Impact of ensemble size on the assessment of model climate signal

Z.X. Li

Laboratoire de Météorologie Dynamique du CNRS,
Ecole Normale Supérieure, 24 rue Lhomond,
75231 Paris cedex 05, France

Abstract: The atmospheric general circulation model, developed at the Laboratoire de Météorologie Dynamique, is forced by the observed sea surface temperature for the period 1979-94. An ensemble of eight simulations is realized with different initial states. The approach of ensemble simulation permits to study the model's internal variability versus the external forced variability. A statistical attempt is made to evaluate the influence of the ensemble's size on model's reproducibility. As an example, the Northern Hemisphere winter circulation and its connection to the tropical sea surface temperature are examined.

1. Introduction

Climate simulations realized with an atmospheric General Circulation Model (GCM) are mainly a boundary-forced problem, the forcing being sea surface temperature (SST) and sea ice distribution. However, recent results obtained by, among others, Harzallah and Sadourny (1995), Stern and Miyakoda (1995), and Barnett (1995) demonstrate that model internal variability can be large and it obscures sometimes the good detection of climate signal. Although the cause of this internal variability is not fully determined, it seems that ensemble approach can overcome the difficulty, since we can then enhance the signal/noise ratio. In the scientific community, ensemble approach is more and more employed although the need on computing resource is largely increased. One important question is thus the following: what is the minimum number of runs necessary to overcome the model internal variability and to study, with confidence, the model's response to external forcing ? It is clear that the answer depends upon the signal to study and the method used to detect the signal.

For a climate problem, both signal and noise levels are unknown a priori. However when a large number of simulations are realized, we can make statistical test to check if a climate signal is significant. An alternative method is the Monte-Carlo technique, i.e., we increase systematically the ensemble size and search for the point where the result becomes stable. This saturation point is precisely the optimum number of realizations to assess the signal.

In this paper, we present some results of the Monte-Carlo approach. The studied signal is the interannual scale signal of El Niño / South Oscillation (ENSO), but the methodology can be easily extended to decadal variability. Section 2 describes the model used and the climate signal to study. Section 3 presents the impact of ensemble size on the assessment of climate signal.

2. Model and climate signal

The model used in this study (LMD-Z, version 1) is derived from the LMD standard GCM (Sadourny and Laval 1984). But the code has been rewritten with more physical consistency, and organized in a more flexible manner for utilization. In the standard model used in LMD, the meridional discretization was regular in sines of the latitude. This made the model to have a poor resolution for high latitudes and thus the simulated atmospheric circulation for these regions tended to loss some reality. That is why the model's grid points in the present study are chosen to be regularly distributed in the longitude-latitude coordinates with resolution of $5° \times 4°$. The model has 11 vertical layers, unevenly spaced to allow a better resolution in the boundary layer. The time step is 6 minutes, however, the contributions of physical parameterizations are evaluated only every 30 minutes.

The model was integrated from January 1979 to December 1994. We used the observed monthly SST and sea-ice distributions based on the COLA-CPC analyses (Reynolds 1988) as the boundary condition of the model. The ensemble of our simulations contains 8 members. The initial state for the first run was from a 6-month simulation forced by the climatological SST, while the initial state for each of the subsequent realizations were assigned with the same values as those at the last time step of the preceding integration. The present study hence entails altogether $16 \times 8 = 128$ years of model integration. Figure 1 shows the SST anomalies of the Niño-3 region (5°S-5°N and 92°W-150°W) which provides a good proxy (Niño-3 SST index) for El Niño events. Note that a low-pass filtre has been applied before drawing Fig. 1 in order to keep only the variability with time period longer than 8 month. Several warm and cold episodes occurred over the experiment period. We can note the El Niño events of 1982/1983, 1987, 1991/1992 and the most pronounced La Niña event of 1988.

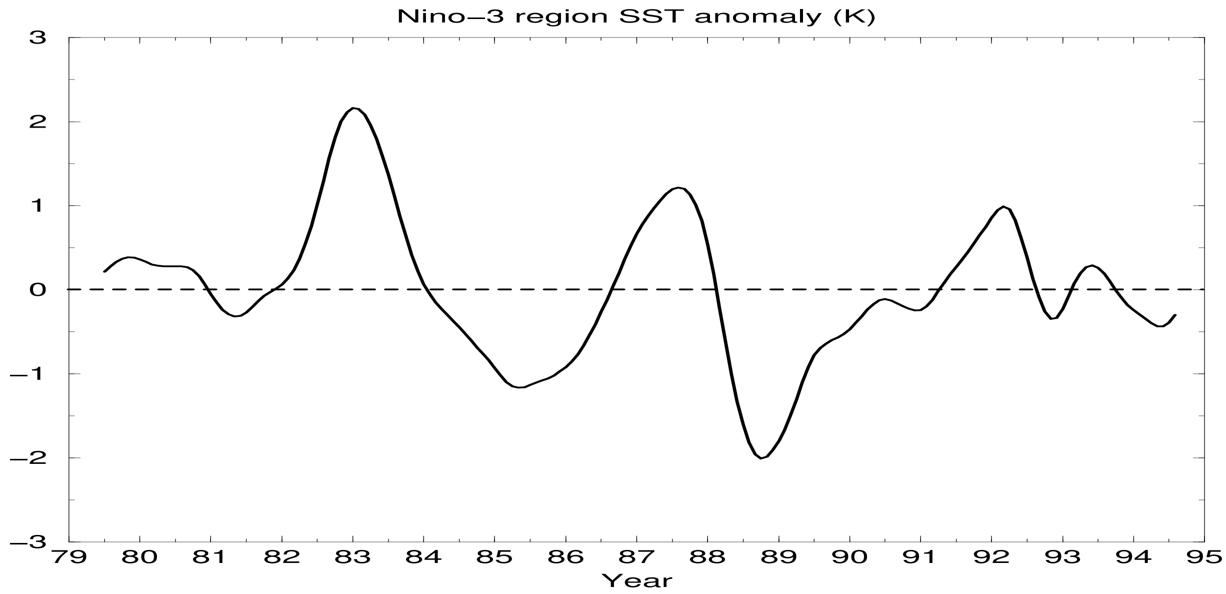

FIG. 1: Filtered sea-surface-temperature anomalies (°C) of the Niño-3 region (equatorial Pacific: 5°S-5°N and 92°W-150°W). The labels of abscissa are plotted at the beginning of each year.

The climate signal that we would like to study is the teleconnection between the tropical Pacific SST and the Northern Hemisphere winter circulation. The latter can be well represented by the 500-hPa geopotential height. We will show, in the rest of the section, that this signal is well captured in the ensemble mean of our simulations. Comparison will be performed against NCEP re-analysis.

The teleconnection detection method is the SVD (Singular-Value-Decomposition) analysis, which is a powerful technique for identifying the dominant coupling modes between two data fields (Bretherton et al. 1992; Wallace et al. 1992; Lau and Nath 1994, among others). Note that the SVD technique is here applied to the cross-correlation matrix, i.e., the anomaly fields have been normalized by the local standard deviations in prior to the SVD analysis. Furthermore, in order to take into account the spatial inhomogeneity of the grid, the anomaly fields are also normalized by $\cos\phi$, where ϕ is the latitude of the grid.

The left three panels of Fig. 2 show the leading SVD mode obtained by using the **ensemble-mean** signal of the model. The middle three panels are from the NCEP re-analysis of the same period (1979-94). We should emphasize that the NCEP data have been objectively interpolated to the model's grid before the statistics is applied, so that the analysis was done in the same manner for both model and NCEP data. The top and middle contour panels give the heterogeneous correlation maps for the height and SST anomalies respectively, and the bottom panels their time coefficients for the leading mode. For the tropical Pacific SST, it is a typical El Niño structure for both observation and simulation. The resemblance between the model and the re-analysis is also good for the 500-hPa geopotential height, both of which show the same zonally-coherent structure in the tropics and several centers of action in the extratropics. The dominant structure is the PNA pattern, but other teleconnection patterns are also superimposed on this leading structure. Several discrepancies should be noted: the negative centers over North of Japan and North Atlantic are absent in simulation. Compared to the observation, a pair of centers over East and central Asia is eastward shifted to China and a negative center of the PNA over Alleutian is shifted westward.

To make sure that the obtained SVD structures are not statistically spurious results, we also made EOF analysis separately for each of the two fields (tropical SST and 500-hPa geopotential height). The leading EOF structures (not shown here) are very close to the leading SVD structures.

3. Impact of ensemble size

We now perform, once again, the SVD as in Section 2. But instead of using the ensemble-mean data, we combined the eight simulations (16 years each) together to yield a long single simulation of 128 years. So the total signal is now analyzed. The leading SVD structure is shown in Fig. 2c. The comparison with Fig. 2a reveals that the spatial structures are roughly the same, but the correlation is generally weaker. The comparison with NCEP re-analysis (Fig. 2b) is less satisfactory in Fig. 2c than in Fig. 2a. It is clear that the internal model variability modifies the results quantitatively, but still the results are qualitatively similar and remain robust.

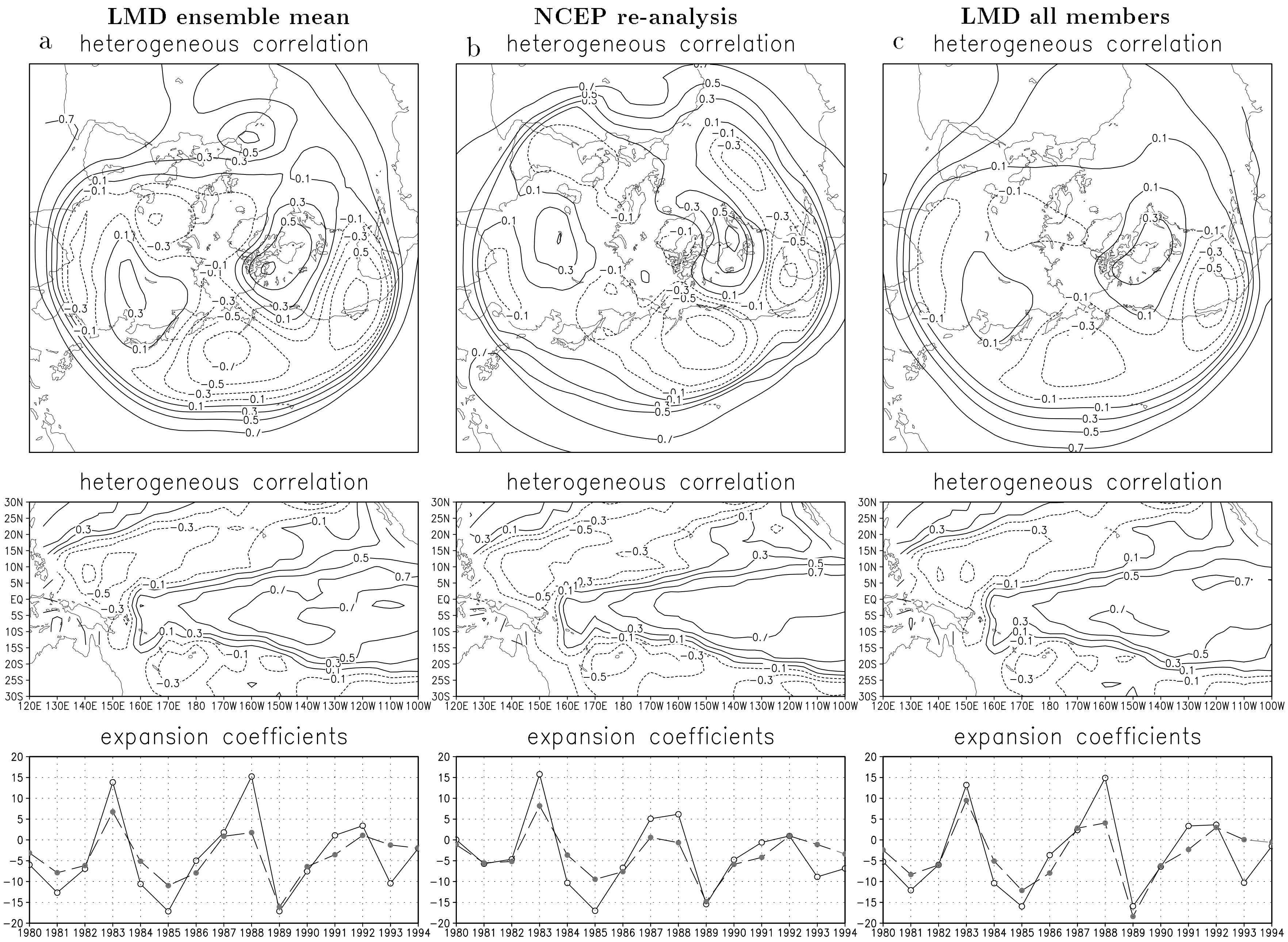

FIG. 2: Leading SVD modes for respectively the model's ensemble mean (left three panels), NCEP re-analysis (middle three panels) and all ensemble members (right three panels). The top contour maps are for the Northern Hemisphere 500-hPa geopotential height, and middle ones the tropical Pacific sea surface temperature. The bottom curves show the expanded time coefficients.

We now evaluate the impact of ensemble size. Our approach is a Monte-Carlo method, already used by Cheng et al. (1995) and Richman (1986) for a similar purpose. We repeat the SVD analysis to subsets of the eight simulations. We can construct eight classes of the subsets based on the eight individual runs. For the subset n, the number of possible combinations is C_8^n. For convenience, the structure of the subset 8 (it is the total ensemble) will be referred to as master structure (Fig. 2c). The structures issued from other subsets will be called perturbed structures. We can now define formly a congruence coefficient to measure the resemblance between the master structure and the perturbed ones:

$$g = \frac{\sum(m_i p_i)}{[(\sum m_i^2)(\sum p_i^2)]^{1/2}}$$

where m_i and p_i denote elements of the respective (master or perturbed) structure vectors. The form of this coefficient is similar to a correlation coefficient, as the possible values range from $+1$ (perfect agreement) through 0 (completely no relationship) to -1 (perfect agreement if sign reversed). As stated in Richman (1986), the congruence coefficient is a better measure than the pattern correlation coefficient in the context of quantitatively assessing the goodness-of-match of loading vectors since the former measures the similarity between two patterns without removing their respective means. However the congruence coefficient is biased towards a higher value than the correlation coefficient. According to Richman (1986) and the references therein, it is reasonable to attach the following goodness-of-match labels to specific ranges of absolute values: 0.98 to 1 (excellent match), 0.92 to 0.98 (good match), 0.82 to 0.92 (borderline match), less than 0.82 (poor match).

TABLE 2. Number of combinations whose congruence coefficient is in the specific match range: excellent, good, borderline and poor. The percentages in parenthesis indicate the ratios over the total number of combinations which is shown in the last column.

	excellent $g \geq 0.98$	good $g \geq 0.92$	borderline $g \geq 0.82$	poor $g < 0.82$	total
subset 1			3(37%)	5(63%)	8
subset 2		8(29%)	18(64%)	2(7%)	28
subset 3		40(71%)	16(29%)		56
subset 4	9(13%)	56(80%)	5(7%)		70
subset 5	23(41%)	33(59%)			56
subset 6	21(75%)	7(25%)			28
subset 7	8(100%)				8

Table 2 shows the results of our calculation. We can see that, for a single realization, five out of the eight cases are under the borderline. However for a combination of five simulations, all the congruence coefficients are in the good or excellent range. We could conclude that at least 5 or 6 simulations are necessary to obtain a robust assessment of the teleconnection between the tropical SST anomalies and the 500-hPa height anomalies over the entire Northern Hemisphere. The time period of our study (1979-94) includes roughly three ENSO cycles, our conclusion may be then translated into another form, i.e., at least fifteen ENSO cycles are necessary for a robust identification of associated teleconnection patterns in the Northern Hemisphere. Here we used implicitly the assumption that the climate system is an ergodic one, without any significant lower-frequency modulation of the ENSO variability.

To demonstrate that the minimum number of realizations is signal-dependent, we repeated the same analysis as above, but this time it was based on the relation between tropical Pacific SST and 200-hPa velocity potential between 45°S and 45°N. We find no single case in which the congruence coefficient falls below the borderline (Table 3). A robust detection of the ENSO signal in the velocity potential can be obtained based on just three (or even two) realizations. This result is consistent with the strong coupling in the tropics between the SST and atmospheric convective anomalies associated with ENSO.

TABLE 3. Same as in Table 2, but for the 200-hPa velocity potential over the region of 45°S to 45°N instead of the 500-hPa geopotential height over the whole Northern Hemisphere.

	excellent $g \geq 0.98$	good $g \geq 0.92$	borderline $g \geq 0.82$	poor $g < 0.82$	total
subset 1	1(13%)	4(50%)	3(37%)		8
subset 2	13(46%)	14(50%)	1(4%)		28
subset 3	44(79%)	12(21%)			56
subset 4	65(93%)	5(7%)			70
subset 5	56(100%)				56
subset 6	28(100%)				28
subset 7	8(100%)				8

4. Conclusion

By using an atmospheric GCM forced through the observed SST from January 1979 to December 1994, we have completed an ensemble experiment with eight members. Our results show that a single model simulation is inadequate and often misleading for studying the interannual climate variability, since the internal model variability (climate noise) is in general larger than the forced variability (climate signal), except for the tropical strip. In general, the length of simulation and the number of ensemble depend upon two factors: the signal-to-noise ratio and the degrees of freedom contained in the experiment. If a signal is absent, it can not be detected with a simulation of even infinite length. On the other hand, if the number of degrees of freedom (the number of ENSO cycles in our case) is small, the signal's statistic significance is not warranted.

By using a Monte-Carlo approach, we evaluated the influence of the internal model variability on the external climate signal and showed that fifteen ENSO cycles are the minimum length of simulation to assess the teleconnection between the tropical Pacific SST and the 500-hPa circulation for the whole Northern Hemisphere. However, this minimum length can be reduced if we are only interested in tropical signals. The Monte-Carlo approach used here can be generalized to other tests of climate reproducibility.

Acknowledgments. This work is partly supported by the Environment Program of the European Commission and the French national program on climate dynamics (PNEDC).

REFERENCES

Barnett, T. P., 1995: Monte Carlo Climate Forecasting. *J. of Climate*, **8**, 1005-1022.

Bretherton, C. S., C. Smith, and J. M. Wallace, 1992: An intercomparison of Methods for Finding Coupled Patterns in Climate Data. *J. of Climate*, **5**, 541-560.

Cheng, X., G. Nitsche, and J. M. Wallace, 1995: Robustness of low-frequency circulation patterns derived from EOF and rotated EOF analyses. *J. of Climate*, **8**: 1709-1713.

Harzallah, A., and R. Sadourny, 1995: Internal versus SST-forced atmospheric variability as simulated by an atmospheric general circulation model. *J. of Climate*, **8**, 474-495.

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 global atmosphere-ocean system. *J. of Climate*, **7**, 1184-1207.

Reynolds, R. W., 1988: A real-time global sea surface temperature analysis. *J. of Climate*, **1**, 75-86.

Richman, M. B., 1986: Rotation of principal components. *J. of Climatology*, **6**: 293-335.

Sadourny, R., and K. Laval, 1984: January and July performance of the LMD general circulation model. *New Perspectives in Climate Modelling*, A. Berger and C. Nicolis, Eds., Elsevier, 173-198.

Stern, W., and K. Miyakoda, 1995: Feasibility of Seasonal Forecasts Inferred from Multiple GCM Simulations. *J. of Climate*, **8**, 1071-1085.

Wallace, J. M., C. Smith, and C. S. Bretherton, 1992: Singular Value Decomposition of Wintertime Sea Surface Temperature and 500-mb Height Anomalies. *J. of Climate*, **5**, 561-576.