Tellus (2005), 57A, 575–588
Printed in Singapore. All rights reserved

TELLUS

Reduction of systematic errors by empirical model correction: impact on seasonal prediction skill

By A. GULDBERG[1]*, E. KAAS[1], M. DÉQUÉ[2], S. YANG[1] and S. VESTER THORSEN[1],
[1]*Climate Research Division, Danish Meteorological Institute, Lyngbyvej 100, DK-2100 Copenhagen Ø, Denmark;*
[2]*Météo-France, Centre National de Recherches Météorologiques, 42 Avenue Coriolis, F-31057 Toulouse Cedex, France*

(Manuscript received 17 February 2004; in final form 25 November 2004)

ABSTRACT

Recent studies indicate that the atmospheric response to anomalies in the lower boundary conditions, e.g. sea surface temperatures, is strongly dependent on the atmospheric background flow. Since all general circulation models have long-term systematic errors it is therefore possible that the skill in seasonal prediction is improved by reducing the systematic errors of the model. In this study sensitivity experiments along this line are made with an empirically corrected dynamical model for which the systematic errors are reduced substantially and the dynamical variability has become more realistic than for the original model. As a measure of seasonal prediction skill, correlation of temporal anomalies between modelled and observed data has been determined. The corrected model shows improved skill in the Southern Hemisphere in general—on average a 20–30% improvement for the Southern Hemisphere compared with the original model. In the Northern Hemisphere skill is improved in some areas, but in other areas the skill of the original model is better. On average there is no improvement for the Northern Hemisphere. Also, pattern correlations have been determined for the following areas: the Northern Hemisphere, the Southern Hemisphere, the tropics and Europe. The general picture is that the two model versions are very similar in the Northern Hemisphere and in the tropics. For Europe the results of the two models are rather different, but no model can be said to be better than the other. In the Southern Hemisphere it is again seen that the correlations are higher for the corrected model than for the original model.

1. Introduction

Seasonal prediction has been an area of increasing interest in recent years; most recently predictions on a monthly timescale have also come into focus (Vitart, 2004). An important method for obtaining predictions on monthly and seasonal timescales is the use of general circulation models (GCMs) for producing ensemble forecasts (e.g. Palmer et al., 2000; Branković and Palmer, 2000; Derome et al., 2001). Important factors affecting the skill of ensemble predictions are the ensemble size, to reduce the unpredictable climate noise (e.g. Déqué, 1997; Kumar and Hoerling, 2000; Kharin et al., 2001), and the quality of the GCM determined, for instance, by the systematic errors of the model and its response to lower boundary conditions (e.g. Kumar et al., 1996; Kharin and Zwiers, 2001). Recent studies indicate that the response of an atmospheric model to anomalies in the lower boundary conditions is strongly dependent on the background flow (Peng et al., 1997; Peng and Whitaker, 1999) and therefore errors in the model climatology could have an impact on the response. As forcing from the lower boundary, e.g. sea surface temperature (SST) or sea ice forcing, can be responsible for a substantial part of the model variance (Derome et al., 2001), it is likely that the skill of ensemble seasonal forecasts will be improved if the systematic errors of the model can be reduced.

It is to be expected that Rossby wave energy dispersion will be simulated more realistically when the basic state of the model is closer to the observed climate. Since such energy dispersion is crucial for the remote response to lower boundary conditions (Branstator, 1983) seasonal prediction skill may benefit from an improved model climatology. On the other hand there is no *a priori* guarantee that an empirically corrected model with small systematic errors in the flow dynamics will be better for seasonal predictions. This is because convection and release of latent heat is a fundamental process by which the atmosphere reacts to SST forcing. Thus, if the physical parametrization of such processes is inadequate reduced systematic errors may not necessarily lead to improved seasonal forecast skill. Furthermore if the spatial resolution of the model is too coarse it is not possible to correctly simulate, for example, the interaction between the dynamics and the physical parametrization (Williamson, 2002).

*Corresponding author.
e-mail: ag@dmi.dk

Sausen and Ponater (1990) reduced systematic errors of a GCM by adding forcing terms (constant in time) based on statistical diagnosis of the average drift of the prognostic variables. Although correction terms were applied only to the zonal mean temperature, an overall improvement of the model climate was seen.

D'Andrea and Vautard (2000) proposed a methodology for reducing systematic errors of a quasi-geostrophic baroclinic atmospheric model. The method consists of estimating the initial tendency errors of the model by a variational assimilation procedure and then correcting the model equations by adding a flow-dependent parametrization of the initial tendency error. The model shows substantial reduction of systematic errors and improvement of both high- and low-frequency variability when empirically corrected in this way.

In the present study the model is corrected by adding an estimate of the model initial tendency errors. The method for obtaining the tendency errors is based on the nudging technique (Jeuken et al., 1996). The nudging residuals obtained during the assimilation are the estimates of the model initial tendency errors. This technique for obtaining tendency errors was also used by (Kaas et al., 1999) with the aim of tuning the parametrization of unresolved scale interactions. The systematic errors of the model are substantially reduced when the model is modified by adding seasonally varying long-term mean values of these tendency error estimates to the prognostic equations, and once improved in this way the model is found to be suitable for testing the expected impact of the reduced systematic errors on the skill of predictions on a seasonal timescale.

Section 2 of this paper describes the general circulation model and the procedure for empirically correcting the model. Section 3 contains a description of the experiments that were performed in order to compare the original model with the empirically corrected model. In Section 4 the results of these experiments are described and in Section 5 the results are discussed and conclusions are drawn.

2. The models used

2.1. The standard model

The general circulation model used is a climate model derived from ARPEGE/IFS (Déqué et al., 1994), version 2. This model is a spectral primitive equation model and the equations are solved using the transform method (Eliasen et al., 1970), Eulerian advection and a three-level semi-implicit time-stepping scheme. The model has 31 sigma-pressure hybrid vertical levels defined as in the ECMWF reanalysis data, ERA-15 (Gibson et al., 1997). The spectral horizontal resolution is T42.

2.2. The empirically corrected model

The empirically corrected model is created by correcting for empirically estimated initial tendency errors of the standard model in the equations governing the time development of the prognostic variables. For a given prognostic variable, Ψ, the tendency error is defined as $(\delta\Psi/\delta t)_{\mathrm{O}} - (\delta\Psi/\delta t)_{\mathrm{M}}$, where the subscript O denotes the instantaneous tendency in the observations. (Note that tendency errors are often defined with an opposite sign to the definition used here.) In practical use analysis data will serve as observations. Subscript M denotes the temporal tendency determined by the model when given the observed atmospheric state vector. To build an empirically corrected model the tendency error (or residual), R, has to be added to the model equations in order for the model to follow the observations, since $(\delta\Psi/\delta t)_{\mathrm{O}} = (\delta\Psi/\delta t)_{\mathrm{M}} + R$.

It is a complicated task to determine realistic tendency residuals, and several methods have been suggested. Klinker and Sardeshmukh (1992) for example, used a set of one time step forecasts for determining averages of tendency residuals, whereas Schubert and Chang (1996) used analysis increments for estimating the forcing errors of a model. Machenhauer and Kirchner (2000) used slow normal-mode data assimilation for obtaining the systematic initial tendency errors. In situations where the model in question is not the model that has been used for producing the analysis data (which is the case in this study), one should be aware that the estimates obtained for the tendency residuals might be polluted due to initial dynamical gravity noise and processes related to moisture spin-up; the consequence being that it is difficult to isolate the true tendency residual.

In the present study the technique described in Jeuken et al. (1996) has been applied, since this technique implies a reassimilation of analysis data as done here. The technique, called nudging, is a simple 4-D assimilation of the reference data, i.e. the analysis data, where the prognostic model variables are relaxed towards the reference data:

$$\Psi(t+\Delta t) = \Psi^{*}(t+\Delta t) + 2\Delta t\,\frac{\Psi^{\mathrm{REF}}(t+\Delta t) - \Psi^{*}(t+\Delta t)}{\tau}. \tag{1}$$

The upper index $*$ indicates the preliminary prognostic variable just before nudging and the upper index REF denotes the reference variable towards which the model is relaxed. Δt is the length of the time step used in the model and τ is the relaxation time. Equation (1) applies to a model using a three-level time-stepping scheme because of the factor $2\Delta t$ on the right-hand side. The ECMWF reanalysis data (ERA-15) (Gibson et al., 1997) are used as the reference data towards which the model is relaxed. These data are available every 6 h in T106, L31 resolution. In order to use them for nudging they have to be truncated to the horizontal resolution used in the ARPEGE model (in this case T42), and interpolated in the vertical to the orography-adjusted hybrid levels. As the relaxation towards the reference data is done at every time step (in this case every 15 min) a cubic spline interpolation is used to obtain reference data at intermediate times.

The term $[\Psi^{\mathrm{REF}}(t + \Delta t) - \Psi^{*}(t + \Delta t)]/\tau$ is the estimate of the tendency residual, because it is the term ensuring that the evolution of the model is close to the observations. Note that only in the case $\tau = 2\Delta t$ is the term equal to the tendency residual, assuming the reference state is known at every model time step. For larger values of τ one gets only approximations to the initial tendency residuals. This also follows from eq. (1) as only in the case $\tau = 2\Delta t$ will the assimilation model follow the reference state exactly. In practice, when using the nudging technique the relaxation time τ has to be chosen carefully since the estimate of the tendency residual is dependent to a certain extent on τ. If τ is too small noise and spin-up problems play an important role, but if τ is too large the model does not follow the reference data set closely enough to obtain a reasonable estimate of the tendency residual. Normally, the same value of τ is not used for all the prognostic variables. (For a discussion of these problems see Jeuken et al. (1996).)

In order to obtain estimates of the tendency residuals an assimilation run was done for each of the 14 winters from 1979–1980 to 1992–1993. The assimilation was started on 1 November and ended on 31 March. The variables assimilated were temperature, vorticity, divergence, logarithm of surface pressure, surface temperature and surface soil moisture. The variables were nudged in spectral space, except for the two fields surface temperature and surface moisture for which the assimilation was done in grid point space. Δt was 15 min. The τ used was not the same for all the variables. The actual values were: 24 h for temperature, 6 h for vorticity, 48 h for divergence, 24 h for logarithm of surface pressure, 48 h for surface temperature and 48 h for surface soil moisture. The choice of these values for τ was decided after several test experiments. The test experiments were trial-and-error experiments where for selected 1-month periods the obtained mean monthly tendency residuals were reinjected into the model equations. The modified model was then run for the same 1-month periods and the monthly mean compared with the ERA-15 data for the month in question. The chosen set of τ values was the set from the collection of tested sets that led to the best reproduction of the ERA data.

It is reasonable to nudge divergence with a rather large value of τ since this field is rather noisy and not well observed, while it is important that the observed value of vorticity is followed closely during the assimilation, so for this variable a small value for τ was used. For all the prognostic variables gradually larger values for τ were used at the seven highest levels because at these levels the ERA data are not as reliable as in the rest of the atmosphere and because of the large dynamical variability in the stratosphere. For some months a very cold winter stratosphere developed when the same τ was used on all vertical levels in the 1-month test experiments mentioned above. The agreement with the ERA-15 data was much better in these cases if the level dependence of τ was introduced. The atmospheric humidity field was not nudged. The reason for this is that processes related to humidity can act on a timescale smaller than 6 h (the temporal resolution of the reference data set in this case) through threshold processes like condensation and deep convection. To avoid moisture spin-up problems it therefore makes more sense to let the model develop its own humidity field. This is especially important in the tropics where convection plays a large role.

Monthly averages of the tendency residuals determined by this method were used as correction terms, not as constant values for each month, but for each day a linear interpolated value obtained from the two closest monthly means was used. When adding these correction terms it turned out that a cold winter stratosphere developed. This problem was remedied by making yet another assimilation. In this second assimilation the monthly means of the tendency residuals from the first assimilation were used as correction terms in the model equations, as just described, and the temperature was then relaxed weakly towards the ERA-15 temperature. τ was 48 h, with gradually larger values on the seven highest levels as in the first assimilation. Also in this second assimilation it is important to nudge the highest levels weakly for the same reasons as for the first assimilation. The final estimate of the temperature residual used as the correction term in the experiments described below was then the sum of the residuals obtained in these two assimilations.

The question can be asked whether large temperature errors remain after the first assimilation, because the GCM used in generating the ERA-15 data set has itself strong temperature tendency errors. Figure 51 in Kållberg (1997) shows the analysis increments of zonal and temporal means of temperature—defined as the difference between the initialized analyses and the +6 h first guess forecasts with the GCM used for the ERA-15 data. The analysis increments show a warming in large parts of the troposphere with maxima at 0.2 K and a cooling in the Northern Hemisphere polar stratosphere with a maximum around 0.2 K in annual mean (which might be even larger in the winter mean). This means that there are errors in either the observed data or the model used for the ERA-15 data, but as explained in Kållberg (1997) this mismatch between the observed and predicted temperatures cannot be attributed unequivocally to either data or model problems; Kållberg (1997) mentions biases in the 1D-Var retrievals and systematic errors in the model parametrizations of radiative cooling and latent heat release as likely candidates for causing the mismatch.

Monthly averages of the residuals were obtained for each of the months November to March and in Fig. 1a the zonal mean of the temperature residual determined from the first assimilation averaged over the last 14 Januaries in the ERA-15 period is shown as a function of model levels. (Approximate pressure levels based on a reference surface pressure of 1013.22 hPa are added on the right vertical axis.) Figure 1b shows the contribution from the second assimilation to the zonal mean of the temperature residual and Fig. 1c shows the sum of the

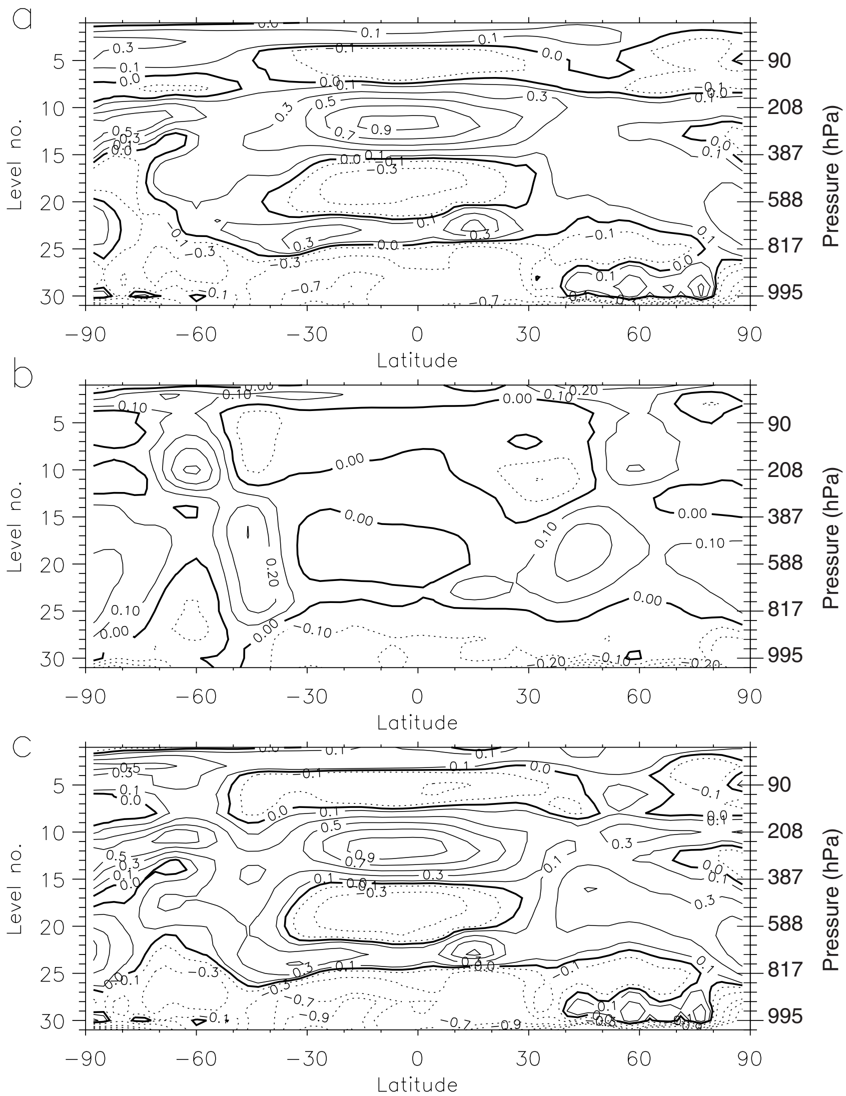

Fig. 1. Fourteen year average of the temperature residual for January (unit K d^{-1}) for ARPEGE, version 2. Negative contour lines are dashed. (a) Contribution to the zonal mean of the residual obtained by the first assimilation. Contour interval 0.2 K d^{-1}. (b) Contribution to the zonal mean of the residual obtained by the second assimilation. Contour interval 0.1 K d^{-1}. (c) Zonal mean of the total residual determined as the sum of the contributions from the two assimilations. Contour interval 0.2 K d^{-1}.

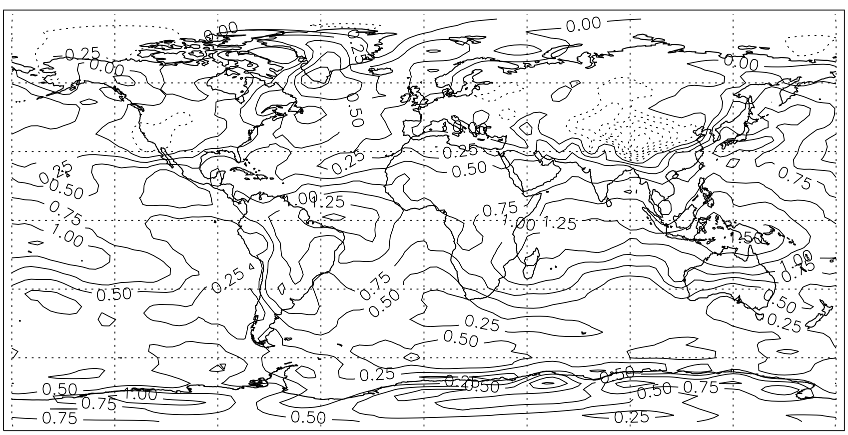

Fig. 2. Geographical distribution of the January mean of the total temperature residual at model level 12 (∼270 hPa). Negative contour lines are dashed.

contributions from the two assimilations. The horizontal distribution at vertical level number 12 (corresponding to approximately 270 hPa) of the January mean of the temperature residual (the sum of the contributions from the two assimilations) is shown in Fig. 2.

It is seen from Fig. 1a that the upper troposphere is dominated by positive values of the residual, meaning that the model needs to be heated in that area in order to follow the reference data set. The residual is negative almost everywhere in the lower troposphere. The contribution from the second assimilation (Fig. 1b) is generally small, but larger contributions are seen in the mid-latitude troposphere. Even though the contribution is small, it might be important as it reduces the stratospheric cold bias. The sum of the two contributions (Fig. 1c) shows the same pattern as the contribution from the first assimilation: the upper troposphere is dominated by positive values and the lower troposphere by negative values. From Fig. 2 it is seen that model forcing errors as large as 1.0 K d^{-1} and larger are evident in the South Pacific intertropical convergence zone (ITCZ) and the tropical Indian and Atlantic ocean, areas with large convective activity. The large forcing errors in these areas could be due to the rather coarse resolution of the model used, meaning that the processes involved are not well resolved, or to an inadequate parametrization of convection or radiation or cloudiness.

When relaxing the model towards the observations it is done for all scales in the T42 resolution model, and since time interpolation of the observed data is performed because the observed data are only available every 6 h this means that there could be a contribution from small-scale features generated by the model's physics that are not in the observations. But as the correction terms used are monthly means averaged over 14 yr, and we assume that these inconsistencies are random, the contribution from these is likely to be negligible.

3. Experiments

A set of seasonal prediction experiments has been performed with the standard version of the model and with the empirically corrected version. The empirically corrected version is defined by adding a correction term based on monthly means of the tendency residuals averaged over the 14 yr used for obtaining these tendency residuals. As mentioned, instead of using a constant correction term for each month, for each day a linear interpolated value obtained from the two closest monthly means is used. Nine-member ensemble forecasts were made for each of the 14 boreal winters from 1979–1980 to 1992–1993. The experimental set-up used here is similar to the set-up defined in the PROVOST project (Palmer et al., 2000): The nine forecasts were initiated on the 22, 23, . . . , 30 November, respectively and ended at the end of February. Observed SSTs from the ERA-15 data were used in these runs. The results of the experiments are compared with the ERA-15 data for the 14 winters.

4. Results

4.1. Systematic errors

The long-term mean systematic errors shown in this section are calculated as the December–January–February (DJF) mean averaged over the 14 boreal winters from 1979–1980 to 1992–1993, and averaged over the nine members of the ensembles, minus the DJF average of the ERA-15 climatology determined for the same period.

Figure 3 shows the systematic errors for the fields' zonal mean of temperature (Figs. 3a and 3b) and zonal mean of zonal wind (Figs. 3c and 3d). For the zonal mean temperature it is seen that the control model (the standard ARPEGE, version 2) has small biases in the lower troposphere whereas in the upper troposphere and in the stratosphere there are large biases. The model is too warm in the lower stratosphere in the tropics and in the rest of the stratosphere it is too cold. In the upper troposphere the model is too cold—in this area an extra positive forcing is added when the model is run in corrected mode (as seen in Fig. 2 level 10–15). The forced model (the empirically corrected version of ARPEGE, version 2) exhibits an overall reduction in systematic errors of the zonal mean temperature. The model is still somewhat too cold in the stratosphere at the North Pole but the systematic errors are generally very small compared with the control model. For the zonal mean of the zonal wind the picture is similar. The large biases in the upper troposphere, and especially in the stratosphere, in the control model, are substantially reduced in the forced model.

Figure 4 shows the systematic errors of the 500 hPa height field for the control and the forced model. In the control model the largest errors are seen over the North Atlantic–European area with negative values and over the North Pacific area with a dipole structure. This error pattern is typical for many GCMs, reflecting that the model is too zonal south of the error minima. In the Southern Hemisphere the largest errors are seen over Antarctica and over the mid-latitude ocean. The 500 hPa geopotential height of the forced model is in much better agreement with the observed data—a substantial reduction of the systematic errors is seen almost everywhere.

Table 1 lists the root mean square (RMS) of the systematic errors averaged over the Northern and Southern Hemispheres for mean sea level pressure, 500 hPa geopotential height and 850 hPa temperature for the control model and for the forced model. In the Northern Hemisphere a reduction of more than 50% of the error is seen for mean sea level pressure and 500 hPa geopotential height. A substantial improvement is also seen in the Southern Hemisphere.

Since the prognostic equation for humidity is not corrected, it is important to make sure that the hydrological cycle is kept reasonable in the corrected model version, and it turns out that for the precipitation field the pattern and the size of the systematic errors are quite similar for the two model versions (not shown).

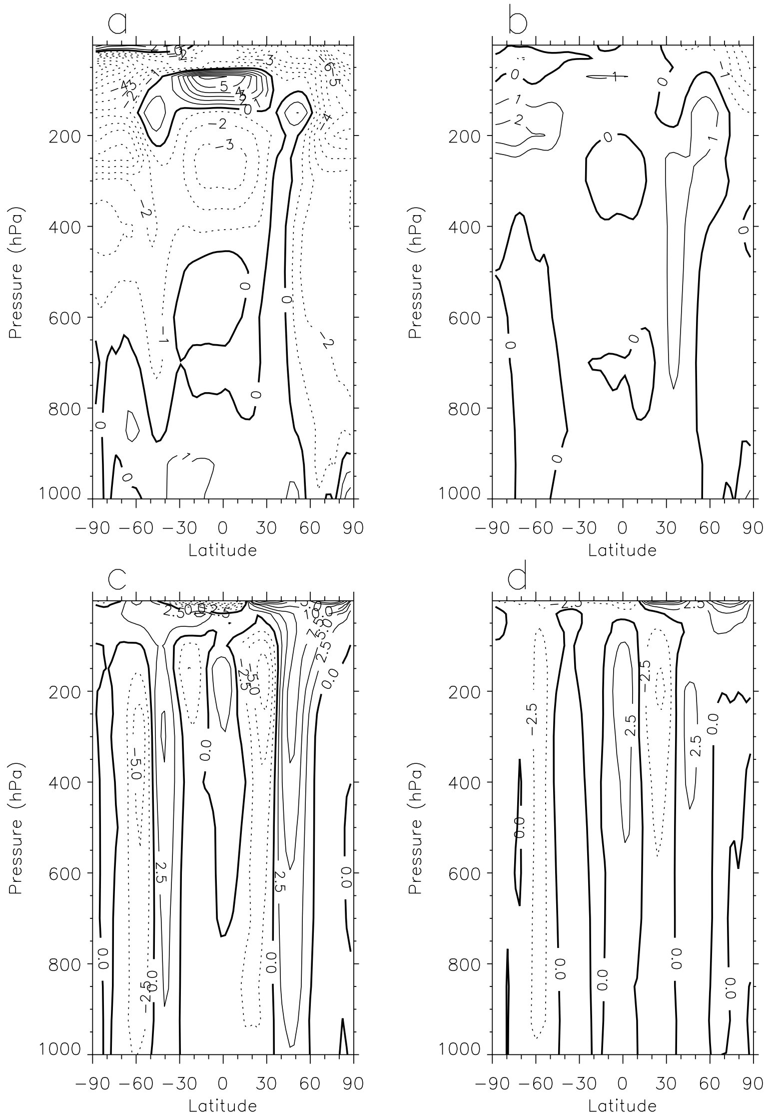

Fig. 3. Mean winter (DJF) systematic error of (a) zonal mean temperature for the control model (unit, K), (b) zonal mean temperature for the empirically corrected model (unit, K), (c) zonal mean of the zonal wind for the control model (unit, m s^{-1}) and (d) zonal mean of the zonal wind for the empirically corrected model (unit, m s^{-1}).

As pointed out in Branković and Palmer (2000), the model error should be less than the signal that is to be predicted, which in this case means that the model error should be less than the magnitude of the observed interannual variability. Figure 5 shows the ratio of the magnitude of the systematic error of the 500 hPa height and the corresponding standard deviation of the analysis anomalies for the winter season for the period 1979–1980 to 1992–1993 for the control and the forced model. This relative error has reduced considerably in the corrected model and is less than 1 almost everywhere except over Antarctica, where the two models are comparable.

4.2. Variability

It is important when forcing the model with empirically determined forcing terms as described above that the variability of the model is not destroyed and therefore both the high- and low-frequency variability of the control and of the forced model are determined. Figure 6 shows the systematic error of the standard deviation of the band-pass filtered (2.5–6 d) 500 hPa height field for the control and the forced model. It is seen that in general the high-frequency variability of the forced model is in better agreement with the reanalysis data than the control model; in particular the Northern Hemisphere storm tracks are more realistic in the forced model. Only in a very few areas is the high-frequency variability better represented in the control model. Figure 7 shows the systematic error of the standard deviation of the low-pass filtered (>10 d) 500 hPa height field and also in the case of low-frequency variability there is a better agreement between the forced model and the ERA-15 data than between the control model and the ERA-15 data almost everywhere. Exceptions are the northernmost parts of the Greenland Sea and Barents Sea

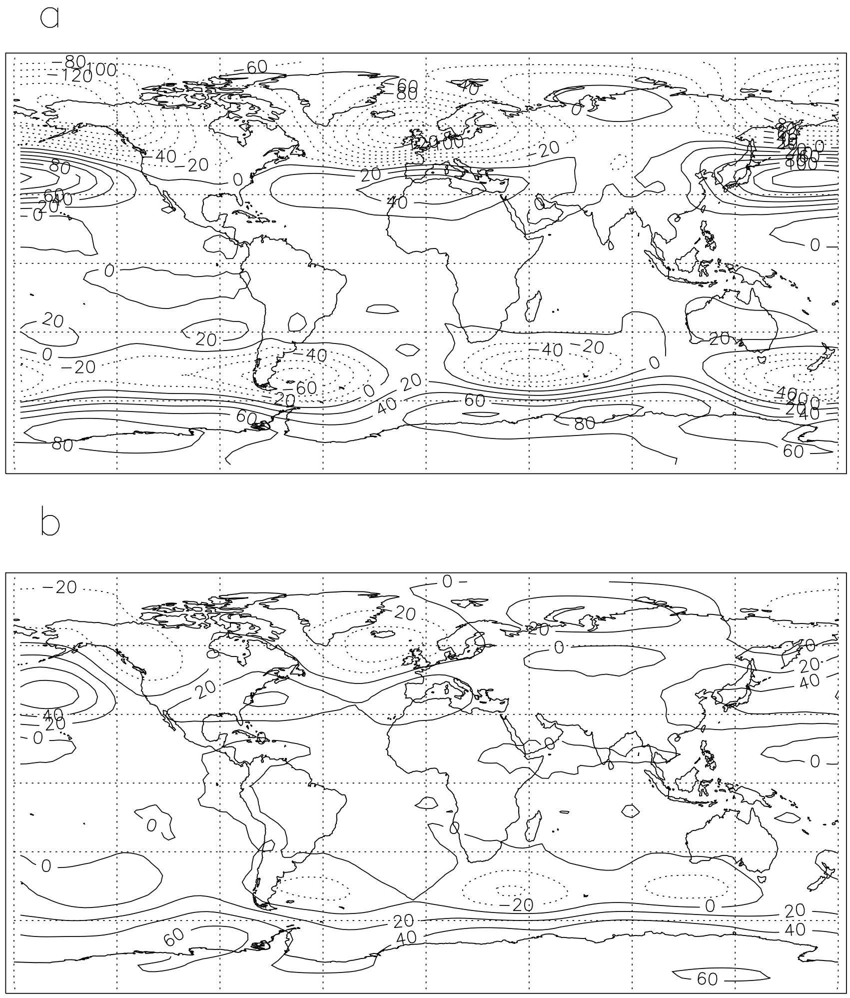

Fig. 4. Mean winter (DJF) systematic error of the 500 hPa height (unit, m) for (a) the control model and (b) the empirically corrected model.

where the low-frequency variability has become even higher, and the southernmost part of the Indian Ocean where the low-frequency variability is even less in the forced model than in the control model.

Table 1. RMS of systematic errors for mean sea level pressure (MSLP), 500 hPa height field (500 Z) and 850 hPa temperature (850 T) as averages over the Northern (NH) and Southern Hemisphere (SH) (units noted in brackets). Results are shown for the control and the empirically corrected version of ARPEGE, version 2.

Field	Area	Control	Forced
MSLP (hPa)	NH	4.0	1.4
MSLP (hPa)	SH	3.1	2.6
500 Z (m)	NH	48.0	19.5
500 Z (m)	SH	30.6	19.1
850 T (K)	NH	1.4	0.8
850 T (K)	SH	1.0	0.5

In conclusion it can be noted that as the ability of the models to make seasonal predictions is studied a realistic description of the variability is important, and in general the description of the variability on the different scales is in better agreement with observations in the empirically forced model than in the control model.

4.3. *Forecast skill*

Anomaly correlation is used as a measure of the ability of the two model versions to make seasonal predictions. Figures 8 and 9 show the geographical distribution of the temporal correlation, over the 14 yr covered by the available data, between model and observed winter season anomalies for two fields. For each of the two models (the control and the forced) the ensemble mean is used in the calculation of the correlations. As the predictions for the DJF season have been made from initial conditions on 22, 23, . . . , 30 November they are zero-lead predictions, and therefore the skill of the models is influenced by

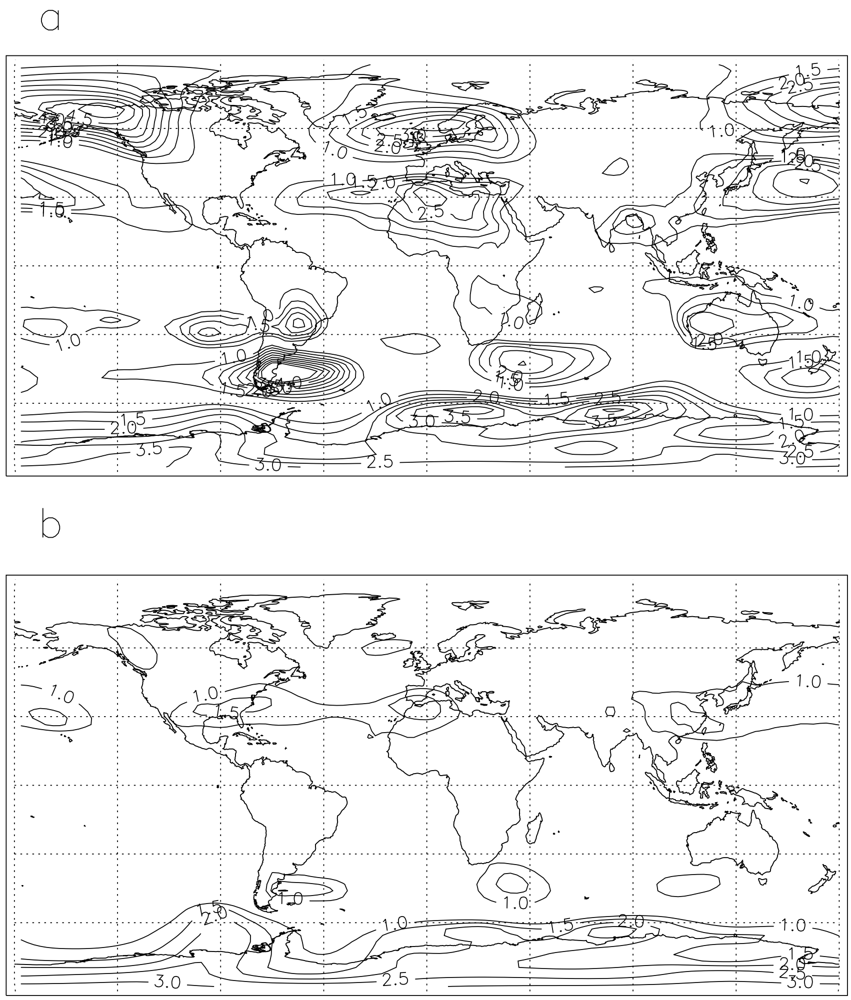

Fig. 5. Mean winter (DJF) ratio of the magnitude of the systematic error of the 500 hPa height and the standard deviation of ERA-15 anomalies for the 500 hPa height for (a) the control model and (b) the empirically corrected model.

both lower boundary forcing (observed SSTs were used) and initial conditions. As observed SSTs were used and as the estimate of the tendency error for the season in question is included in the correcting term used in the forced model, these seasonal prediction experiments are so-called potential predictability experiments.

Figures 8a and b show the temporal correlation between the anomalies of the control run and the ERA-15 data and the forced run and the ERA-15 data respectively, for mean sea level pressure. For both models the well known picture of relatively high correlations in the tropics and smaller and even negative correlations in the extratropics is seen. Comparing the results for the empirically forced model with the results for the control model, it is seen that there is a general increase in the correlation in the Southern Hemisphere for the forced model. Only in very few areas is the correlation larger for the control model than for the forced model. In the Northern Hemisphere, on the other hand, there is no general improvement. In some areas, like most of the Atlantic–European area, the forced model has the highest correlation, whereas in other areas, for example Central Asia, the correlation of the control model is the highest. Correlations larger than 0.53 are significant at the 5% level or better in a Student's *t*-test.

Figure 9 is similar to Fig. 8, but here correlations for the 850 hPa temperature field are shown. In this case the correlation pattern is somewhat more noisy, but again some general improvement is seen in the Southern Hemisphere when the forced model is compared with the control model.

In Table 2 the average of the time correlations for the Northern Hemisphere and for the Southern Hemisphere is listed for each of the two model versions for the fields of mean sea level pressure, 500 hPa height and 850 hPa temperature. For all of these three fields the average correlation for the Southern Hemisphere is larger for the forced model than for the control model. In the Northern Hemisphere the two model versions are comparable.

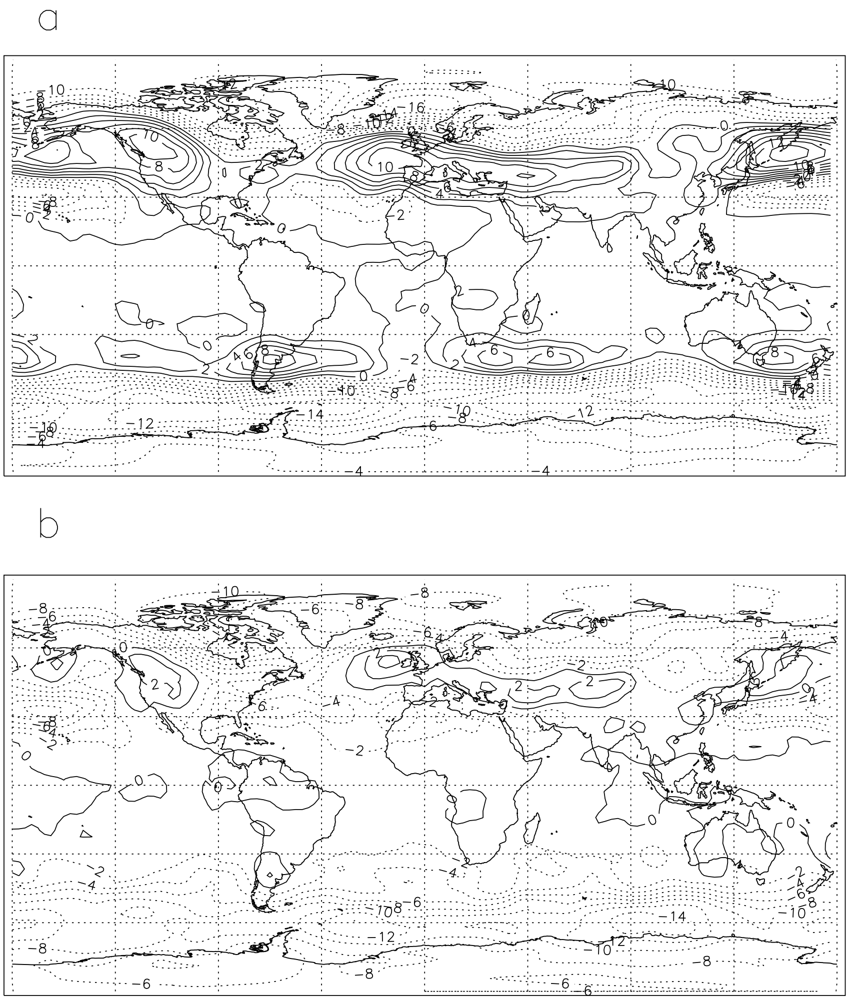

Fig. 6. Mean winter (DJF) systematic error of the standard deviation of the band-pass filtered (2.5–6 d) 500 hPa height (unit, m) for (a) the control model and (b) the empirically corrected model.

Figures 10 and 11 show pattern correlations for the winter season for each year in the period covered by these experiments for the temperature at 850 hPa and the 500 hPa height field, respectively. Correlations are shown for the following areas: Northern Hemisphere (20°N–80°N), Southern Hemisphere (80°S–20°S), Europe (35°N–75°N, 13°W–43°E) and the tropics (30°S–30°N). For each field and for each area the figures show the correlation between the control model and the ERA-15 data (full lines), the forced model and the ERA-15 data (dotted lines) and the correlation between the two models (dashed lines). The correlations between the two models and the ERA-15 data are characterized by rather high interannual variability for both fields and for all areas. For the tropical area, where seasonal predictions are largely influenced by the SSTs, the two models give almost identical results and the correlation between the two models is very high. For the European area the results for the two models are quite different, but it cannot be concluded that one model is better than the other. For the Northern Hemisphere as a whole there is quite large agreement between the two models and the correlation between the two models is relatively high. The results for the Southern Hemisphere deviate from the results for the other areas. Except for a few years where the correlation between the two models and the ERA-15 data is very similar, the correlation between the forced model and the ERA-15 data is higher than the correlation between the control model and the ERA-15 data. The Southern Hemisphere is also characterized by a high correlation between the two models.

So as already seen for the temporal correlation, in the case of pattern correlation the results obtained with the empirically corrected model are also improved in the Southern Hemisphere as compared with the results obtained with the control model.

Note, that although the forced version of the model is improved considerably with regard to both systematic errors and variability compared with the control model, the anomaly correlation between the two models is still high and higher than the correlation with the ERA-15 data for either of the models.

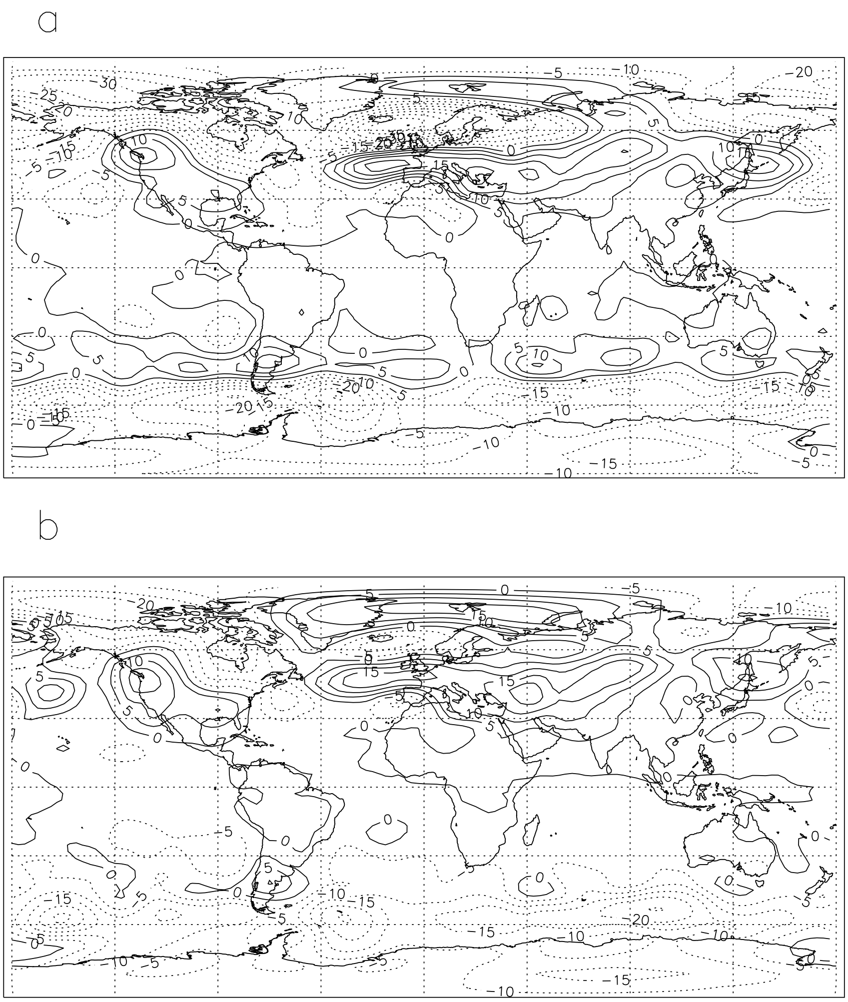

Fig. 7. Mean winter (DJF) systematic error of the standard deviation of the low-pass filtered (>10 d) 500 hPa height (unit, m) for (a) the control model and (b) the empirically corrected model.

5. Discussion and conclusion

The general circulation model ARPEGE, version 2 has been used in a study of the impact of model systematic errors on the skill of ensemble forecasts for seasonal predictions. The model has been modified by including empirically obtained forcing terms, and this corrected version of the model has been compared with the standard ARPEGE. The forcing terms were determined as estimates of the tendency errors of the model by using the nudging technique to assimilate the ECMWF reanalysis data (ERA-15).

The empirically corrected model has small systematic errors and they are reduced considerably compared with the systematic errors of the standard model. Also the variability on different scales of the corrected model is in better agreement with observations than the variability of the standard model.

Calculations of temporal and pattern correlations of anomalies show that the corrected model with an improved description of the background flow has a somewhat higher seasonal forecast skill in the Southern Hemisphere in general, whereas there is no general improvement in the Northern Hemisphere, although the relative reductions of systematic errors are largest in the Northern Hemisphere.

It should be emphasized that in the procedure for correcting the model only the most simple linear approach has been followed by using the 3-D climatology of tendency errors. This climatology of tendency errors is independent of any time variations such as the El Niño–Southern Oscillation (ENSO) cycle. Situations where the spatial structure of the tendency errors reflect, for example, the spatial structure of the tropical deep convection, which for example in the Pacific depends strongly on the ENSO cycle, may not be properly accounted for and could cause problems when attempting to make seasonal predictions for El Niño and La Niña years. A more sophisticated, non-linear approach, such as the flow-dependent correction method used by D'Andrea and Vautard (2000), is needed to take these problems into account.

As the systematic errors of the corrected model are indeed reduced considerably compared with the standard version, it seems likely that in order to obtain better skill scores by improving the

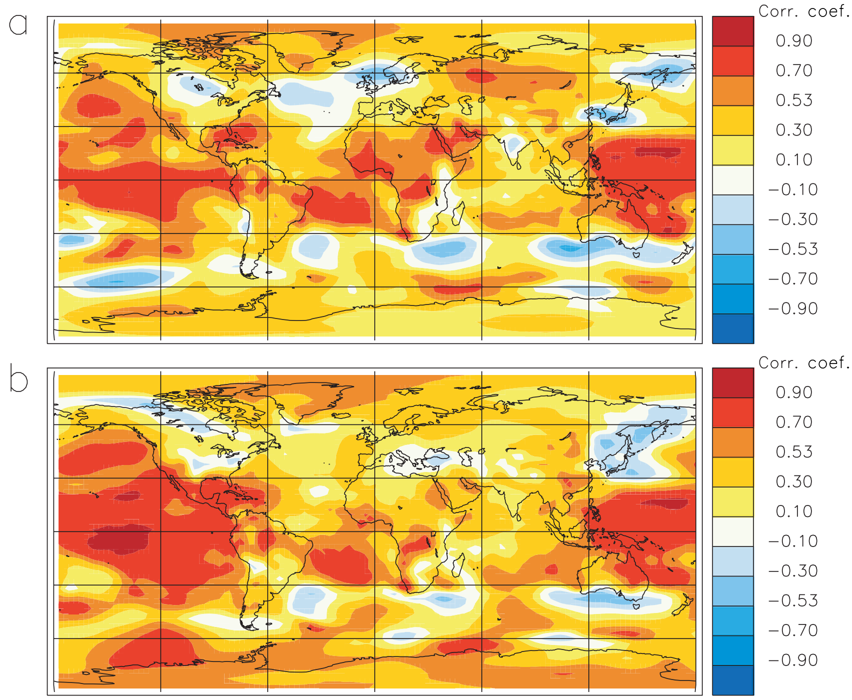

Fig. 8. Temporal anomaly correlation between ERA-15 winter (DJF) average and nine-member ensemble forecasts for mean sea level pressure for (a) the control model and (b) the empirically corrected model. Correlations larger than 0.53 are significant.

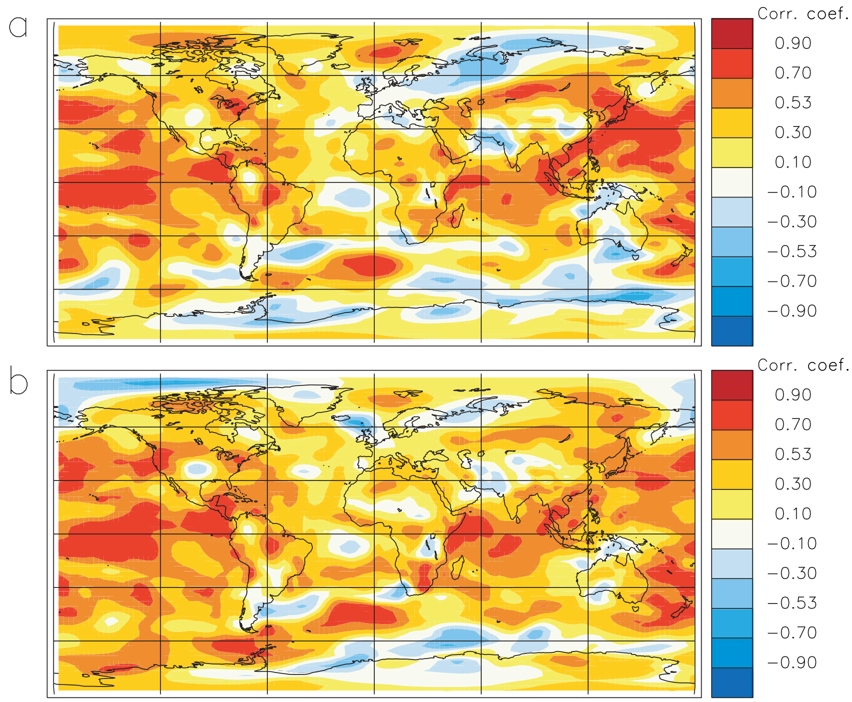

Fig. 9. Temporal anomaly correlation between ERA-15 winter (DJF) average and nine-member ensemble forecasts for the 850 hPa temperature for (a) the control model and (b) the empirically corrected model. Correlations larger than 0.53 are significant.

model along the lines described in this paper a more realistic description of the model variability is needed—although the variability is already better described in the corrected model version. It is likely that this could be obtained by correcting the model by adding a flow-dependent parametrization of the initial tendency errors as demonstrated by D'Andrea and Vautard (2000) for a quasi-geostrophic model.

Yang and Anderson (2000) have corrected a coupled ocean–atmosphere GCM by obtaining the systematic initial tendency error for ocean temperature and using this tendency error as the

Table 2. Temporal anomaly correlations for mean sea level pressure (MSLP), 500 hPa height field (500 Z) and 850 hPa temperature (850 T) as averages over the Northern (NH) and Southern Hemisphere (SH). Results are shown for the control and the forced version of ARPEGE, version 2.

Field	Area	Control	Forced
MSLP	NH	0.43	0.43
MSLP	SH	0.37	0.48
500 Z	NH	0.48	0.47
500 Z	SH	0.42	0.51
850 T	NH	0.40	0.42
850 T	SH	0.31	0.41

correction term. The corrected model is used for studying ENSO forecast skill and some improvement has been achieved. Yang and Anderson (2000) suggest that further improvement could be obtained by correcting not only ocean temperature but all the prognostic variables in the coupled system and add that such a procedure is technically more difficult as it is necessary to impose balance requirements when correcting all the variables. Using long-term mean values as determined here by the nudging technique is one way of determining balanced initial tendency errors.

The experiments made with ARPEGE, version 2 show evidence that an improved description of the background flow leads to higher skill of the model when used for seasonal predictions. On the other hand the improvement is not striking in spite of

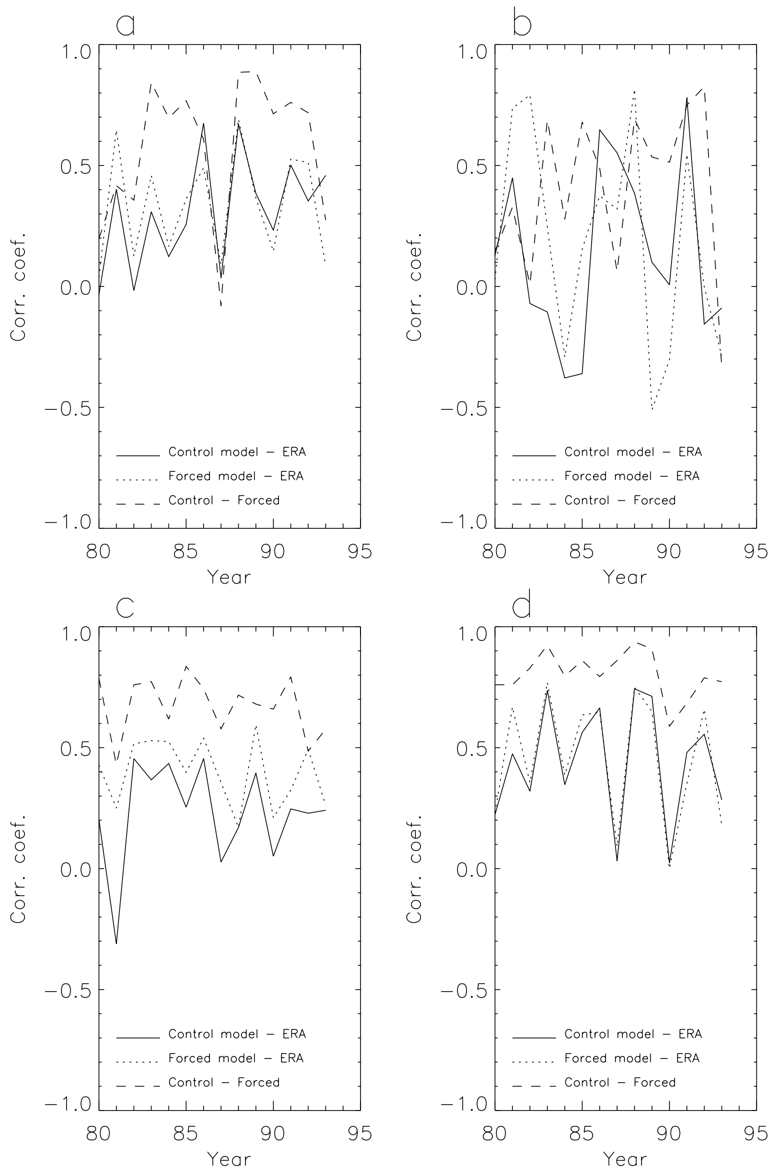

Fig. 10. Pattern anomaly correlation between ERA-15 winter (DJF) average and nine-member ensemble forecasts for the years 1979–1980 to 1992–1993 for 850 hPa temperature. Full line, correlation between control model and ERA-15 data; dotted line, correlation between the empirically corrected model and ERA-15 data; dashed line, correlation between control model and the empirically corrected model. The correlations are shown for different areas: (a) the Northern Hemisphere, (b) Europe, (c) the Southern Hemisphere, (d) the tropics.

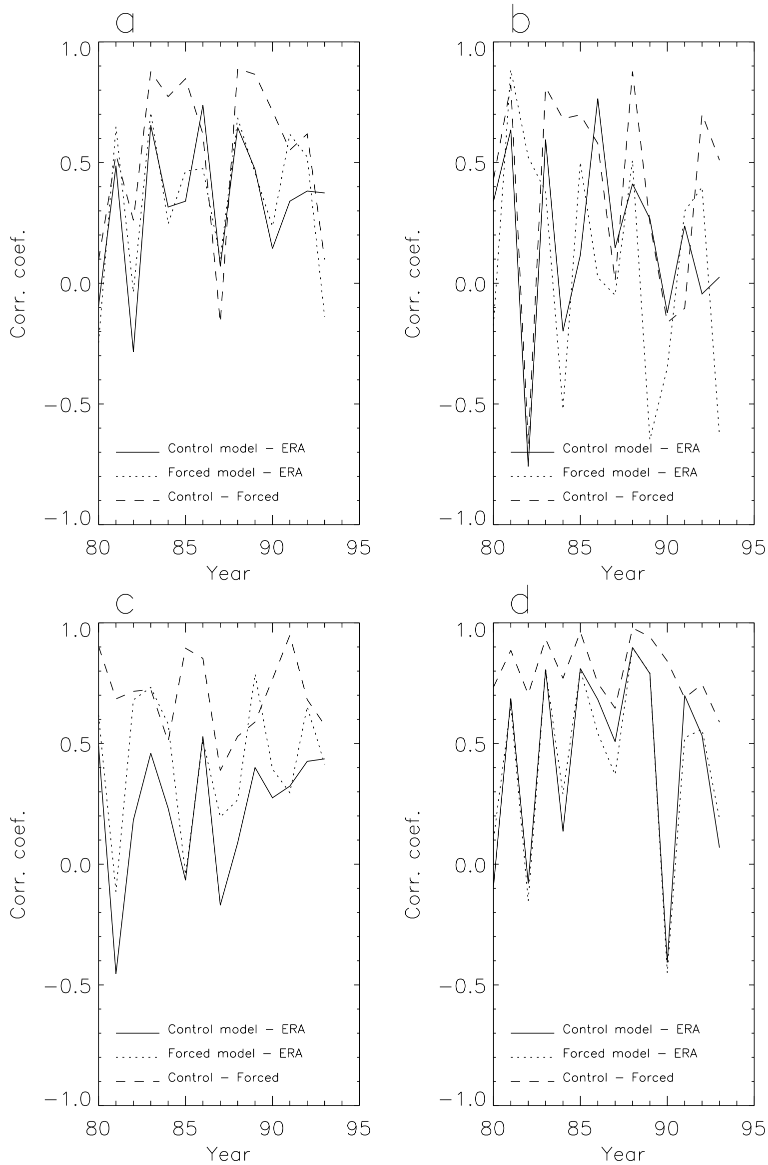

Fig. 11. Same as Fig. 10, except that results are shown for the 500 hPa height.

a substantial reduction in the systematic errors. As mentioned above, it is likely that a further improved description of the model variability by the use of flow-dependent forcing terms (D'Andrea and Vautard, 2000) is needed in order to see a larger influence on the seasonal prediction skill scores. Also more sophisticated techniques for obtaining initial tendency errors, like the methods developed by Machenhauer and Kirchner (2000) and D'Andrea and Vautard (2000), may lead to improvements. The experience from the experiments described here is that a good description of the stratosphere is essential, and that an even better described model stratosphere might be of importance.

There is a general improvement of predictive skill in the Southern Hemisphere but not in the Northern Hemisphere. One reason for this could be that the low-frequency variability in general is higher in the Northern Hemisphere and that low-frequency noise makes is difficult to improve skill in the Northern Hemisphere. Despite the large reductions in systematic errors in the northern extratropics there is no improvement in predictive skill. This could suggest that improving model climatology is simply not sufficient to improve model predictability for the northern extratropics, and it may not be possible due to chaos.

It should be mentioned that the seasonal prediction experiments have been performed for the same period as was used for obtaining the tendency errors, showing the potential of the method. The next step would be to make the seasonal prediction experiments for a different time period using independent data.

The purpose of this study has been to investigate whether reduced long-term model systematic errors would lead to improved skill when using the model for seasonal prediction. An empirical correction method has been used to obtain a model with reduced systematic errors, but it should be emphasized that in general model improvements in terms of, for example, better parametrization schemes are preferable to empirical corrections like the one used here.

6. Acknowledgments

The many beneficial discussions with Bennert Machenhauer and his valuable comments on the manuscript are highly appreciated. The work was financially supported by the European Commission's 4th Framework Programme, "Environment and Climate Research Programme" under contract number ENV4-CT97-0497 ("POTENTIALS").

References

Branković, Č. and Palmer, T. N. 2000. Seasonal skill and predictability of ECMWF PROVOST ensembles. *Q. J. R. Meteorol. Soc.* **126**, 2035–2067.

Branstator, G. 1983. Horizontal energy propagation in a barotropic atmosphere with meridional and zonal structure. *J. Atmos. Sci.* **40**, 1689–1707.

D'Andrea, F. and Vautard, R. 2000. Reducing systematic errors by empirically correcting model errors. *Tellus* **52A**, 21–41.

Déqué, M. 1997. Ensemble size for numerical seasonal forecasts. *Tellus* **49A**, 74–86.

Déqué, M., Dreveton, C., Braun, A. and Cariolle, D. 1994. The ARPEGE/IFS atmosphere model: a contribution to the French community climate modelling. *Clim. Dyn.* **10**, 249–266.

Derome, J., Brunet, G., Plante, A., Gagnon, N., Boer, G. J. and co-authors 2001. Seasonal predictions based on two dynamical models. *Atmos. Ocean* **39**, 485–501.

Eliasen, E., Machenhauer, B. and Rasmussen, E. 1970. *On a Numerical Method for Integration of the Hydrodynamical Equations with a Spectral Representation of the Horizontal Fields*. Report no 2, Institute for Theoretical Meteorology, Copenhagen University, Copenhagen.

Gibson, J. K., Kållberg, P., Uppala, S., Hernandez, A., Nomura, A. and co-author 1997. *ERA-15 Description*. ECMWF Re-analysis Project Report Series, Volume 1. ECMWF, Reading.

Jeuken, A. B. M., Siegmund, P. C., Heijboer, L. C., Feichter, J. and Bengtsson, L. 1996. On the potential of assimilating meteorological analyses in a global climate model for the purpose of model validation. *J. Geophys. Res.* **101**, 16 939–16 950.

Kållberg, P., 1997. *Aspects of the Re-Analysed Climate*. ECMWF Reanalysis Project Report Series, Volume 2. ECMWF, Reading.

Kaas, E., Guldberg, A., May, W. and Déqué, M. 1999. Using tendency errors to tune the parameterisation of unresolved dynamical scale interactions in atmospheric general circulation models. *Tellus* **51A**, 612–629.

Kharin, V. V. and Zwiers, F. W. 2001. Skill as a function of time scale in ensembles of seasonal hindcasts. *Clim. Dyn.* **17**, 127–141.

Kharin, V. V., Zwiers, F. W. and Gagnon, N. 2001. Skill of seasonal hindcasts as a function of the ensemble size. *Clim. Dyn.* **17**, 835–843.

Klinker, E. and Sardeshmukh, P. D. 1992. The diagnosis of mechanical dissipation in the atmosphere from large-scale balance requirements. *J. Atmos. Sci.* **49**, 608–627.

Kumar, A. and Hoerling, M. P. 2000. Analysis of a conceptual model of seasonal climate variability and implications for seasonal prediction. *Bull. Am. Meteorol. Soc.* **81**, 255–264.

Kumar, A., Hoerling, M. P., Ji, M., Leetmaa, A. and Sardeshmukh, P. 1996. Assessing a GCM's suitability for making seasonal predictions. *J. Climate* **9**, 115–129.

Machenhauer, B. and Kirchner, I. 2000. Diagnosis of systematic initial tendency errors in the ECHAM AGCM using slow normal mode data-assimilation of ECMWF reanalysis data. *CLIVAR Exchanges*, no. 18, 9–10. International CLIVAR Project Office, Southampton.

Palmer, T. N., Branković, Č. and Richardson, D. S. 2000. A probability and decision-model analysis of PROVOST seasonal multimodel ensemble integrations. *Q. J. R. Meteorol. Soc.* **126**, 2013–2033.

Peng, S., Robinson, W. A. and Hoerling, M. P. 1997. The modeled atmospheric response to midlatitude SST anomalies and its dependence on background circulation states. *J. Climate* **10**, 971–987.

Peng, S. and Whitaker, J. S. 1999. Mechanisms determining the atmospheric response to midlatitude SST anomalies. *J. Climate* **12**, 1393–1408.

Sausen, R. and Ponater, M. 1990. Reducing the initial drift of a GCM. *Beitr. Phys. Atmosph.* **63**, 15–24.

Schubert, S. and Chang, Y. 1996. An objective method for inferring sources of model error. *Mon. Weather Rev.* **124**, 325–340.

Vitart, F. 2004. Monthly forecasting. *ECMWF Newsletter* **100**, 3–13.

Williamson, D. L. 2002. Time-split versus process-split coupling of parameterizations and dynamical core. *Mon. Weather Rev.* **130**, 2024–2041.

Yang, X. and Anderson, J. L. 2000. Correction of systematic errors in coupled GCM forecasts. *J. Climate* **13**, 2072–2085.