Reduction of Model Systematic Error by Statistical Correction for Dynamical Seasonal Predictions

HENRIK FEDDERSEN

Danish Meteorological Institute, Copenhagen, Denmark

ANTONIO NAVARRA

IMGA–CNR, Bologna, Italy

M. NEIL WARD

CIMMS, University of Oklahoma, Norman, Oklahoma

(Manuscript received 16 February 1998, in final form 18 August 1998)

ABSTRACT

Singular value decomposition analysis (SVDA) is used to analyze an ensemble of three 34-yr general circulation model (GCM) simulations forced with observed sea surface temperature. It is demonstrated how statistical postprocessing based on the leading SVDA modes of simulated and observed fields, primarily precipitation, can be applied to improve the skill of the simulation. For a given limited prediction region, the GCM has the potential to nonlinearly transform the SST information from around the globe and produce a dynamic solution over the region that can be statistically corrected to account for such features as systematic shifts in the location of anomaly dipoles. This paper does not address the separate question of whether the current generation of GCMs contain information above that which could be extracted using linear statistical relationships with SST.

For precipitation, examples are drawn from skillful tropical regions, as well as the moderate-to-low skill Pacific–North American and North Atlantic–European regions. Skill averaged across the analysis domain, as measured by the mean anomaly correlation, is notably improved by the statistical postprocessing in almost all situations where there is at least some real skill in the raw model fields. Postprocessing based on leading canonical correlation analysis (CCA) modes has been compared to postprocessing based on leading SVDA modes. The two methods show small differences, but neither one of the methods can be claimed to do better than the other. A third method, which is based on the leading empirical orthogonal functions of the simulations, has been tested on examples of tropical rainfall where it is shown to also be successful, but with skill generally a little below that based on SVDA or CCA modes.

The statistical postprocessing appears to have the greatest potential to improve skill for a variable like precipitation, which can have particularly strong anomaly gradients. Application of the postprocessing to large-scale atmospheric fields of 500-hPa geopotential height and sea level pressure produced smaller skill improvements relative to the skill of the raw model output.

1. Introduction

Multivariate statistical techniques such as singular value decomposition analysis (SVDA; Bretherton et al. 1992) and canonical correlation analysis (CCA; Barnett and Preisendorfer 1987) have over the last decade become very popular ways to extract information about dominant coupled modes in the climate system. A typical application has been analysis of links between observed sea surface temperature (SST) and atmospheric large-scale circulation (Wallace et al. 1992; Zhang et al. 1996; Peng and Fyfe 1996). Both SVDA and CCA give as ''output'' sequences of coupled modes that can be used to form the basis of an approximate expansion of the original time series. The coupled modes are ranked according to the relative covariance or correlation between the expansion coefficients, similar to the way in which EOFs of a time series of one anomaly field are ranked according to the relative amount of explained variance.

The leading patterns from analyses of pairs of observed SST and observed atmospheric fields have been compared to the leading patterns from analyses of simulations by atmospheric general circulation models forced by observed SST in order to study how the at-

Corresponding author address: Dr. Henrik Feddersen, Danish Meteorological Institute, Lyngbyvej 100, DK-2100 Copenhagen, Denmark.
E-mail: hf@dmi.dk

mospheric circulation is related to SST (e.g., Lau and Nath 1994). Barnett and Preisendorfer (1987) showed how leading CCA patterns could be used to formulate a set of predictive equations so that one field could be specified from the other. By applying CCA to time-lagged fields of observed monthly SST, sea level pressure, and United States surface temperature, they showed how to make predictions of monthly U.S. surface temperature. The same ideas have been used extensively by Barnston (1994) and Shabbar and Barnston (1996) to make seasonal predictions of surface temperature and precipitation for North America, and Barnston and Smith (1996) applied CCA to predict regional surface temperature and precipitation from global SST.

A different approach has been taken by Ward and Navarra (1997) who applied SVDA to simultaneous fields of GCM simulated precipitation and observed precipitation, respectively, in order to extract information about errors in the model response to SST forcing. Graham et al. (1994) also noted the potential of applying coupled pattern techniques to GCM integartions. One of the problems when assessing the skill of a model simulation is that the model is likely to geographically shift regions of significant variability and even a slight error of this type can result in a substantial drop in skill scores when skill is based on the perfomance at individual grid boxes. The results for SVDA applied to seasonal rainfall by Ward and Navarra (1997) were more promising than those of Renwick and Wallace (1995) for weather forecasts, where discrepancies between model and observed SVDA modes were generally small.

This paper aims to further develop SVDA of simulated versus observed fields on interannual timescales and use the ideas of Barnett and Preisendorfer (1987) to make a ''prediction'' of the observed field from the corresponding simulated field; that is, we use the leading SVDA modes to calculate an adjustment of the GCM simulation for a limited region. The methodology is demonstrated through examples that show the results of SVDA and subsequent adjustment of precipitation fields for both high-skill, tropical regions and moderate-to-low skill, extratropical regions (North America and Europe). For the latter regions, we also apply the analyses and adjustments to fields of surface temperature, 500-hPa geopotential height, and mean sea level pressure. All analyses are repeated using CCA instead of SVDA in order to provide a comparison between the two methods. A simpler alternative is to replace the SVDA (or CCA) by EOF decompositions of the simulated and observed fields, respectively, and to base the adjustment of the GCM simulation on the leading EOF modes. This is compared to both SVDA- and CCA-based adjustments for precipitation in tropical regions. The cross-validated skill of the adjusted fields is quantified using Linear Error in Probability Space (LEPS) skill scores (Ward and Folland 1991; Potts et al. 1996) and mean anomaly correlations (Déqué and Royer 1992) and the skill is compared to the skill of the direct model output.

If atmospheric anomaly fields (such as precipitation) were linearly related to the SST forcing, we would simply recover this linear relation by applying the SVDA or CCA adjustments to the model simulations and there would be no need for the GCM. However, the conceptual model we have is that the GCM nonlinearly transforms the SST information from around the globe and produces a set of solutions over a particular region. When these solutions are systematically biased in their spatial anomaly patterns, for example, by geographically shifting a rainfall anomaly dipole relative to observations, we correct the systematic errors using SVDA (or CCA). However, it is beyond the scope of this paper to assess the different question of whether this particular GCM actually contains any skill that is additional to that achievable from simple linear statistical relationships between the predictand field and SST.

As pointed out by Cherry (1996), SVDA and CCA modes are likely to include an undesired fit to random variations in the analyzed fields, especially for fields with many spatial degrees of freedom, which results in high correlations between the leading mode time series—correlations that do not necessarily reflect the actual relation between the two fields. In order to obtain more realistic time series correlations, we apply a jackknife procedure by which SVDA or CCA is repeatedly applied to fields from which data from one year is excluded. This procedure is described in detail in the appendix.

Following a description of the data and model simulations in section 2, examples of discrepancies between observed and model modes in the northeastern Brazil and Pacific–North American (PNA) sectors are given in section 3. Section 4 describes the postprocessing methodologies to generate improvements upon the direct GCM output, while section 5 reviews the skill improvements that result from the SVDA-, CCA-, and EOF-based techniques. A discussion of the results and conclusions are given in section 6.

2. Data

The simulations are generated by the atmospheric general circulation model ECHAM4 (Roeckner et al. 1996). An ensemble of three integrations (horizontal resolution T30, 19 vertical levels) has been analyzed. Each integration was started from different initial atmospheric conditions, but all were forced by identical, observed SST (GISST2.2 from the Hadley Centre), over the 34-yr period 1961–94 (Moron et al. 1998).

Observations compose globally gridded ($5° \times 5°$) monthly means of 500-hPa geopotential height and mean sea level pressure for the period 1964–89 (Oort 1983), globally gridded ($3.75° \times 2.5°$) monthly precipitation [land only; Hulme (1994)] for the period 1961–94, and globally gridded ($5° \times 5°$) near-surface air temperature [land only; Jones et al. (1986)] also for 1961–94.

In order to reduce the skewness of precipitation distributions, we analyze everywhere the square root of precipitation rather than precipitation itself (Stephenson et al. 1999).

Rather than trying to fill in gaps in observed precipitation or surface temperature using some kind of interpolation, a grid box is completely omitted from the analysis if more than 25% of the time series of observations is missing for that grid box. If data is only occasionally missing, the gaps are simply filled in with climatological values.

3. Pattern analysis

a. Method

SVDA and CCA are thoroughly described in Bretherton et al. (1992) and in Barnett and Preisendorfer (1987), and so in the following the methodologies will only be described to the extent that is needed to introduce a notation that can be referred to in the later sections.

In both SVDA and CCA, sequences of coupled modes, that is, patterns and associated time series, are calculated that can be used to form an expansion of the time series of the two fields being analyzed (called left and right fields); that is, in our case

$$\mathbf{z}_{\text{sim}}(t) \approx \sum_j u_j(t)\mathbf{g}_j, \tag{1}$$

$$\mathbf{z}_{\text{obs}}(t) \approx \sum_j v_j(t)\mathbf{h}_j, \tag{2}$$

where the simulated and observed anomaly fields, $\mathbf{z}_{\text{sim}}(t)$ and $\mathbf{z}_{\text{obs}}(t)$, are normalized to have mean zero and standard deviation one at each grid point. The patterns $\mathbf{g}_j$ and $\mathbf{h}_j$ and the time series or expansion coefficients $u_j(t)$ and $v_j(t)$ follow directly from both SVDA and CCA. For SVDA, the patterns are the singular vectors and the expansion coefficients are projections of the simulated and observed fields onto the singular vectors that follow from the singular value decomposition of the cross-covariance matrix.

For CCA it is usually necessary to reduce the spatial dimensions of the simulated and observed fields. This is readily done by applying CCA to a limited number of principal components as described in Barnett and Preisendorfer (1987). The number of retained principal components should represent a compromise between sufficient amounts of explained variance and spatial dimensions significantly less than the temporal dimension.

The relative importance of the individual modes is given by the agreement between the pairs of associated expansion coefficients as determined by the covariance (for SVDA) or correlation (for CCA) between them. However, these covariances and correlations can be misleading as the modes almost inevitably include an undesired fit to random variations in the atmospheric fields, particularly if the number of spatial degrees of freedom greatly exceeds the number of temporal degrees of freedom (Wallace et al. 1992). A simple way to give a better estimate of the interrelation between the coupled modes is to calculate the correlation between the so-called cross-validated expansion coefficients as described in the appendix.

In Ward and Navarra (1997) the individual members of the ensemble of simulations were stacked to give one long time series of the simulated field. However, as also pointed out by these authors, this gives SVDA patterns that are identical to the patterns resulting from SVDA of the observations and the ensemble mean of the simulations. In the following we shall use the ensemble mean approach.

b. Examples of discrepencies between model and observed modes

We consider as a first example precipitation in northeast Brazil in the rainy season February–April (FMA). Precipitation in northeast Brazil in this season is known to be strongly linked to tropical Atlantic as well as tropical Pacific SST (Hastenrath et al. 1984; Ward and Folland 1991; Hastenrath and Greischar 1993), which is utilized in real-time experimental forecasts and regularly published in the Climate Prediction Center's (CPC) Experimental Long-Lead Forecast Bulletin.

Atmospheric GCMs, including ECHAM4, forced by observed SST have been shown to simulate northeast Brazilian rainfall with good skill (see, e.g., dynamical forecasts in the CPC Experimental Long-Lead Forecast Bulletin). However, a comparison of the first pair of SVDA heterogeneous correlation patterns [i.e., temporal correlations in each grid point between $\mathbf{z}_{\text{sim}}(t)$ and $v_1(t)$ and between $\mathbf{z}_{\text{obs}}(t)$ and $u_1(t)$] indicates that the simulated precipitation anomaly dipole (between northeastern Brazil and the region to the north) is shifted southeast in the model relative to the observed precipitation anomaly dipole (Fig. 1). EOF analyses and point correlations (not shown) confirm this difference between simulated and observed precipitation patterns.

The expansion coefficients associated with the first SVDA mode for northeast Brazilian FMA rainfall (Fig. 2a) show good agreement both within the simulated ensemble (low spread) and between the ensemble mean and observations. The correlation between the expansion coefficients for the ensemble mean and for the observations is 0.81 for the first SVDA mode, which accounts for 64% of the total squared covariance. The fraction of variance of one field (simulated or observed) explained by the expansion coefficient time series of the other field (VARF) is 16% for simulated rainfall and 15% for observed rainfall. VARF can be computed as the spatial average of the squared correlations in the grid points of the heterogeneous correlation pattern (Lau and Nath 1994).

The cross-validated expansion coefficients (Fig. 2b) are almost indistinguishable from the original expansion

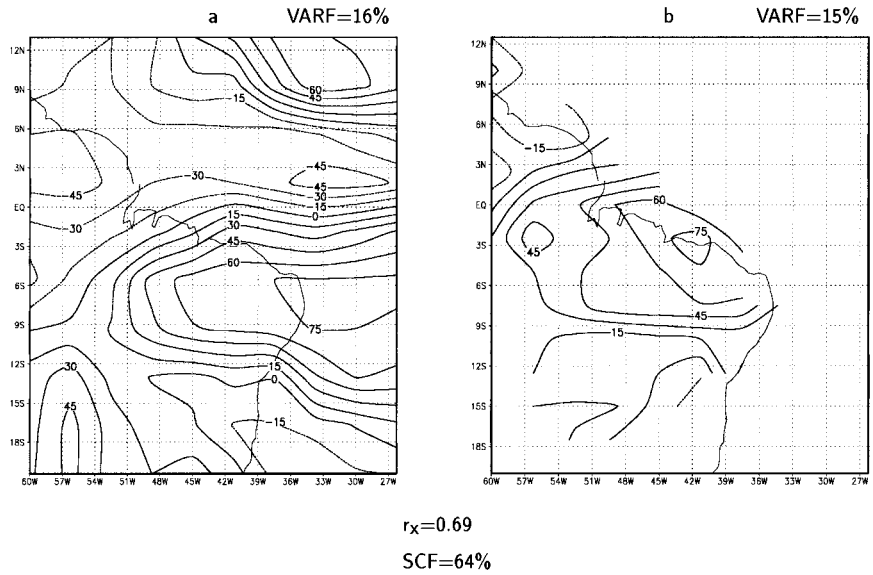

FIG. 1. Leading pair of SVDA heterogeneous correlation patterns for precipitation in northeast Brazil in FMA: (a) simulations and (b) observations. Here, r_x is the correlation between the corresponding cross-validated expansion coefficients, SCF is the squared covariance fraction between simulated and observed precipitation explained by the leading SVDA mode, and VARF is the fraction of variance of simulated or observed precipitation explained by the leading mode expansion coefficient time series of the opposite field.

coefficients, but the correlation between ensemble mean and observations drops slightly from 0.81 to 0.69 for the first mode. For the secondary SVDA modes, the effect of the cross-validation procedure becomes clear. Correlations for the first four SVDA modes that together account for 87% of the total squared covariance are summarized in Tables 1 and 2 (note that the total squared covariance is based on the original time series, no attempt has been made to compute a ''cross-validated covariance''). Clearly, the interrelations between the secondary coupled modes are largely overestimated in non-cross-validated SVDA. The picture is similar for CCA. We note here, and it will be seen in later results as well, that the cross-validation sometimes finds lower modes (such as numbers 3 and 4 in Table 2) to be more skillful than higher modes (such as number 2 in Table 2). This surprising result requires further investigation but is not fundamental to the aims of the present paper.

The fact that the first simulated SVDA mode correlates so well with the corresponding observational mode, even when cross-validated, suggests that statistical postprocessing of simulated rainfall could improve the skill of the model in the area where the SVDA patterns in Fig. 1 differ. We shall return to examples of postprocessed precipitation in section 5b.

As a second example consider the first pair of SVDA patterns for North American precipitation in winter (January–March; Fig. 3). Simulations as well as observations show approximate dipole structures, but there are distinct differences between the two patterns, notably around the Great Lakes. The maximum near the Great Lakes in the SVDA pattern in Fig. 3b (observations) is missing in the SVDA pattern in Fig. 3a (simulations), which suggests that precipitation is incorrectly simulated in this area. EOF analysis (not shown) suggests that the dominant observed precipitation pattern is only partly reproduced by the ECHAM4 model [and composite analyses in Livezey et al. (1997) seem to indicate a similar result for the National Centers for Environmental Prediction atmospheric GCM]. Near the Great Lakes the simulated precipitation is missing a center of variability, but as the observed precipitation here correlates well with the expansion coefficient of the first mode of simulated precipitation, SVDA is able to pick up the precipitation variability near the Great Lakes [pointwise correlation analyses (not shown) indicate that this is primarily through anticorrelation between observed precipitation near the Great Lakes and simulated precipitation off the western United States].

As is the case for northeast Brazilian rainfall, the expansion coefficients for North American JFM precipitation show both low spread and good agreement between observations and the ensemble mean of the simulations (correlation coefficients of 0.79 without cross-validation and 0.61 with cross-validation for the first mode; Fig. 4), which at this point suggests a possible improvement in simulation skill after statistical postprocessing; see section 5b.

There are cases where we find no agreement between simulations and observations, neither spatial nor temporal, for example, for precipitation in Europe during the transition seasons. In these cases the cross-validated

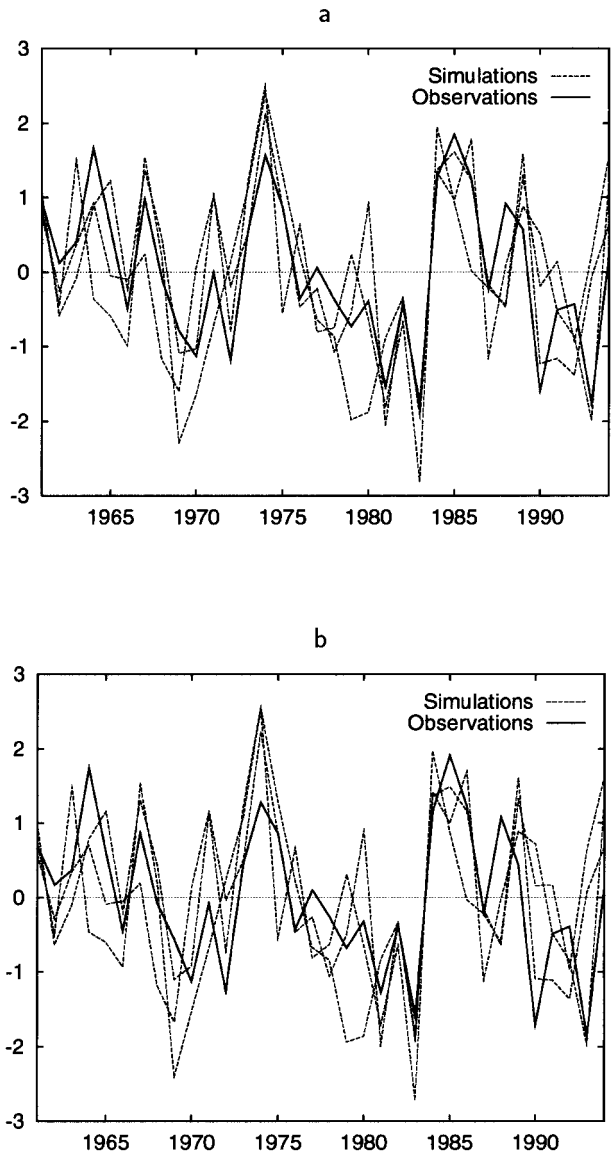

FIG. 2. (a) Normalized expansion coefficient time series for the first SVDA mode for precipitation in northeast Brazil in FMA. The correlation between observations and the ensemble mean of the simulations is $r = 0.81$. (b) Normalized cross-validated expansion coefficients. The correlation between observations and the ensemble mean of the simulations is $r_x = 0.69$.

TABLE 1. Correlations between expansion coefficients for observations and simulations (individual ensemble members as well as ensemble mean) of FMA precipitation in northeast Brazil for the first four SVDA modes.

Mode	Ensemble members			Ensemble mean
1	0.73	0.70	0.68	0.81
2	0.60	0.49	0.48	0.62
3	0.57	0.40	0.55	0.67
4	0.71	0.38	0.40	0.74

TABLE 2. As in Table 1 but for cross-validated expansion coefficients.

Mode	Ensemble members			Ensemble mean
1	0.64	0.62	0.53	0.69
2	−0.22	−0.23	−0.56	−0.46
3	0.51	0.18	0.17	0.34
4	0.58	0.18	0.33	0.53

expansion coefficients show large spread and the correlation between the ensemble mean of the simulations and observations is practically zero. In such cases where the GCM shows no skill at all, either because the model is very poor or because the model tries to simulate a quantity that is essentially unpredictable due to the chaotic nature of the atmosphere, we have no reasons to believe that postprocessing can improve the skill of the model simulations.

4. Adjusting the model simulations

As shown in the previous section there are examples where SVDA suggests that the model is either producing centers of variability in geographical positions slightly off the corresponding observed positions or the model is missing some centers of variability. Such errors may

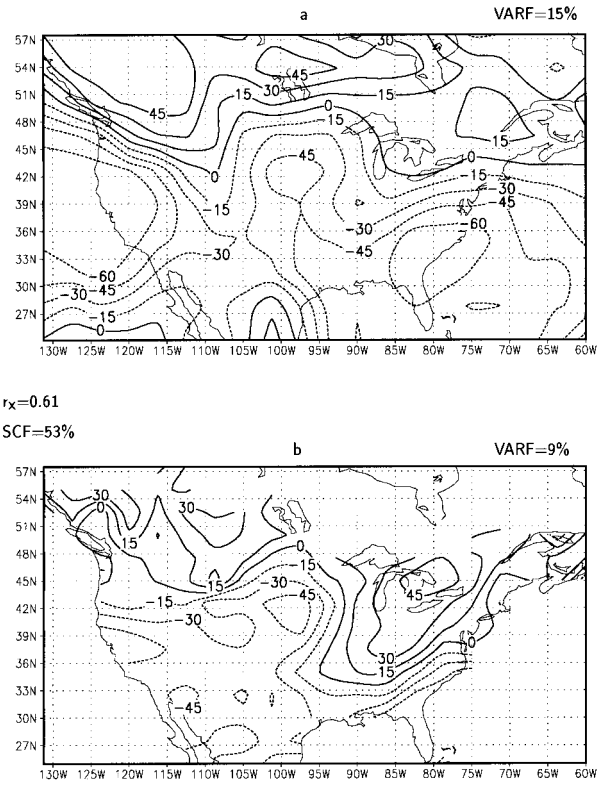

FIG. 3. Leading pair of SVDA heterogeneous correlation patterns for precipitation in North America in JFM: (a) simulations and (b) observations.

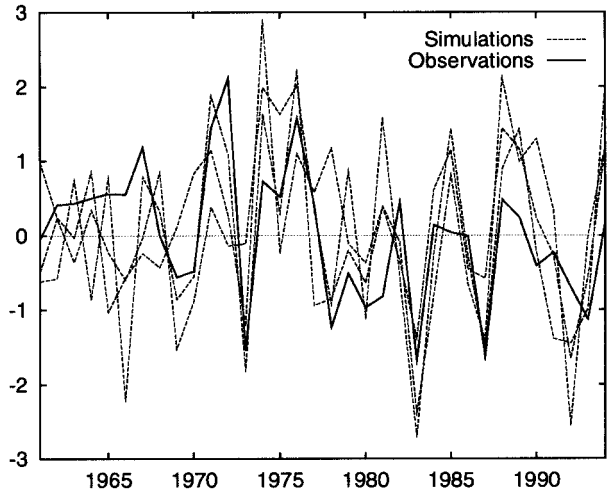

FIG. 4. Normalized cross-validated expansion coefficient time series for the first SVDA mode for precipitation in North America in JFM. The correlation between observations and the ensemble mean of the simulations is $r_x = 0.61$.

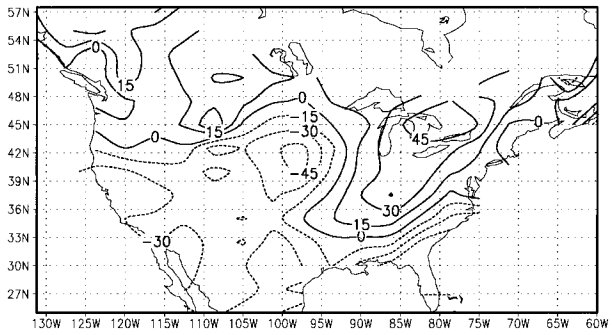

FIG. 5. Leading SVDA heterogeneous correlation pattern of adjusted simulation of JFM precipitation in North America (cf. Fig. 3).

degrade the skill of the model if the skill score is based on gridpoint comparisons between simulations and observations as is very common, for example, using correlation coefficients or LEPS skill scores. The relatively high correlations between cross-validated expansion coefficients of simulations and observations of the leading SVDA modes support the interpretation that the model responds correctly in time (to the extent indicated by the values of the correlations between the cross-validated expansion coefficients) to external forcing, that is, to variations in SST. Following this line of thought, it should be possible to use the leading SVDA (or CCA) modes to postprocess the model simulations in a way that will make the leading SVDA (or CCA) correlation patterns of the observations and postprocessed simulations agree better and also increase the model skill scores. The idea is to use linear regression to specify the right field (observations) from the left field (simulations) analogously to the way described in Barnett and Preisendorfer (1987); that is, the adjusted simulation is expressed as linear combinations of the left expansion coefficients $u_j(t)$ as

$$\mathbf{z}_{\text{adj}}(t) = \sum_j \mathbf{A}_j u_j(t). \tag{3}$$

Exactly which modes, j, to include in the sum in Eq. (3) is a nontrivial problem. The correlations between the cross-validated expansion coefficients provide a good indication of ''good'' and ''bad'' modes. For example, for northeast Brazil, SVDA modes 1, 4, and possibly mode 3 should be included, whereas mode 2 should not be included (see Table 2). Cross-validated hindcast experiments show that this choice of modes in Eq. (3) is indeed a good choice, but as the selection of modes is done a posteriori, the skill of the adjusted simulation is likely to be overestimated. We note, however, that if the adjustment is based on more than one SVDA mode, the effect of a specific mode is weighted according to the fraction of squared covariance accounted for by that mode; that is, for northeast Brazil where the first SVDA mode is very dominant, the higher-order modes have limited influence on the adjustment. In the next section a so-called double cross-validation procedure is outlined in which the modes are selected on an a priori basis. An alternative is a priori to decide to include a fixed number of modes in Eq. (3), for example, the first five modes.

The elements A_{ij} of the vectors $\mathbf{A}_j$ in Eq. (3) are determined by minimizing

$$e_i = \sum_{t=1}^{T} \left[z_{\text{obs},i}(t) - \sum_j A_{ij} u_j(t) \right]^2 \tag{4}$$

for all grid points i. For CCA, the orthogonality of u_i and u_j enables an analytic solution, $\mathbf{A}_j = \rho_j \mathbf{h}_j$, where ρ_j is the correlation between the jth pair of expansion coefficients (the jth canonical correlation), whereas for SVDA, $\mathbf{A}_j$ must in general be determined numerically.

The adjusted output at an independent time $T + 1$ follows from Eq. (3) where $u_j(T + 1)$ is given by the projection

$$u_j(T + 1) = \mathbf{z}_{\text{sim}}(T + 1) \cdot \mathbf{l}_j. \tag{5}$$

Here $\mathbf{l}_j$ is the left weight vector that for SVDA is simply the left pattern $\mathbf{g}_j$. For CCA the weight vectors are given by products of the covariance matrices and the CCA patterns (Bretherton et al. 1992).

Figure 5 shows the leading heterogeneous correlation pattern of adjusted JFM precipitation for North America obtained from SVDA of pairs of observed and adjusted precipitation. This pattern agrees much better with the leading heterogeneous correlation pattern of observed precipitation (Fig. 3b) than with the leading pattern of directly simulated precipitation (Fig. 3a).

Instead of basing it on coupled SVDA (or CCA) modes, the statistical postprocessing can be based on EOF modes of the simulated field. The methodology is as described above with $u_j(t)$ now denoting principal components of the simulated field, that is, the simulated

anomaly field projected onto the leading EOFs. The left weight vectors in Eq. (5) simply become EOFs of the simulated field. This method is tested on tropical rainfall in the next section. Note that with this method we still validate the adjusted field against the full observed field. If we had made validations against the leading EOFs and principal components of the observed field, we could probably achieve better skill scores as the leading principal components are generally more predictable (Branstator et al. 1993).

5. Verification of the adjustment methods

a. Practical issues and skill scores

In order to avoid overestimated skill scores, the postprocessed simulations are cross-validated (Michaelsen 1987); that is, SVDA (or CCA) is repeatedly applied to data from which one year is excluded and an adjustment is made for that year. By repeating for all years, we obtain a full set of adjusted simulations that can be verified against observations.

As mentioned in the previous section, the selection of which modes to include in the sum in Eq. (3) can be tricky. A simple solution to the problem is a priori to decide to include the first N modes, typically between 1 and 5. However, tests have shown that this approach is not always optimal. Sometimes a specific mode has a negative effect on the skill of the adjustment in which case the adjustment ideally should be done without that mode. The predictive skill of the individual modes can be assessed a priori by determining the correlations between the cross-validated expansion coefficients. The skill of the adjusted simulations should still be determined using cross-validation; that is, in each step of the cross-validation procedure, cross-validated expansion coefficients are computed in order to select which modes to include in the adjustment; hence, the procedure is termed ''double cross-validation'' (Kaas et al. 1996). The procedure is outlined below for adjustment of a simulated field at time t_0 (it may help to think of t_0 as being ''now'' and all other times as being prior to now).

1) Compute the cross-validated expansion coefficients from data from which t_0 data is excluded. 2) Find correlations between the cross-validated expansion coefficients and decide which modes to include in the adjustment [include a mode if the corresponding correlation exceeds a given threshold that we somewhat arbitrarily choose as 0.25 for precipitation and surface temperature (34 yr of data) and 0.30 for 500-hPa height and sea level pressure (26 yr of data)]. 3) Do SVDA (or CCA), still with t_0 data excluded. 4) Apply the adjustment to the simulated t_0 field using the selected modes.

By repeating for all t_0, we get a full set of adjusted simulations to verify. Although this procedure can be rather time consuming, as it is necessary for a dataset of length N in time to do $N(N - 1)$ SVD (or canonical correlation) analyses, it provides a semiobjective selection procedure with which specific modes can be included or excluded.

A problem with the selection of ''good'' modes in the cross-validation procedure arises when the order of the SVDA (or CCA) modes is not robust. We identify the modes by the covariance or correlation between the expansion coefficients rather than by, for example, the associated patterns, but occasionally—and particularly for short time series—there is not a one-to-one correspondence between mode patterns and the order of covariance or correlation for the repeated SVD (or canonical correlation) analyses in the cross-validation procedure. Thus, the adjustments may occasionally be based on modes that are not skillful, in which case the skill of the adjusted fields are likely to be destroyed.

Unless stated otherwise, we choose to always include the first SVDA or CCA mode in the adjustment and base selection of secondary modes on the correlation between the cross-validated expansion coefficients.

A skill score that is commonly used to measure seasonal predictability is the linear correlation in time in each grid point between predictions (or simulations) and observations (Barnston 1994). Correlation coefficients are independent of variability; that is, the correlation coefficient will not tell whether, for example, a prediction is near climatology in absolute value. Particularly for cross-validated correlation skill scores such predictions will occasionally result in local high negative correlations (Barnston and Van den Dool 1993). The LEPS skill score, which gives more weight to extreme predictions and less weight to near-climatological predictions, does not have so much of a problem with high negative skill scores for near-climatological predictions. As one of the properties of SVDA- (and CCA-) based specifications or predictions is that they are near climatology in areas where the leading correlation patterns are near zero, we prefer to use LEPS skill scores to validate the postprocessing.

Linear regression techniques that have least squares fit requirements, including the SVDA- (and CCA-) based adjustments described in the previous section, almost inevitably lead to loss of variability in absolute terms; that is, the adjusted field will always stay close to climatology. Thus, some sort of inflation, whereby the variability of the adjusted field is increased, can be desirable. The standard inflation technique is to multiply the adjusted values (i.e., regression-based values in each grid box) by the ratio between the standard deviation of the observations and the standard deviation of the adjusted values. However, as standard inflation quite likely would turn near-climatological, low-variance, low-skill predictions into nonclimatological, low-skill predictions, it makes practical sense to leave grid boxes with no real skill uninflated, while concentrating the inflation on more skillful grid boxes. Therefore we choose to weight the inflation factor such that in grid boxes where the uninflated adjusted field shows rela-

tively large variability, the values are fully inflated whereas in grid boxes of little variability, the inflation is strongly damped.

If we denote by $\tilde{z}_{k,\mathrm{adj}}$ the inflated value of the adjusted field in grid point k, we have (recalling that observations are normalized before the adjustment)

$$\tilde{z}_{k,\mathrm{adj}} = \frac{z_{k,\mathrm{adj}}}{\sigma_{k,\mathrm{adj}}} w(\sigma_{k,\mathrm{adj}}/\sigma_{\mathrm{max,adj}}), \tag{6}$$

where $\sigma_{k,\mathrm{adj}}$ is the standard deviation of the (uninflated) adjusted field in grid point k, $\sigma_{\mathrm{max,adj}} = \max_k(\sigma_{k,\mathrm{adj}})$, and $w(x)$ is an s-shaped weight function that is close to zero for $x = 0$ and close to one for $x = 1$. We choose the function

$$w(x) = \frac{1}{2}\left\{1 + \tanh\left[a\left(x - \frac{1}{2}\right)\right]\right\}, \tag{7}$$

which for our choice of a, $a = 6.9$, satisfies $w(0) \approx 0.001$ and $w(1) \approx 0.999$.

In order to calculate LEPS skill scores, we need an estimate of the cumulative probability distribution in each grid point of the observed field. We have chosen to base all LEPS scores shown in the following on normal distributions after tests showed only negligible differences between LEPS scores based on normal distributions and LEPS scores based on purely empirical distributions in each grid point, even for precipitation. The relation between LEPS skill scores and correlation coefficients is discussed in Barnston (1992). For moderate skill, LEPS scores are generally somewhat smaller than correlation coefficients.

In order to summarize the effect of the adjustments, we shall represent skill for an entire region by the so-called mean anomaly correlation (Déqué and Royer 1992; Déqué 1997). The mean anomaly correlation, which gives more weight to strong anomalies than an average of anomaly correlation coefficients, is calculated from three mean covariances as

$$\mathrm{MAC} = \frac{C_{SO}}{(C_{SS}C_{OO})^{1/2}}, \tag{8}$$

where, for example, the mean covariance C_{SO} is defined as

$$C_{SO} = \langle\overline{(S_k(t) - \overline{S}_k)(O_k(t) - \overline{O}_k)}\rangle. \tag{9}$$

Here S denotes the simulated (or adjusted) field and O denotes the observed field. The fields are not standardized as more weight should be given to areas of high variability. The brackets indicate a spatial average and the overbar indicates a time average; C_{SS} and C_{OO} are found in a similar manner.

The mean anomaly correlations before and after the adjustment give an indication of the average effect of the adjustment for the entire area under consideration. Note that the variance of the simulated and adjusted fields are in general different as the adjusted fields are inflated to match the variance of the observed fields. In order to test whether the difference in variance is important, mean anomaly correlations for the raw model output was compared with mean anomaly correlations for inflated model output (inflated to match the variance of the observed fields). It was found that this inflation had virtually no effect on the mean anomaly correlations, so all mean anomaly correlations shown in the following for raw model output are based on uninflated fields.

b. Precipitation

For northeast Brazil, the LEPS skill score map of the unadjusted simulated precipitation (Fig. 6a) shows an area of high skill in the Nordeste region. Figures 6b and 6c show cross-validated LEPS skill score maps of SVDA-adjusted simulations; in Fig. 6b the adjustment is based on SVDA mode 1 only, whereas in Fig. 6c additional modes are included using the double cross-validation technique; that is, in each step of the cross-validation procedure the modes are included in the adjustment according to the value of the correlations between the cross-validated expansion coefficients. The three skill score maps show that the SVDA-adjusted simulation is more skillful than the direct model output along most of the north coast and that almost all of the skill is associated with the first SVDA mode. The mean anomaly correlation increases from 0.23 to 0.36 following the adjustment.

When the first SVDA mode is very dominant, as in the example of precipitation in northeast Brazil, the LEPS skill score map of the adjusted simulation will in general resemble the leading SVDA heterogeneous correlation pattern of observed precipitation in the sense that high correlations (positive or negative) translate to skill maxima whereas near-zero correlations translate to near-zero skill.

For JFM precipitation in North America, the SVDA-based adjustment results in increased skill almost everywhere. Figure 7 shows LEPS skill score maps before and after adjustment. Note the substantial improvement in skill in the area around the Great Lakes where the leading SVDA patterns (Fig. 3) indicated that the model was in error. The mean anomaly correlation increases from 0.13 for the raw model output to 0.22 for the adjusted simulation.

The effect of adjusting the simulated precipitation has also been calculated for the tropical regions that were studied in Moron et al. (1998), see Fig. 8, to get a general impression of the postprocessing performance. Results are summarized in Fig. 9. The generally relatively high skill tropical situations are contrasted with extratropical results for North America and Europe in Fig. 10. In Figs. 9 and 10, CCA (following Barnett and Preisendorfer 1987) is applied to the first six principal components of the simulated precipitation and the first six principal components of the observed precipitation. For the tropical regions (Fig. 9) postprocessing generally

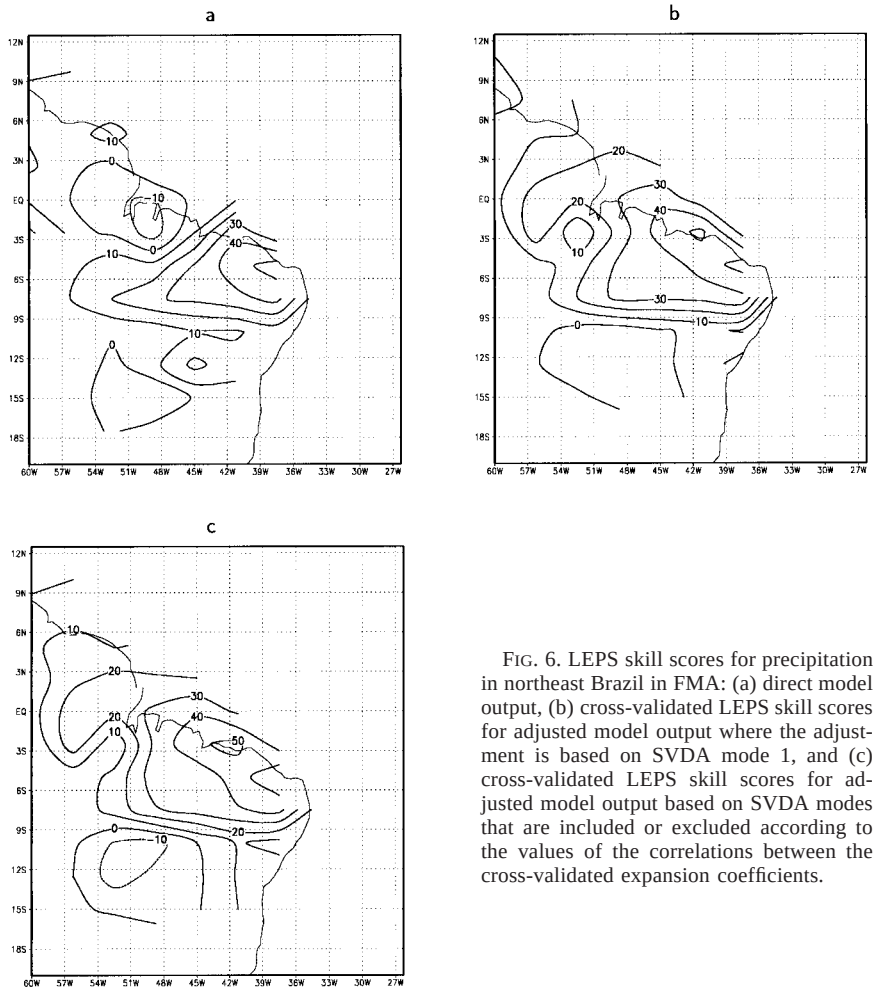

FIG. 6. LEPS skill scores for precipitation in northeast Brazil in FMA: (a) direct model output, (b) cross-validated LEPS skill scores for adjusted model output where the adjustment is based on SVDA mode 1, and (c) cross-validated LEPS skill scores for adjusted model output based on SVDA modes that are included or excluded according to the values of the correlations between the cross-validated expansion coefficients.

improves the model performance, that is, the mean anomaly correlation skill scores for both SVDA- and CCA-adjusted precipitation are higher than for the direct model output in all examples (Fig. 10). One exception is for tropical South America where the mean anomaly correlation skill score drops from 0.2 for the direct model output to near zero for the CCA-based adjustment. The reason for the failure of the CCA-based adjustment is closely linked to the robustness of the order of the CCA modes in the cross-validation procedure. A closer examination shows that CCA mode 1 is robust and that an adjustment based on this mode only does not lead to a drop in skill (Fig. 11). But the secondary modes become mixed and destroy the skill of the adjusted simulation.

For both North America and Europe the model simulations are moderately skillful in winter, but with no clear skill in the other seasons before postprocessing. Application of the postprocessing leads to increased mean anomaly correlation skill scores for both regions in winter and for North America also in spring and summer (Fig. 10). The mean anomaly correlation for SVDA-adjusted precipitation in North America in the July–September season (Fig. 10a) is remarkably high compared to the mean anomaly correlation of the direct model output. The cross-validated expansion coefficients for SVDA mode 1 shows that the ensemble mean of the simulation agrees well with observations (correlation coefficient of 0.63), though the large spread within the ensemble suggests more experiments are needed for a firm conclusion.

In order to eliminate the problems with selection of secondary modes in the double cross-validation procedure, we have looked at postprocessing based only on the first, dominant mode in the tropical rainfall examples. Here, almost all skill of the postprocessed simulations was associated with the first mode (northeast Brazil is an illustrative example; see Fig. 6). The coupled patterns of the leading SVDA or CCA mode are usually similar to the leading EOFs of the simulated and observed fields, so postprocessing based on the leading EOFs can be expected to give results that are similar to those obtained by postprocessing based on the leading SVDA or CCA mode. Figure 11 shows that in terms of

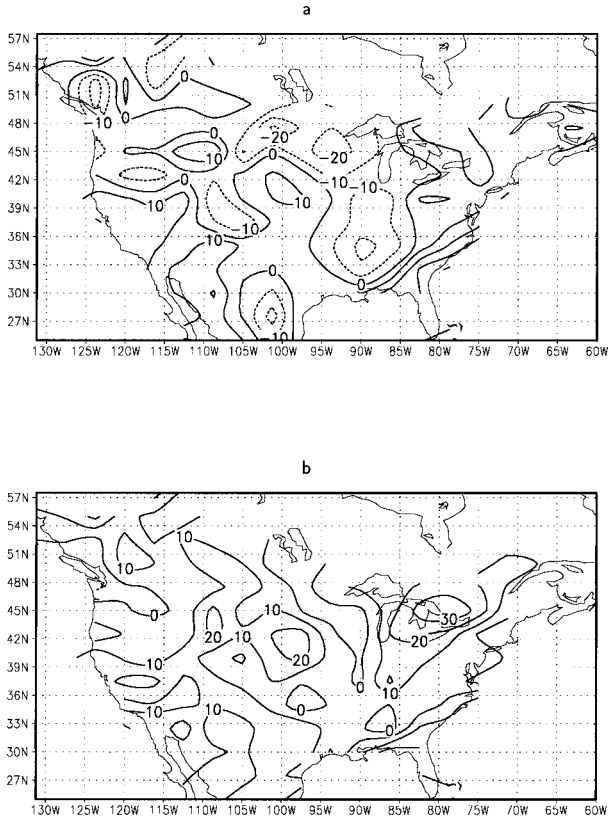

FIG. 7. LEPS skill scores for precipitation in North America in JFM: (a) direct model output and (b) cross-validated LEPS skill scores for SVDA-adjusted model output.

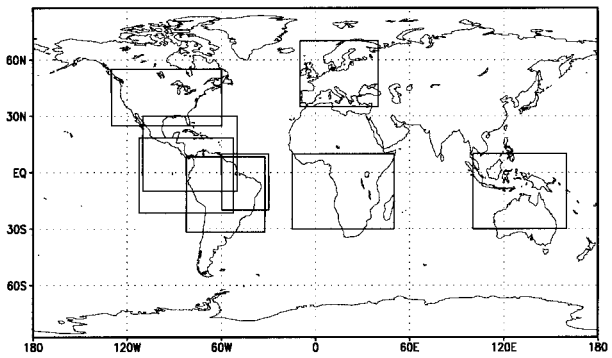

FIG. 8. Regions in which the effect of SVDA- and CCA-based statistical postprocessing of precipitation is tested.

mean anomaly correlation skill scores this appears to be the case for postprocessing of tropical rainfall, but the level of skill for the EOF-based postprocessing is in most cases slightly below that of SVDA- and CCA-based postprocessing. We also note that SVDA mode 1 mostly is slightly more skillful than CCA mode 1, but the secondary CCA modes are often more skillful than the secondary SVDA modes (cf. Fig. 9). In general, however, all three methods work well for the tropical rainfall examples and the differences between the mean anomaly correlation skill scores for the three methods are in most cases very small.

c. The 500-hPa geopotential height and mean sea level pressure

To provide a contrast to fields of precipitation, this section presents results for large-scale dynamic fields of 500-hPa geopotential height and mean sea level pressure for the extratropical Pacific–North American (17°–68°N, 160°–60°W) and North Atlantic–European (25°–70°N, 60°W–35°E) regions. Compared to precipitation in these regions, the dynamic fields are simulated more skillfully by the model, and the effects of statistical postprocessing do not have such a clear positive impact on skill averaged over the whole domain (Fig. 12).

During the transition seasons the model simulates both fields very poorly in the two regions, and SVDA- or CCA-based adjustments do not improve the performance. In winter (JFM) and summer (July–September, JAS) the mean anomaly correlation skill scores (Fig. 12) show examples of both (slightly) improved and (seriously) degraded performance following postprocessing. We have analyzed the ''bad'' cases in order to understand why the method sometimes fails.

An example where both SVDA- and CCA-based adjustments fail to improve the model performance is mean sea level pressure in the PNA region in JAS (Fig. 12a). Although we find that the first pair of SVDA correlation patterns are similar, the actual correlations (the contour values) are small, as is the correlation between the cross-validated expansion coefficients of the first SVDA mode (not shown). Without the temporal agreement between simulations and observations, the postprocessing cannot be expected to improve the skill of the model.

The reason for the failure of the CCA-based adjust-

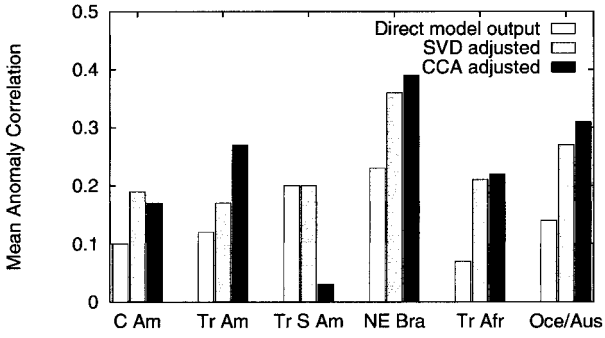

FIG. 9. Mean anomaly correlation skill scores before and after SVDA- and CCA-based statistical postprocessing for precipitation in Central America (10°S–30°N, 110°–50°W) in JAS, tropical America (20°S–20°N, 110°–50°W) in MAM, tropical South America (30°S–10°N, 80°–30°W) in DJF, northeast Brazil (20°S–10°N, 60°–30°W) in FMA, tropical Africa (30°S–10°N, 15°W–50°E) in JFM, and Oceania–Australia (30°S–10°N, 100°–160°E) in JFM.

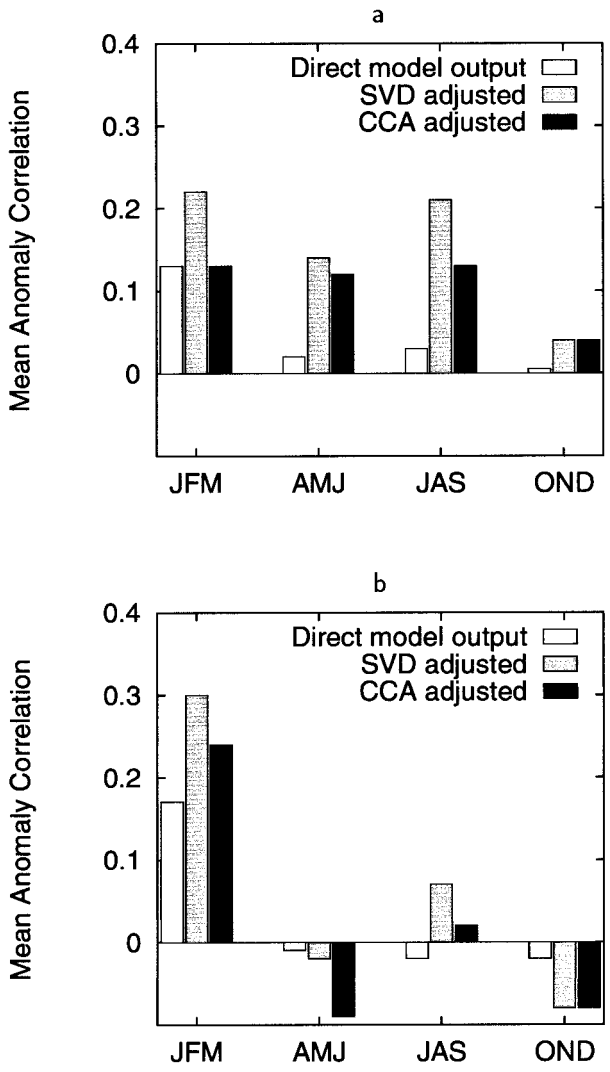

FIG. 10. Mean anomaly correlation skill scores before and after SVDA- and CCA-based statistical postprocessing for different seasons for precipiation in (a) North America and (b) Europe.

ment of mean sea level pressure in the North Atlantic–European region in JFM (Fig. 12b) is the same as was noted for the CCA-based adjustment of precipitation in South America: The order of the CCA modes does not remain the same when data from one year is replaced by data from another year in the double cross-validation procedure and the adjustment fails when the individual modes are mixed up. This mixing of the modes even affects mode number 1 and is reflected in the correlation ($r_x = -0.18$) between the cross-validated expansion coefficients for the first CCA mode. For the cross-validated SVDA expansion coefficients, where the modes do not get mixed, the corresponding correlation is $r_x = 0.46$.

d. Land surface temperature

When adjusting surface temperature, care must be taken to include only simulated land surface temperatures as the SST is prescribed with observed values and thus is not a result of the model simulation. Summaries of the effect of the adjustments in winter and summer are included in Fig. 12. Note that the model simulates land surface temperature, whereas the observations are for near-surface air temperature. The postprocessing can be expected to include a mapping from surface to near-surface temperature if there are systematic differences between the two types of temperature, both in mean, variance, and variability patterns.

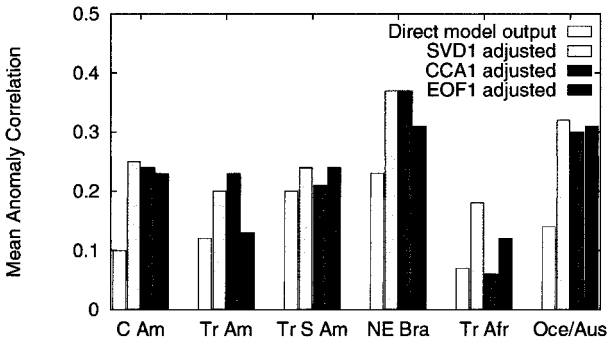

FIG. 11. As in Fig. 9 but for statistical postprocessing based only on the leading mode and including also an EOF-based adjustment.

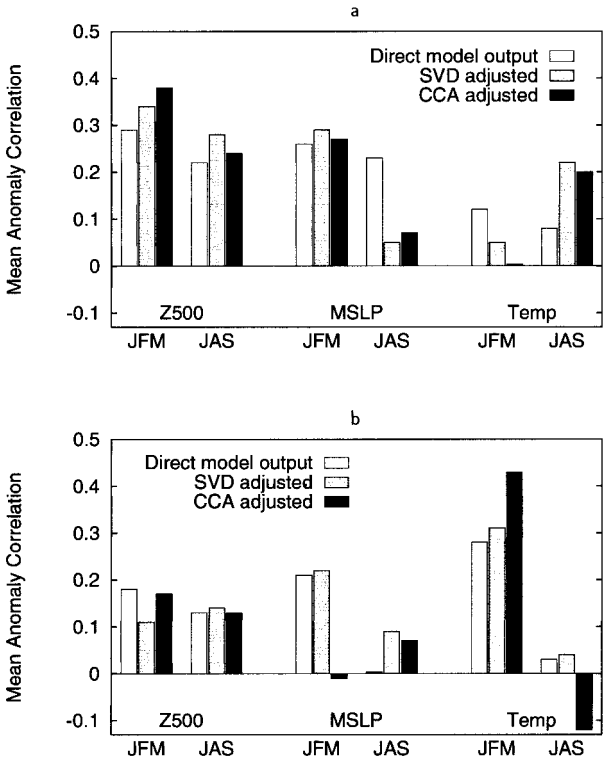

FIG. 12. Mean anomaly correlation skill scores before and after SVDA- and CCA-based statistical postprocessing in winter (JFM) and summer (JAS) for 500-hPa height, mean sea level pressure, and surface temperature in (a) the Pacific–North American region and (b) the North Atlantic–European region.

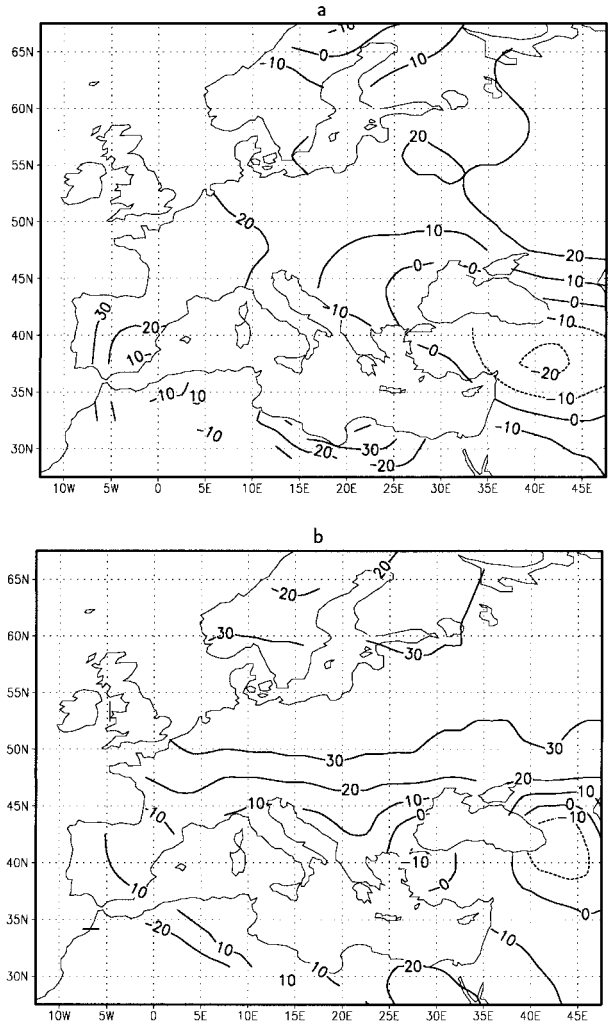

FIG. 13. LEPS skill scores for surface temperature in Europe in JFM: (a) direct model output and (b) cross-validated LEPS skill scores for CCA-adjusted model output.

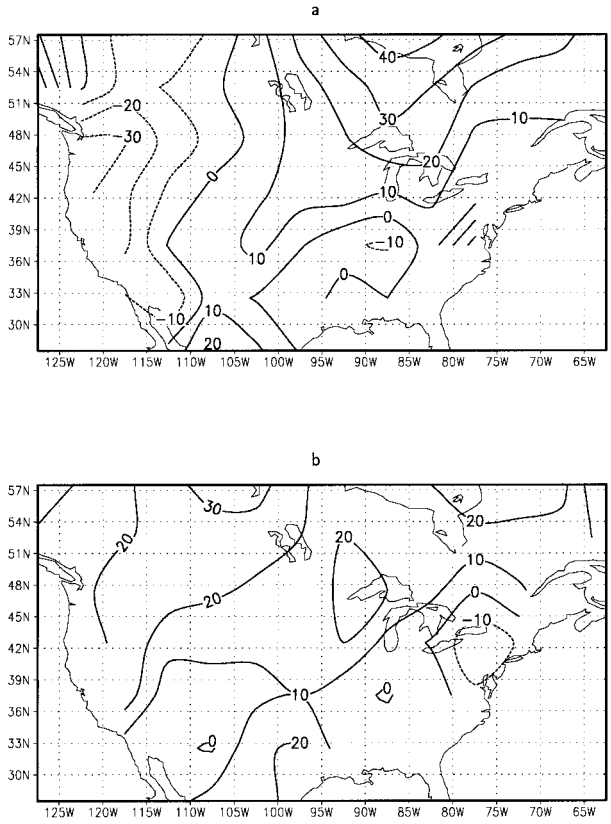

FIG. 14. LEPS skill scores for surface temperature in North America in JAS: (a) direct model output and (b) cross-validated LEPS skill scores for SVDA-adjusted model output.

The model simulates European surface temperature in JFM relatively skillfully (Fig. 12b), which is consistent with the relatively skillful simulation of precipitation in Europe in the same season. The CCA-based adjustment improves skill almost everywhere, particularly in the northeastern part of the region (Fig. 13), and many parts attain LEPS skills of over 20%. A closer examination of the applied adjustment reveals that CCA mode number 3 is the important mode, whereas the first two modes do not correlate when cross-validated. CCA mode number 3 (not shown) shows a north–south dipole that closely resembles the first, dominant EOF (not shown), which explains more than 50% of the observed variance. This highlights a potential problem with this type of CCA, where the high-order CCA modes can be composed of unstable low-order EOFs that explain little variance. The cross-validation procedure overcomes this defficiency, enabling us to identify lower-order CCA modes that have genuine skill.

A skill increase following postprocessing is also observed for North American surface temperature (Fig. 12a) in JAS. The first pair of SVDA heterogeneous correlation patterns (not shown) suggests that the model has problems with the variability in the northwestern part of the region, which is reflected in the LEPS skill score maps (Fig. 14): The model simulates surface temperature in the western part of the continent very poorly, whereas the adjusted simulation in this area in general is much more skillful. In JFM the leading modes do not agree in time when cross-validated (r_x = 0.01 for SVDA, r_x = 0.18 for CCA). Consequently, postprocessing fails to improve the model performance. No skill is observed for either of the two regions during spring and autumn.

6. Discussion and conclusions

In recent years, there has been an increasing interest in seasonal forecasting (Palmer and Anderson 1994). The efforts have gone in two directions: empirical methods (Barnett and Preisendorfer 1987; Barnston 1994; Ward and Folland 1991; Vautard et al. 1996) and dy-

namical modeling (Stern and Miyakoda 1995; Kumar et al. 1996; Brankovic and Palmer 1997), but only limited work has been done on combining the two approaches (Sarda et al. 1996). Presently, the skill of empirical methods easily matches or is even better than the skill of the much more complicated and CPU time consuming dynamical models (Barnston and Smith 1996) where skill is limited by a number of different types of systematic errors. In this paper we have focused on geographical errors in simulated climate variability on interannual timescales, and we have tested the potential for correcting such errors by statistically postprocessing the model output.

By applying SVDA (or CCA) to simultaneous pairs of simulated and observed fields in cases where the dynamical model is skillful (but far from perfect), we are able to extract information about geographical errors in the simulation from the differences between the pair of leading SVDA (or CCA) patterns. It is obvious that geographically shifted model fields can result in a degradation of skill when the skill scores are based on the performance in individual grid boxes, particularly for fields that contain small-scale geographical variations such as precipitation or surface temperature where minor geograhical errors are sufficient to degrade skill significantly.

One of our aims has been to test the extent to which statistical postprocessing based on linear regression of the leading SVDA (or CCA) modes can be used to increase the skill of the model. We adjust the model output in such a way that the leading SVDA or CCA patterns for pairs of adjusted and observed fields become nearly identical. Loosely speaking, the aim is to adjust the simulation so that the agreement in space, which is assumed limited by model errors that geographically shift or otherwise corrupt the leading modes of variability relative to observations, is increased to match the agreement in time between the leading SVDA or CCA modes.

It is important to stress that if the dynamical model does not possess some genuine skill, then our statistical correction will not work. If the model is validated in individual grid boxes, for example, using LEPS or anomaly correlation skill scores, we have demonstrated that the model skill may not show up without the statistical postprocessing. If we were interested in knowing a model skill score that is independent of a linear transformation of a simulated field, then we would need to use an invariant correlation or root-mean-square error made using the Mahalanobis metric (Stephenson 1997).

For precipitation, in tropical as well as extratropical regions, postprocessing was found to improve the model performance in almost all situations where the dynamical model was at least moderately skillful, which would be of potential value for dynamical seasonal forecasting. For large-scale dynamic fields of 500-hPa height and sea level pressure, we did not find a similar unambiguous improvement in performance. It is likely that the postprocessing performs better on fields that can have strong anomaly gradients, such as precipitation, but as postprocessing of the dynamic fields were only tested in moderate-to-low skill extratropical regions, we hesitate to conclude that the method is not suitable for these fields.

In a couple of cases the statistical postprocessing failed and destroyed whatever skill the raw model output possessed. This can happen if the adjustment unintentionally includes unskillful modes. Unskillful modes are coupled modes that are the result of covarying random fluctuations (noise). Example 2 in the appendix demonstrates how the left and right field time series can be highly correlated for random (and hence unskillful) SVDA and CCA modes, and additional experiments have shown that such highly correlated unskillful modes also can exist when the noise is added to a weak covarying signal. When correlations (or covariances) for the unskillful modes exceed those for more skillful modes we have cases where unskillful modes become leading modes.

A selection procedure that is based on the correlation between cross-validated expansion coefficients effectively retains the skillful modes and filters out the unskillful modes. In one case (CCA-based adjustment of JFM surface temperature in Europe), the first two modes were both unskillful while virtually all skill was carried by CCA mode number 3. The order of the modes was robust throughout the cross-validation in this example, so the two leading unskillful modes were ignored in every step of the cross-validation and the adjustment resulted in improved skill. However, if the singular values or canonical correlations of some of the leading coupled modes are near degenerate, the order of the modes is likely to get mixed in the course of the cross-validation procedure so that adjustments occasionally will be based on unskillful modes. The shorter the training period is for the statistical correction procedure, the more likely the mode mixing is. When this mode mixing affects the leading mode, the postprocessing will almost inevitably fail.

In the skillful examples we found that usually the first coupled mode accounts for more than 50% of the squared covariance. With such a dominant first mode, the overall effect of postprocessing will be closely linked to the skill associated with this mode. If skillful, the secondary modes can contribute positively to the skill in the less skillful areas, but sometimes at the expense of the high skill in the more skillful areas. In some cases we found that the effect of the postprocessing was simply to geographically shift maxima of skill. Still, we consider this a positive effect if the skill maxima of the adjusted field coincide with areas of high observed variability. In such cases we also observe an increase in the mean anomaly correlation skill score. A possible way to overcome the problem of decreased skill scores in some of the grid boxes following statistical correction would be to specify an intelligent combination of adjusted and raw output from the model, which

would further increase the potential for practical use of SVDA- or CCA-based postprocessing in seasonal forecasting.

A comparison between postprocessing based on leading SVDA and CCA modes shows only small differences, and neither one of the methods can be claimed to do better than the other. For tropical rainfall we also tested postprocessing based on leading EOF modes. This method also works well, but with skill generally slightly below that based on SVDA or CCA modes.

One aspect that is not taken into account is the spread between individual ensemble members. Seasonal forecasts are conveniently presented as probability forecasts where the probabilities are based on the different behavior of the members of an ensemble of simulations. Preferably, a statistically corrected forecast should also be presented in terms of probabilities. The method that has been presented in this paper was only applied to ensemble means (in which the ensemble spread is only taken indirectly into account in the sense that high spread generally results in near-climatological ensemble means). Further work is needed in order to make proper use in the statistical corrections of the information, which is contained in the spread between ensemble members.

Acknowledgments. The gridded precipitation observations were provided by M. Hulme and the model runs were carried out by CINECA. The work was partly funded by the Commision of the European Union (Contract ENV4-CT95-0109).

APPENDIX

Correlations between SVDA and CCA Time Series

In a recent paper Cherry (1996) demonstrates how SVDA of two spatially and temporally uncorrelated fields results in high correlations between the expansion coefficients of the leading modes. Using a number of constructed examples with a varying degree of common signal between the left and right fields, he finds that the highest correlations arise when the two fields are spatially uncorrelated and have no common signal. For two fields with no common signal, SVDA correlations tend to reflect the number of spatial degrees of freedom. The more spatial degrees of freedom, the more highly correlated linear combinations for SVDA to exploit. The same remarks apply to CCA.

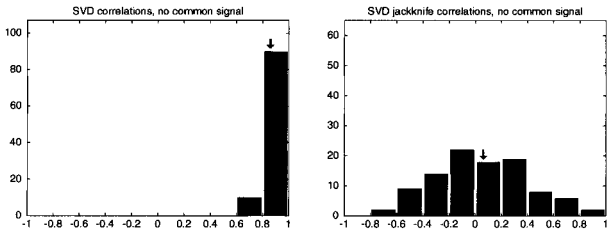

FIG. A1. Histograms of the distribution of 100 correlations between the leading SVDA mode expansion coefficients for two random fields with no common signal (example 1). (left) Correlations between the usual expansion coefficients. (right) Correlations between cross-validated expansion coefficients. The arrow indicates the average correlation.

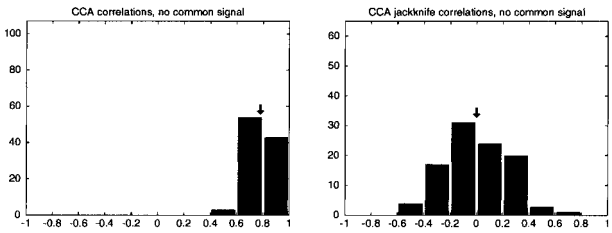

FIG. A2. As in Fig. A1 but for CCA.

Such high SVDA or CCA correlations are unfortunate as they give the false impression that the two analyzed fields are statistically linked. However, if one tries to specify one field from the expansion coefficients of the other field, as done in Barnett and Preisendorfer (1987) for CCA, it becomes evident that the two fields are not related. As always, it is important that the specification is done on data that is independent of the dataset (the training set) that is used in the calculation of the linear specification model. The performance of the specification is commonly validated using a jackknife or cross-validation technique in which data from one (or more) points in time systematically are withheld from the dataset, a specification model is derived from the remaining part of the dataset, and the specification is tested on the withheld data (Michaelsen 1987).

This cross-validation technique has motivated the introduction of the following SVDA and CCA cross-validated time series or expansion coefficients. The correlations between the cross-validated leading mode expansion coefficients represent the statistical link between the two analyzed fields much better than the correlations between the usual SVDA and CCA expansion coefficients.

In order to calculate cross-validated expansion coefficients, we do the following. 1) Withhold data from one point in time, say t_0, from the left and right data fields and apply SVDA or CCA to the remaining data and derive the so-called left and right weight vectors

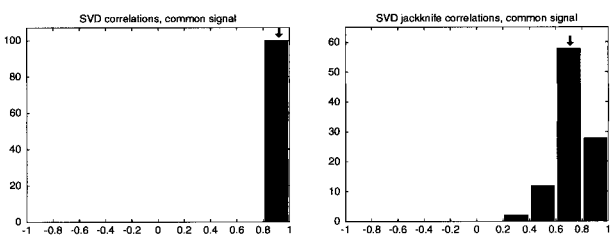

FIG. A3. As in Fig. A1 but for two fields with a strong common signal (example 2).

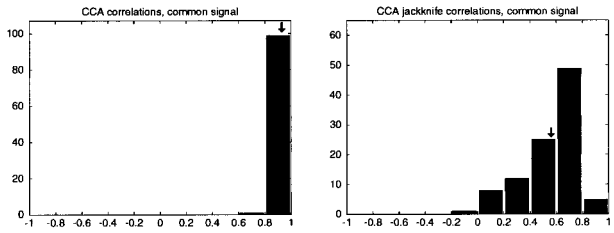

FIG. A4. As in Fig. A3 but for CCA.

(Bretherton et al. 1992), $\mathbf{l}_j'(t_0)$ and $\mathbf{r}_j'(t_0)$, where the prime denotes that no t_0 data is used in the calculation (for SVDA the left and right weight vectors are simply the left and right singular vectors). 2) Project the withheld left and right t_0-anomaly fields onto the left and right weight vectors, respectively, to obtain cross-validated expansion coefficients for t_0, $u_j'(t_0)$, and $v_j'(t_0)$. 3) Repeat this procedure for all t_0 in order to obtain pairs of time series $u_j'(t)$ and $v_j'(t)$.

The correlations between the cross-validated leading mode expansion coefficients are compared to the correlations between the usual leading mode expansion coefficients in two constructed examples in the following. In both examples we have 16 points in space (grid points or stations) for both left and right field and we consider time series of length 20.

Example 1: All points are spatially and temporally uncorrelated. The samples are normally distributed and standardized to zero mean and standard deviation one.

Example 2: A deterministic signal, 10 sech$(k - 8)$ sin$(\pi n/10)$ ($k \sim$ space, $n \sim$ time), has been added to the random signal in example 1 for both the left and right fields.

SVDA and CCA have both been applied to 100 realizations of each of the two examples [following Barnett and Preisendorfer (1987), CCA is applied to the first five principal components of both the left and the right field]; distributions of correlations between the usual leading mode expansion coefficients as well as between the cross-validated leading mode expansion coefficients are shown in Figs. A1–A4, which clearly illustrate how the cross-validation procedure eliminates the undesired high correlations between uncorrelated fields while retaining high correlations between highly correlated fields. The drawback is that instead of doing SVDA or CCA only once, we have to do N SVD or canonical correlation analyses, where N is the length in time of the datasets.

Note that we have not addressed the caveat discussed in Newman and Sardeshmukh (1995) where it is demonstrated that SVDA only in very special circumstances is able to fully recover a known linear relationship between two fields. This is a different problem, which is not handled by the cross-validation procedure.

REFERENCES

Barnett, T. P., and R. Preisendorfer, 1987: Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. *Mon. Wea. Rev.,* **115,** 1825–1850.

Barnston, A. G., 1992: Correspondence among the correlation, RMSE, and Heidke forecast verification measures: Refinement of the Heidke score. *Wea. Forecasting,* **7,** 699–709.

——, 1994: Linear statistical short-term climate predictive skill in the Northern Hemisphere. *J. Climate,* **7,** 1513–1564.

——, and H. M. van den Dool, 1993: A degeneracy in cross-validated skill in regression-based forecasts. *J. Climate,* **6,** 963–997.

——, and T. M. Smith, 1996: Specification and prediction of global surface temperature and precipitation from global SST using CCA. *J. Climate,* **9,** 2660–2697.

Brankovic, C., and T. Palmer, 1997: Atmospheric seasonal predictability and estimates of ensemble size. *Mon. Wea. Rev.,* **125,** 859–874.

Branstator, G., A. Mai, and D. Baumhefner, 1993: Identification of highly predictable flow elements for spatial filtering of medium- and extended-range numerical forecasts. *Mon. Wea. Rev.,* **121,** 1786–1802.

Bretherton, C. S., C. Smith, and J. M. Wallace, 1992: An intercomparison of methods for finding coupled patterns in climate data. *J. Climate,* **5,** 541–560.

Cherry, S., 1996: Singular value decomposition analysis and canonical correlation analysis. *J. Climate,* **9,** 2003–2009.

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

——, and J. F. Royer, 1992: The skill of extended-range extratropical winter dynamical forecasts. *J. Climate,* **5,** 1346–1356.

Graham, N. E., P. Barnett, R. Wilde, M. Ponater, and S. Schubert, 1994: On the roles of tropical and midlatitude SSTs in forcing interannual to interdecadal variability in the winter Northern Hemisphere circulation. *J. Climate,* **7,** 1416–1441.

Hastenrath, S., and L. Greischar, 1993: Further work on the prediction of northeast Brazil rainfall anomalies. *J. Climate,* **6,** 743–758.

——, M.-C. Wu, and P.-S. Chu, 1984: Towards the monitoring and prediction of north-east Brazil droughts. *Quart. J. Roy. Meteor. Soc.,* **110,** 411–425.

Hulme, M., 1994: Validation of large-scale precipitation fields in general circulation models. *Global Precipitation and Climate Change,* M. Desbois and F. Desalmand, Eds., NATO ASI Series, Springer-Verlag, 387–405.

Jones, P. D., S. C. B. Raper, R. S. Bradley, H. F. Diaz, P. M. Kelly, and T. M. L. Wigley, 1986: Northern Hemisphere surface air temperature variations, 1851–1984. *J. Climate Appl. Meteor.,* **25,** 161–179.

Kaas, E., T.-S. Li, and T. Schmith, 1996: Statistical hindcast of wind climatology in the North Atlantic and northwestern European region. *Climate Res.,* **7,** 97–110.

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

Lau, N.-C., and M. J. Nath, 1994: A modeling study of the relative roles of tropical and extratropical SST anomalies in the variability of the global atmosphere–ocean system. *J. Climate,* **7,** 1184–1207.

Livezey, R. E., M. Masutani, A. Leetmaa, H. Rui, M. Ji, and A. Kumar, 1997: Teleconnective response of the Pacific–North American region atmosphere to large central equatorial Pacific SST anomalies. *J. Climate,* **10,** 1787–1820.

Michaelsen, J., 1987: Cross-validation in statistical climate forecast models. *J. Climate Appl. Meteor.,* **26,** 1589–1600.

Moron, V., A. Navarra, M. N. Ward, and E. Roeckner, 1998: Skill and reproducibility of seasonal rainfall patterns in the tropics in ECHAM-4 GCM simulations with prescribed SST. *Climate Dyn.,* **14,** 83–100.

Newman, M., and P. D. Sardeshmukh, 1995: A caveat concerning singular value decomposition. *J. Climate,* **8,** 352–360.

Oort, A. H., 1983: Global atmospheric circulation statistics, 1958–1973. NOAA Prof. Paper 14, 180 pp. [Available from Geo-

physical Fluid Dynamics Laboratory/NOAA, P.O. Box 308, Princeton, NJ 08542.]

Palmer, T. N., and D. L. T. Anderson, 1994: The prospects for seasonal forecasting—A review paper. *Quart. J. Roy. Meteor. Soc.*, **120,** 755–793.

Peng, S., and J. Fyfe, 1996: The coupled patterns between sea level pressure and sea surface temperature in the midlatitude North Atlantic. *J. Climate,* **9,** 1824–1839.

Potts, J. M., C. K. Folland, I. T. Jolliffe, and D. Sexton, 1996: Revised ''LEPS'' scores for assessing climate model simulations and long-range forecasts. *J. Climate,* **9,** 34–53.

Renwick, J. A., and J. M. Wallace, 1995: Predictable anomaly patterns and the forecast skill of Northern Hemisphere wintertime 500-mb height fields. *Mon. Wea. Rev.,* **123,** 2114–2131.

Roeckner, E., and Coauthors, 1996: The atmospheric general circulation model ECHAM-4: Model description and simulation of present-day climate. Rep. 218, Max-Planck Institut für Meteorologie, Hamburg, Germany, 90 pp. [Available from MPI, Bundestrasse 55, Hamburg, D-20146 Germany.]

Sarda, J., G. Plaut, C. Pires, and R. Vautard, 1996: Statistical and dynamical long-range atmospheric forecasts: Experimental comparison and hybridization. *Tellus,* **48A,** 518–537.

Shabbar, A., and A. G. Barnston, 1996: Skill of seasonal climate forecasts in Canada using canonical correlation analysis. *Mon. Wea. Rev.,* **124,** 2370–2385.

Stephenson, D. B., 1997: Correlation of spatial climate/weather maps and the advantages of using the Mahalanobis metric in predictions. *Tellus,* **49A,** 513–527.

——, K. R. Kumar, F. J. Doblas-Reyes, J.-F. Royer, F. Chauvin, and S. Pezzulli, 1999: Extreme daily rainfall events and their impact on ensemble forecasts of the Indian monsoon. *Mon. Wea. Rev.,* in press.

Stern, W., and K. Miyakoda, 1995: Feasibility of seasonal forecasts inferred from multiple GCM simulations. *J. Climate,* **8,** 1071–1085.

Vautard, R., C. Pires, and G. Plaut, 1996: Long-range atmospheric predictability using space–time principal components. *Mon. Wea. Rev.,* **124,** 288–307.

Wallace, J. M., C. Smith, and C. S. Bretherton, 1992: Singular value decomposition of wintertime sea surface temperature and 500-mb height anomalies. *J. Climate,* **5,** 561–576.

Ward, M. N., and C. K. Folland, 1991: Prediction of seasonal rainfall in the north Nordeste of Brazil using eigenvectors of sea-surface temperature. *Int. J. Climatol.,* **11,** 711–743.

——, and A. Navarra, 1997: Pattern analysis of SST-forced variability in ensemble GCM simulations: Examples over Europe and the tropical Pacific, *J. Climate,* **10,** 2210–2220.

Zhang, Y., J. M. Wallace, and N. Iwasaka, 1996: Is climate variability over the North Pacific a linear response to ENSO? *J. Climate,* **9,** 1468–1478.