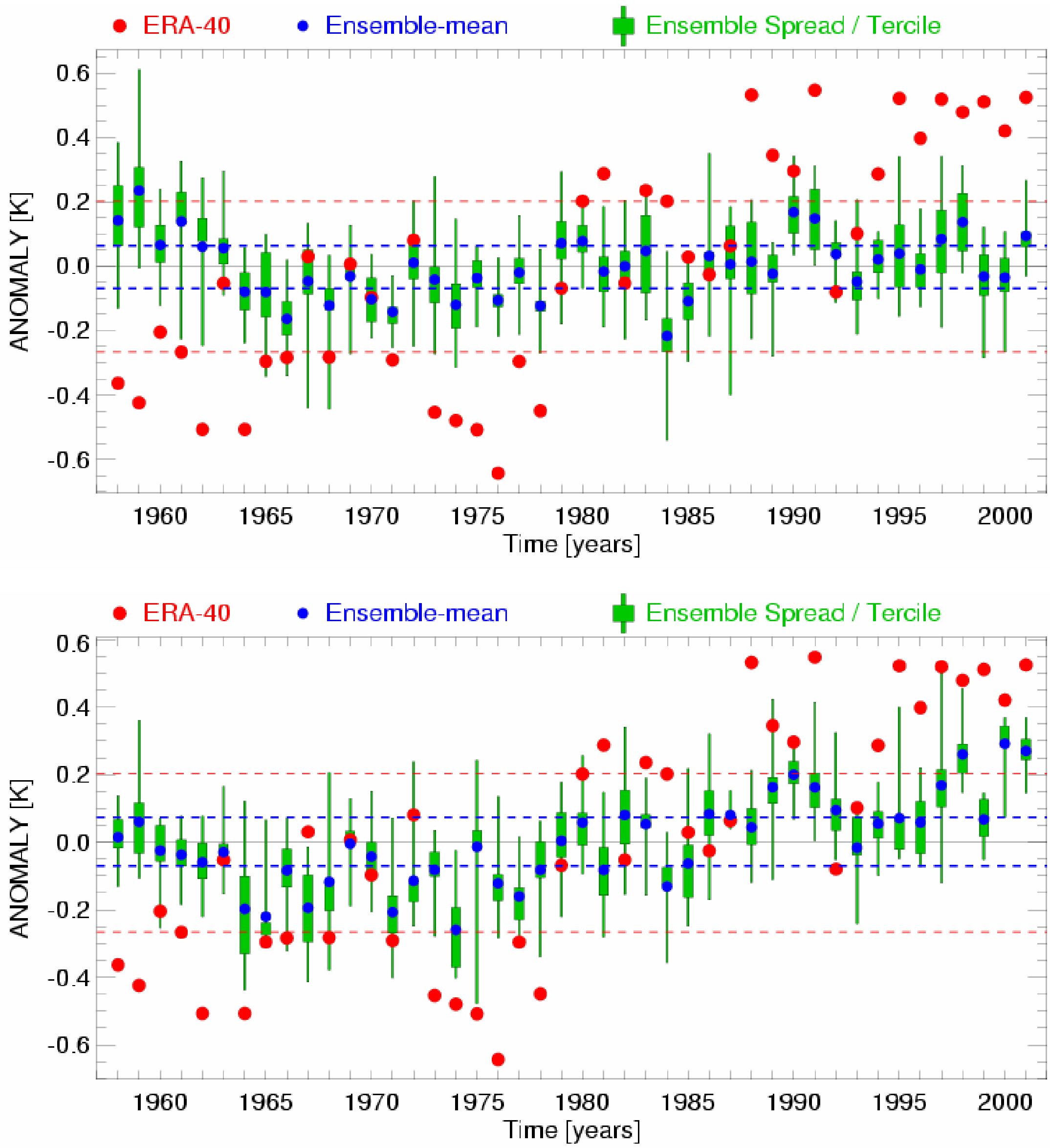

Figure 1: Global average late summer (May start date, 3-month lead, August-to-October) 2-metre temperature for the 9-member ensemble experiments a) CONS and b) VARI. ERA-40 and ensemble-mean values are depicted with red and blue dots, respectively. The box-and-whisker symbols show the range of ensemble values, with each whisker representing a third of the members. The red and blue dashed lines show the climatological terciles for ERA40 and the model, and illustrate the underestimation of the model variability.

Table 1: Correlation (Corr) between the ensemble mean and the reference and ranked probability skill score (RPSS, a measure of forecast quality for probabilistic forecasts that estimates the accumulated quadratic distance in probability space between the observed and forecast probability distribution function) for tercile categories for global-average (G), land-only (L) and ocean-only (O) 2-metre temperature time series of the CONS and VARI experiments. Values significantly different from zero with a 99% confidence level are shown in bold. Several lead and averaging times are considered for the simulations starting on the first of May and November: first month of the integration (May, M, and November, N), 1-month lead (June-to-August, JJA, and December-to-February, DJF) and 3-month lead (August-to-October, ASO, and February-to-April, FMA) seasonal averages.

		1st May start (M)	Average 2-4 May start (JJA)	Average 4-6 May start (ASO)	1st month Nov start (N)	Average 2-4 Nov start (DJF)	Average 4-6 Nov start (FMA)
CONS	Corr	(G) **0.79** (L) **0.63** (O) **0.86**	**0.52** **0.30** **0.71**	0.29 0.13 **0.39**	**0.76** **0.59** **0.60**	**0.46** **0.46** 0.25	0.25 **0.33** 0.13
VARI	Corr	(G) **0.81** (L) **0.71** (O) **0.81**	**0.79** **0.70** **0.80**	**0.68** **0.58** **0.69**	**0.79** **0.66** **0.77**	**0.67** **0.56** **0.50**	**0.58** **0.60** **0.45**
CONS	RPSS	(G) **0.47** (L) **0.18** (O) **0.48**	**0.10** 0.01 **0.22**	**0.09** -0.06 -0.07	**0.21** **0.16** **0.20**	**0.10** **0.09** -0.11	-0.05 **0.06** -0.11
VARI	RPSS	(G) **0.58** (L) **0.31** (O) **0.57**	**0.41** **0.25** **0.39**	**0.38** **0.25** **0.25**	**0.25** **0.24** **0.16**	**0.27** **0.19** **0.14**	0.04 **0.23** **0.05**

Enfield and Mestas-Núñez (1999) applied complex empirical orthogonal function analysis to a global SST data set with the El Niño-Southern Oscillation signal removed and found that the first mode contains the global warming signal. To test this hypothesis in our experiments, empirical orthogonal function (EOF) analyses of global and land-only 2-metre temperature and SST have been performed. The covariance matrix was used in the eigenvalue problem. The first (second) principal component (PC) of global 2-metre temperature in summer (winter) shows a rising trend while the second (first) PC correlates strongly with the variability in the central tropical Pacific. While the leading couple of EOF patterns resemble those of ERA40 in both experiments, the rank of the summer pair is inverted, pointing at the weaker warming signal in the experiments. As above, the VARI experiment performs better than the CONS one for global, land and ocean EOFs. For example, the correlation of the 3-month lead, November start date ensemble-mean leading PC of global 2-metre temperature for CONS and VARI is 0.12 and 0.70, respectively.

The previous results suggest that the improvement in the simulation of surface air and sea temperature when using realistic GHG concentrations is mainly due to the better representation of the long-term trends. A parametric approach has been used to estimate the long-term trends using a linear fit to the ensemble mean and ERA40 anomalies. The slope coefficient has been divided by the standard deviation of the ensemble and ERA40 detrended anomalies. The coupled model used here represents the temperature trends in the first month of the integration both at the surface and in the lower troposphere. However, the relative amplitude of the simulated trend is severely underestimated with increasing the integration time. Figure 2 shows the standardized slope for ERA40 and both experiments for the 3-month lead ASO 850 hPa temperature. ERA40 shows a positive trend over the tropical band, except for some continental areas. Cooling trends are found

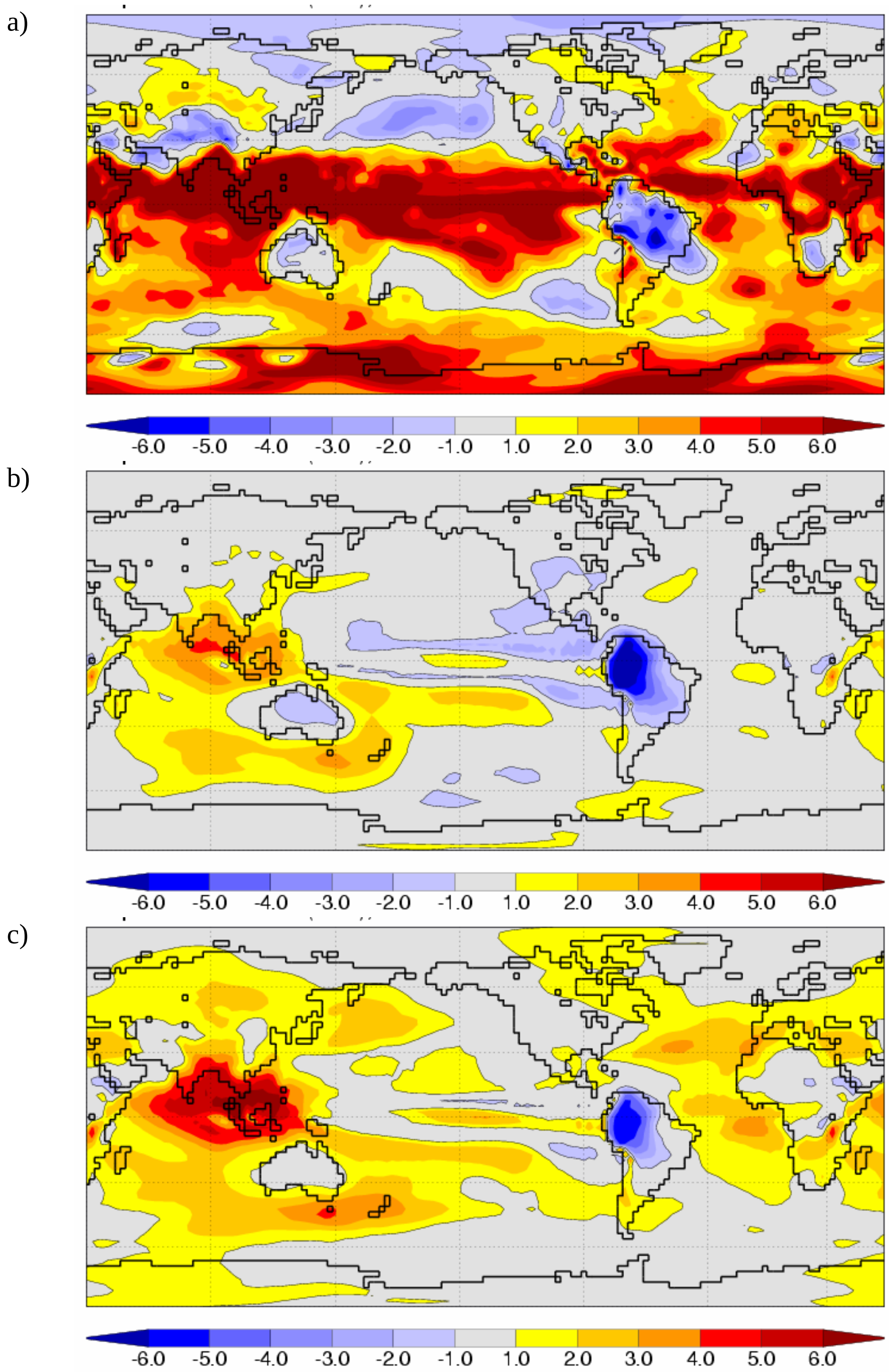

Figure 2: Linear trend divided by the detrended standard deviation for the 3-month lead August-to-October seasonal-average 850 hPa temperature for a) ERA40, and b) the CONS and c) VARI experiments over the period 1958-2001.

over some extratropical areas. The CONS experiment not only underestimates the size of the trends, but also shows an excessive cooling trend over the tropical Pacific that agrees well with a decrease in net surface solar radiation (not shown). These errors are substantially alleviated in the VARI experiment, as shown in Figure 2c. The improvement is linked to more realistic trends of both surface and top of the atmosphere net longwave radiation (not shown). The increase in the amplitude of the temperature trend of the VARI experiment with regard to the CONS experiment agrees with a more realistic seesaw in the geopotential height trends between the tropics and the extratropics (not shown), as described in Zhu et al. (2002). Similar results are found for the winter season, the near-surface temperature and other lead times. Some authors have found that the previous temperature trends are almost in phase opposition in the middle troposphere (e.g., Agudelo and Curry, 2004), with slight cooling over the tropics and warming over the midlatitudes, especially the southern Pacific (Kim et al., 2005). However, both experiments considered here reveal an excess of tropical warming at 500 hPa, in particular the VARI experiment, an error present since the first month of the integrations. Further aloft, at 200 hPa, the experiments agree again with the ERA40 trends, with strong warming over the tropics (stronger for the VARI experiment) and some (underestimated) cooling over midlatitudes.

A better representation of global temperature or the long-term trends in ensemble forecasts does not guarantee an increase in forecast quality. However, given that seasonal forecast skill tends to be relatively low, even marginal improvements can prove very useful. Figure 3 shows the Brier (BSS), reliability (BSS_rel) and resolution (BSS_res) skill scores (Jolliffe and Stephenson, 2003) for seasonal average 2-metre and 850 hPa temperature computed over the Northern and Southern Hemisphere (left column) and the tropical band (right column) for several start dates, events and lead times. The skill scores for CONS are plotted versus those of VARI. The numbers on top of each panel are the percentage of skill scores for which VARI is better than CONS. They clearly show that forecast quality in the former experiment is better than in the second one, especially for the Southern Hemisphere and the tropics. The improvement is the result of a substantial increase in the resolution component of the Brier score, in particular for 850 hPa temperature. Interestingly, this is achieved simultaneously improving the reliability. This is a highly relevant result in dynamical forecasting as, given an appropriate method, ensemble predictions can be calibrated to produce reliable forecasts (thus increasing the reliability skill score), while no method can increase the resolution of the forecasts, unless additional information is merged into the prediction (Doblas-Reyes et al., 2005; Stephenson et al., 2005). Some marginal improvements are also found for predictions of 500 hPa geopotential height.

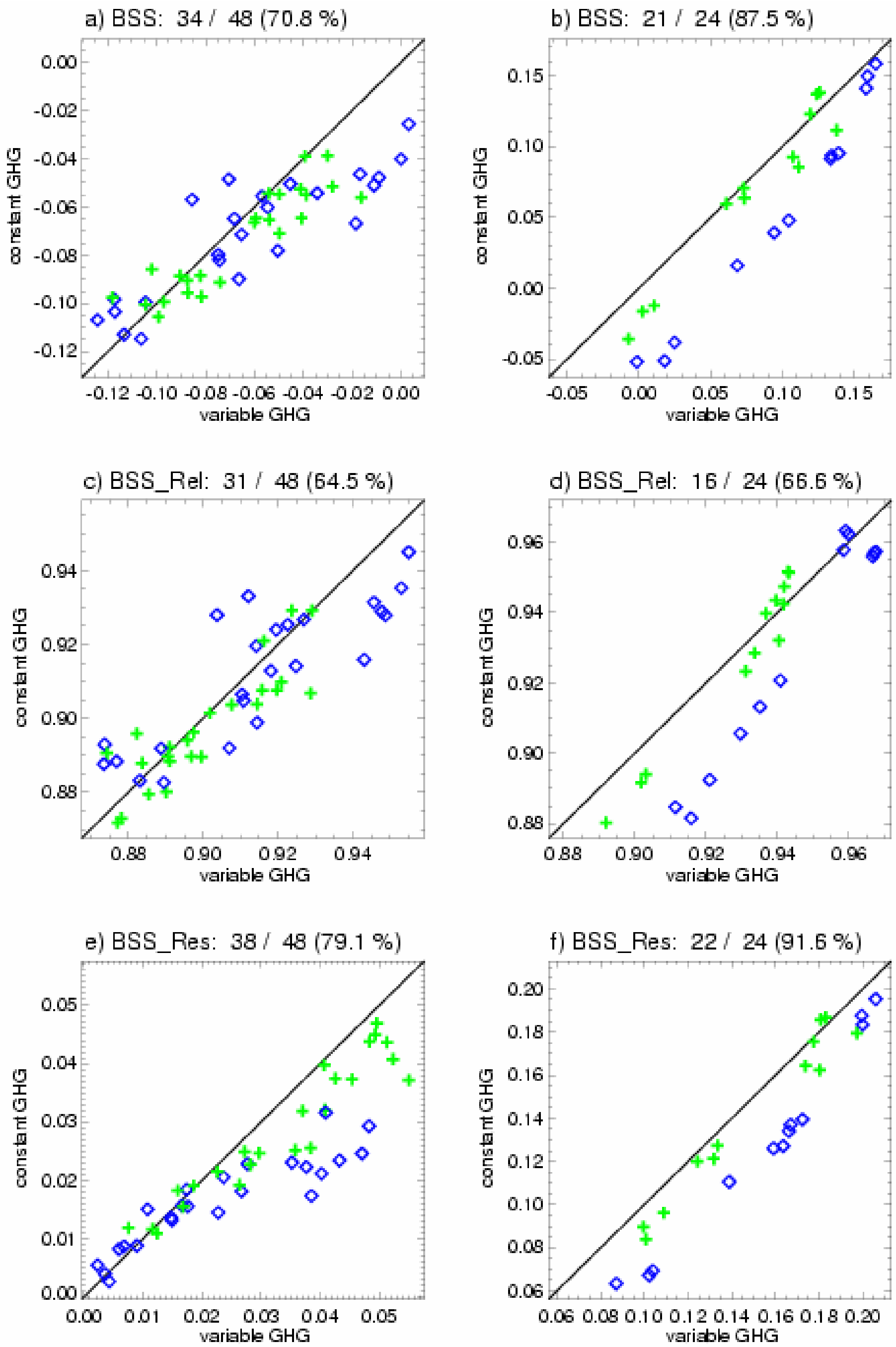

Figure 3: Scatter plots of Brier skill score (BSS, top panels), resolution skill score (BSS_Res, central panels) and reliability skill score (BSS_Rel, bottom panels) of the CONS versus the VARI experiments. BSS is a measure of forecast quality for probabilistic forecasts of binary events and estimates the quadratic distance in probability space between the observed and forecast probabilities. BSS_rel and BSS_res are obtained from a decomposition of the BSS and measure the ability of the forecast system to issue probabilities that, on average, match the observed frequency of the event and the ability of a reliable system to discriminate between events and non-events, respectively. The results are for seasonal average 2-metre (green crosses) and 850 hPa (blue diamonds) temperature computed over the extratropics (Northern and Southern Hemisphere, left column) and the tropical band (30°N-30°S, right column) for several start dates (May and November), events (anomalies above the mean, above 0.43 times and below -0.43 times the standard deviation) and for 1- and 3-month lead time.

4. Discussion and conclusions

This paper has focused on the impact of GHG concentrations in sets of seasonal forecasts performed with coupled models. It has been shown that a set of 44-year 6-month long seasonal ensemble forecasts performed with GHG concentrations updated annually display 1) more realistic variability in temperature and 2) better forecast quality than a standard seasonal ensemble forecast experiment with fixed concentrations.

Interannual variability does not seem to be significantly affected by the annually updated GHG concentrations. Instead, long-term trends become more realistic. The lack of long-term trends in surface and tropospheric temperatures in the CONS experiment (that mimics the operational seasonal ensemble forecast model in which GHG concentrations are kept constant for all the simulations) indicates that realistic initial conditions do not constrain this long-term variability beyond the first month of the simulation. Using realistic GHG concentrations partially alleviates this problem, which is an additional indication of the impact on climate of anthropogenic changes in atmospheric composition. Zhao and Dirmeyer (2004) did not find a significant impact of varying GHG concentrations in seasonal integrations. The disagreement with their results, might reside in the ocean-atmosphere interaction, which is missing in their experiment. The relevance of ocean-atmosphere coupling in seasonal experiments has already been discussed for other processes, such as the simulation of monsoon precipitation (Wang et al., 2005) or the ocean forcing of midlatitude atmospheric circulation (Bretherton and Battisti, 2000).

The improvement in probabilistic forecast quality is mostly due to an increase in the ability to reliably discriminate between events and non-events, measured in terms of resolution. Calibration and combination of single-model forecasts (Doblas-Reyes et al, 2005; Stephenson et al., 2005) are methods seriously considered by operational seasonal forecasting centres to improve forecast quality. The methods used to calibrate and combine forecasts require long time series of past forecasts. Hence, correctly reproducing past climate variability in coupled simulations is a basic requisite for the production of well-calibrated skilful seasonal predictions.

While improved seasonal forecasts are linked to progress in coupled modelling, ocean and land-surface initialization, ensemble generation and other fundamental aspects, we have demonstrated that an adequate treatment of the atmospheric composition can improve simulations in a seasonal forecast framework, increasing simultaneously the forecast quality of seasonal-average temperature predictions. The experiments discussed here illustrate the possible benefits of combining the knowledge of modelling communities involved in seasonal prediction and in the assessment of anthropogenic climate change. We expect in a near future to carry out, within the framework of the EU-funded ENSEMBLES project, similar experiments with different coupled models, as well as to assess the relevance of changes in other atmospheric components, such as anthropogenic or volcanic aerosols.

Acknowledgement

This work was supported by the ENSEMBLES project (GOCE-CT-2003-505539). The authors wish to thank ECMWF staff and consultants for their support. The feedback of some colleagues (M. Balmaseda, T. Jung, T. Stockdale, A. Weisheimer) has significantly improved the quality of the manuscript. Special thanks go to M. Liniger for his useful comments and for checking the data quality.

References

Agudelo, P. A., and J. A. Curry (2004), Analysis of spatial distribution in tropospheric temperature trends. *Geophys. Res. Let*., **31**, doi:10.1029/2004GL020818.

Bretherton, C. S., and D. S. Battisti (2000), An interpretation of the results from atmospheric general circulation models forced by the time history of the observed sea surface temperature distribution, *Geophys. Res. Let.*, **27**, 767-770.

Cai, M., and E. Kalnay (2005), Can reanalysis have anthropogenic climate trends without model forcing? *J. Climate*, **18**, 1844-1849.

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

Enfield D. B. and A. M. Mestas-Nuñez (1999), Multiscale variabilities in global sea surface temperatures and their relationships with tropospheric climate patterns, *J. Climate*, **12**, 2719-2733.

Graham, N.E. (1994), Decadal-scale climate variability in the tropical and North Pacific during the 1970s and 1980s: observations and model results, *Climate Dyn*., **9**, 135-162.

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.

IPCC (Intergovernmental Panel on Climate Change) (2001), *Climate change*, edited by J. T. Houghton, L. G. Meira Filho, B. A. Callender, N. Harris, A. Kattenberg, and K. Maskell, Cambridge Univ. Press, New York.

Jolliffe, I. T. and D. B. Stephenson (2003), *Forecast verification: A practitioner's guide in atmospheric science*, Wiley, New York.

Kim, D., N. Nakashiki, Y. Yoshida, K. Maruyama, and F. O. Bryan (2005), Regional cooling in the South Pacific sector of the Southern Ocean due to global warming, *Geophys. Res. Let*., **32**, L19607, doi:10.1029/2005GL023708.

Meehl, G. A., W. M. Washington, C. A. Ammann, J. M. Arblaster, T. M. L. Wigley, and C. Tebaldi (2004), Combinations of natural and anthropogenic forcings in Twentieth-Century climate, *J. Climate*, **17**, 3721-3727.

Michaelsen, J. (1987), Cross-validation in statistical climate forecast models, *J. Climate Appl. Meteorol*., **26**, 1589-1600.

Palmer, T. N. et al. (2005), Development of a European multi-model ensemble system for seasonal to inter-annual prediction (DEMETER), *Bull. Amer. Meteor. Soc*., **85**, 853-872.

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.

Uppala, S. et al. (2005), The ERA-40 reanalysis, *Quart. J. Roy. Meteor. Soc*., in press.

Wang, B., Q. Ding, X. Fu, I.-S. Kang, K. Jin, J. Shukla, and F. J. Doblas-Reyes (2005), Fundamental challenge in simulation and prediction of summer monsoon rainfall, *Geophys. Res. Let*., **32**, L15711, doi:10.1029/2005GL022734.

Zhao, M., and P. A. Dirmeyer (2004), Pattern and trend analysis of temperature in a set of seasonal ensemble simulations, *Geophys. Res. Let*., **31**, L11202, doi:10.1029/2003GL019298.

Zhu, J., S. Wang, and X. Zhang (2002), Global warming mode of atmospheric circulation, *Atmos. Sci. Let*., doi: **10**.1006/asle.2002.0052.