Pattern Analysis of SST-Forced Variability in Ensemble GCM Simulations: Examples over Europe and the Tropical Pacific

M. Neil Ward
IMGA–CNR, Bologna, Italy, and Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma

Antonio Navarra
IMGA–CNR, Bologna, Italy

(Manuscript received 15 April 1996, in final form 9 October 1996)

ABSTRACT

An ensemble of atmospheric general circulation model (GCM) simulations with prescribed sea surface temperature (SST) generates a rich dataset. The main aim here is to advocate and demonstrate an approach to skill and reproducibility based on spatial anomaly patterns. Benefits and applications of this type of analysis include the efficient extraction of the model’s forced variability, guidance on systematic errors in the model’s response to SST forcing, clues to physical mechanisms, and a basis for model output statistics for seasonal forecasting. Some of the possible statistical techniques are illustrated, though the aim is not to provide an exhaustive comparison of the different spatial analysis techniques available. The examples are taken from an ensemble of three GCM integrations forced with observed SST through 1979–88. Boreal summer examples are given for the tropical Pacific and Europe, providing a contrast of a high and a low skill situation, respectively.

1. Introduction

There are now many examples of atmospheric general circulation models (GCMs) forced with the observed time-varying sea surface temperature (SST) (Lau 1985; Latif et al. 1990; Kitoh 1991a, b; Gates 1992; Graham et al. 1994; Lau and Nath 1994; Stern and Miyakoda 1995; Harzallah and Sadourny 1995; Zwiess 1995; Davies et al. 1996; Kumar et al. 1996). Usually, the experiments will be repeated starting from different initial atmospheric conditions, building an ensemble of simulations. The ensemble is needed to isolate the SST-forced component of variability, from the variability internal to the model (Palmer 1993; Barnett 1995). This paper aims to clarify some analysis issues and methods for ensemble simulations, advocating an approach based on spatial anomaly patterns. Examples are included from a new set of GCM integrations to illustrate the techniques.

It is important to assess if the GCM, forced with the evolving observed SST, simulates correctly the observed climate anomalies in each season. Success in this will be referred to as simulation skill. Assessment of simulation skill at each model grid box gives an overly pessimistic view of the model’s ability to simulate large-scale climate anomalies. Assessment of model variability averaged over larger areas is more sensible, es-
pecially for a quantity like precipitation because this removes some of the variability that is generated by smaller-scale processes internal to the atmosphere. A further problem is that the model will likely shift geographically regions of remote teleconnections, so comparison of identically located model and observed regions may be overly pessimistic as well. In this paper we use singular value decomposition analysis (SVDA) (Bretherton et al. 1992; Renwick and Wallace 1995) to identify model and observed regions of covariation. This provides guidance on how to average climate anomalies for investigation of model simulation skill and also yields information on the patterns of model response to SST. Furthermore, by matching observed teleconnection patterns to model teleconnection patterns, the technique may identify useful information and skill in the model that otherwise would remain hidden. Graham et al. (1994) also used a coupled pattern analysis (canonical correlation analysis, CCA) between model and observed data; the concept and its applications are here explored in some detail using SVDA and in the context of the analysis of ensembles.

It is also important to assess the extent to which the ensemble members cluster together. This is commonly referred to as reproducibility and is used as an indication of potential seasonal forecast skill from SST forcing (e.g., Stern and Miyakoda 1995; Rowell et al. 1995). Following on from the work of Madden (1976) and Zwiers (1987), techniques have been developed to assess reproducibility for a given location (Stern and Miyakoda 1995; Rowell et al. 1995). Again, as for simulation skill, assessment of the variability at each grid box may be overly pessimistic. Variability averaged over a larger scale is desirable to get an indication of the reproducibility of the large-scale climate processes. Harzallah and Sadourny (1995) used the principal components of the ensemble mean anomaly to define the atmospheric patterns in their model that appeared to be SST forced. Their work is developed here by focusing on the model anomaly patterns that are reproducible among ensemble members and assessing statistical significance.

Boyle and Sengupta (1993) performed a common principal component analysis (Flury 1988) between observations and the output from four GCM integrations for 1979–88 made with different models. Their analysis sought the common spatial anomaly structures that existed in the models, without regard to when the structures occur in time and thus without regard to skill and reproducibility. One difference here is that model simulation skill (section 3) and model reproducibility (section 4) are targeted and isolated separately by using two different types of analysis. The forced pattern technique discussed in section 4 can also be used to identify the reproducible patterns among different models, though the examples in this paper identify the reproducible patterns among three runs with the same model.

The examples were chosen to illustrate the techniques in two very different situations. One example considers European summers. Previous work (Folland et al. 1988; Ward 1992, 1994) suggests that there may be a modest tropical-forced component to European summer variability. A contrasting example is the central and western tropical Pacific (CWTP) in boreal summer. Here many studies have shown a strong SST-forced component to seasonal atmospheric anomalies, including Bjerknes (1966, 1972), Rasmusson and Carpenter (1982), Meehl (1987), and Deser and Wallace (1990). The examples analyze precipitation since this is an important predictive variable for seasonal forecasting, and through its association with diabatic heating, it is also a useful diagnostic on the functioning of the tropical climate system.

2. Data and GCM

a. Observed rainfall

Rainfall data were taken from Hulme (1994) on a 2.5° lat × 3.75° long grid, built from all available land station monthly means using a Thiessen polygon system. Datasets for July–August (JA) were formed by summing the monthly grid box values, as detailed in Rowell et al. (1995).

b. Observed outgoing longwave radiation

Outgoing longwave radiation (OLR) data at the 2.5° lat × 2.5° long monthly resolution from the National Oceanic and Atmospheric Administration (NOAA) are formed into a 5° lat × 5° long July–September (JAS) dataset. The data are used here as a proxy for tropical rainfall (Gruber and Krueger 1984; Arkin and Ardunay 1989; Chelliah and Arkin 1992; Waliser et al. 1993).

c. Model results

The ECHAM4 GCM (Roeckner et al. 1992, 1996) has been integrated at T30 resolution over the years 1979–88, forced with observed SST (Reynolds and Marisco 1993). Three integrations were made. The first integration (run 1) was initialized with model climatological atmospheric conditions provided by the Max-Planck-Institut (E. Roeckner 1994, personal communication). The second and third integrations (run 2, run 3) were started from the conditions prevalent in run 1 on 1 January 1980 and 1 January 1981, respectively. This procedure was simply employed to generate different but dynamically stable initial conditions for the second and third runs.

3. Patterns to maximize simulation skill

a. Method

European summers can be of very contrasting character depending on whether the midlatitude atmosphere
over European longitudes takes on a strongly zonal westerly flow or a predominantly blocked pattern. Many of these anomalous summers are associated with an eastward extension of anomalies from the traditional Atlantic centers of the North Atlantic Oscillation (Walker and Bliss 1932; van Loon and Rogers 1978; Barnston and Livezey 1987). The model simulations may be generating an approximation to blocked summers and westerly summers in the same years as observed, but the model and observed seasonal atmospheric anomaly patterns may have centers repeatedly and systematically offset so as to mask the skill when the model is verified grid box by grid box. It may therefore be useful to search for ways to average model and observed variability over the domain of study to yield maximum model-observed agreement. Choosing the quantity covariance to measure the agreement, the problem becomes a coupled pattern SVDA (Bretherton et al. 1992) between model and observed fields. The first SVDA mode averages model and observed variability in a way that gives the maximum covariance possible between a model and observed time series. The lower-order modes describe additional model-observed covariation but may quickly degenerate into over fitting to noise, so for this paper just SVDA is analyzed. Variations on this approach would use other coupled pattern analysis methods such as CCA (Barnett and Preisendorfer 1987; Graham et al. 1994) in which the criterion for maximizing agreement between model and observations would be different. The SVDA is attractive for this application because it requires both a good correlation between the derived time series and a good fraction of the model and observed variability to be utilized since covariance is being maximized.

To apply the technique to JA variability over Europe, the cross-covariance matrix is calculated with run 1, run 2, and run 3 making 30 yr of model JA rainfall anomaly fields over Europe for the data matrix on the left-hand side (lhs; see Bretherton et al. 1992), while observed fields of JA rainfall anomalies for 1979–88 form the right-hand side (rhs) data matrix. The observed fields are repeated three times so observations for year $i$ always match up with a model simulation for year $i$. Without loss of generality, the data were standardized, so every grid box has a mean of zero and a standard deviation of 1. This gives equal weight to each grid box, which in the applications here was judged desirable.

The singular value decomposition of the model-observed cross-covariance matrix ($C_{mv}$) yields

$$C_{mv} = USV',$$

(1)

where the columns of $U$ contain the modes associated with the lhs (model fields) and the columns of $V$ contain the modes associated with the rhs (observed fields).

When the ensemble mean of the standardized values are used as the lhs data matrix, it is easy to show that $C_{mv}$ is the same as above (multiplied by a small degrees of freedom adjustment). Thus, an SVDA between the ensemble mean of the standardized anomalies and observations yields identical patterns to the analysis of all three ensemble members. The only difference is the apparent variance explained in the model data (see section 3c).

The time series ($T_m$, $T_v$) associated with the model and observed modes are calculated:

$$T_m = D_m U$$

(2a)

$$T_v = D_v V,$$

(2b)

where $D_m$ and $D_v$ are the standardized model (lhs matrix) and observed (rhs matrix) data.

b. Model verification over Europe and the central and western tropical Pacific

The first SVDA model mode of European summer rainfall (Fig. 1a) contains a center of action through southern Europe. There are weak negative weights to the north of the United Kingdom and Germany (analyses over larger domains show these negative weights stretching west across the central Atlantic). The corresponding observed SVDA mode (Fig. 1b) also has a center of action through southern Europe, shifted west from the model’s center by about 10$^9$ longitude. In northern and central Europe, the negative weights in the observed pattern are much more prevalent than in the model mode. The observed mode is interpreted as a description of the contrasting rainfall anomaly patterns observed in westerly summers (wet in the north, dry in the south) and strongly blocked summers (dry in the north, wet in the south). The observed mode in Fig. 1b is very similar to the first eigenvector of JA precipitation calculated for 1949–88 (not shown).

There are two possible interpretations to the SVDA. First, it is possible that the SVDA is a fit to random noise (Cherry 1997) and the GCM has zero simulation skill over this region. Statistical significance of the mode is discussed in section 3d. The second interpretation is that the GCM mode describes the contrasting states that the GCM tends to take when blocked compared to when westerly type summers are observed. This implies that the amplitude of the simulated GCM mode (Fig. 1c) can be used to skillfully specify rainfall anomalies associated with the observed mode.

In the above example the forced signal was expected to be weak. The CWTP (Fig. 2) provides a contrasting example. Theory predicts a strong SST-forced component to variability in the tropical Pacific (Lindzen and Nigam 1987; Neelin and Held 1987), and many studies have demonstrated a close association between SST and atmospheric circulation in the region related to El Niño–Southern Oscillation (Bjerknes 1966, 1972; Rasmusson

---

1 The method is called SVDA because it uses the singular value decomposition of the cross-covariance matrix between two sets of spatial fields.
Fig. 1. SVDA between model and observed July–August precipitation over Europe. (a) First model pattern (the singular vector with weights multiplied by 100). Dark shading for regions >12, light shading <−12. (b) Same as (a) but for the first observed pattern and (c) SVDA time series for the three model runs (solid) and observed (dashed).

Fig. 2. SVDA between July–September model precipitation and July–September observed outgoing longwave radiation over the central and western tropical Pacific. (a) First model pattern (the singular vector with weights multiplied by 100). Dark shading for regions >6, light shading <−6. (b) Same as (a) but for the first observed pattern and with shading reversed to aid comparison with (a). (c) SVDA time series for the three model runs (solid) and observed (dashed).

and Carpenter 1982; Meehl 1987; Ropelewski and Halpert 1987, 1989; Kiladis and van Loon 1988; Kiladis and Diaz 1989; Deser and Wallace 1990). The domain in Fig. 2 has been shown (Ward and Hoskins 1996) to have a particularly strong local association in JAS between observed SST and near-surface divergence. The GCM precipitation over the region is verified using an SVDA with observed OLR (Fig. 2). The first model SVDA pattern (Fig. 2a) has strongest weights near the equator around 170°E. The weights become negative in the western Pacific but generally remain weak. The corresponding observed mode (Fig. 2b) resembles the model mode very accurately in the eastern node of the pattern (weights are of course reversed in sign since OLR is negatively correlated with rainfall). In the western Pacific west of 150°E, the observed mode has strong weights especially near 0°–10°S, with sign in opposition to the eastern part of the pattern, as found in the near-surface divergence analysis of Ward and Hoskins (1996). The absence of the strong weights in the model’s western Pacific suggests that the model’s mode may lack variability in this region. The negative weights in the model mode appear to cut through the model’s South Pacific convergence zone, rather than extending westward across Indonesia as in the observed mode. The model and observed time series agree almost perfectly (Fig. 2c and Table 1), so the deficiency in the model’s mode would not be a serious problem for specifying western Pacific rainfall variations associated with this first mode. Models are often verified in this region using the sea level pressure Tahiti minus Darwin Southern Oscillation index (SOI) (Walker and Bliss 1932; Troup 1965; Trenberth 1976). The GCM ensemble mean JAS SOI reproduces the observed JAS SOI very accurately (correlation = 0.89) (observed data from R. Allan 1995, personal communication). The SVDA in Fig. 2 may offer useful additional information about model variability in the SOI region.

Composite analyses and point correlation analyses (not shown) all support the existence in ECHAM4 of the anomaly patterns found by the SVDA technique in Figs. 1 and 2. Such a check is necessary to be confident in the patterns since the technique may yield unstable
The squared covariance and squared covariance fraction (Bretherton et al. 1992) do not indicate the fraction of total variance explained by the models. This can be easily computed, since the orthogonality of the SVDA patterns enables the original data to be recovered from the SVDA modes and time series using Eq. (2). As in the family of empirical orthogonal function (EOF) methods (Jolliffe 1986), a percentage of unique recovered (or explained) variance can be associated with each of the observed and model modes (Table 1). For the European example, 25.3% of the observed variance is recovered from the SVDA pattern (Fig. 1b). When the SVDA is performed using all three runs, the first model pattern accounts for 13.2% of the total model variance. When the ensemble mean anomaly is used in the analysis, the SVDA model pattern remains identical but now accounts for 29.4% of the ensemble mean model variance. Calculating the ensemble mean anomaly has therefore removed some of the variance that was internal to the model atmosphere and that had no covariance with observations. The good balance between the variance explained in the model and observed data (29.4% vs 25.3%) can be an indication of a genuine coupled mode. When there is no genuine coupling, often the first SVDA mode is made up of the first EOF of, for example, the model fields (large fraction of model variance explained), while the matching mode for the observed data would be dictated by the covariance constraint, so that it may actually explain very little variance of the observed data.

The correlation between the ensemble mean model series and observed series for SVD1 is 0.83 (Table 1). In a least squares regression sense, the model series specifies $0.83^2 = 68.9\%$ of the variance in the observed SVD1 series. Since the observed mode represents 68.9% of that variance, the total percentage of observed variance specified by the model SVD1 mode is $f_{v1} = 0.689 (0.253)100 = 17.4%$:

$$f_{v1} = r^2 \frac{\sigma^2_{v1}}{\sigma^2_{v_{\text{tot}}}} (100),$$

where $\sigma^2_{v1}$ is the variance of the first observed SVDA time series, $\sigma^2_{v_{\text{tot}}}$ is the total variance of the observed dataset, and $r$ is the correlation between the model (ensemble mean) and observed SVDA time series. When standardized data are used for the SVDA, the statistic $f$ is equivalent to the mean of the squared correlations in the heterogeneous correlation map, as used by Lau and Nath (1994) to measure the strength of association in a coupled mode. The CWTP model precipitation SVD1 mode specifies $f_{v1} = 32.3\%$ of observed OLR variability. In the CWTP, the lower-order SVDA modes are more likely to represent genuine skill than in the European example since a high component of forced variability is expected in regions like the tropical western Pacific.

c. Explained variance

The correlation between the ensemble mean model and observed fields. The first column is the July–August model and observed rainfall over Europe and the second is the July–September model rainfall and observed outgoing longwave radiation over the central and western tropical Pacific (CWTP). Corr is the correlation between the observed SVDA time coefficients and the model SVDA time coefficients.

| Corr (run 1 vs observed) | 0.60 | 0.94 |
| Corr (run 2 vs observed) | 0.83 | 0.96 |
| Corr (run 3 vs observed) | 0.86 | 0.97 |
| Corr (ensemble mean vs observed) | 0.83 | 0.96 |
| Squared covariance | 200.9 | 4152.6 |
| Squared covariance fraction | 44.0% | 68.4% |
| Variance explained: Observed | 25.3% | 35.1% |
| Variance explained: Model total | 13.2% | 22.7% |
| Variance explained: Model ensemble mean | 29.4% | 38.1% |

d. Statistical significance

The SVDA patterns (Figs. 1a,b and 2a,b) were defined to maximize covariance between the model and observed SVDA time series (Figs. 1c and 2c). A Monte Carlo simulation has been used to assess if the covariance found between the series is significantly greater than would be expected by chance. One approach (test 1) is to randomly scramble the order of the 30 GCM integrations and repeat the SVDA. When the order is scrambled, a result is rejected if any of the years remain in their same location since this would make the test conservative in situations where there is a genuine model-observed covariance in the results. Scrambling the 30 GCM integrations makes the assumption that all integrations are independent. Reproducibility among ensemble members compromises this assumption. An al-

---

**Table 1.** Statistics for the first mode in SVDAs between model and observed fields. The first column is the July–August model and observed rainfall over Europe and the second is the July–September model rainfall and observed outgoing longwave radiation over the central and western tropical Pacific (CWTP). Corr is the correlation between the observed SVDA time coefficients and the model SVDA time coefficients.
ternative (test 2) is to test the SVDA between the ensemble mean anomaly and the observations.

Monte Carlo test 1 and test 2 have been applied to the Europe and CWTP examples, using 500 random realizations in each test. The squared covariance for SVD1 in the original model-observed SVDA's (Table 1) is located in the distribution of squared covariances from the Monte Carlo simulations. The CWTP example is clearly significant, with none of the simulations in either test 1 or test 2 beating the actual squared covariance. In test 1 for the Europe example the actual squared covariance (200.9) was beaten by 15% of the random SVDA's. In test 2, the actual squared covariance was beaten by 33% of the random SVDA's. Assessed in these ways, further GCM experiments are needed for statistical significance. However, many of the Monte Carlo SVDA modes will contain model-observed patterns that are clearly unrelated. Ideally, a test is needed that combines explained covariance with a physical feasibility assessment of the two associated patterns. In this spirit, Hense and Romer (1995) used anomaly fields predicted by a simplified linear model of the Tropics to assist in the interpretation of GCM response to SST.

A further way to explore the reality of the modes is to make a cross-validation, repeatedly removing 1 yr from the SVDA and making the SVDA on the remaining 9 yr. The first SVDA mode pattern pair proved extremely stable for the CWTP example (mean pattern correlation among the model modes was 0.95 and among the observed modes was also 0.95). In the European example, the basic features of the first mode were successfully maintained in the cross-validation, but the results did confirm that 10 yr is insufficient to establish the details of the modes with reliability in such low skill situations (mean pattern correlations were 0.84 for the model modes and 0.87 for the observed modes).

4. Patterns to maximize reproducibility

a. Method

Estimating the ratio of internal to externally forced model variability has become an important issue in the analysis of ensemble GCM integrations with prescribed SST (Stern and Miyakoda 1995; Zwiars 1995; Rowell et al. 1995). The ratio calculated is sensitive to the scale and method of averaging the model data. Averaging data over regions of generally coherent model response to SST will remove some spatial variability generated by internal atmospheric processes and lead to higher estimates of the forced component. In the limit, consider projecting the ensemble members onto a base pattern that is chosen to yield time series (one for each member) whose covariances sum to a maximum. Also, consider the simplest example where there are just two ensemble members: run 1 and run 2. A coupled SVDA (Bretherton et al. 1992) between run 1 and run 2 yields a pattern for run 1 and a pattern for run 2 that produce a pair of time series with maximum covariance (SVD1). To identify a single base pattern, run 2 is copied and placed below run 1 on the lhs and run 1 is copied and placed below run 2 on the rhs. Now the SVD1 pattern for the lhs and rhs is identical; it is referred to as forced pattern 1 (FP1).

Consider calculating FP1 in a pair of 10-yr integrations. For an analysis of a specific season, each run has 10 time points (one value per year). FP1 projects onto the ensemble members to yield a pair of time series (the FP1 time coefficients) with maximum covariance. To see this, note that the SVD1 time series, which are now of length 20, can be written as

$$T_{L(1-10)} = D_F$$
$$T_{L(11-20)} = D_F$$
$$T_{R(1-10)} = D_F$$
$$T_{R(11-20)} = D_F$$

where $T_{L(1-10)}$ are time coefficients 1–10 for the lhs, $T_{L(11-20)}$ are time coefficients 11–20 for the lhs, subscript R refers to rhs time coefficients, $D$ is the matrix of standardized data for run i, and $F$ is the first column vector (FP1) returned by the SVDA and maximizes the covariance between $T_L$ and $T_R$. From Eq. (4), it follows that

$$\text{cov}(T_L, T_R) = \frac{\text{cov}(D_F, D_F) + \text{cov}(D_F, D_F)}{2},$$

assuming that $D_F^2 = D_F = 0$. This assumes no systematic difference in a grid box mean value among the ensemble members, which can be expected for the population. Equation (5) can be made valid for the sample if climatological values appropriate to each individual ensemble member are used in the transformation of the raw data into anomalies. From Eq. (5) it follows that maximizing $\text{cov}(T_L, T_R)$ must also maximize $\text{cov}(D_F, D_F)$, which is the covariance between the FP1 time coefficients of the two ensemble members over the analysis period.

Generalization to larger ensemble sizes shows that all possible combinations of the members must be placed opposite each other in the lhs and rhs data matrices and then the matrices copied beneath each other. For example, Table 2 shows the composition for three ensemble members. To calculate the FPs over Europe in JA using the three GCM runs for 1979–88, the lhs and rhs matrices for calculating the cross-covariance matrix contain 60 rows, while the number of model grid boxes in the analysis domain defines the number of columns in each matrix.

---

2 The patterns must be identical but may be of opposite sign. An opposite sign mode is associated with negative reproducibility since the base pattern is isolating negative covariance among ensemble members.
TABLE 2. Composition of the left-hand side (lhs) and right-hand side (rhs) data matrices for an FP analysis with three ensemble members.

<table>
<thead>
<tr>
<th>lhs</th>
<th>rhs</th>
</tr>
</thead>
<tbody>
<tr>
<td>run 1</td>
<td>run 2</td>
</tr>
<tr>
<td>run 2</td>
<td>run 3</td>
</tr>
<tr>
<td>run 1</td>
<td>run 3</td>
</tr>
<tr>
<td>run 2</td>
<td>run 1</td>
</tr>
<tr>
<td>run 3</td>
<td>run 2</td>
</tr>
<tr>
<td>run 3</td>
<td>run 1</td>
</tr>
</tbody>
</table>

Inspection of the lhs and rhs matrices in Table 2 shows that their cross-covariance matrix $C$ is made up of elements $c_{ij}$ that are an equally weighted combination (the average multiplied by a constant) of all possible covariances between grid box $i$ of one ensemble member and grid box $j$ of another ensemble member. That is, for a given element $c_{ij}$, the value in the matrix $C$ is

$$c_{ij} = \frac{1}{n(n-1)} \sum_{k=1}^{n} \sum_{l=1, l \neq k}^{n} c_{kl} / 2n(N-1),$$

where $n$ is the ensemble size and $c_{ij}$ is the covariance between grid box $i$ in ensemble member $k$ and grid box $j$ in ensemble member $l$. The global mean of 0 is used in the covariance calculations, and the last term in Eq. (6) is a degrees of freedom adjustment for each element, where $N$ is the number of time points in each run. This adjustment simply multiplies $C$ by a constant and so does not affect the SVDA patterns. The matrix $C$ is symmetric, with $c_{ij} = c_{ji}$, so the patterns yielded by a singular value decomposition could also be found by calculating the eigenvectors of $C$. FP1 has been calculated for Europe and CWTP (Figs. 3 and 4), and their interpretation is discussed further in section 4b.

An alternative way to identify the model patterns that are SST forced is to calculate the EOFs of the ensemble (ENS) mean anomaly fields (ENS−EOF). The analysis below shows why the ENS−EOF patterns are different from the FPs for a finite sample size. The ENS−EOF1 pattern is defined to yield a set of time coefficients $z_e$ with maximum variance. For simplicity, consider an ensemble of just two members. The time coefficients $z_e$ can be written as a combination of the anomalies in run 1 ($D_1$) and run 2 ($D_2$):

$$z_e = \frac{1}{2} \sum_{i=1}^{N} a_{i} d_{i1} + \frac{1}{2} \sum_{i=1}^{N} a_{i} d_{i2} = z_1 + z_2,$$

where $N$ is the number of grid boxes in the analysis and $a_i$ are the ENS−EOF weights (spatial loadings), and $z_1$ and $z_2$ can be viewed as time coefficients for each of the two runs. Now, since ENS−EOF1 maximizes the variance of $z_e$, this is equivalent to maximizing

$$\text{var}(z_1 + z_2) = \text{var}(z_1) + \text{var}(z_2) + 2r[\text{var}(z_1)\text{var}(z_2)]^{1/2},$$

where $r$ is the correlation between $z_1$ and $z_2$. In contrast, FP1 is derived to just maximize the last term in Eq. (8) (the covariance). Equation (8) shows that the ENS−EOFs give less weight on maximizing reproducibility. The FPs and the ENS−EOFs are closely related. It is easy to show that if the condition $l \neq k$ is removed from Eq. (6), the patterns that result from matrix $C$ are now identical to the ENS−EOFs. In the FP framework, removing the condition $l \neq k$ is like adding an exact
rerun of every ensemble member to the FP analysis; so in Table 2 an extra row is added for each of the three runs, with run $i$ opposite run $i$, inflating apparent reproducibility. The adjustment in Harzallah and Sadourny (1995) to make an unbiased estimate of the ensemble mean variance makes their ENS–EOFs almost identical to the FPs. The adjustment effectively tries to remove the internal atmospheric variance and covariance that enters $C$ through the cross-multiplication of run $i$ with run $i$. This is an example of the noise contamination problem discussed in Bretherton et al. (1992). For an infinite sample size, the contribution from internal variance to the ensemble mean tends to zero, and the FPs and ENS–EOFs become equivalent.

The (variance adjusted) ENS–EOFs were used by Harzallah and Sadourny (1995) to represent the forced variability. The above analysis interprets the meaning of the ENS–EOFs in terms of the concept of reproducibility. The approach (using either FPs or variance adjusted ENS–EOFs) appears to be an efficient way to study reproducibility in models and establish regions of statistical significance (see next section).

### b. Forced patterns over Europe and the central and western tropical Pacific

The FP1 of European summer rainfall (Fig. 3a) is very similar to the pattern that maximized model simulation skill (Fig. 1a). This is an encouraging result, showing that the model’s most reproducible pattern has turned out to be the pattern that best matches the observed variability. The strength of the forced signal enables ENS–EOF1 (not shown) to be almost identical to FP1. Consistent with the analysis in the previous subsection, the time series associated with FP1 have a slightly higher mean covariance than do the time series calculated from the ENS–EOF1 pattern (Table 3).

To assess if the reproducibility associated with FP1 (Fig. 3b) is statistically significant, the 30 model experiments have been randomly scrambled 500 times (see section 3d). The FP analysis was repeated on the 500 random realizations, and none of the squared covariances was greater than the result in Table 3. Thus, reproducibility in the FP over Europe is highly statistically significant ($p < 0.2\%$). During the Monte Carlo experiment, opposite sign patterns are often found for the lhs and rhs, implying a base pattern that yields negative covariance among ensemble members. When constructing the distribution of squared covariances, those FPs associated with opposite sign patterns are given a zero-squared covariance.

As in the model-observed SVDA (section 3c), the FPs can be used to reconstruct the original data, and a fraction of the total variance can be associated with FP1. For the European example, FP1 reconstructs 14.3% of the total variance in the 30 model JA fields. The time series associated with FP1 (Fig. 3b) can be subjected to an analysis of variance (Rowell et al. 1995) to determine an estimate of the common (SST-forced) variance compared to the total variance in the three series. This fraction ($f_{\text{SST}}$) is 79.8%.

The FP1 in the CWTP (Fig. 4a) also matches up with the pattern that maximized model skill (Fig. 2a). Though ENS–EOF1 is a very similar pattern (not shown), the ENS–EOF1 time series for each member again show the slightly lower covariance (Table 3) expected from the statistical theory [Eq. (8)]. The reproducibility of the FP, as measured by the squared covariance, is highly statistically significant (not beaten by any of the Monte Carlo random realizations). The SST-forced variability in the FP1 time series (Fig. 4b) is extremely high ($f_{\text{SST}} = 96.7\%$). FP1 here reconstructs just 23.4% of the total variance, yet represents a large fraction (72.9%) of the total squared covariance. There must therefore be some variance at the gridbox scale that is not reproducible, whereas the large-scale pattern extracted by FP1 is almost perfectly reproducible.

### 5. Discussion and conclusions

Spatial analysis methods have been applied to an ensemble of GCM integrations with prescribed SST. We have advocated the analysis of large-scale spatial anomaly patterns for studying model skill and reproducibility because they smooth out some spatial variations generated by processes internal to the atmosphere and enable a better focus on the forced component of climate variability. The techniques have efficiently extracted the large-scale forced signal from the ensemble. Applying such analyses can be expected in principle to make a modest reduction to the ensemble size that is needed for identifying simulation skill in long integrations and for making seasonal forecasts in shorter integrations.

The first issue addressed was model simulation skill. A coupled SVDA between model and observed fields identifies the model and observed base patterns that yield time series with maximized covariance. The SVDA patterns describe the optimum geographical shift and smoothing to match model and observed interannual

---

**Table 3. Statistics for the first FP and ENS–EOF over Europe and CWTP.**

Mean covariance and correlation for the FP are the mean values calculated using the three ensemble member FP time series (Figs. 3b, 4b). For ENS–EOF, three time series were created by projecting each ensemble member onto the ENS–EOF1 pattern.

<table>
<thead>
<tr>
<th></th>
<th>Europe</th>
<th>CWTP</th>
</tr>
</thead>
<tbody>
<tr>
<td>FP: Mean covariance</td>
<td>13.37</td>
<td>77.52</td>
</tr>
<tr>
<td>FP: Mean correlation</td>
<td>0.892</td>
<td>0.975</td>
</tr>
<tr>
<td>FP: Variance of ensemble</td>
<td>14.39</td>
<td>78.36</td>
</tr>
<tr>
<td>time series</td>
<td>144.64</td>
<td>501.54</td>
</tr>
<tr>
<td>FP: Squared covariance</td>
<td>48.9%</td>
<td>72.9%</td>
</tr>
<tr>
<td>FP: Squared covariance fraction</td>
<td>13.19</td>
<td>77.46</td>
</tr>
<tr>
<td>ENS – EOF: Mean covariance</td>
<td>0.863</td>
<td>0.971</td>
</tr>
<tr>
<td>ENS – EOF: Mean correlation</td>
<td>14.58</td>
<td>78.44</td>
</tr>
</tbody>
</table>
variability. Differences between the model SVDA patterns and observed SVDA patterns may describe systematic errors in the model's response to SST, information of use to model developers and also to seasonal forecasters who wish to statistically correct the output from existing models. Regressions between model SVDA time series and verifying gridbox SST are already being applied to correct seasonal forecasts of SST (S. Zebiak 1996, personal communication), while CCA is being explored in the same manner for atmospheric seasonal forecasts (R. Livezey and T. Smith 1997; manuscript submitted to J. Climate.) Only the first coupled mode has been analyzed in this paper, but it is clear that the lower modes will in most instances contain further useful model-observed covariation. Future work could investigate the number of significant modes by applying the spatial degrees of freedom concept (e.g., Wallace et al. 1991).

The second issue addressed was the reproducibility among the ensemble members. The cross-covariance among all ensemble members is used to construct a cross-covariance matrix for decomposition. The first eigenvector of the matrix is the spatial pattern that yields a set of time series (one for each of the ensemble members) whose covariances sum to a maximum. The pattern is referred to as the model's first forced pattern (FP1), based on the assumption that the similarity between the ensemble members derives from a common response to the same varying SST (and sea-ice) forcing used in the experiments. No attempt is made to suggest whether FP1 is in fact an internal mode of the atmosphere whose probability of occurrence is modified by the SST forcing. FP1 may have no relation at all to internal atmospheric modes, or it may be a version of an internal mode distorted by the SST forcing. Note also that interactions with the land surface can also influence the character of the atmospheric modes in the model.

In other integrations, the forcing may take alternate forms, such as observed time-varying land surface features. The FP method also yields lower-order forced patterns, but their (statistical) significance will require careful assessment. For example, the lower-order modes may contain reproducibility that is not part of a coherent recurring pattern but rather results from unique features of the SST-forcing in given years. Such modes may have statistically significant temporal covariance associated with them, whereas the associated spatial mode has little physical meaning.

To place the above approaches in a broader context, Table 4 summarizes the combination of possible cross-covariance matrices that can be calculated. Subsets or averages of subsets of the matrices are subjected to singular value decomposition (or alternatively eigenanalysis if the matrix is symmetric) to yield various types of analyses:

1) The SVDA between model and observations in section 3 decomposed the average (expected value) of the matrices $C_{11}$, $C_{22}$, and $C_{33}$ (or equivalently $C_{22}$, $C_{33}$, and $C_{32}$). Estimating the expected value is a delicate issue, and the analyses in sections 3 and 4 concentrated on statistics over the sample. Monte Carlo simulations of small samples suggest the estimates used are not seriously biased, though for the FP analysis covariance estimates may be improved by centering each ensemble member relative to its own mean prior to analysis.

2) If each run is from a different model, it may be useful to modify (1) and allow, for each observed SVDA mode, a corresponding mode that has a different pattern for each model. Derivations similar to those used for Eqs. (4) and (5) show that this can be achieved by decomposing the subset $C_{11}$, $C_{22}$, and $C_{33}$, where the matrices are stacked to form a single larger matrix.

3) To get, for each mode, a single base pattern that maximizes cross-covariance among runs, the average of $C_{11}$, $C_{12}$, $C_{21}$, $C_{22}$, $C_{13}$, and $C_{23}$ is decomposed (the FP analysis in section 4).

4) The inclusion of $C_{12}$, $C_{22}$, and $C_{33}$ to the subset of matrices being averaged in (3) yields a cross-covariance matrix whose decomposition is equivalent to an EOF analysis of the ensemble mean anomaly. The variance adjustment in Harzallah and Sandourny (1995) makes their ensemble mean EOFs almost identical to the FPs.

5) If each run is from a different model, it may be useful to modify (3) so that each mode can have a different base pattern for each model. Derivations similar to those for Eqs. (4) and (5) can be used to show the combination of cross-covariance matrices to stack (the combination does not include the covariance of each run with itself).

6) If the nine model cross-covariance matrices are stacked exactly as in Table 4, then analysis (5) becomes a standard combined principal component analysis (Bretherton et al. 1992) for the three model runs. For studying reproducibility, this suffers analogous problems to those of the ensemble mean EOF.

---

Table 4. Possible cross-covariance matrices for three GCM runs and a set of verifying observations. $R_i$ is the GCM run $i$, and $C_{ij}$ is the cross-covariance matrix between run $i$ and observations.

<table>
<thead>
<tr>
<th>$C_{ij}$</th>
<th>$C_{ij}$</th>
<th>$C_{ij}$</th>
<th>$C_{ij}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_{11}$ = R1 vs R1</td>
<td>$C_{12}$ = R1 vs R2</td>
<td>$C_{13}$ = R1 vs R3</td>
<td>$C_{1v} = R1$ vs OBS</td>
</tr>
<tr>
<td>$C_{21}$ = R2 vs R1</td>
<td>$C_{22}$ = R2 vs R2</td>
<td>$C_{23}$ = R2 vs R3</td>
<td>$C_{2v} = R2$ vs OBS</td>
</tr>
<tr>
<td>$C_{31}$ = R3 vs R1</td>
<td>$C_{32}$ = R3 vs R2</td>
<td>$C_{33}$ = R3 vs R3</td>
<td>$C_{3v} = R3$ vs OBS</td>
</tr>
<tr>
<td>$C_{1v}$ = OBS vs R1</td>
<td>$C_{2v}$ = OBS vs R2</td>
<td>$C_{3v}$ = OBS vs R3</td>
<td>$C_{vv} = OBS$ vs OBS</td>
</tr>
</tbody>
</table>
because it includes the covariance of each run with itself.

An ensemble of three GCM integrations forced with the observed SST for 1979–88 was used to illustrate techniques (1) and (3). One set of examples concentrated on the model’s simulation of precipitation during Boreal summer over Europe. Previous research indicated a modest forced component may be expected. Using the FP and ensemble mean EOF methods, a pattern of July–August precipitation anomaly was identified over Europe with good reproducibility. Monte Carlo tests showed the pattern’s reproducibility, as measured by the covariance among the ensemble member FP time coefficients, to be significant at <0.2% (better than 1 in 500). The FP was weighted mainly through southern Europe.

It is an encouraging result if the model’s FP1 is very similar to the model base pattern that best maximizes model simulation skill. For the European summer example this seemed to be true because in an SVDA between model and observed precipitation, the model’s SVD1 mode is very similar to FP1. The observed SVD1 mode resembles the known first mode of atmospheric variability in summer in the region: a contrast between blocked summers (dry in the north, wet in the south) and westerly summers (reversed rainfall anomaly pattern). The SVDA suggests a tendency for the model and observed modes to respond to a given SST forcing and produce similar mode amplitudes. Results like this are of potential value to seasonal forecasting. The GCM, forced with a predicted SST field, can be used to specify the rainfall anomaly across Europe based on the amplitude of the predicted model mode. A Monte Carlo test did not show the strength of the association between the model mode and observed mode to be statistically significant. The example was presented here to illustrate the techniques, and the proof of genuinely useful skill in this region with ECHAM4 awaits further model experiments and empirical analysis.

The FP analysis of July–September precipitation in the CWTP identified a highly reproducible pattern, as expected for this region. The SVDA between model precipitation and observed OLR identified modes that have near-perfect temporal agreement. However, the model mode contained much less variability in the western tropical Pacific than the observed mode, suggesting the model’s response to SST in this region may be rather different from observed. Such information is useful for model developers.

Acknowledgments. The authors are grateful to Kiku Miyakoda and Mike Richman for useful discussions, to Erich Roeckner and MPI for cooperation in running the GCM, to Vincent Moron and John Janowiak for assistance with the OLR data, and to the referees for constructive comments. The work was supported under the EU DICE project EVSV-CT94-0538.

REFERENCES


Kiladis, G. N. and H. van Loon, 1988: The Southern Oscillation. Part VII: Meteorological anomalies over the Indian and Pacific sec-


