SHORT CONTRIBUTIONS

Comparison of a sequence of model generated 500 mb-topographies with climate

By HANS VON STORCH, *Meteorologisches Institut der Universitaet Hamburg, Bundesstrasse 55, 2 Hamburg 13, F.R. Germany*

ABSTRACT

Based on a sample of 11 observed Januaries, a 95% probability region for the hemispheric 500 mb topography was estimated in order to see whether a 75 d sequence of model-generated fields (starting from realistic initial conditions) shows the tendency to move away from climate with time or not. It turns out that such a tendency does not exist, but 20% of the simulated states lie outside the 95% probability region, indicating that sometimes significantly unrealistic states are simulated.

1. The statistical frame

The primary goal of numerical experiments with GCMs is to reproduce the statistical behavior of the atmosphere in the expectation that the detailed structure of the model calculation may also occur in nature. The question then arises whether the simulated states show the tendency to move away from real climate. We propose a solution to this problem by means of a statistical test with given risk using the statement "the model generated state does not occur in nature" as an alternative. Since the simplest attempt to make a series of univariate tests does not give a risk (Chung and Fraser, 1958), a multivariate comparison is necessary.

In order to have a probability model with moderate dimensions, some, say n, amplitudes of 500 mb topographies in a spectral representation are considered. The set of amplitudes of the spectral components possible in nature is our "climate ensemble" ε. Assuming the probability distribution on ε to be normal, the isofaces of the corresponding probability density function are surfaces of ellipsoids centered around a mean vector μ, with shapes given by a covariance matrix S. For each probability $p \in (0, 1)$ there is a unique ellipsoid $E(p)$ containing a random $x \in \varepsilon$ with probability $P(x \in E(p)) = p$. Further on, the function g defined by $g(x): = (x - \mu)' \, S^{-1}(x - \mu)$ is χ_n^2-distributed with n degrees of freedom, and $x \in E(p)$ is true if and only if $g(x)$ is lesser than the p-fractile of χ_n^2.

The algorithm to decide whether a model generated state belongs to the climate ensemble is simply: After fixing subjectively a significance level $\tilde{p}$, usually $\tilde{p} = 0.95$, the model generated state x is said to be admissible in the climate ensemble, when $x \in E(\tilde{p})$ is true. Then a fault probability ("risk") of $1 - \tilde{p}$ results, if $x \in \varepsilon$ and $(1 - \tilde{p})*100\%$ of all states occurring in nature will fail the test. The risk is valid only for the application to an individual field. When a sequence of (daily) fields is investigated, no risk can be calculated. What remains is that a field failing the test lies outside $E(\tilde{p})$, and that, on an average, $(1 - \tilde{p})*100\%$ of atmospheric fields will fail the test.

Since the true mean vector μ and the covariance matrix S are unknown, they have to be estimated by atmospheric data. For this purpose we assume that the year-to-year sequence of a single month defines a stationary climate, such that daily data measured in, e.g. January can be used to estimate a density function for a January ε, which is sufficient, since we concentrate on January simulations only. The data used are daily DWD analyses of Januaries 1967–1977.

It is not possible to give a proof or even a statistical test to decide whether the observed states satisfy the assumption of normality. We have performed for each single amplitude a Lilliefors test (Lilliefors, 1967) based on an independent subsample and obtained the result that the 95% test is failed by less than 5% of the spectral components, pointing to the admissibility of the assumption of normality.

2. The GCM data

Both the atmospheric and the model-generated 500 mb height fields are expanded into the first 30 EOFs given by Rinne and Karhila (1979). We use EOFs because they represent generally more information than spherical harmonics (Savijärvi, 1978). The GCM employed for this study is the Hamburg University model (Roeckner, 1979). With its 3-layer/2.8° version, two experiments with January conditions and realistic initial states were performed. The first, denoted by A, gave unsatisfying results due to a model defect and was therefore terminated at day 50. The improved version, experiment B, gave, after a crude inspection, acceptable results and was integrated up to day 75. Unfortunately, the numerical procedure for getting the 500 mb heightfield in the model differed from that applied in the DWD analyses, resulting in different hemispheric mean heights. This fact is, however, not essential for the model's performance as only gradients of the heightfield are dynamically relevant. But it has consequences on the amplitudes of the first EOFs, while the higher modes are not affected. For this reason we restrict ourselves to EOFs $11 \leqq k \leqq 30$.

3. Results

3.1. First guess

In order to make the test less conservative, the number of degrees of freedom were reduced by means of the "bad" experiment A: We assume that the components being endangered to show an unrealistic behavior will be those that are in experiment A either too small or too great. Therefore, for all EOFs an univariate analysis to determine whether their amplitudes in experiment A lie in the corresponding 95% probability interval,

was done. This analysis showed that the EOFs 1, 16–21 and 30 had unusual amplitudes. As already mentioned, EOF 1 should not be taken into account. Therefore, we took as a first guess the EOFs 16–21 and 30, i.e. $n = 7$.

3.2. Multivariate analysis

For each day of the two integrations A and B the number g was calculated daily. The resulting curves and—as a straight horizontal line—the 95% fractile ($= 14 \cdot 1$) are shown in Fig. 1. As might be expected, experiment A departs from the climate ensemble. After about 15 days g leaves the 95% probability ellipsoid, after 25 days $E(99\%)$ and the final state lies very far off $E(99.9\%)$. Due to the prechecking of the data, the daily tests for A were unfair. But the same test based on all EOFs 11–30 produced essentially the same development, indicating that the increase of g is not caused by the restriction onto the 7 degrees of freedom.

The g statistic for B remains most of the integration time within $E(95\%)$. Especially between the 19th and the 47th day g oscillates about the expectation value 7 of g, indicating that the states are quite realistic. Only on 15 days ($= 20\%$) are states found that occur in nature in less than 5% of the time. This rate of rejection of the daily null hypothesis is about four times expectation (5% $\cong$ 4 d), provided the simulated states stay all the time in ε. Therefore, we believe that some of these rejections are reliable, especially the three longer periods 12 d–14 d, 48 d–51 d, 55 d–58 d. The first peak at 3 d may be due to an initial inbalance.
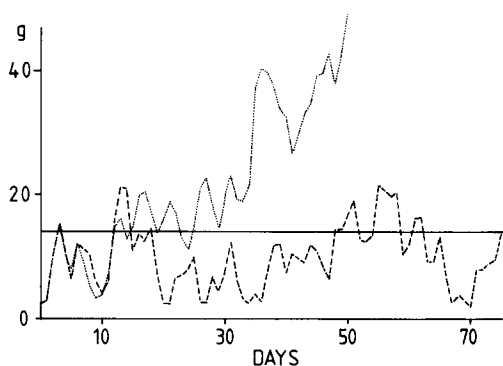

Fig. 1. Temporal development of the statistic g (defined in 1) for 7 degrees of freedom. The dotted line shows experiment A, the dashed one experiment B. The 95% fractile is given by the straight horizontal line.

3.3. *Univariate detailed analysis*

After having seen that some of the states simulated during experiment B are rare ones, a closer inspection with an univariate analysis for all amplitudes $k = 1$–30 was done. This yields too large amplitudes of EOF 1 compared (univariate) with what is usual in nature. Though statements concerning EOF 1 are not reliable, its behavior may be taken as an indication that the mean meridional gradient of the 500 mb heightfield is poorly simulated, because EOF 1 essentially describes this gradient. The last three rejection intervals of the multivariate test coincide with those of EOF 16, indicating that this EOF causes the large value of g. The rejection during day 12–14 can be explained by EOF 18. Both EOF 16 and EOF 18 correspond to about zonal wave number 5,

suggesting that the unrealistic behavior of the model may be due to an exaggerated baroclinic activity.

REFERENCES

Chung, J. H. and Fraser, D. A. S. 1958. Randomization tests for a multivariate two-sample problem. *J. Am. St. Ass. 53*, 529–535.

Lilliefors, H. W. 1967. On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *J. Am. St. Ass. 62*, 399–402.

Rinne, J. and Karhila, V. 1979. Empirical orthogonal functions of 500 mb height in the northern hemi-

sphere determined from a large data sample. *Quart. J. Roy. Met. Soc. 105*, 873–884.

Roeckner, E. 1979. A hemispheric model for short range numerical weather prediction and general circulation studies. *Beitr. Phys. Atm. 52*, 262–286.

Savijärvi, H. 1978. Verification and storing with empirical orthogonal functions. ECMWF internal report *18*.