Tellus (1995), 47A, 955–973
Printed in Belgium – all rights reserved

TELLUS
ISSN 0280–6495

Simulated annealing and weather regimes classification

By A. HANNACHI* and B. LEGRAS, *Laboratoire de Météorologie Dynamique, École Normale Supérieure, 24, rue Lhomond, 75231 Paris Cedex 05, France*

(Manuscript received 19 July 1994; in final form 16 December 1994)

ABSTRACT

A new approach based on the travelling salesman problem and simulated annealing is used to classify weather maps and to associate them with weather regimes. The usual classification methods are able to identify the preferred patterns of large-scale atmospheric flow. They fail, however, in characterizing the boundaries of various regimes which are generally a result of the geometrical constraints imposed by the chosen method. The method is aimed at overcoming this difficulty and helping to define the boundaries between weather regimes in order to study the characteristics of and transitions between these regimes. Weather regimes are derived from a series of 37 winters of 700 hPa geopotential height observations using the algorithm developed by Vautard (1990). The new approach is then applied to assign each observation to its corresponding weather regime. It is shown that the method provides results which agree with the literature and helps to establish recurrence in agreement with observations. We then, speculate on the relationship which may link the boundaries of weather regimes to the transitions between these regimes.

1. Introduction

It is well-known from observations that the large scale atmospheric flow tends to remain quasi-stationary for considerable periods around large scale flow patterns called *weather regimes* which are now fairly well documented. Kimoto (1989) defined weather regimes using local maxima of a multivariate probability density function and found large scale flow patterns that appeared repeatedly and persisted beyond synoptic time scales. Legras et al. (1987), Mo and Ghil (1988) and Molteni et al. (1988, 1990) used clustering algorithms to gather atmospheric flow patterns. In a related study Dole and Gordon (1983) defined weather regimes by a persistent anomaly of geopotential height, i.e., as those situations where the absolute value at a key point exceeds a given threshold value for some minimum period of time. In spite of the different methods, large similarities exist in fact between the patterns obtained by various approaches.

The transitions between atmospheric regimes have been studied theoretically by Egger (1981) who used the Fokker-Plank equation for the probability density function of a barotropic system when perturbed by a stochastic forcing. He discusses the frequency of occurrence and of switching of states located in the vicinity of two stationary states named as high and low index, as defined by Charney and DeVore (1979) which have the maximum frequency of occurrence. This probability of switching is linked with the properties of the boundary separating the potential wells of the unperturbed system in the phase space.

Mo and Ghil (1988) investigated a clustering analysis based on the pattern correlations between anomalies of stream function, which correspond to using the cosine of the angle between two maps as clustering criterion. Applying this method to a long stream function time series obtained from a spherical barotropic model, they found 5 clusters among which the 2 largest exhibit blocking and zonal flow patterns respectively. They found that there is no direct transition between these two

* Corresponding author. Present affiliation: Department of Meteorology, Edinburgh University, Edinburgh EH9 3JZ, UK.

regimes. Molteni et al. (1990), using maximum of probability density function, found 6 clusters representing regimes in the northern extratropical winter circulation. The transition frequencies between the clusters are dependent on the amplitude of the perturbation.

In this study, weather regimes are defined as in Vautard (1990) and identified as quasi-stationary solutions of the large scale geopotential tendency equation. A quasi-Newtonian algorithm was used to find out these solutions by minimizing a tendency function as explained in Section 2. The search is carried out within the phase space given by the first 9 dominant spatial modes or Empirical Orthogonal Functions (EOFs) of the sample covariance matrix of the height anomalies over the N. Atlantic. The data consists of 37 years of 700 hPa geopotential routine analysis field from NMC (Vautard, 1990).

Usual classification methods tend to form clusters with smooth boundaries between patterns of large scale atmospheric flow (Legras et al., 1988). This means that the different classes are bounded by polygonal boxes or by spherical surfaces (see Subsection 3.4). For example when using a similarity criterion based on the correlation between anomaly maps of streamfunction, as done in Mo and Ghil, 1988, the clusters are conical. There is not any reason, however, to expect such simple boundaries from the nature of atmospheric flows. Clusters of any shape are indeed expected, and in particular, the boundaries may be fractal. Unlike the classical methods where the shape of clusters is fixed by the search method itself, we are motivated here to look for an agglomeration in which cluster shapes depend widely on the nature of the data set itself. This is a natural way to perform the agglomeration, especially when the data set is a result of a temporal evolution process, like those encountered in meteorology.

For this purpose, we seek an agglomeration of the neighbouring states following some preferred routes between neighbouring points. We use a minimization principle to search for the shortest closed trajectory in an Euclidian phase space spanned by the first few representative EOFs using a simple and effective simulated annealing algorithm to solve the travelling salesman problem (TSP) for the classification. Some of the characteristics of weather regimes are obtained from this study using Markov chain, and a visualization on low dimensional space of the boundary regimes is derived from this analysis.

This manuscript is organized as follows; in the following section, we review the definition and computation of the weather regimes as in Vautard (1990). The third section will discuss the problems encountered with previous methods of meteorological classification. In Section 4, we introduce in detail, the procedure of clustering which leads to the TSP problem (Appendix A) and the application of the simulated annealing algorithm to solve the problem. We perform some tests based on clustering and dynamics and compare our results with the results of k-mean method in Section 5. In Section 6, we apply our method to the atmospheric data set and we describe the regime boundaries and transition properties. Conclusions and further discussions are offered in the final section.

2. Definition and derivation of weather regimes

The evolution equation of the large-scale pattern, as formulated by Vautard and Legras (1988), requires a separation of the observed motion into complementary phase spaces of large (L) and small (S) scales. The separation can be performed on the basis of an EOF analysis. Under this hypothesis, we write;

$$\frac{dL}{dt} = X(L, S), \tag{1}$$

where $X(L, S)$ is a functional in the $L - S$ phase space representing the dynamical system of the large-scale atmospheric evolution and contains mixed interactions between large and small scales. Since we seek an equation for the large-scale flow only, the dependence of X upon S occurs through a conditional probability average. As a consequence of this closure hypothesis, the evolution vector field X of (1) is not known analytically, and we rather use the following estimation from the data set (Vautard, 1990);

$$T(L) = \left\langle \left. \frac{dL'}{dt} \right|_{L'=L} \right\rangle, \tag{2}$$

where $\langle\ \rangle$ is a conditional average of the temporal variation of the large-scale flow L over observable

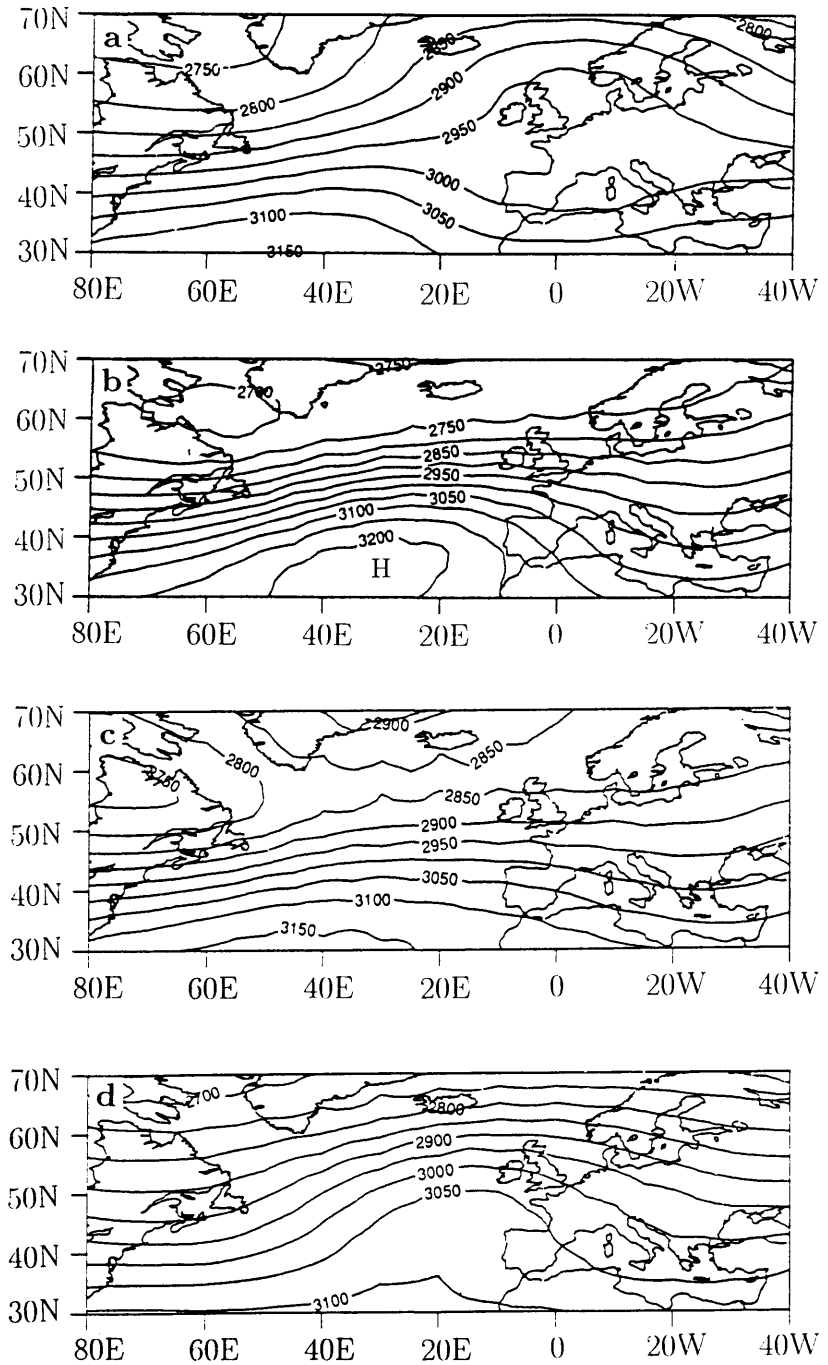

Fig. 1. Geopotential height (*m*) of the 4 weather regimes obtained by the equilibration method; the Block (BL) (a), the Zonal (ZL) (b), the Greenland-Anticyclone (GA) (c) and the Atlantic Ridge (AR) (d). The contour interval in (c) is 50 m.

atmospheric states. If this average is weighted using the ergodic measure of the attractor, then it is an ensemble average. The quasi-stationary solutions or regimes are then obtained as the large scale flows satisfying;

$$T(L) = 0. \tag{3}$$

Vautard (1990) used the NMC data set, composed of 37-winter series of 700 hPa geopotential maps over the Atlantic area (30° N–70° N, 80° W–40° E). The data set is filtered using a low pass filter to remove transient fluctuations of time scales less than 10 days and sampled every 2 days to avoid redundancy and then projected onto the first nine EOF's which explain 94% of the total low-pass variance. Notice that the first three EOFs yet explain 55% of the low-pass variance. As the observations only provide a sampling of the attractor, further averaging is required to define the statistical tendencies. A differentiable radial weighting function $\phi(x)$, is used. It is a bell function with $\phi(0) = 1$ and decreasing to zero rapidly as x increases.

The statistical tendency $T(L)$ of (2), for any given flow L, is then;

$$T(L) = \frac{\sum_{i=2}^{n-1} \phi(\|L - L_i\|)(L_{i+1} - L_{i-1})}{2\tau \sum_{i=2}^{n-1} \phi(\|L - L_i\|)}, \tag{4}$$

where $\tau = 1$ day and n is the length of the data set one day apart. The norm $\|L - L_i\|$ is the distance between the large scale atmospheric flow L (a point in phase space) and the large-scale data point L_i. The width (or radius) of the weighting function ϕ, is chosen such that about 100 neighbours are used (with non-vanishing weight) on the average in the definition of the statistical tendency. This radius will be referred to as the influence sphere and will be considered latter. Here the tendency $T(L)$ given by (4) depends smoothly on L, so that eq. (3) can be solved using quasi-Newton optimization algorithm. Each solution (or regime) is accompanied by its influence sphere, which is simply the ball centered on the weather regime with the width of the weighting function as diameter. The ensemble of points (observations within the phase space) falling inside the influence sphere are the *kernel* of the corresponding regime and will be considered latter. There is a kind of one-to-one correspondence between the weather regimes and their corresponding kernels in the sense that the latter constitute a small ball of relatively high density around the weather regimes. The 4 solutions obtained are briefly described here (for further details see Vautard, 1990).

(i) The blocking regime (BL hereafter): typical European blocking pattern, bringing Easterlies over western Europe (Fig. 1a).

(ii) The zonal regime (ZO hereafter): flow is more parallel with the jet extending across the Atlantic (Fig. 1b).

(iii) The Greenland anticyclone (GA hereafter): identified by a strong positive anomaly sitting over Greenland, whereas the southern part of the pattern is more zonally symmetric (Fig. 1c). Unlike the blocking regime but like the zonal regime, this flow structure allows synoptic perturbations to traval across the Atlantic towards Europe.

(iv) The Atlantic ridge (AR hereafter): the prominent feature of this regime is a positive anomaly extending over the Atlantic and western coast of Europe, showing a strong ridge over the eastern Atlantic, bringing over Europe moist cold air advected from the north Atlantic (Fig. 1d).

3. Weather regimes classification and boundaries

Classification originates from the field of pattern recognition and constitutes a class of problems that can often be easily solved by human beings, but are very hard to solve by computers (Aart and Korst, 1990). To overcome this difficulty, there has been considerable interest in connection network models, using Boltzmann machines to solve these problems (Aart and Korst, 1990).

Clustering methods are widely used in pattern recognition (Diday and Simon, 1976) and data analysis in many fields ranging from archeology to economics and meteorology (Gordon, 1981). In some cases, as in meteorology, where the temporal evolution process is naturally present, data can be ordered in time, but this ordering is generally not exploited.

The usual methods attempt to identify centroids and to form groups which accumulate around these centroids. The boundaries of these groups are usually strongly constrained by the method itself rather than by the distribution of data. Their

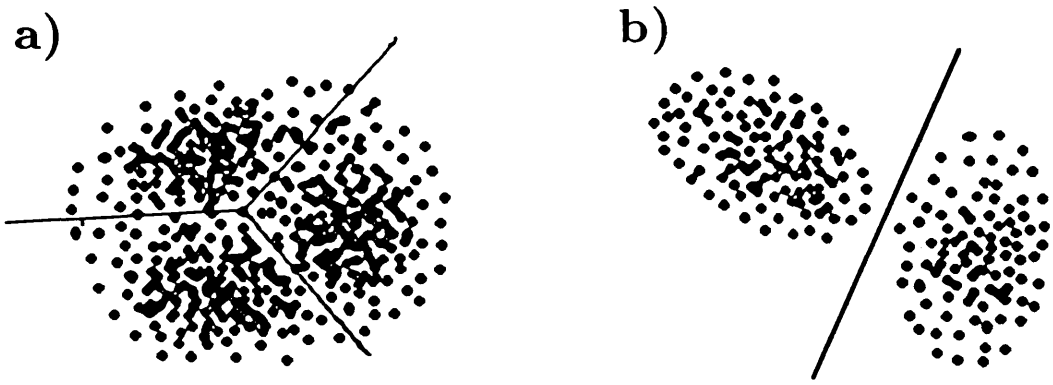

Fig. 2. Example of clusters with simple boundaries. (a) elliptic clusters separated by a hyperplane. (b) clusters with elliptic-polygonal boundaries.

shape, depending on the metric used, is generally spherical or ellipsoidal if the metric is quadratic, which is usually the case. These special kinds of shapes are defined following the principal axes of variance, and the different classes appear as a portion of this very simple shape and the boundary between them is given in general by some portion of a smooth hypersurface, often a plane in 2D case (Figs. 2). Although widely used, these methods pose several problems, when the real shapes of the clusters are not of this simple kind. Figs. 3 shows examples of clusters which are not of this simple kind. We are interested here to look for a method which depends directly on the distribution of data rather than on the geometrical constraints fixed by the method itself.

The atmosphere varies according to very complicated physical and dynamical processes. Boundaries between weather regimes are likely to be complicated, and may be fractal. This can be seen for example from the link which may exist between the regimes and the atmospheric attractor. Lorenz (1963) has shown in fact, that a simplified model of atmospheric convection, may exhibit an attractor with fractal topology, that is usually called strange attractor. Chaos is present at some scales, but organized structures may also be present at other scales (see, e.g., Zaslavsky et al., 1992). It is a common conjecture that these organized structures could be obtained as solutions of some general minimization process (see, e.g., Shepherd, 1990). As the trajectory of motion minimizes the time, there are other trajectories which minimize other criteria. In phase space, although the historical

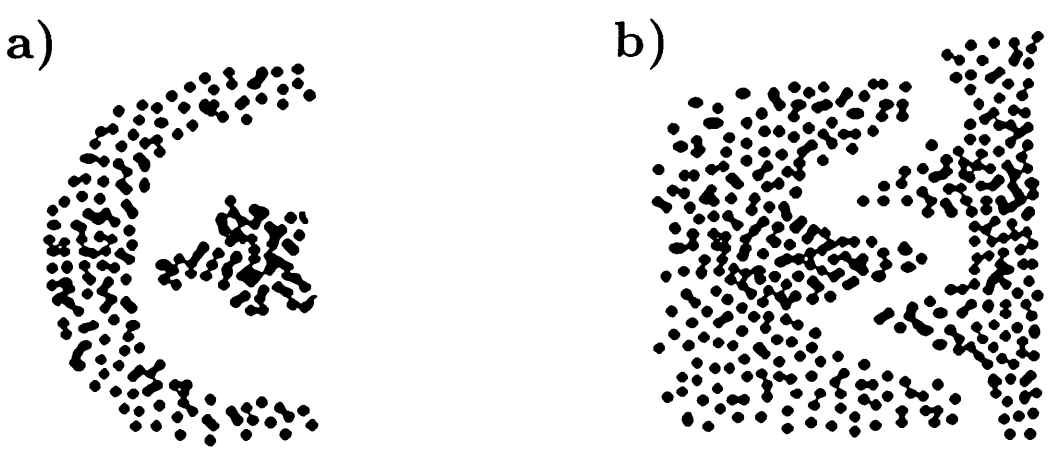

Fig. 3. Examples of clusters which may not be easily distinguishable by usual methods.

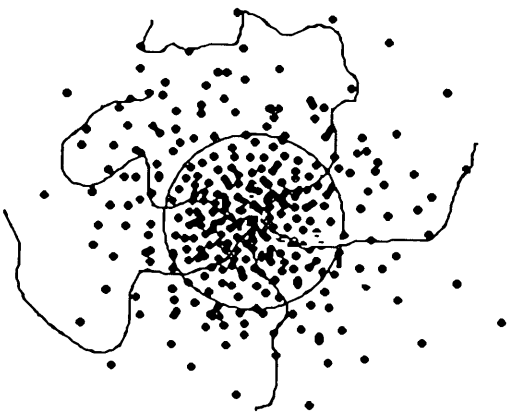

Fig. 4. Schematic representation of the mode of aggregation of atmospheric states following neighbouring states in the EOF phase space. The kernel of the weather regime is represented by the set of points inside the circle. The center of the circle (black square) represents the weather regime. Atmospheric states are represented by the points. The curves represent routes between neighbouring states.

trajectory determines the transitions, it is not the only possible route connecting the points of a time series. Thus, our first principle is to classify the atmospheric states according to some preferred routes in phase space. In order, we observe that the kernels of weather regimes are often an agglomeration defining a high density area. We will look for an aggregation of atmospheric states around these weather regimes, following routes which have the minimum spatial separation between neighbouring states (Fig. 4).

This requires us to explore many routes for optimization. To take into account some symmetry principle of the problem, we are looking for the closed minimized route passing through all points of the data set in phase space. The optimization criterion is formulated with respect to the Euclidian metric within the EOF space. This is the well known travelling salesman problem (TSP) (Appendix A). The application to classification is now explained in the next section.

4. Aggregation of atmospheric states around the regimes

4.1. The TSP algorithm applied to the atmospheric data set

To apply the TSP to our case, we consider the atmospheric data set described earlier as a set of points described by the first N EOF coefficients in an N-dimensional phase space. These points constitute the "cities" for the TSP which search the shortest closed path passing through these points (see Appendix A for the formulation of this problem). To solve this problem we use a simulated annealing algorithm which draws its name from an analogy with crystal processing. The basic idea is that the energy E of a crystal in thermal equilibrium at temperature T, has a Boltzmann probability distribution:

$$P(E) \sim \mathrm{e}^{-E/kT}, \tag{5}$$

where k is the Boltzmann constant. The algorithm minimizes the energy E using this probability distribution for energy jumps (Appendix B).

4.2. The complete algorithm for the TSP

We describe briefly the different steps of the algorithm.

(i) For a given ordering of the points in phase space (cities) labeled by $i = 1, 2, ..., n$, with coordinates, e.g., (x_i, y_i) if $N = 2$, the energy E is defined as:

$$E^2 = \sum_{i=1}^{n} ((x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2). \tag{6}$$

(ii) Generate some initial configuration, e.g., the temporal trajectory, and calculate its corresponding energy E_0.

(iii) Choose the initial temperature T.

(iv) Change the configuration (the path) by choosing a set of moves (Lin and Kernighan, 1973) (Fig. 5) and calculate the new energy E_1 and energy variation $\Delta E = E_1 - E_0$. Then;

- if $\Delta E \leqslant 0$: the new configuration is accepted;
- if $\Delta E > 0$: this configuration is accepted with probability $\mathrm{e}^{-\Delta E/kT}$. For this later condition, Metropolis et al. (1953) propose to generate a random number a uniformly distributed between 0 and 1, and to check if $a \leqslant \mathrm{e}^{-\Delta E/kT}$, otherwise the new path is rejected (see, also, Kirkpatrick et al., 1983).

(v) Decrease the temperature linearly (see Appendix B) and go to (iv).

The simulated annealing as a minimization algorithm is advantageous compared to other

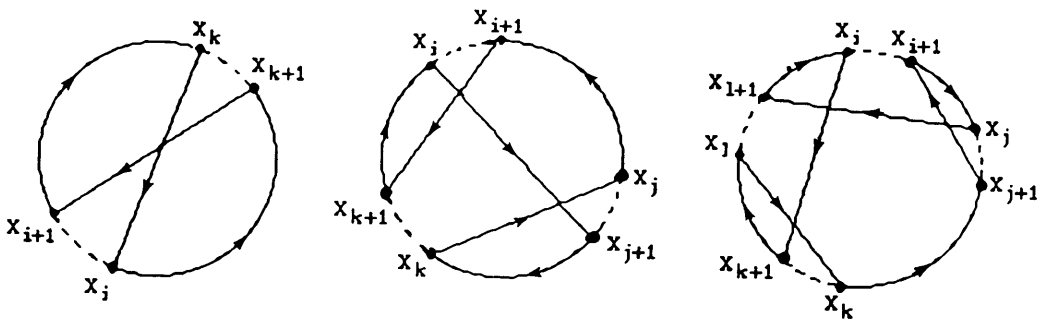

Fig. 5. A set of three kinds of moves for the configuration.

deterministic algorithms. It is well-known, for instance, that the Newton algorithm converges when it starts from points within a small neighbourhood of a minimum point. If this latter is a local minimum (which is in general the case) the algorithm converges to this point and cannot reach the global minimum (Fig. 6a). On the other hand, the annealing algorithm is based on a relaxation principle expressed by (5). This allows the algorithm to jump from a local minimum to other ones and eventually to reach the global minimum (Fig. 6b) under the condition stated above.

4.3. *Aggregation of atmospheric states around the regimes*

The aggregation of atmospheric states around weather regimes is conducted from the kernels of these regimes. We denote by Ker_i the kernel of the ith regime composed of all points falling inside the influence sphere of the ith regime. This small set of points characterizes accurately the regime.

We denote, also, by (x_i^k) the set of points of the kernel Ker_i. An observed state x is assigned to the ith regime if and only if there exists a point x_i^k in the kernel Ker_i which is nearer to x than all other points of all other kernels, where the distance between atmospheric points is now computed following the shortest closed path passing through all observed states (see below). In other words, we write:

$$x \in \text{regime } i \Leftrightarrow \exists x_i^k \in \mathrm{Ker}_i / \delta(x, x_i^k) \leqslant \delta(x, x_j^l), \qquad \forall j, l, \tag{7}$$

where δ is the distance (see below), defined (Fig. 7) after solving the travelling salesman problem over

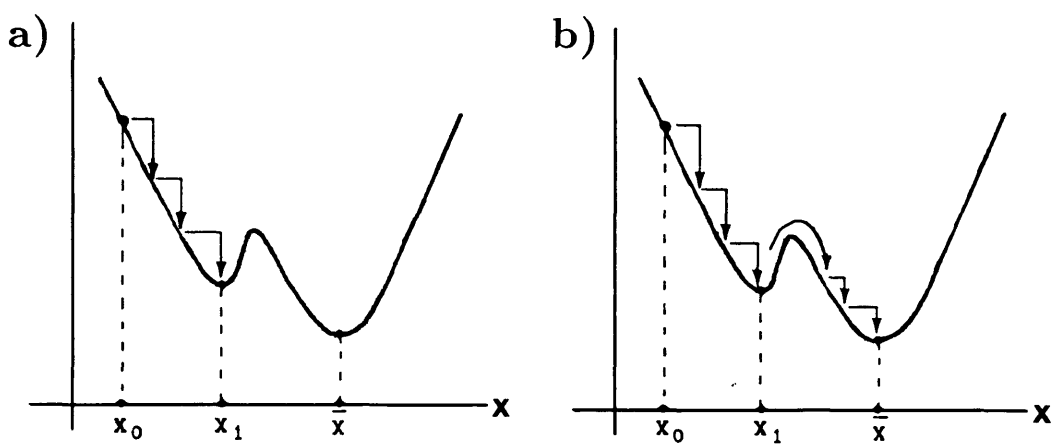

Fig. 6. Comparison between Newton algorithm without jumping (a), and the annealing algorithm with the possibility of a jump (b).

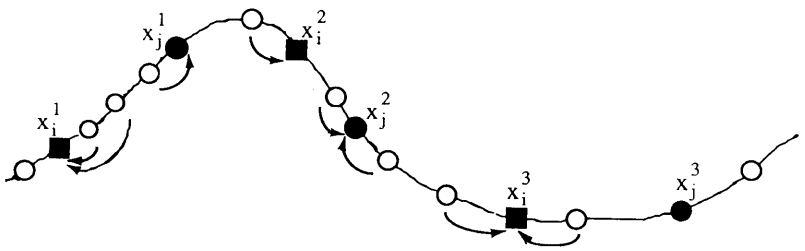

Fig. 7. Mode of aggregation of atmospheric states around weather regimes following the minimal trajectory. Dark elements constitute the points of different kernels of the regimes, with a different shape for each regime. Here, we have represented only two shapes. The atmospheric states not classified are represented by white circles.

the whole set of points of the observed atmospheric states in phase space.

Relation (7) means that, an observed state x in phase space is assigned to the ith regime if and only if it is closer to the kernel Ker_i (as a set of points) of this regime than any other kernel, which can be formally written as:

$$x \in i\text{th regime} \Leftrightarrow \delta(x, \mathrm{Ker}_i) \leqslant \delta(x, \mathrm{Ker}_j), \quad \forall j, \quad (8)$$

where

$$\delta(x, \mathrm{Ker}_\alpha) = \min\{\delta(x, x_\alpha^k), x_\alpha^k \in \mathrm{Ker}_\alpha\}. \quad (9)$$

The metric δ between the observed points is defined as follows. If d denotes the Euclidian metric and if the points in the final configuration are enumerated as 1, 2, ..., n (along the minimal route) then the new metric δ is defined as follows.

If we designate by

$$A = d(x_k, x_{k+1}) + \cdots + d(x_{l-1}, x_l)$$

and

$$B = d(x_l, x_{l+1}) + \cdots + d(x_{n-1}, x_n) + d(x_n, x_1) + \cdots + d(x_{k-1}, x_k)$$

for $k \leqslant l$. Then the new metric δ between x_k and x_l is given by;

$$\delta(x_k, x_l) = \min\{A, B\}. \quad (10)$$

It can be seen immediately that this defines a distance over all data points.

5. Test including clustering and dynamics

Before applying our classification method on the 37 years of NMC winter data and the weather regimes mentioned earlier, we will compare it with other methods encountered in this field.

5.1. *Comparison with clustering tests*

There exists in the literature many different methods of cluster analysis which lead generally to an ellipsoidal shapes separeted by hyperplanes. Although the same kernels are generally obtained, the partitioning of the data set strongly depends on the metric used. For example, if one uses the L_∞ norm, that is:

$$\|x\|_\infty = \max_i |x_i|, \quad (11)$$

then the limiting hyperplanes of iso-norm bounds parallelepipedic classes in phase space.

To be consistent, any method must be compared with other known techniques. Here, we use as a reference the so called "k-mean method", often used in classification analysis. The basic k-mean clustering attempts to find a classification which minimizes the sum of the squares within the cluster. The sum of the squared difference of each observation about its assigned cluster mean is used as the criterion for the assignment:

$$\phi = \sum_{i=1}^{k} \sum_{j=1}^{p} \sum_{m=1}^{n_i} \omega_{v_{im}} \delta^{v_{im,j}} (x_{v_{im,j}} - \overline{x_{ij}})^2, \quad (12)$$

where we use:

v_{im}: row index of the mth observation in the ith cluster.

n_i: number of elements of the ith group.

ω: number of elements of the ith group.

δ: 0 if the jth variable on observation v_{im} is missing, 1 otherwise.

$\overline{x_{ij}}$: average of the available observations for variable i over group j.

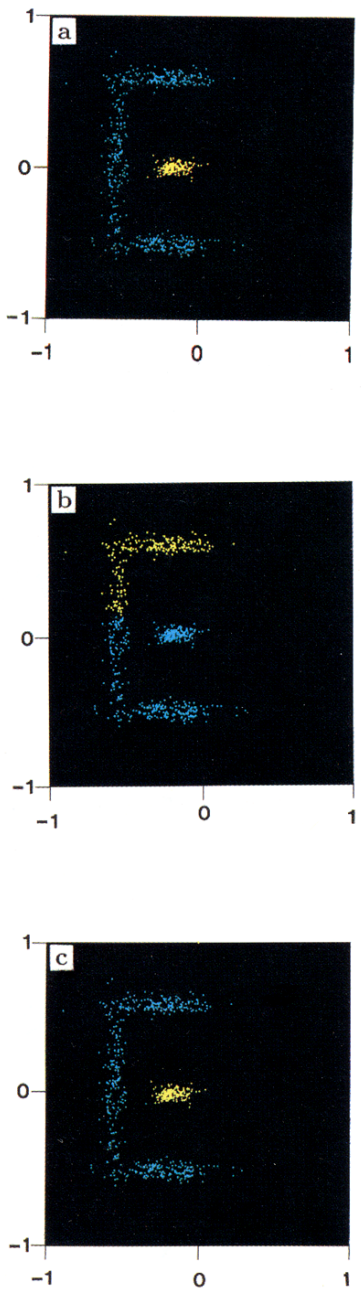

Fig. 8. Comparison between the present method and the k-mean method. (a) Two different chosen distributions. (b) The obtained clusters by the k-mean method. (c) The obtained clusters by the present method.

In the standard k-mean algorithm, observations are transferred from one cluster to another when this operation decreases the sum of squared differences within the cluster. The algorithm stops when a pass through the entire data set does not produce any new transfer. The algorithm computes k means (centroids) Euclidian metric clusters for an input matrix-data starting with initial estimates of the k clusters means.

The comparison with our method has been first done on a planar distribution generated from the superposition of three identical axisymmetric Gaussian distributions centered at different locations for which the k-mean method is known to give good results. Starting from the same kernels, the two methods are in good agreement (Fig. 9 and 10 of Hannachi, 1991). In another test where the shapes of the Gaussian distributions have been changed, it can be seen (Fig. 11 and 12 of Hannachi, 1991) that one cluster was badly estimated by the k-mean while the present gives better result. Fig. 8 shows the result of a test where one cluster is surrounded by another one. In this case, the k-mean method fails completely to recognize the distribution because it attempts to form clusters separeted by hyperplanes while our method gives the correct partition.

5.2. *Test including dynamics*

We consider now a case which is closer to our representation of atmospheric dynamics, that is a simple stochastic dynamical system with a potential function possessing several wells indicating the existence of preferential stagnation situations (or regimes) of the system. Such systems have been used as ideal models to explore the transitions between weather regimes by Sutera (1980) and Egger (1981).

We consider the following two dimensional stochastic dynamical system:

$$\mathrm{d}X(t) = -\nabla V(X)\,\mathrm{d}t + \varepsilon^{1/2}\,\mathrm{d}W_t, \tag{13}$$

where $V(X)$ is a potential with three axisymmetric wells and the perturbation W_t is a Brownian process with amplitude given by the diffusion coefficient ε. The potential must be smooth enough in order to ensure the existence and the uniqueness of the solution of (13). We first define the profile of a single well v as:

$$v(x)=\begin{cases}\alpha e^{1/(x^2-a^2)} & \text{if } |x|<a \\ 0 & \text{otherwise,}\end{cases} \tag{14}$$

where α and a are positive parameters respectively equal to 20 and 0.5 in our test.

Then we consider the potential which combines three such wells respectively centered on the vertices (0, 0), (0, 2a) and (a, $a\sqrt{3}$) of an equilateral triangle. In addition, we confine the solution to the vicinity of these three wells by adding a quadratic potential with vertex at the center of the triangle. The final expression of the potential V is given by:

$$\begin{aligned}V(x,y)&=v(x^2+y^2)+v((x-2a)^2+y^2)\\&\quad+v((x-a)^2+(y-a\sqrt{3})^2)\\&\quad+b\left((x-a)^2+\left(y-\frac{a}{\sqrt{(3)}}\right)^2\right),\end{aligned} \tag{15}$$

where b is a positive parameter chosen equal to 0.25.

The probability density P of the Markovian stochastic process $X_\varepsilon(t)$, solution to (13) satisfies the Fokker-Plank equation:

$$\frac{\partial P}{\partial t}=\frac{\varepsilon}{2}\nabla^2 P+\nabla V\cdot\nabla P+P\nabla^2 V. \tag{16}$$

It is well-known that the steady solution of (16) is:

$$P(X)=ce^{-2V(X)/\varepsilon}, \tag{17}$$

where c is a normalization constant.

By the law of large numbers, the solution $X_\varepsilon(t)$ to (13) tends to $X_0(t)$ when ε goes to zero, with $X_0(t)$ being the solution of the unperturbed system of (13). For small enough ε, the system stays for a long time within one well before jumping to the others.

We are rather interested in situation for which the solution cannot be trapped within one well for a long period and often crosses the potential barrier which separates the wells. After several trials we have chosen the value 0.38 for ε. The data set of points resulting from numerical integration of the stochastic system is shown in Fig. 9a, while the trajectory of motion of the Markovian particle is presented in Fig. 9b, which exhibits a dark area on the boundary between the centers of the wells, corresponding to the preferred transition paths between regimes. This trajectory is taken as initial state for the TSP for which the final trajectory is

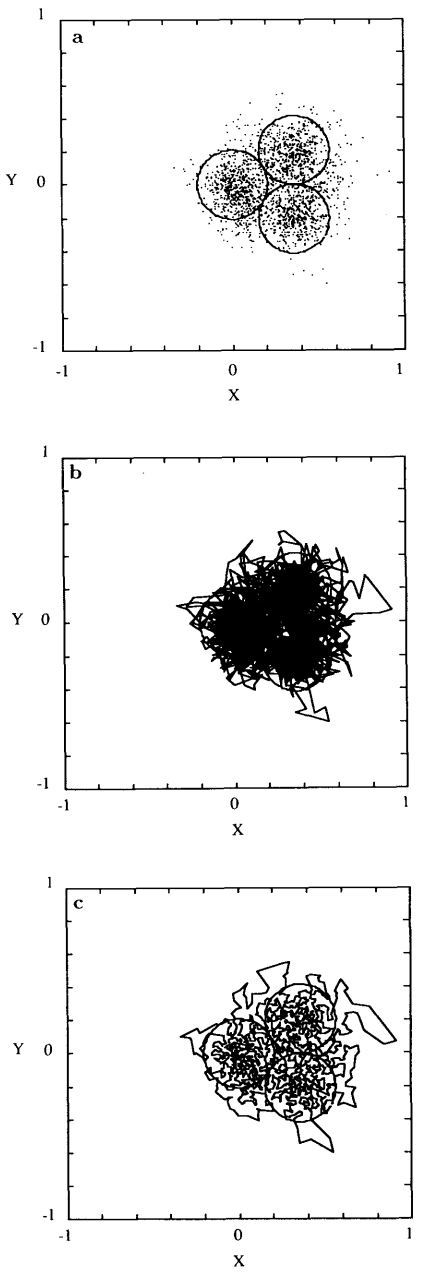

Fig. 9. The output of the dynamical Markovian model. (a) The data set of 2000 points given by numerical integration of the stochastic dynamical system. The circles indicate the boundaries of the three potential wells. (b) The trajectory of the corresponding Markovian particle. (c) The final trajectory of the TSP applied on the preceding data set.

presented in Fig. 9c. After applying our classification procedure with kernels located at the center of the different potential wells, we obtain the final clusters shown in Fig. 10. We note from Fig. 10 that:

(i) we are able to distinguish clearly the different wells of the potential;

(ii) the irregular boundaries of the attractive sinks are due to the noise introduced in the dynamical system; this irregularity affects the smooth boundaries of the attractive wells of the unperturbed dynamical system shown as dotted circles in Fig. 10;

(iii) the clusters are connected by branches which are the preferred transition paths of the stochastic dynamical system.

Now, we take into account specific effects that we want to address with our method, namely the possibility of more complicated boundaries. We assume here that the canonical variables in which the Markovian system (13) is formulated are not directly observable but rather that we observe quantities which are nonlinear functions of these canonical variables. This is equivalent to a deformation of Fig. 9a. More precisely, we have applied successively two one-to-one maps. The first one is defined by:

$$(X, Y) = (x, (1 + \beta x)\, y), \tag{18}$$

where (x, y) and (X, Y) are respectively the old and new Cartesian coordinate system. The second transformation is defined in the polar coordinate system by:

$$(\mathscr{P}, \Theta) = (\rho, \theta + \gamma \sin \theta) \tag{19}$$

where $(\mathscr{P}, \Theta)$ and (ρ, θ) are respectively the old and new polar coordinate system. The constants β and γ are chosen to be equal to 0.8 and 1 respectively.

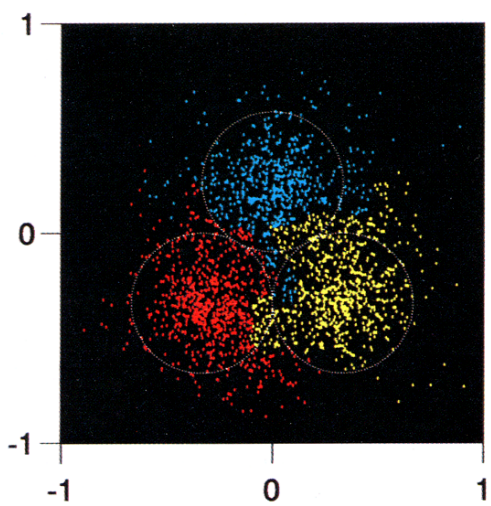

Fig. 10. Detection of the three potential wells by the present method.

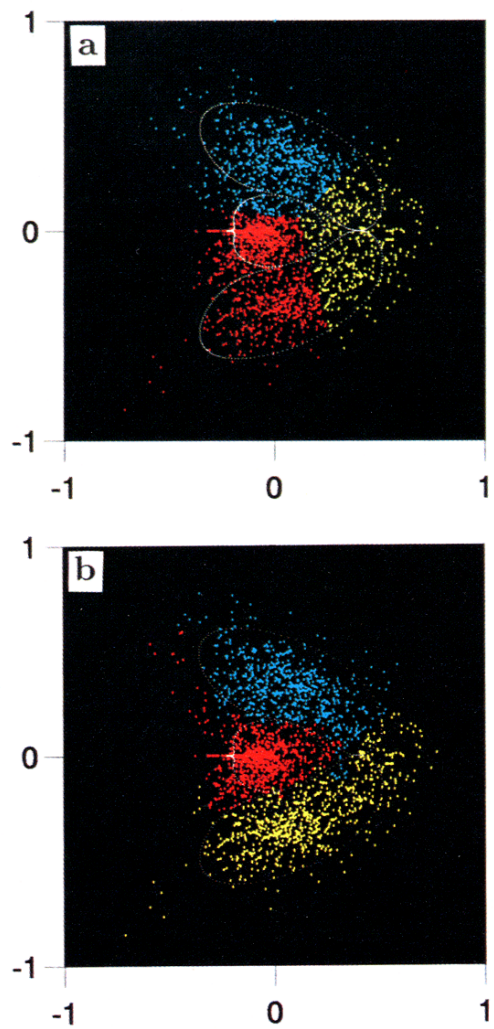

Fig. 11. Comparison between the k-mean method (a) and the present method (b) for detecting the transformed potential wells.

After transforming the data set shown in Fig. 9a using successively (18) and (19), we have applied the k-mean and our method to the new data set using circular kernels with centers located respectively at the points (0, 0), (0.165, −0.32) and (0.075, 0.36). The resulting classification is shown in Fig. 11a where it is easily seen that it does not recover the transformed potential wells. In fact, the k-mean algorithm has ejected the centers far from their initial locations and the classes with their boundaries have changed drastically. Now the result of our method is shown in Fig. 11b where one sees that the boundaries of the transformed potential wells are, this time, well approximated. We believe that this test demonstrates the utility of our method to determine cluster boundaries once a good guess of cluster centers has been obtained.

6. Application to weather regimes

The data set used here is that presented in Section 2 from which equilibrated weather regimes have been obtained. We apply the TSP algorithm to the projected atmospheric sample onto the phase space defined by the first few leading EOFs. Fig. 12 shows the decrease of "annealing energy" as a function of "annealing temperature" when

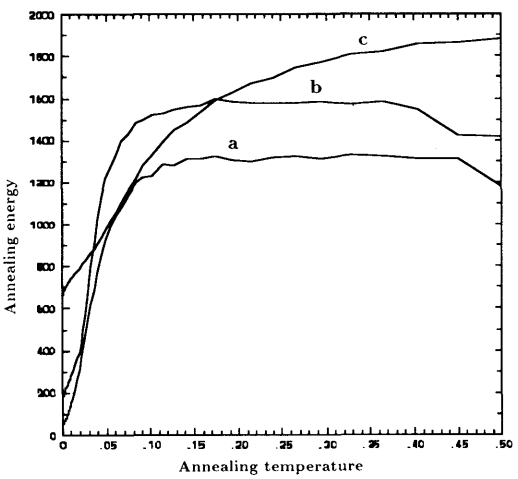

Fig. 12. The decrease of annealing energy (length of the trajectory) with respect to the annealing temperature (control parameter) during the annealing process respectively in the two first EOFs plane (curve a), three first EOFs space (curve b) and in the nine first EOFs space (curve c).

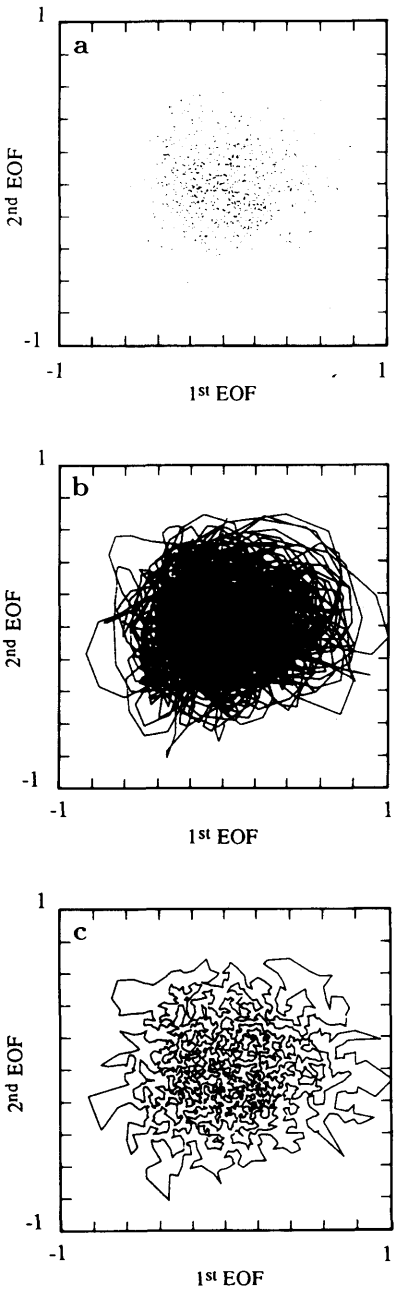

Fig. 13. Representation of the winter geopotential data set in the two first EOFs plane. (a) Projection of the principal components of the geopotential field. (b) Temporal trajectory of the principal components (taken as initial trajectory for the TSP). (c) The final TSP trajectory (in the two first EOFs space.)

using the first two, three and nine EOFs. It can be seen that the rate of change of energy (when approaching the final state) is much less for nine EOFs than for two and three EOFs. Thus the TSP algorithm becomes less efficient when the space dimension increases (Lawler et al., 1985).

The initial state for the TSP algorithm is taken to be the temporal trajectory. The data set projected onto the first two EOFs space is shown in Fig. 13a, for which the time trajectory is presented in Fig. 13b. The minimized trajectory relative to the two first EOFs space is shown in Fig. 13c. It is obtained after decreasing the temperature until the value of the energy (i.e., length of the trajectory) becomes constant. The first three retained EOFs explain more than 55% of the low frequency variance but do not capture, however, the complete structure of the atmospheric attractor. The use of the first three EOFs phase space to pursue our analysis is a compromise between the effectiveness of the algorithm and the embedding dimension.

6.1. *Weather regime boundaries*

Here, we classify the atmospheric data set into four clusters assuming we know the kernels of the four regimes BL, ZO, GA and AR defined in Section 2.

In the nine-dimensional phase space associated to the first nine EOF coefficients, the four regimes (or solutions) form a tetrahedron which spans a three-dimensional subspace that we will refer to as the "weather regimes frame". Each observed state of the atmosphere can be projected onto this three-dimensional space. The origin is fixed at the AR regime and the trihedral formed with the three other regimes is orthonormalized by the Gram-Schmidt procedure. The transformed data are then renormalized between -1 and $+1$.

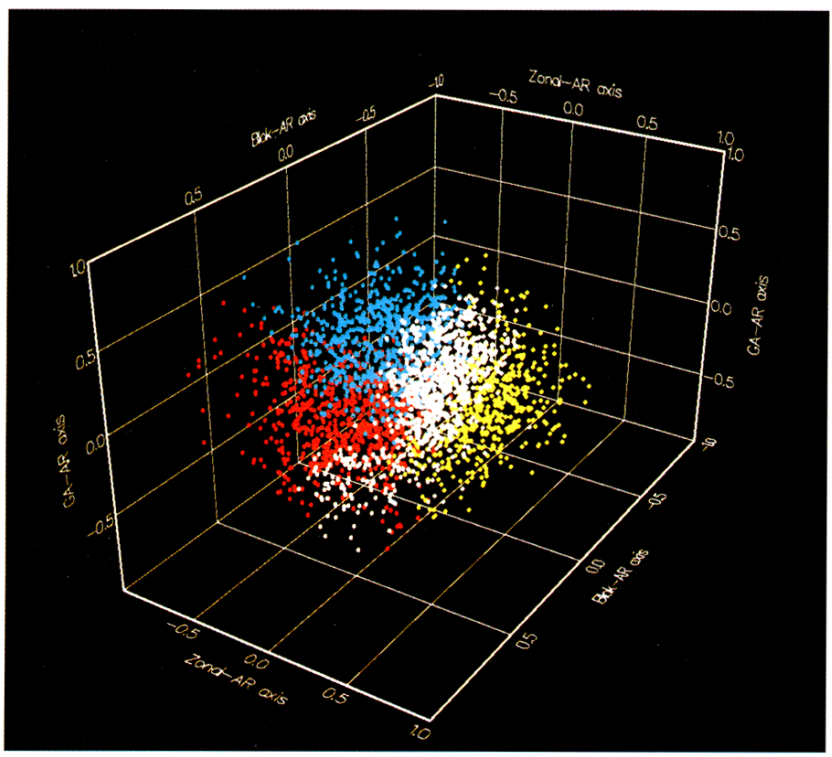

Fig. 14. Projection of the 4 weather regime clusters on the three dimensional space (trihedral) defined by the four weather regimes: the red for BL-cluster, the yellow for the ZL-cluster, the blue for the GA-cluster, and the white for the AR-cluster. Data are renormalized between -1 and $+1$. The two horizontal axes are respectively ZL–AR axis and BL–AR axis. The vertical one is GA–AR axis.

After solving the TSP in the first three EOFs phase space equipped with an Euclidian metric, we aggregate the atmospheric states to the kernels following the algorithm described in Subsection 5.1. Fig. 14 shows the 4 clusters projected onto the weather regimes frame, identified with different colors, which exhibit irregular boundaries.

Although this is a preliminary study of weather regime boundaries, it provides, however, a new viewpoint and is a step towards a study of realistic, possibly "fractal", weather regime boundaries. One can see for example that in this frame the regime clusters are distributed irregularly with respect to the data set. This fact will introduce preferred transitions between regimes and inhibit them (i.e., reduced transition probability) between others as it will be seen in Subsection 6.2.

Fig. 15 shows the projection of the AR-cluster, taken as an example, onto the three different faces of the weather regimes trihedral. None of these three projected groups exhibit a simple shape with circular or polygonal boundaries.

Notice that the sampling interval was chosen in agreement with the decorrelation time of the filtered signal and the sampling theorem (Koopmans, 1974). The main difficulty is to avoid oversampling that would trigger a solution of TSP that is just the historical order (at least locally). Increasing the sampling interval, while keeping the same number of points should not, however, modify the boundaries. This can be seen as a consequence of the embedding theorem of Takens.

6.2. *Weather regime transitions*

Boundaries between weather regimes are important because they determine the transitions between these regimes. The classification and the history of the system can be used to study the transitions between winter weather regimes under the basic assumption of a Markovian process. Within this framework, a homogeneous discrete Markov chain with states denoted by $(e_1, e_2, ...)$ is a sequence of random variables $(X_i)_{i=1,2,...}$ with the following conditional probability properties.

(i) Markov property

$$\Pr(X_{n+m} = e_k / X_1 = e_1, ..., X_m = e_m) = \Pr(X_{n+m} = e_k / X_m = e_m). \quad (20)$$

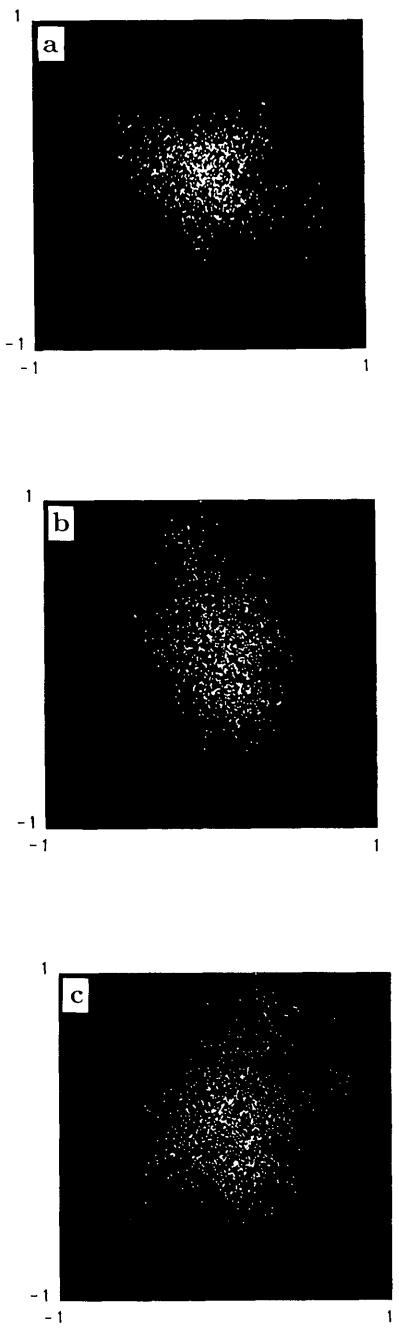

Fig. 15. Projection of the AR-cluster on the three faces defining the weather regimes frame: (a) (AR–BL, AR–ZL) face. (b) (AR–ZL, AR–GA) face. (c) (AR–GA, AR–BL).

(ii) Stationarity

$$\Pr(X_{n+m} = e_j / X_m = e_i) \text{ is independent of } m. \quad (21)$$

The transition probability is then the matrix $P = (p_{ij})$ defined as:

$$p_{ij} = \Pr(X_{n+1} = e_j / X_n = e_i). \quad (22)$$

The actual transitions between weather regimes are calculated by counting only transitions between distinct weather regimes, i.e., the transition probability p_{ij} from the ith to the jth weather regime is given by:

$$p_{ij} = \frac{v_{ij}}{\mu_i}, \quad (23)$$

where v_{ij} is the number of cases for which the temporal trajectory goes from the ith to the jth regime cluster at successive times:

$$v_{ij} = \sum_{m=1}^{N-1} (1 - \delta_{ij})\, 1_{(X_m = e_i,\, X_{m+1} = e_j)}, \quad (24)$$

and

$$\mu_i = \sum_j v_{ij}, \quad (25)$$

where δ is the Kronecker symbol and $1_{(X_m = e_i, X_{m+1} = e_j)}$ equals 1 if the atmospheric state (observation) at time m belongs to the ith regime cluster and to the jth regime cluster at time $m+1$, and vanishes otherwise.

Fig. 16 gives the main transition paths when the minimization process for the TSP is carried out using the first three EOF analyses. Note that the two main transitions are leading to the AR regime while the two main predecessors of the BL regime are the GA and AR regimes. Then, from this analysis one can say that the transition from the ZO to the AR regime is a preferred one, in agreement with the results of Vautard (1990). The 1st and the 3rd principal transitions observed in Fig. 16 are the same as the preferred transitions in Vautard (1990) while the second and the fifth transitions, observed here as main transitions are found in Vautard (1990) as preferred transitions to successors or from predecessors. The main difference with Vautard (1990) is that in our case the transition path from the BL to the AR regime emerges while it does not exist in his study.

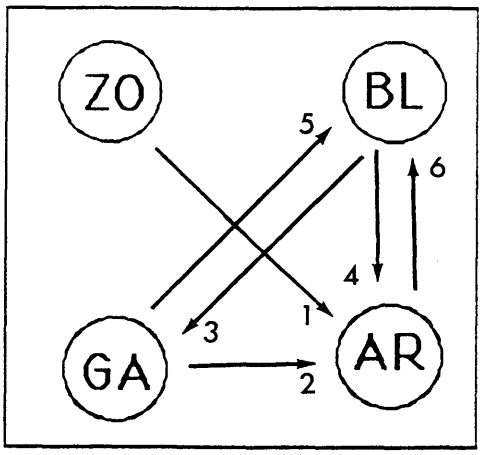

Fig. 16. Diagram of the main significant transitions between weather regimes indicated following their order (see text) using the first three EOFs for the TSP minimization.

Note that from our transition diagram, there is no direct transition from the zonal regime to the blocking regime. The flow has to pass through the Atlantic ridge regime. This has been also observed by Mo and Ghil (1988) who found that, in spite of the dominance of clusters 1 and 2 in their analysis, identified respectively as zonal and blocking, there are no direct transitions between these regimes. Instead to go from cluster 1 to cluster 2, the flow has to pass through other clusters. This observation is also in agreement with the results drawn out by Molteni et al. (1990) in which the number of transitions between their cluster 3, which can be compared to our ZL regime, and their cluster 4, which can be regarded as the equivalent of our BL regime, is small and not significant. They also obtained very few transitions between cluster 4 and cluster 2 (equivalent to our GA regime).

Again our transition diagram shows that there is no successions of the order of transition between any two regimes. This shows a sort of asymmetry in the transitions between weather regimes. This conclusion has been in fact pointed out by Molteni et al. (1990) who found an asymmetric behaviour in the transitions between regimes with low and high amplitude of planetary waves.

The diagram of Fig. 16 shows good agreement with the locations of the clusters (Fig. 14). In fact,

Tellus A/47/5:2-5

if we consider only the main transitions, one can see that the strongest transition observed in Fig. 16 is from ZO to AR. This result is due to the fact that the ZO cluster has as dominant neighbour, that is the AR regime, with which it shares a very large boundary, while there is not an extended contact with the other clusters. The 2nd important boundary in order is that observed between the GA and the AR clusters (Fig. 14) which explains the 2nd transition (Fig. 16). The same reasoning applies to other transitions. It is worth mentioning that the transitions are not symmetric.

In addition to the above transitions which are well known synoptically from daily weather maps, there are others like the 4th and the 5th (Fig. 16), which are not documented elsewhere. They may be due to projection effects and considered as spurious. It is shown also within the Markovian framework using the Markovian stationary probability (Hannachi, 1991) that the AR is the most persistent regime in agreement with the observations (Egger, 1981).

7. Conclusion and discussion

In this study, we have established a classification of atmospheric states of the mid-latitude geopotential based on the four winter regimes identified by Vautard (1990), the European blocking (BL), the zonal regime (ZO), the Greenland anticyclone (GA) and the Atlantic ridge (AR). Our motivation was to look for a classification procedure which, unlike previous studies, produces more realistic boundaries that depend on the distribution of the data set rather than on the classification method itself. In this purpose, we have considered an aggregation of neighbouring states following some preferred paths in phase space obtained by solving the travelling salesman problem (TSP) for our dataset of 700 hPa NMC geopotential height projected onto a three-dimensional space spanned by the weather regimes. We have tested our method by comparing it with the k-mean method using artificial data, and it is shown that it gives better results by comparison with the true results.

We have applied this method to the atmospheric data set and the obtained classification was used to study weather regimes transitions. We have obtained similar results to Vautard (1990) in the sense that the primary transitions in our case are from ZL to AR and from BL to GA which are found to be preferred in his study. The two other main transitions from GA to AR and from GA to BL are also in agreement with his study where in fact he found that AR and BL are preferred successors for the GA. The main difference, however, is that the main transition from BL to AR in our case is not a preferred one in his study.

The non existence of a direct transition between ZL and BL, where the flow has to pass through AR to get from ZL to BL, has been found also by Mo and Ghil (1988). We have also obtained an agreement with Molteni et al. (1990), where we note that there is an asymmetric behaviour in the transitions between weather regimes. Synoptically non obvious transitions, like those from BL to AR and from GA to BL which are probably spurious in our analysis may be due to the projection effect related to the embedding dimension of the atmospheric attractor.

Notice that our definition of weather regimes is not intrinsic. It depends on filtering synoptic structures, that is putting a scale separation in the time-space domain. Since this cannot be performed on purely rigourous basis for atmospheric flows, a sensitivity of the classification upon the number of the retained EOFs may exist. It is generally observed, however, that a hierarchical structure is obtained as the number of retained modes grows. Here the number of clusters is fixed by the study of the equilibrated regimes. Our agglomeration algorithm is only effective when the density of data points is large enough. We found that this condition is satisfied by retaining more than three components but that the number of local minima of the minimized functional increases very fast as the number of modes increases, thus leading to very slow decrease of the Gibbs energy and an ambiguous final classification. Using three components is a compromise between the effectiveness of the algorithm and the embedding dimension.

This study provides a simple approach to the fact that the atmospheric weather regimes are likely to possess a complicated boundary in phase space, that may even be a fractal. It would be very useful to further characterize the regime boundaries in physical terms (requiring to extend the analysis to three-dimensional dataset) and to investigate the potential for seasonal forecasting of large-scale atmospheric flows.

8. Acknowledgments

The authors are grateful to Professor M. Ghil and Dr. R. Vautard for their valuable discussions. We are also grateful to Dr. K. Haines for helpful discussions and improving the original manuscript. Thanks also to an anonymous reviewer for helpful comments. This study is based in part on A.H's Ph.D dissertation and was supported partly by the Centre National de la Recherche Scientifique (CNRS) and partly by UK NERC Grant GR3/8274.

9. Appendix A

The travelling salesman problem (TSP)

The travelling salesman problem as formulated in linear programming can be described as follows.

Let $n+1$ cities (points in the plane) be enumerated 0, 1, ..., n. The traveller starts from city 0, must pass through all cities only once, and return to city 0. The distribution of cities is known by the table of their distances c_{ij}. The goal is to seek the route of shortest distance.

If we want to solve this problem using only permutations, we need to choose the best one among a choice of $n!$ possibilities. As soon as n exceeds a few tens, this approach becomes untractable in practice and it is well known also, that the TSP belongs to the family of $N-P$ hard problems.

The integer programming version of the problem can be formulated by introducing the integer variables:

$$x_{ij} = \begin{cases} 1 & \text{if the traveller leaves from city } i \text{ to } j \\ 0 & \text{otherwise.} \end{cases} \qquad \text{(A1)}$$

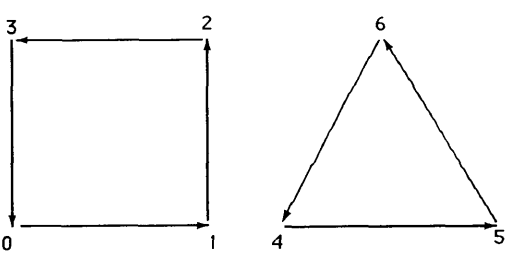

Fig. 17. Example of figure not connected but satisfying (A2.2) and (A2.3).

To get the whole set of equations, we introduce also the "connecting" variables u_i $(i=1, 2, ..., n)$ (Achmanov, 1981). Then, the following system is obtained that describes entirely our problem:

$$\begin{cases} L = \min \sum_{i=0}^{n} \sum_{j=0}^{n} c_{ij} x_{ij} & \text{(A2.1)} \\ \sum_{i=0}^{n} x_{ij} = 1; \quad j = 1, 2, ..., n & \text{(A2.2)} \\ \sum_{j=0}^{n} x_{ij} = 1; \quad i = 1, 2, ..., n & \text{(A2.3)} \\ u_i - u_j + n x_{ij} \leqslant n - 1; \quad i, j = 1, 2, ..., n & \text{(A2.4)} \\ x_{ij} \in \{0, 1\}; \quad i, j = 0, 1, ..., n & \text{(A2.5)} \end{cases}$$

Eqs. (A2.2) and (A2.3) state that the traveller reaches and leaves the jth city only once. These two conditions are not sufficient to take into account the connection of the route. For example, the case shown in Fig. 17 satisfies (A2.2) and (A2.3) but is not connected. The conditions (A2.4) ensure the connection of the route (Achmanov, 1981).

The simplex method, which is very popular for minimization problems cannot be used easily here, because the TSP as, stated above, is an integer linear programming problem. Other methods like the cut-method (Achmonov, 1981) are only applicable only for a small number of cities. It is worth mentionning that when the TSP is solved by the simulated annealing algorithm following the steps presented in Subsection 4.2, there is no need to the connecting variables because we start our algorithm by explicitely choosing a closed path (connected by choice) and then apply different moves, according to Fig. 5, which are also connecting moves. The final path will be then connected. The simulated annealing used here is presented below.

10. Appendix B

The simulated annealing algorithm

This method which was first introduced by Metropolis et al. (1953) and refined by other authors (see, e.g., Lin and Kernighan, 1973) is based on a physical model of crystallization (William et al., 1986). The method described here uses a combination of the original physical model and the Kernighan algorithm. The main idea is

that when heating a crystal and then cooling it very slowly, thermal mobility of the atoms is lost gradually and the crystal gets to the state of its minimum absolute energy while a fast cooling would lead to a local minimum with higher energy (and the presence of many defaults in the crystal structure). This is the annealing.

A physical system in thermal equilibrium at temperature T has its energy E distributed following the Boltzmann probability law given by (5). Even at low temperature, there is always a "chance", albeit very small, of the system being in a high energy state. The system is assumed to change its configuration from energy E_1 to energy E_2 with probability:

$$p = \min[1, e^{-(E_2 - E_1)/kT}]. \tag{B1}$$

If $E_2 < E_1$, p is equal to one which means that if the transition to a lower energy state is possible, it is certain.

We now show how the simulated annealing can be used to minimize nonlinear functions in multidimensional spaces. The value of the function at a given point x, is here assumed to be the energy of the system in the state x.

We then define a control parameter T which is treated as the temperature of the system. The associated probability measure at temperature T is given by the Gibbs law:

$$G_T(dx) = \frac{e^{-E(x)/kT}}{\int_\Omega e^{-(-E(y))/kT}\, dy}\, dx, \tag{B2}$$

where the summation is performed over the set Ω of possible states of the system. If we define the solution of the minimization problem as

$$\Omega_{\min} = \{x \in \Omega, /\forall y \in \Omega; E(x) \leqslant E(y)\}, \tag{B3}$$

we have the important result

$$\lim_{T \to 0} G_T(\Omega_{\min}) = 1. \tag{B4}$$

Eq. (B4) means that when the temperature is decreasing, the system concentrates around $\Omega_{\min}$. Strictly speaking (see below), this result holds when the limit is taken for infinite time.

To solve the problem of minimizing E, Metropolis et al. (1953) proposed the following construction.

(i) Choose a symmetric transient Markov chain (i.e., with symmetric probability transition).

(ii) This chain is used to construct a process X_m over Ω by giving X_0 and the induction rule:

- For every state X_m, associate a random state Y_{m+1} obtained by the transition matrix P of the Markov chain.
- If the variation of energy $\Delta E_m = E(Y_{m+1}) - E(X_m)$ is negative, one accepts this displacement (i.e., $X_{m+1} = Y_{m+1}$).
- If ΔE_m is positive, one accepts this displacement with a probability equals to $e^{-\Delta E_m/kT}$.

Such a process satisfies:

$$\forall x \in \Omega, \quad \lim_{m \to \infty} P_T(X_m = x) = G_T(x); \tag{B5}$$

using (B4) and (B5) we get:

$$\lim_{T \to 0} \lim_{m \to \infty} P_T(X_m \in \Omega_{\min}) = 1. \tag{B6}$$

More precisely, we have the following theorem (Geman and Geman, 1984).

Theorem. If for every $k > N$, we have:

$$T_k > \frac{C}{\log(k)}; \tag{B7}$$

then:

$$\lim_{m \to \infty} P(X_m \in \Omega_{\min}) = 1, \tag{B8}$$

where $C = n\,\Delta$, n is the dimension of the problem, and Δ is the variation of energy over Ω. This important theorem provides a guidance for decreasing the temperature. In practice, however, it is not possible to use a logarithmic law at low temperature and we are bound to use a decreasing law of the type $T_k = T_0/k$.

REFERENCES

Aart, E. and Korst, J. 1990. *Simulated annealing and Boltzmann machine.* John Wiley and Sons, New York.

Achmanov, S. 1981. *Programmation linéaire.* Edition Mir, Moscou.

Charney, J. G. and DeVore, J. G. 1979. Multiple flow equilibria in the atmosphere and blocking. *J. Atmos. Sci.* **36**, 1205–1216.

Diday, E. and Simon, J. C. 1976. *Clustering analysis in communication and cybernetics digital pattern recognition*, ed.: K. S. Fu. Springer-Verlag, Berlin, 47–94.

Dole, R. M. and Gordon, D. N. 1983. Persistent anomalies of the extratropical northern hemisphere winter circulation. Geophysical distribution and regional persistent characteristics. *Mon. Wea. Rev.* **111**, 1567–1586.

Egger, J. 1981. Stochastically driven large scale circulation with multiple equilibria. *J. Atmos. Sci.* **38**, 2606–2618.

Geman, S. and Geman, D. 1984. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.* **6**, 721–741.

Gordon, A. D. 1981. *Classification: methods for the exploratory analysis of multivariate data.* Chapman and Hall, London.

Hannachi, A. 1991. *Régimes de temps: dynamique et statistique.* PhD Thesis, Université Pierre et Marie Curie, Paris.

Kimoto, M. 1989. *Multiple flow regimes in the northern hemisphere winter.* PhD Thesis, University of California, Los Angeles, USA.

Kirkpatrick, S., Gelatt, C. D. and Vecchi, M. P. 1983. Optimization by simulated annealing. *Science* **220**, 671–680.

Koopmans, L. H. 1974. *The spectral analysis of time series.* Academic Press, New York.

Lawler, E. L., Lenstra, J. K., Rinnooy, A. H. and Shmoys, D. B. 1985. *The travelling salesman problem.* John Wiley and Sons, New York.

Legras, B., Despont, T. and Piguet, B. 1987. Cluster analysis and weather regimes. *Proceedings of the ECMWF Workshop on the Nature and prediction of extra-tropical weather systems.* ECMWF, Shinfield Park, Reading, 123–149.

Lin, S. and Kernighan, B. W. 1973. An effective heuristic algorithm for the travelling salesman problem. *Operat. Res.* **21**, 498–516.

Lorenz, E. N. 1963. Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130–141.

Mo, K. C. and Ghil, M. 1988. Cluster analysis of multiple planetary flow regimes. *J. Geophys. Res.* **93**, 927–952.

Molteni, F., Sutera, A. and Tronci, N. 1988. EOFs of the geopotential eddies at 500 mb in winter and their probability density distributions. *J. Atmos. Sci.* **45**, 3063–3080.

Molteni, F., Tibaldi, S. and Palmer, T. N. 1990. Regimes in wintertime circulation over northern extratropics. (I) Observational evidence. *Quart. J. R. Met. Soc.* **116**, 31–67.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. 1953. Equation of state calculation by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092.

Shepherd, T. G. 1990. Symmetries laws and Hamiltonian structures in geophysical fluid dynamics. *Adv. in Geophys.* **32**, 287–338.

Sutera, A. 1980. On stochastic perturbations and large term climate behaviour. *Quart. J. R. Met. Soc.* **107**, 137–151.

Vautard, R. 1990. Multiple weather regimes over the north Atlantic: Analysis of precursors and successors. *Mon. Wea. Rev.* **118**, 2056–2081.

Vautard, R. and Legras, B. 1988. On the source of mid-latitude low-frequency variability. Part II: Non linear equilibration of weather regimes. *J. Atmos. Sci.* **45**, 2845–2867.

William, H. P., Brian, F. P., Saul, A. T. and William, T. V. 1986. *Numerical recipes.* Cambridge University Press.

Zaslavsky, G. M., Sagdeev, R. Z., Usikov, D. A. and Chernikov, A. A. 1992. *Weak chaos and quasi-regular patterns.* Cambridge University Press.