Tellus (1985), *37A*, 361–377

The extra-tropical atmospheric response to El Niño events—a multivariate significance analysis

By HANS VON STORCH, *Meteorologisches Institut der Universität Hamburg, Bundesstrasse 55, 2000 Hamburg 13, FR Germany* and HARALD A. KRUSE, *Max Planck Institut für Meteorologie, Bundesstrasse 55, 2000 Hamburg 13, FR Germany*

(Manuscript received October 2, 1984; in final form February 18, 1985)

ABSTRACT

Up to now, a well-defined and consistent image of the extra-tropical atmosphere's response to intense El Niño warmings of the sea surface was missing—in contrast to the tropical area near the forcing source. We attribute this to the currently usual inadequate analysis methods which lead to a misinterpretation of multidimensional signals in the presence of large random fluctuations. In this paper, we demonstrate how the multi-variate significance test procedures, reviewed herein, serve to detect the El Niño signal in the GCM atmosphere of the ECMWF as well as in the observed circulation in 1982/83. The results of the rigorous statistical tests are illustrated by stability and univariate considerations and are "synoptically" compared with respective response patterns generated by a few other GCMs.

It is concluded that a stable and highly significant extra-tropical response in terms of 500 mbar height to an El Niño warming exists. It is found that the January 1983 observed and the GCM generated patterns are highly coherent. This signal covers the 270° sector of longitude from the North-East Pacific to Eurasia. The model response to a cold SST anomaly is marginally significant, less stable, and different in structure, so that a strong non-linearity outside the Pacific region is evident. Furthermore, we hypothesize that an intense El Ninõ event induces an intensification of the north-westerly Atlantic daily variability.

1. Introduction

General Circulation Models (GCMs) are powerful tools to estimate the impact of a changed external forcing on the atmospheric circulation. A quite popular example of a changed external boundary condition is the warming of the equatorial east Pacific known as El Niño event (cf. reviews by O'Brien (1978) or Cane (1983)). In the beginning, climate change experiments were built up simply by two integrations, one with "normal" and one with "anomalous" forcing. The mean difference of the two integration results was considered to be the "signal" produced by the forcing. At the beginning of the 70s, it was realized that the difference in the two fields is not at all entirely due to the forcing, but that part of it is random due to the inherent natural variability of the model and real atmosphere. Thus, the difference field is composed from the true signal plus (generally large) noise patterns (Warshaw and Rapp, 1972; Leith, 1972; Chervin et al., 1974; Chervin and Schneider, 1976a, b). To distinguish between signal and noise, standard statistical tests were performed at each grid point separately ("univariate approach").

Since the late 1970s, warnings were issued (Hasselmann, 1979b; Storch, 1982; Hayashi, 1982; Livezey and Chen, 1983) that a series of simultaneous univariate tests cannot yield any assessment of the multidimensional problem:

(I) *Is there any significant change in the circulation of the model atmosphere?*

To overcome the univariate deadlock, a number of strategies were proposed. A popular one is simply to ignore the problem. Another one is to regard those large-scale patterns as significant which exhibit locally large *t*-statistic values and coincide with the climate anomaly patterns one is most interested in.

The criticism of the widely used univariate approach is directed towards its inability to allow a decision with a given error probability of whether an estimated difference between the undisturbed observed or GCM simulated climate and its disturbed counterpart is due to noise or does at least partly reflect a real signal. However, the point-by-point consideration may be used successfully to gain further insight into the characteristics of the difference after it was found to be multivariately significant (Section 2).

Two classes of serious statistical approaches have evolved. The first concerns the rate of "local" rejections of the null hypothesis that "no local signal exists" (Storch, 1983; Livezey and Chen, 1983; Veysseire, 1983; Zwiers, 1983). If the distribution of the number of erroneous local rejections is known, a threshold value of this number can be fixed. If the actual number exceeds this threshold, at least one local rejection is correct with a known error probability. The other class of procedures is based upon an *a priori* reduction of the degrees of freedom by reducing the number of parameters describing the pattern properties and on a subsequent multivariate test, either parametric (e.g. Hasselman, 1979b; Hayashi, 1982; Storch and Roechner, 1983a; Hannoschöck, 1984) or non-parametric (Storch and Roeckner, 1983b; Preisendorfer and Barnett, 1983). We will discuss this strategy in some detail in Section 2. A nice review of the 1983 state of the art is given by Livezey (1983).

The first purpose of this paper is to demonstrate that the multivariate methods of the latter class are well developed such that they can be successfully used for practical questions. To do so, we study the midlatitudinal response of the ECMWF GCM (in its T21-L15 version) to a prescribed El Niño type SST anomaly in terms of northern hemisphere 500 mb height. We find a significant response pattern characterized by an intense low north of the SST anomaly and two or three stably located midlatitudinal highs downstream (Sections 3 and 4).

After having found a significantly changed model circulation, the question arises:

(II) *Is the significant model response consistent with observation?*

Up to now, this question was treated in a subjective way by the visual side-by-side comparison of the GCM-generated signal and appropriate maps based upon observations, e.g. composite or correlation maps (e.g. Blackmon et al., 1983; Shukla and Wallace, 1983; Cubasch, 1983).

The second purpose of the present paper is to apply the same objective methods utilized for problem (I) to show the consistency of the detected GCM-simulated El Niño response with an observed midlatitudinal atmospheric anomaly pattern, namely that of January 1983. This pattern was found to be significantly untypical (Storch, 1984) and appeared contemporaneously with a very intense El Niño event (Quiroz, 1983) (Section 5).

Thus, we shall show that the application of the multivariate significance test methods successfully gives confident answers to the two questions raised (I) and (II), namely:

(I) an intense El Niño SST anomaly induces a significant and stable pattern of 500 mb height anomalies in the entire northern hemisphere of the GCM atmosphere;

(II) the significance GCM response pattern and the significant observed anomaly pattern in winter 1982/83 are coherent in an objectively defined sense.

In what follows, we refer to the numeration (I) and (II), respectively, when the problems (I) or (II) are addressed.

Further, we consider the question of whether the extra-tropical response is a linear function of the SST anomaly (Section 5). The discussion on this topic did not stand on a firm basis as long as the response patterns were not clearly separated in the true signal and random fluctuations. Lastly, a discussion is initiated on the behaviour of the transient activity which is loosely related to the stationary patterns, and which has not gained any attention so far.

2. The methodical framework

Since the quantities considered are given as fields, multivariate techniques have to be used. Multivariate tests are likely to fail to detect an existing signal if too many (noisy) parameters (e.g. grid-point values, Fourier coefficients) are used to describe the circulation patterns (Hasselmann, 1979b). A very simple example is given in

Appendix A. Thus, an *a priori* reduction of the number of parameters is necessary. Possible methods to compress the data are:

(A) averaging with respect to spatial coordinates, eventually over limited areas of special interest (successfully used by Storch and Roeckner (1983a, b)).
(B) spectral expansion of the spatial structures; functions for the expansion are chosen for reasons either of fast convergence (empirical orthogonal functions, EOF's, successfully used by Storch and Roeckner, 1983a, b) or representativeness of certain spatial scales (e.g., spherical surface harmonics, SSH's, successfully used by Hannoschöck (1984)).
(C) projection onto a set of "guess patterns" or "modes" that are presumed to build up the signal due to dynamical considerations; however, trials with the barotropic and the first baroclinic response mode of a (probably too simple) thermally forced linearized atmospheric model gave unsatisfactory results (Hannoschöck, 1984).
(D) Projection onto a set of "guess patterns" established by means of similar independent observations or numerical experiments. This set of "guess patterns" will often consist of just one element. An example is given below in Subsection 5.2.

Technically, the approaches (B, C, D) consist of the determination of the expansion coefficients by linear regression. (The notion of a "guess pattern" emphasizes that the choice of the expansion series is rather arbitrary from a statistical view point; in terms of regression analysis, one would call them "predictors".) The regression fit may be optimized *a priori* by weighting the guess patterns with the inverse of their natural variance (Hasselmann, 1979b; Hannoschöck, 1984). Practically, this is done using an EOF expansion. However, in case of small or moderate sample sizes, the estimates of the variances in the EOF space are biased (Storch, 1983). As a consequence, the resulting test procedure becomes radical i.e., rejects the null hypothesis too often. A simple example is given in Appendix B. Therefore, we use this elegant optimization method only with reservation.

In the expansions (B, C, D), the optimal truncation of the series cannot always be determined *a priori*. It is therefore necessary to consider a hierarchy of truncations (Hasselmann, 1979a). To obtain correct probability statements, it is inevitable to specify *a priori* the ordering of the hierarchy as well as the selection criterion for the optimal model to be picked from the hierarchy.

Furthermore, approach (A) can be put in the form of a regression technique if a sequence of disjunct areas is considered. In connection with a hierarchy test, this method can be used to estimate "how far a signal reaches downstream".

In order to find the results I and II announced in the Introduction, we perform the compression by means of the following data representations:

Problem I

(a) Averaging of the quasistationary 500 mb height field from 30° to 60° N in order to focus on the midlatitudinal part of the response, and expansion of the resulting function of longitude into a short series of EOFs. (A compression according to method A.)
(b) Expansion into a short series of large-scale spherical surface harmonics (SSHs) expected to represent the global character of the response (Method B).

Problem II

Projection of the GCM-generated fields onto the "observed response", defined as the difference of the mean 500 mb height field of the (El Niño) January 1983 minus that of the (normal) Januaries 1967–72/74–82 (cf. Storch, 1984) (Method D).

The statistical hypotheses are formulated in two ways. If the data compression is performed without reference to dynamical or other *a priori* considerations (A, B), state vectors of expansion coefficients are formed for anomalous and control cases and the null hypothesis (which hopefully will be rejected) states the identity of the expectation vectors. If an *a priori* hypothesis on the response pattern is formulated (C, D), the projection onto the "guess space" is tested for zero length. This amounts in the same null hypothesis as above, namely on the identity of the expectation values of the low-dimensional random vectors of the "control" and "anomaly" cases.

The test statistics are either rank statistics in conjunction with a (randomized) permutation method (Storch and Roeckner, 1983b; Preisendorfer and Barnett, 1983) or a vector square norm with the χ^2 (Hasselmann, 1979b; Storch and Roeckner, 1983a; Hannoschöck, 1984) or the Hotelling distribution (Hayashi, 1982; Hannoschöck, 1984).

It should be stressed that the parameter reduction is necessary in order to have a chance to detect a signal; the more-or-less intuitive *a priori* choice of a suitable guess pattern hierarchy must be based on experience from results independent of the data in question. Under these circumstances, the statistical test procedure yields precise statements concerning state vectors in the low-dimensional parameter space.

An analysed 95% significance means only that the null hypothesis of no response everywhere is rejected with 5% probability of error. It cannot be interpreted as assuring that the estimated averaged response pattern resembles the true response pattern (Hayashi, 1982). However, the univariate approach, which is inappropriate for treatment of the multidimensional significance problem, may be used to ask as to which spatial parts of the pattern might be connected to the significance of the signal in the low-dimensional space. If the disturbed observed or GCM-generated climate exceeds somewhere clearly, or in different experiments repeatedly, the variability of the normal climate, this excess may be interpreted as one detail of the significant response pattern. However, it must be kept in mind that such a procedure yields only arguments of plausibility and stability considerations (see below), which are not supported by the multivariate statistical test performed in the low-dimensional space. However, an objective way to test questions concerning the natural grid point space is the "downstream area sequence" procedure mentioned above in the context of the compression method (A). Another way is the following. Based on one experimental result, one may formulate a hypothesis concerning a few grid points which seem to exhibit special features, and then test this question by using another independent data set which has not been used in formulating the hypothesis.

3. The GCM experiment

In order to find out what kind of anomalous hemispheric circulation pattern is induced by an intense El Niño event, a respective GCM experiment performed by Cubasch (1983) at the ECMWF was studied. The GCM is a spectral model with a triangular truncation of the horizontal structure at total wavenumber 21 with 15 levels in the vertical. An integration covering 10 years with climatologically varying sea surface temperature (SST) is used as the control experiment.

An objective verification with the methods given by Storch and Roeckner (1983b) of the January 500 mb height climatology shows that the model is quite capable of reproducing the observed climate mean states. The north–south difference of the zonal and monthly mean is slightly reduced as compared to the observed quantity. Statistically significant differences between the simulated and the observed climatology (German Weather Service (DWD) analyses, Januaries 1967–83) are: too weak a midlatitudinal quasi-stationary disturbance over the Atlantic and Europe (Fig. 1a); a general underestimation of the mean January zonal variation of stationary (Fig. 1b) and transient (daily) (Fig. 1c) variance. Furthermore, the interannual variability of the January stationary patterns is clearly too weak (Fig. 1d). A detailed paper on a systematic comparison of the 500 mb climatology of the T21 model circulation and the observed circulation has been submitted for publication.

The El Niño experiment consists of 15 winter (December through March) integrations. Besides the 9 "control" winters, 3 winters with a doubled Rasmusson/Carpenter (1981) warm El Niño type standard SST anomaly (see Fig. 2) and 3 with the same anomaly but reversed sign were integrated. The 3 warm anomaly experiments were integrated with the same SST anomaly and differ in their initial conditions only. The same holds for the 3 cold anomaly cases. Thus, 2 classes of experiments are available, each consisting of 3 statistically equivalent parts. The difference of the 2 classes is the sign of the anomaly.

In the tropics, the GCM responds with a well-defined signal, which is summarized by Cubasch (1983) as follows: "The response in the tropical Pacific seems to be strongly correlated with the anomaly and resembles the one obtained ... using a shallow water equation model. The SST anomalies cause a pressure perturbation in the tropics similar to the observed 'southern oscillation'. A major warming ... does ... occur in the region with the largest precipitation increase ...". In midlatitudes, however, the model's response seemed vague and ambiguous. Thus, the hypotheses were

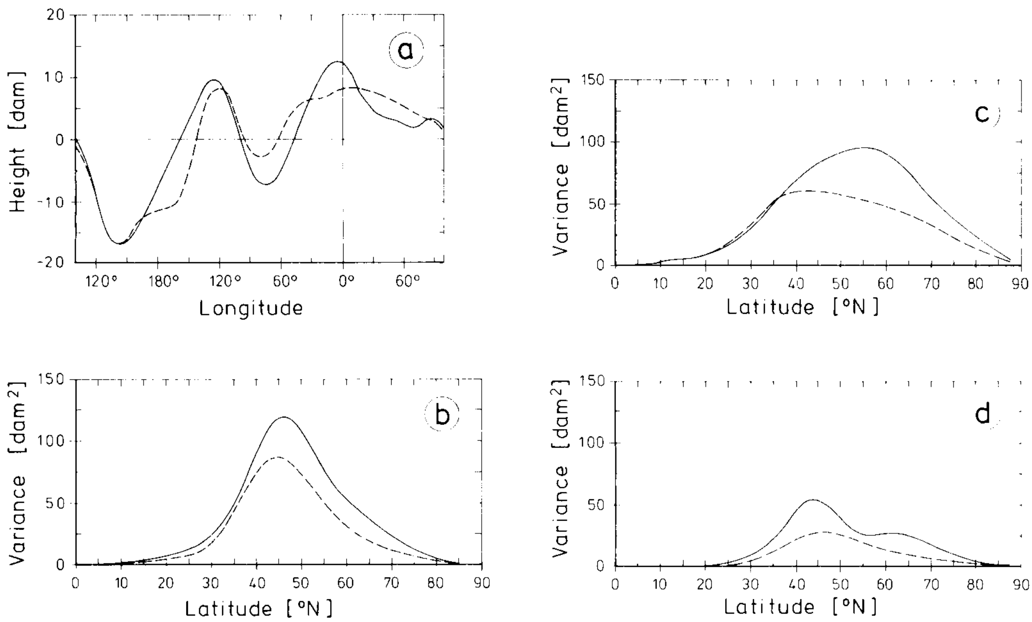

Fig. 1. Comparison of the 500 mb height climatology as simulated by the ECMWF model (dashed curves) and observed (solid curves). The differences are statistically significant. (a) 30°–60° N average of monthly mean. (b) Zonal average of the variance of stationary eddies, weighted with the cosine of latitude φ. (c) Zonal average of the variance of transient eddies, with cos φ weight. (d) Interannual standard deviation of (b).

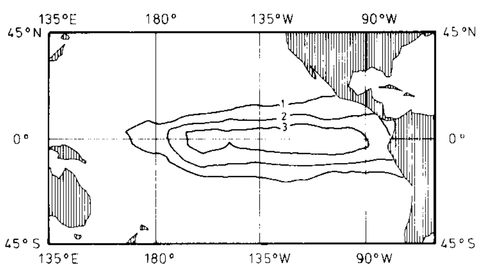

Fig. 2. Intensity and shape of the El Niño SST anomaly used in the ECMWF GCM experiments (adapted from Cubasch, 1983). One group of 3 experiments was integrated with positive values, another group with negative values.

put forward that either the sign of the response pattern is independent of the sign of the SST anomaly or the SST anomaly does not induce a well-defined midlatitudinal circulation anomaly.

In order to demonstrate the merits of the multivariate strategy outlined in Section 2, we reanalyse the extra-tropical part of the model's response in Section 4. We shall show that the mid-latitudinal and almost-hemispheric response to the warm anomaly is highly significant and stable in structure, whereas both of these properties are not found for the cold anomaly.

4. Problem I: Significance of the simulated response pattern

In order to demonstrate the capabilities of the proposed methods, we use 2 different approaches. Firstly, we use data compression Ia (the mid-latitudinal focus, cf. Section 2) in connection with the nonparametric generalized Mann/Whitney test (Storch and Roeckner, 1983b), abbreviated by MW/Ia in the following. Secondly we use a hierarchy of spherical surface harmonic (SSH) truncations (data compression Ib) together with an optimized or standard χ^2 test denoted by CS/Ib (Hannoschöck, 1984). Due to historical reasons, MW/Ia utilizes January data (30 d intervals) and CS/Ib winter data (= 120 d intervals). As will be seen, the conclusions are not affected by this difference. In order to filter out the seasonal trend, the 15-sample mean (9 controls, 3 cold anomalies, 3 warm anomaly samples) for each calendar day was subtracted from the data. Thus, the investigated time series are stationary to first order, and the model's mean state is removed. Consequently, deviations from the 15-sample mean are exclusively displayed in the following figures.

4.1. Results obtained with MW/Ia (midlatitudinal focus)

Data compression Ia consists of a 30°–60° N averaging of 500 mb height and a subsequent EOF decomposition of the zonally distributed function. The EOFs are obtained from the respective observed January 1967–83 data and are thus totally independent of the actual model data sample under consideration. Therefore, we do not have to worry about systematic errors of the variance represented by the EOF's (Storch, 1983). We use a 5-EOF truncation, i.e., the data originally given as a function of longitude are projected onto a 5-dimensional subspace.

4.1.1. The warm anomaly. In the case of the warm anomaly experiment, the test yields a highly significant difference of the means of the 9-sample control set and the 3-sample anomaly set. To analyse the differences, we "leave the statistical sector" and return to the grid space by plotting the 95% band of the control ensemble and the 3 curves simulated by the warm anomaly experiment (Fig. 3). The "95% band" is univariately defined as that interval containing 95% of all 9 control cases at the particular longitude. No statistical assessment is connected with this band, but it visualizes the locations, where the anomalous states deviate extraordinarily from the "normal". As can be deduced from Fig. 3, this occurs at the longitude of the anomalous SST warming (shaded area). The deviations amount to 20, 60 and 85 m. Furthermore, 3 maxima are stably located at about 90° W, at Greenwich longitude, and at 80° E.

4.1.2. The cold anomaly. For the cold SST anomaly, which increases the mean east–west difference of the equatorial Pacific SST to about 10 K, no significant impact on the midlatitudinal atmospheric circulation can be detected in terms of the 30°–60° N average. This is displayed in Fig. 4, which shows the 95% control band together with the 3 cold anomaly curves: The curves leave the 95% band only seldomly and irregularly. The ambiguity of the circulation anomaly at the longitude of the SST anomaly points to the qualification of method MW/Ia.

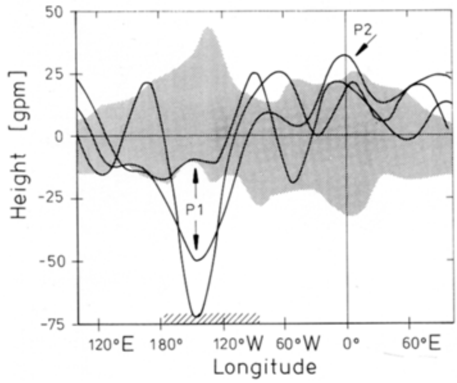

Fig. 3. 30°–60° N average of monthly mean 500 mb geopotential height simulated with the ECMWF GCM. The data are centred around the total emsemble mean. Stippled: 95% band of the control ensemble. Curves: 3 January states forced with the warm SST anomaly. P1 and P2: remarkable features observed in January 1983 during the intense El Niño event (Subsection 5.1). Shaded area: longitudes of the SST warming.

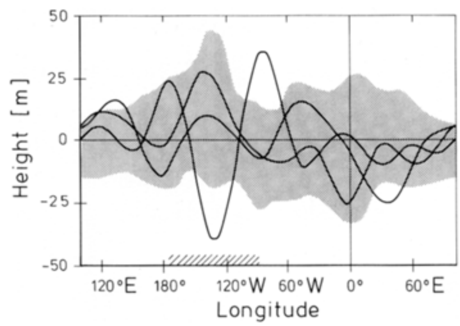

Fig. 4. Same as Fig. 3, but showing 3 states generated in the cold SST anomaly experiment.

4.2. Results obtained with CS/Ib (hemispheric structure)

In order to consider the response structure on the entire extra-tropical hemisphere, we use an SSH hierarchy containing all harmonics up to total wavenumber 7, with a sequence from large to shorter length scales. As the final significant response, we choose that member of the hierarchy which yields maximum skill at the fixed 95% significance level.

The significance (equal to 100% minus the error probability) of this finally selected member is not equal to the fixed significance (here: 95%) of the single member but depends on the significance of the preceding (lower order) members, or more generally, on the selection criterion (Barnett et al., 1981). If all preceding members are at least significant with 95%, then the "sequential" selection rule may be applied, and the final result is significant with more than 95%. If not all preceding members reach the fixed significance level, then only the "non-sequential" selection rule is applicable, and the significance of the final result is less than 95%. Barnett et al. (1981) published tables of the significance of a signal that had been selected from a hierarchy with various rules. They obtained these tables under the constraint of statistically independent predictors. Unfortunately, this assumption does not hold in our cases. Nevertheless, we use these tables in order to get an impression of the magnitude of the actual error probability of our selection.

4.2.1. The warm anomaly. In view of the high significance level obtained in Subsection 4.1, we study each of the 3 winter seasons separately. The χ^2 test is applied to the standard regression fit and to the optimized variance-weighted version.

In 2 cases, both the optimized and the standard regression show up χ^2 statistic values exceeding the 95% (and even the 99%) quantile for all elements of the hierarchy. Thus, the selected member is that with maximum resolution. For the third experiment, only the optimized version yields an exceedance of the 95% quantile, and that for a resolution up to 18 SSHs. This third experiment is that sample with the smallest 30°–60° N mean deviation north of the anomaly, namely about 20 m (cf. Fig. 2).

The selected significant members of the hierarchy are shown in Fig. 5. The differences of one particular anomalously forced mean winter (December–February) 500 mb height field minus the ensemble mean of the respective 9 control 500 fields are plotted. We show the individual response patterns instead of the mean pattern in order to demonstrate the similarity of the patterns. In all 3 cases, the structures of the response patterns are similar, in particular, the sequence of middle and high latitude extrema appears stable within the 270° sector covering the central Pacific, north America, and parts of Eurasia. The dissimilarity of Fig. 5c on the one hand and Fig. 5a and b on the other is mainly due to the difference resolutions (18 and 30 SSH's, respectively).

4.2.2. The cold anomaly. In contrast to the

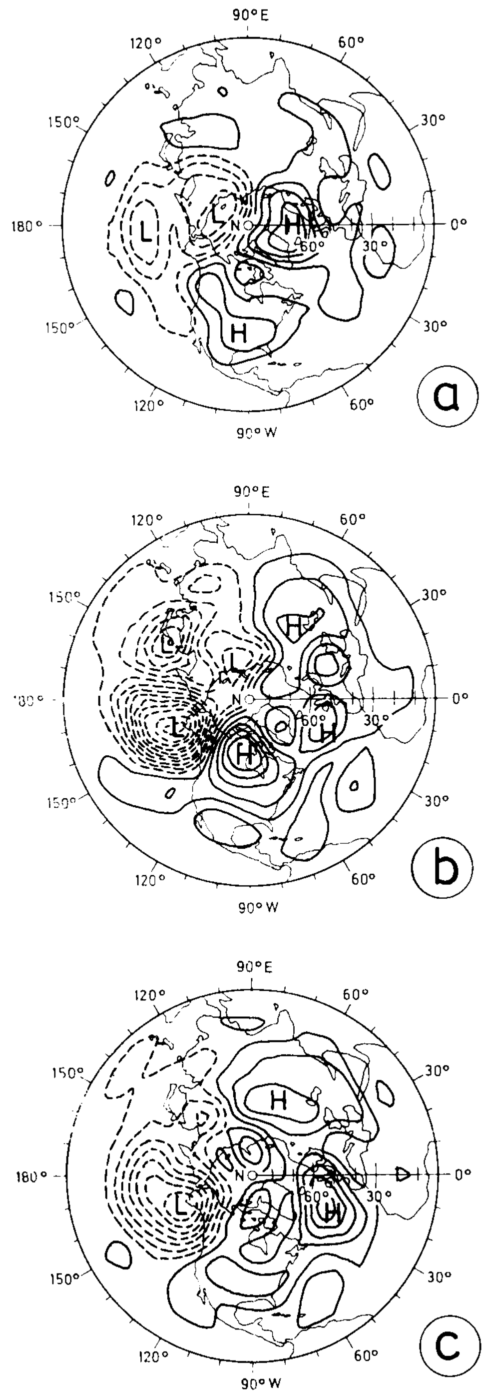

Fig. 5. Significant northern-hemispheric response of the ECWMF GCM to the warm El Niño SST anomaly for 3 different initial conditions. (a) consists of 18 (real) SSHs, (b) and (c) of 30 SSHs. Contour interval 10 m.

warm anomaly experiments, we have found earlier that the cold anomaly induced a multivariately non-significant response in terms of the 30°–60° N average. In order to increase the *a priori* chance of detecting a significant signal in the hemispheric structure, we have to keep the basic time series as long as possible (Leith, 1972). Therefore we consider the average signal of all 3 experiments. The standard regression fit yields members with significance much below 95%, only. The optimized version, however, yields the first 13 members of the hierarchy, each more than 90% significant (Fig. 6). The selected maximum skill member represents a significance greater than 90% (even 99% if the guess vectors were independent). As has been mentioned in the beginning, the probability statements with respect to the optimized regression are subject to errors (the error probability is underestimated). Therefore, we look at the last result with reservation.

In order to get further information on the stability of the response, we considered the 3 winter response patterns individually as we did for the warm case. In the winter case shown in Fig. 7c, the response to the negative anomaly is insignificant for any model order. The other two cold anomaly response patterns (Fig. 7a, b) in the optimized version are significant at about the 85% level according to Barnett et al.'s tables. In the 2 weakly significant cold-anomaly response patterns, coincidence of the structure is found in the 180° sector covering the north-eastern Pacific and the north Atlantic. If we adopt the convention to term the sequence of extrema as "wave trains" (admittedly out of fashion at present), we may say that these wave trains emanate eastward, starting from a

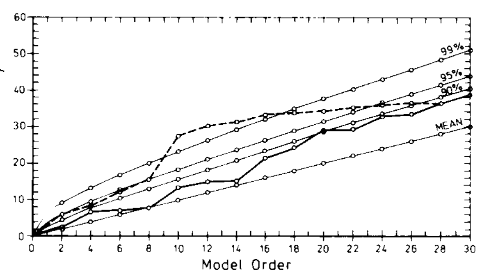

Fig. 6. Vector norm ρ^2 of the simulated average response to the cold SST anomaly as expanded in an SSH series by a standard regression fit (solid line) and by an optimized variance-weighted fit (dashed line). Also shown are the expected values ("mean") and the 90 ... 99% bounds for a zero population value of ρ^2.

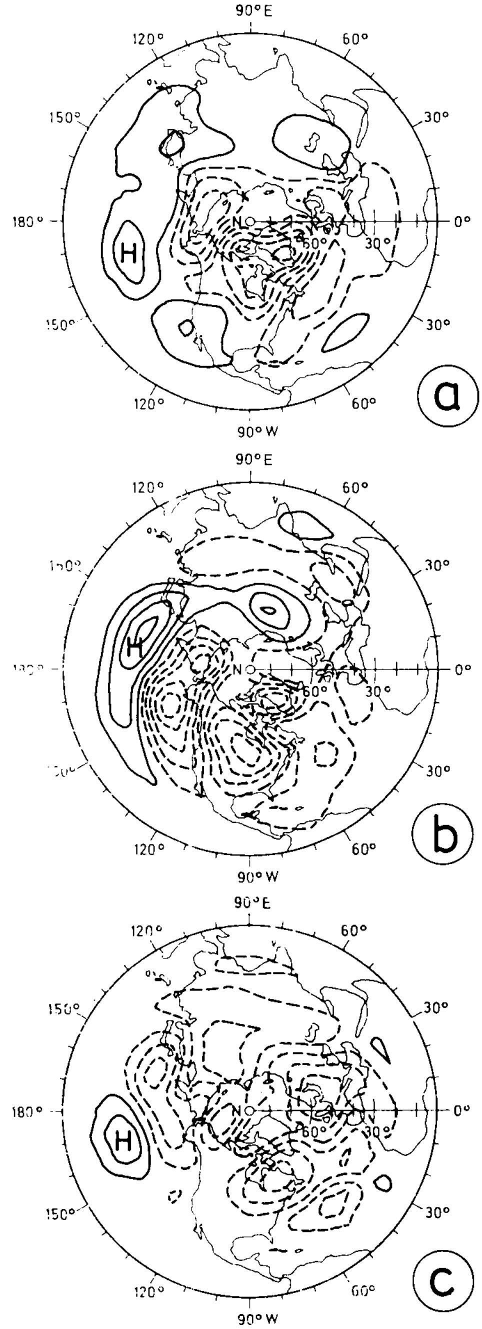

Fig. 7. Northern-hemispheric response of the ECMWF model to the cold SST anomaly for the same 3 initial conditions as in Fig. 5. (a) and (b) are significant according to the variance-weighted regression analysis, with 16 and 20 real SSHs, respectively. For case (c), no significant truncation was found; the full 30-SSH-pattern is shown instead.

centre near the date line with opposite signs for the "warm" and "cold" pattern. Due to the different wavelength, the downstream response has in some locations the same sign in both cases, e.g., over North-East America. Thus, at first glance, the midlatidudinal response close to the anomalous forcing seems to be linear, whereas the more remote is not.

5. Problem II: consistency of the simulated response with observation

5.1. *The observed circulation anomaly in the 1982/83 El Niño winter*

We want to relate the findings of Section 4 to atmospheric reality. Unfortunately, an oceanic cooling comparable with the cold SST anomaly prescribed above has never been observed. Thus, the "cold anomaly experiment" cannot be verified at all. However, a warming similar to that included in the "warm anomaly experiment" has occurred in 1982/83 (Wagner, 1983; Krueger, 1983; Quiroz, 1983; Chen, 1983; Arkin et al., 1983) during the most intense El Niño event ever observed. The SST anomaly in January 1983 is shown in Fig. 8.

Storch (1984) objectively compared the 500 mb geopotential height of this January 1983 with the set of more-or-less normal Januaries 1967–1981 using the data compression Ia and a standard χ^2 test (CS/Ia). He found it significantly untypical with respect to:

(P1) an intense low north of the SST anomaly;
(P2) a strong high at about Greenwhich longitude;
(P3) an intensified wavenumber 3 type structure.

The former two features may easily be identified in Fig. 9.

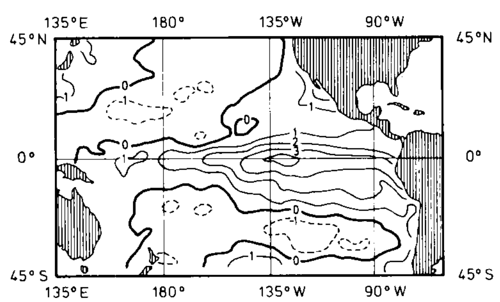

Fig. 8. Equatorial Pacific SST anomaly observed in January 1983 (adapted from Arkin et al., 1983).

In his discussion, Storch (1984) pointed out that a causal relation of (P1) and the occurrence of an El Niño event was already proven earlier by Chiu et al. (1981) and that there are indications that (P2) and (P3) are connected with the east equatorial Pacific warming. However, a final proof that an intense El Niño event will induce, besides (P1), also (P2) and (P3), is still missing and the relationship El Niño —(P2)/(P3) has to be regarded a hypothesis.

In Subsection 5.2, we will use the GCM result analysed in Section 4 as an argument for the validity of this hypothesis. In fact, a visual comparison with Fig. 3 alludes to a coincidence of the extraordinaries of January 1983 and those simulated in the warm SST anomaly experiment.

The second most intense El Niño ever observed took place in 1877/78 (Kiladis and Diaz, 1984). Of course, in the last century, neither upper air nor global measurement was made. However, a mean surface pressure map for January 1878 covering north America, northern Atlantic and western Europe is available (Fig. 10): The dipole structure of a "Greenland/northern Europe low versus Atlantic/western Europe high" (P2) is easily recognizable in both maps, 1878 and 1983. A further common property is the existence of a ridge crossing Canada. The orientation of the 1878 isobars at the north American west coast point to an intensified and/or eastward shifted Aleutan Low. Thus, Kiladis' and Diaz' map may be taken as a support of the above hypothesis.

5.2. *Consistency of the simulated and the observed response*

In order to relate the observed and the simulated extra-tropical response pattern, we simply test the hypothesis that the expectation vector of the simulated response has a component parallel to (and in the same direction as) the observed response vector of 1983. For that purpose, we have to test whether the dot product of these vectors is positive. That is, we use the data compression method (II) in its simplest form, where the observed pattern serves as the single predictor for the simulated pattern. We use a rank statistic together with a permutation test according to Mann and Whitney (cf. Conover, 1971). Thus, procedure MW/II is used.

Application of this test shows that the simulated response to the warm SST anomaly has a compo-

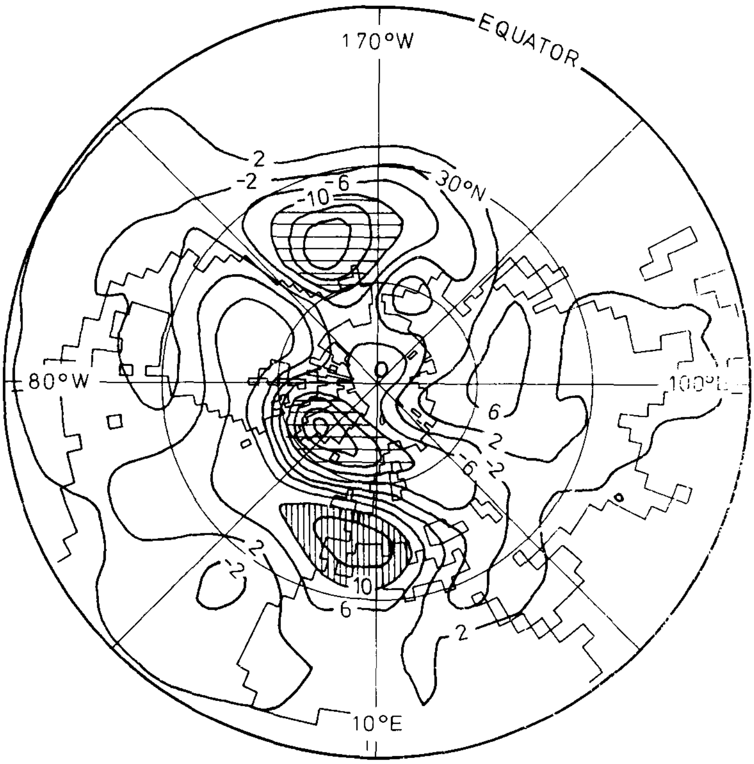

Fig. 9. The mean 500 mg geopotential height anomaly field observed in January 1983. Contour interval 40 m.

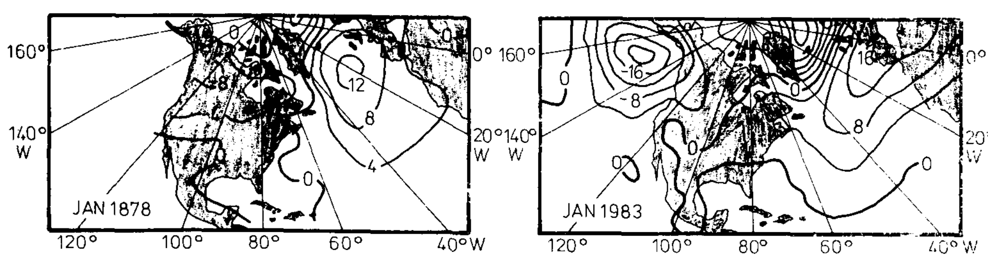

Fig. 10. Observed sea level pressure maps for January 1878 (left) and 1983 (right). Contour intervals 40 m (adapted from Kiladis and Diaz, 1984).

nent in the direction of the observed 1983 signal with 5% error probability, i.e., the visual similarity in the pattern structure is significant. For the cold anomaly, we find no significant result (i.e., no antiparallel component). This is a further indication that the relation between the SST anomaly and the hemispheric response is non-linear.

For the case of the warm anomaly, we know now that there exists a simulated component in the direction of the observed signal. However, Fig. 1 demonstrates that there are considerable differences between the GCMs and the atmosphere's climate, especially in midlatitudes. Thus, it is reasonable to study whether there exists in the GCM simulated data, a component orthogonal to the observation. If not, we will have explained the observed structure completely by the model. We can put this as a precise question: is the dot product of the simulated and the observed response vector equal to the product of the lengths of these vectors?

This hypothesis is to be rejected with as much as 10% error probability. Thus, we may say that indeed there exists an orthogonal part but with weaker significance than for the parallel part.

5.3. *Consistency with response patterns simulated by other GCMs*

A number of similar GCM-experiments have been performed, which we may easily compare visually with the results discussed before. Results obtained with the NCAR climate community model and the GLAS model have been published by Blackmon et al. (1983) and Shukla and Wallace (1983), respectively. Their signals in terms of 500 mb and 300 mb height, are reprinted here as Figs. 12a,b, respectively. The GLAS model's response (Fig. 12a) resembles the observed 1983 pattern (Fig. 9) very nicely with respect to the horizontal distribution within a sector of about 180° degrees downstream of the anomaly. The coincidence with the ECMWF's GCM response (Fig. 11) is striking with respect to (P1), (P2), and the polar depression centred at Greenland and northern Siberia, but over the north American continent there are some differences. On the other hand, the NCAR model generates a signal (Fig. 12b), which is, besides (P1), totally different from the observed one as well as from the GCM signal of the ECMWF and the GLAS model.

As an additional data set, the results of an intense El Niño experiment with the GCM of the French Weather Service (CNRM) (Deque and Royer, 1984) is available to us. Unfortunately, the generated anomaly time series are too short to

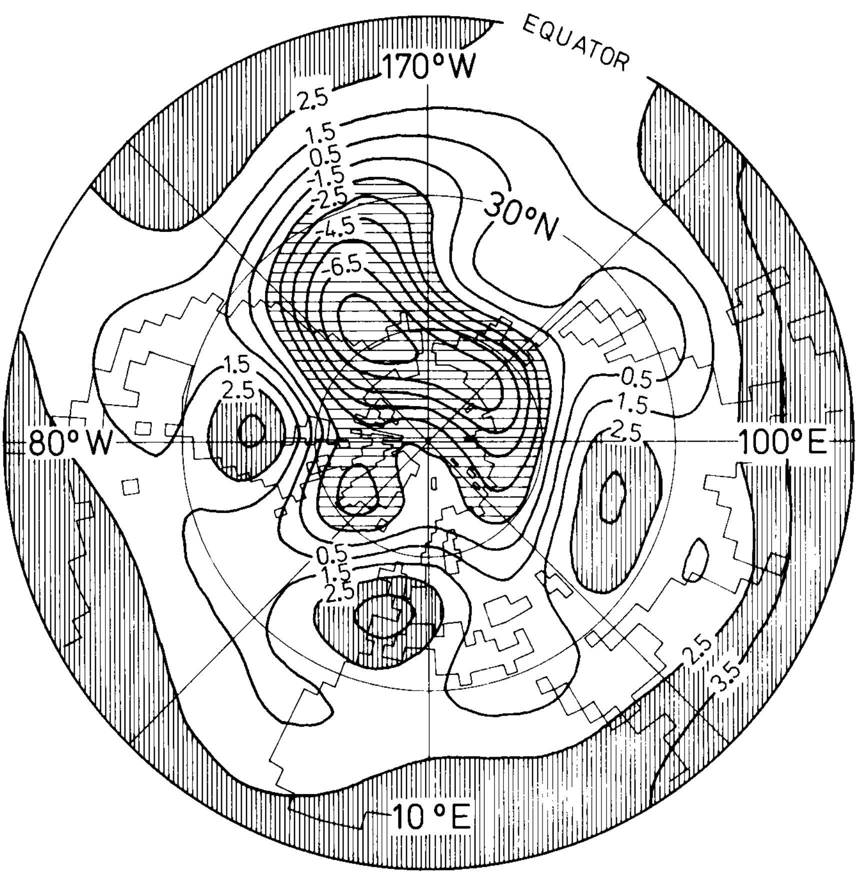

Fig. 11. Three-January mean stationary northern-hemispheric response pattern simulated by the ECMWF GCM in full spatial resolution.

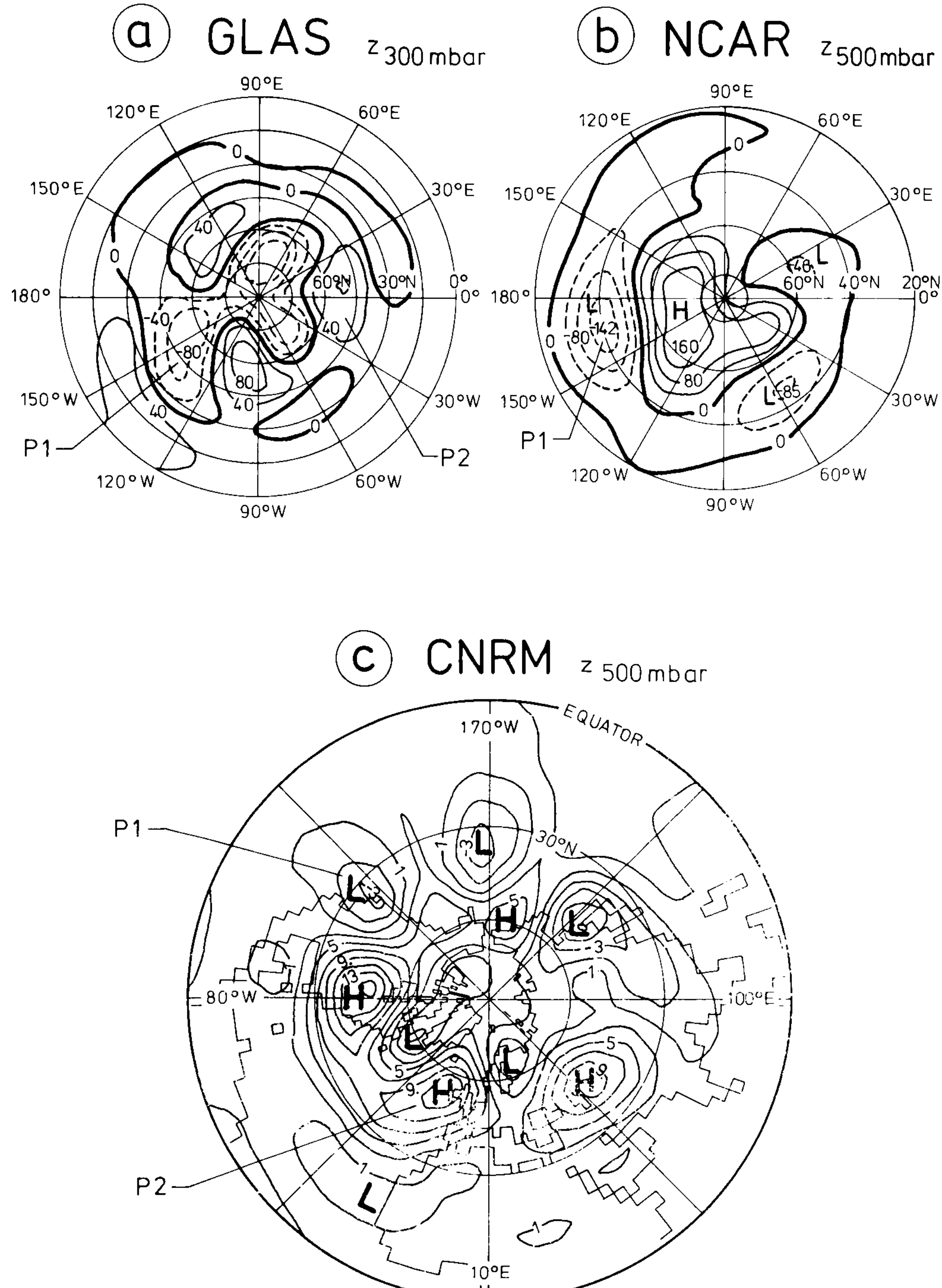

Fig. 12. Mean response patterns generated by other GCMs: (a) GLAS at 300 mb (reprinted from Shukla and Wallace, 1983), (b) NCAR at 500 mb (adapted from Blackmon et al., 1983) and (c) CNRM at 500 mb. P1 and P2 refer to the extraordinaries observed in January 1983 (Subsection 5.1).

detect a significant signal. Nevertheless, we plot the full signal in Fig. 12c. We find a feature P1, which is, however, quite weakly developed, a feature P2, and a general depression in the polar regions with maximum deviations at Greenland and Northern Siberia. Also with respect to north America, the CNRM result resembles the observed and the GLAS result.

6. Transient response

Since we have found significantly disturbed extra-tropical mean flow, a changed intensity of daily variability may be expected. The analysis of the January 1983 midlatitudinal transient activity resulted in a difference from the preceding January's significant at about the 92% level (CS/Ia). This relatively weak global significance is associated in the high-dimensional gridpoint space with an enhanced daily variability over the western midlatitude Atlantic ocean in January 1983. The analysis of the GCM-generated midlatitudinal transient response by means of MW/Ia yields a significant result: an enhancement of the variance owing to daily variability takes place at the midlatitude western Atlantic ocean (about 60° W), see Fig. 13, and this is a common feature of all 3 warm anomaly runs.

Unfortunately, even if we see a similarity with respect to location and sign between the model's response and the January 1983 pattern, we cannot confirm it with statistical confidence. However, an intensified cyclogenesis south of Greenland (Fig. 13) is reasonable because of the anomalous northwesterly flow across north-East Canada apparent in the stationary response of the observed (Fig. 9) and the simulated (Fig. 11) climate. Thus, our study of the transient response to an intense El Niño type SST anomaly yields a hypothesis with respect to the location and the sign. However, it has to be clarified whether this response is similar to the natural one or whether it is an artefact of the model under consideration.

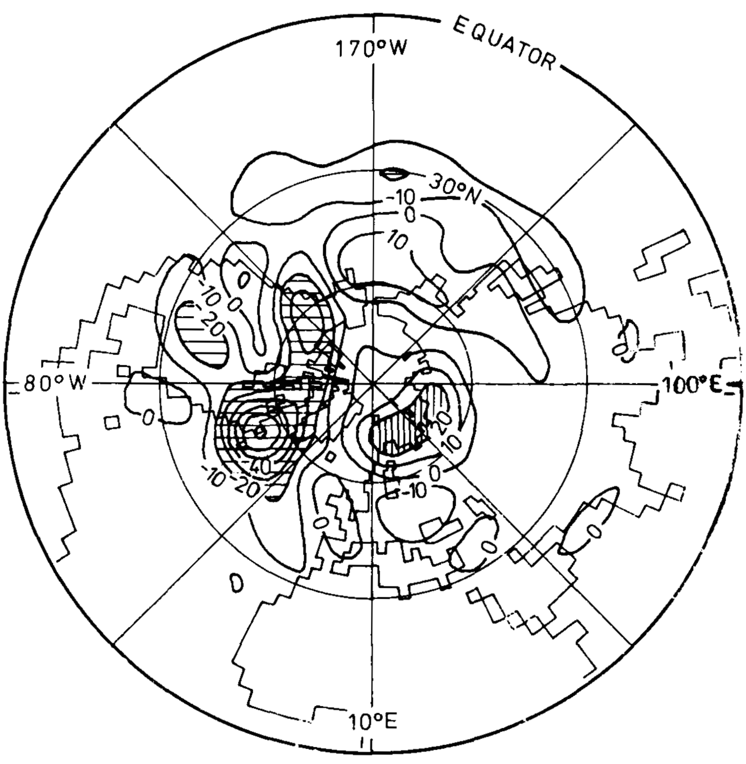

Fig. 13. The northern-hemispheric structure of the transient response (monthly mean of daily variance) generated by the ECMWF GCM.

7. Conclusions

We have reviewed the statistical significance test strategies appropriate for the analysis of multi-component signals in the presence of a relatively large noise background. We have demonstrated the way to apply them by using results of the El Niño responses experiment performed with the ECMWF general circulation model (Cubasch, 1983) and by consideration of observed circulation patterns (cf. Storch, 1984). We have focused our attention on the extra-tropical regions of the northern hemisphere, where univariate assessment methods failed to yield a consistent picture up to now—in contrast to the tropical areas.

We have found that the ECMWF model responds to a warm sea surface temperature anomaly of El Niño type with a significant extra-tropical hemispheric winter-mean pattern. The structure of this response is very stable for the 3 different model winters within a 270° sector of longitude.

The observed mean circulation of January 1983 is significantly outside the range of normal variations. The observation and the ECMWF simulation of the stationary response are visually similar with respect to location and signs of extrema, and this is confirmed by the statistical test. However, the simulated response vector also has a component orthogonal to the observation which is less significant than the parallel part.

Visual comparison of the hemispheric maps from observation and from the ECMWF simulation with 2 other GCM response patterns (GLAS and CNRM) yields as common features the north-east Pacific low, the Greenwich high, and the polar depression. The third "external" GCM (NCAR) has in common with the others only the Pacific low. However, no multivariate significance statements on these 3 latter GCM results are available; thus we do not know, e.g., whether the mutualities are by accident, or whether the differences are due to the GCMs' natural variability.

The experiment with the ECMWF model with a cold SST anomaly of El Niño structure leads to a response that is—if compared with the response to the warm anomaly—much less significant and less stable from model winter to model winter, but is on the average of 3 winters different from zero with weak significance. We conclude that the pressure cell in the north east Pacific reacts approximately linearly, whereas the more remote response is essentially nonlinear.

To the stationary response pattern is connected a changed spatial distribution of the transient activity. Owing to a disadvantageous signal-to-noise ratio and lacking reports on the GCMs' behaviour in this respect, the interpretation of our finding is difficult. Although the significance of the transient response is weaker than the stationary one, it seems physically reasonable.

8. Acknowledgements

We were supported by Drs. G. Fischer, K. Hasselmann and E. Roeckner in critical discussions and helpful comments. Dr. G. Hannoschöck provided us with his signal-to-noise analysis programs (SNAP), Dr. U. Cubasch (ECMWF) with his GCM El Niño experiment data, and Dr. M. Deque (CNRM) with Fig. 12c. The figures were thoroughly drawn by Mrs. M. Grunert. We wish to thank all of them.

9. Appendix A. Necessity of a-priori data compression

A very simple example is given to demonstrate the necessity of an *a priori* reduction of degrees of freedom in order to detect a signal.

Let x be a univariate random variable with vanishing expectation, $E(x) = 0$, and variance 1. Let $z = (x_1, \dots x_p)$ be the p-variate random vector with pairwise independent components x_i identical to x. We want to test by means of the χ^2 test whether one sample of the random vector $y := z + (a, 0, \dots 0)$ leads to the rejection of the (false) null hypothesis $E(y) = 0$. The test statistic is simply

$$\rho^2(y) = yy' = (x_1 + a)^2 + \sum_{i=2}^{p} x_i^2, \tag{A.1}$$

and its expectation

$$E(\rho^2(y)) = E(x^2) + a^2 + p - 1 = a^2 + p. \tag{A.2}$$

If for instance $a = 2$, according to Table 1, the null hypothesis will on average correctly be rejected at the 95% level of significance if $p < 3$. On the other hand, the signal will be detected only seldomly if more than, say $p = 5$ degrees of freedom are considered.

Table 1. *Dependence of the 95% quantile and the expected test statistic (A.2) on the dimension p with* $a = 2$

p	$E(\rho^2(y))$	95% quantile
1	5	3.8
2	6	4.0
3	7	7.8
4	8	9.5
5	9	11.1

10. Appendix B. The optimal signal-to-noise approach

By means of principal vectors (EOFs) and a good "first guess" (*a priori* conception on what really takes place), the probability of detecting a climate change by means of a χ^2 test may be increased considerably (Hasselmann, 1979b, Hannoschöck, 1984). We discuss this method by means of a simple example in order to demonstrate the principal power, and also to display the problems coming from the fact that the EOFs are badly estimated. For that purpose, let v be the true signal and x the random vector describing the undisturbed climate under consideration. Let us assume, without loss of generality, that its covariance matrix X is a diagonal one. Then, the principal vectors are the unit vectors with eigenvalues r identical to the diagonal elements of the covariance matrix. Let the indices be ordered according to the size of the eigenvalues. Furthermore, we need a first guess denoted by $\hat{v}$ with unit length.

We consider the projection b of the signal v on the guess vector $\hat{v}$:

$$b := v'\hat{v}. \tag{B.1}$$

Then, the test statistic $\rho^2(b)$ is given by

$$\rho^2(b) = (v'\hat{v})^2/\mathrm{var}(v'\hat{v}) = b^2/\hat{v}'X\hat{v} = b^2 \sum_i \hat{v}_i^2\, r_i, \tag{B.2}$$

where $\hat{v}_i$ is the ith component of the guess vector $\hat{v}$. The basic idea is to replace $\hat{v}$ by an "optimized" guess $L\hat{v}$ with that positive definite matrix L such that the $\rho^2(b_\mathrm{L})$ is maximum if $v = \hat{v}$. For that, b_L is defined to be $v'L\hat{v}$. It can be shown (Hasselmann, 1979b; Hannoschöck, 1984) that $\rho^2(b_\mathrm{L})$ is a maximum if $L = X^{-1}$ and that

$$\begin{aligned}\rho^2(b_\mathrm{L}) &= (v'L\hat{v})^2/\mathrm{var}(v'L\hat{v})\\ &= b^2(\hat{v}'\,X^{-1}\,\hat{v})^2/\hat{v}'\,L'\,XL\hat{v} = b^2\,\hat{v}'\,X^{-1}\,\hat{v}\\ &= b^2 \sum_i \hat{v}_i^2/r_i.\end{aligned} \tag{B.3}$$

The main difference between (B.2) and (B.3) is that the first is dominated by low-indexed (i.e. generally large-scaled, dynamically more important) principal vectors. On the other hand, the optimized version (B.3) is controlled by high-indexed (i.e. generally small-scaled, dynamically less or not important) components. Unfortunately, when working with real data, the true covariance matrix is unknown. Thus the eigenvalues r_i can only be estimated. However, the standard estimator is biased (Storch, 1983) such that the low-indexed eigenvalues are underestimated and the high-indexed ones considerably overestimated.

In order to demonstrate the consequences, we present a Monte Carlo example. Let the dimension of the problem be 7 and the eigenvalues of the true covariance matrix be 5.0, 2.5, 1.3, 0.6, 0.3, 0.2 and 0.1. Let the guess vector be $\hat{v} = (1,1,1,1,1,1,1)/\sqrt{7}$. If $v = \hat{v}$ we get, with the true eigenvalues,

$$\rho^2(b) = 0.70, \quad \text{and} \quad \rho^2(b_\mathrm{L}) = 3.05. \tag{B.4}$$

As can be seen, the significance becomes considerably increased by the optimization procedure. However, if a sample of size, say 10, is available to estimate the eigenvalues, we get on average the following biases: 18%, 0%, −22%, −27%, −33%, −60% and −70% for the sequence of eigenvalues. Inserting the mean estimated quantities, one gets instead of (B.4)

$$\rho^2(b) = 0.71, \qquad \rho^2(b_\mathrm{L}) = 7.81.$$

Thus, we get a clearly overestimated test statistic by the optimized method, namely by a factor of 2.6. Even if 100 samples are available to describe the system's variability, the factor is still 1.1. To summarize: as long as the sample is not large, we have to expect the optimized test to be radical, i.e., yields too high a test statistic.

REFERENCES

Arkin, P. A., Kopman, J. D. and Reynolds, R. W. 1983. Event quick look atlas: 1982–1983 El Niño/southern oscillation. November 1983 NOAA/National Weather Service, Num. Mod. Centre, Washington D.C. 20 233.

Barnett, T. P., Preisendorfer, R. W., Goldstein, L. M. and Hasselmann, K. 1981. Significance tests for regression model hierarchies. *J. Phys. Oceanogr. 11*, 1150–1154.

Blackmon, M. L., Geisler, J. E. and Pitcher, E. J. 1983. A general circulation model study of January climate anomaly patterns associated with interannual variations of equatorial Pacific sea surface temperatures. *J. Atmos. Sci. 40*, 1410–1425.

Cane, M. A. 1983. Oceanographic events during El Niño. *Science 222*, 1189–1194.

Chen, W. Y. 1983. The climate of Spring 1983—a season with persistent global anomalies associated with El Niño. *Mon. Wea. Rev. 111*, 2371–2384.

Chervin, R. M., Gates, W. L. and Schneider, S. H. 1974. The effect of time averaging on the noise level of climatological statistics generated by atmospheric general circulation models. *J. Atmos. Sci. 31*, 2216–2219.

Chervin, R. M. and Schneider, S. H. 1976a. A study of the response of NCAR GCM climatological statistics to random perturbations: estimating noise levels. *J. Atmos. Sci. 33*, 391–404.

Chervin, R. M. and Schneider, S. H. 1976b. On determining the statistical significance of climate experiments with general circulation models. *J. Atmos. Sci. 33*, 405–412.

Chiu, W. C., Lo, A. Weider jr, D. H. and Fulker, D. 1981. A study of the possible relationship between the tropical Pacific sea surface temperature and atmospheric circulation. *Mon. Wea. Rev. 109*, 1013–1020.

Conover, W. J. 1971. *Practical nonparametric statistics*. Wiley, New York London Sydney Toronto.

Cubasch, U. 1983. The response of the ECMWF global model to the El-Niño anomaly in extended range prediction experiments. Technical Report 38, European Centre for Medium Range Weather Forecast, Reading, Berkshire, UK.

Deque, M. and Royer, J. F. 1984. A general circulation model simulation of the atmospheric response to the January 1983 El Niño. *Tropical Ocean-Atmos. Newsl. 27*, 2–4.

Hannoschöck, G. 1984. A multivariate signal-to-noise analysis of the response of an atmospheric circulation model to SST anomalies. Hamburger Geophysikalische Einzelschriften, A67 (Meteorologisches Institut der Universität Hamburg, Bundesstr. 55, 2000 Hamburg 13, FR Germany).

Hasselmann, K. 1979a. Linear statistical models. *Dyn. Atmos. Ocean. 3*, 501–521.

Hasselmann, K. 1979b. On the signal-to-noise problem in atmospheric response studies. *Meteorology of Tropical Oceans*. Roy. Met. Soc., 251–258.

Hayashi, Y. 1982. Confidence intervals of a climatic signal. *J. Atmos. Sci. 39*, 1895–1905.

Kiladis, G. and Diaz, H. F. 1984. A comparison of the 1982-83 and 1877–78 ENSO events. *Tropical Ocean-Atmos. Newsl. 25*, 7–9.

Krueger, A. F. 1983. The climate of Autumn 1982, with a discussion of the major tropical Pacific anomaly. *Mon. Wea. Rev. 111*, 1103–1118.

Leith, C. E. 1972. The standard error of time average estimates of climatic means. *J. Appl. Meteorol. 12*, 1066–1069.

Livezey, R. E. 1983. Statistical analysis of general circulation model climate simulation, sensitivity and prediction experiments. Proc. II International Meeting on Statistical Climatology, September 26–30, 1983. Lisboa, Portugal, Instituto de Meteorologica e Geofisica, Rua C. Aeroporto, 1700 Lisboa.

Livezey, R. E. and Chen, W. Y. 1983. Statistical field significance and its determination by Monte Carlo techniques. *Mon. Wea. Rev. 111*, 46–59.

O'Brien, J. J. 1978. El Niño—an example of ocean-atmosphere interactions. *Oceanus 21*, 40–46.

Preisendorfer, R. W. and Barnett, T. P. 1983. Numerical model—reality intercomparison tests using small sample statistics. *J. Atmos. Sci. 40*, 1884–1896.

Quiroz, R. S. 1983. The climate of the 'El Niño' winter of 1982/83—a season of extraordinary climate anomalies. *Mon. Wea. Rev. 111*, 1685–1706.

Rasmusson, E. and Carpenter, T. 1981. Variations in tropical SST and surface wind fields associated with the southern oscillation/El Niño. *Mon. Wea. Rev. 110*, 354–384.

Shukla, J. and Wallace, J. M. 1983. Numerical simulation of the atmospheric response to equatorial Pacific sea surface temperature anomalies. *J. Atmos. Sci. 40*, 1613–1630.

Storch, H.v. 1982. A remark on Chervin-Schneider's algorithm to text significance. *J. Atmos. Sci. 39*, 187–189.

Storch, H.v. 1983. Statistical aspects of empirical orthogonal functions based on small sample sizes. Proc. II International Meeting on Statistical Climatology, September 26–30, 1983, Lisboa, Portugal, Instituto de Meteorologica e Geofisica, Rua C. Aeroporto, 1700 Lisboa.

Storch, H.v. 1984. An accidental result: the mean 1983 January 500 mb height field significantly different from its 1967–81 predecessors. *Beitr. Phys. Atmos. 57*, 440–444.

Storch, H.v. and Roeckner, E. 1983a. Methods for the verification of general circulation models applied to the Hamburg University GCM. Part I: test of individual climate states. *Mon. Wea. Rev. 111*, 1965–76.

Storch, H.v. and Roeckner, E. 1983b. On the verification of January GCM simulations. Proc. II International Meeting on Statistical Climatology, September 26–30, 1983, Lisboa, Portugal, Instituto de Meteorologica e Geofisica, Rua C. Aeroporto, 1700 Lisboa.

Veysseire, J.-M. 1983. A study of the statistical significance of the response of a spectral general circulation model to a modification of solar radiation. Proc. II International Meeting on Statistical Climatology, September 26–30, 1983, Lisboa, Portugal, Instituto de Meteorologica e Geofisica, Rua C. Aeroporto, 1700 Lisboa.

Wagner, A. J. 1983. The climate of Summer 1982. A season with increasingly anomalous circulation over the equatorial Pacific. *Mon. Wea. Rev.* *111*, 590–601.

Warshaw, M. and Rapp, R. R. 1972. An experiment on the sensitivity of a global circulation model. *J. Appl. Meteorol.* *12*, 43–49.

Zwiers, F. W. 1983. Evaluating climate change experiments. Proc. II International Meeting on Statistical Climatology, September 26–30, 1983, Lisboa, Portugal, Instituto de Meteorological e Geofisica, Rua C. Aeroporto, 1700 Lisboa.