

REVUE DE STATISTIQUE APPLIQUÉE

MICHEL DEQUE

Étude de la persistance d'un champ météorologique

Revue de statistique appliquée, tome 31, n° 3 (1983), p. 39-56

http://www.numdam.org/item?id=RSA_1983__31_3_39_0

© Société française de statistique, 1983, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

ETUDE DE LA PERSISTANCE D'UN CHAMP METEOROLOGIQUE

Michel DEQUE

Météorologie Nationale, 77 rue de Sèvres, 92100 Boulogne, France

RESUME

Lorsqu'on est en présence d'une série de valeurs numériques non indépendantes on peut mesurer la persistance et rechercher les cycles privilégiés en étudiant le diagramme d'autocorrélation. En météorologie et plus particulièrement en circulation générale atmosphérique on a affaire à des champs. Pour une échéance donnée on peut décomposer un champ en différentes composantes décorrélées dont la persistance décroît de la première à la dernière. Leur mode de construction est analogue à celui des Composantes Principales mais au lieu de maximiser la variance, on cherche à maximiser l'autocorrélation temporelle. Diverses applications de ce type de décomposition sont menées sur une série de 10 ans de champs de géopotentiel à 500 mb simulés dans une expérience de circulation générale effectuée au Centre Européen de Prévisions Météorologiques à Moyen Terme.

I. INTRODUCTION

La persistance joue un rôle important en météorologie car elle sert de référence pour une prévision. Cependant, toutes les variables n'ont pas la même persistance. Le géopotentiel à 200 mb évoluera beaucoup plus lentement que la nébulosité. Lorsqu'on a affaire à une variable unidimensionnelle (précipitations en une station) on peut étudier la persistance par le diagramme d'autocorrélation temporelle. Plus les coefficients d'autocorrélation d'ordre k ρ_k sont proches de 1 plus la série est persistante. En outre l'apparition de ρ_k proche de 1 alors que les valeurs de ρ_i pour $i < k$ étaient inférieures met en évidence un cycle de période k . Lorsqu'on a affaire à une variable multidimensionnelle et si la dimension est grande, il est hors de question d'étudier les composantes une à une, surtout si elles ne sont pas indépendantes. Une expérience de circulation générale atmosphérique a été réalisée au Centre Européen de Prévisions Météorologiques à Moyen Terme (CEPMMT) (Voir André 81 — Rilat et Thépaut 82) pour modéliser le comportement de l'atmosphère pendant 10 ans.

Même en se restreignant au champ de géopotentiel à 500 mb qui résume le mieux la circulation générale de l'atmosphère, il y a 2048 composantes (c.a.d 2048 points sur une grille de 32 latitudes et 64 longitudes) fortement corrélées entre elles. On a recours classiquement à une décomposition en Composantes Principales de ce champ (Der Mégréditchian et Mante 1981, Lau 1981).

L'analyse en Composantes principales de cette expérience a été présentée dans Volmer *et al.* 1982. Il se trouve que les premières Composantes Principales sont plus persistantes que les suivantes mais il s'agit là d'une propriété météorolo-

gique et non statistique (il suffit de traiter les champs dans le désordre chronologique : les Directions Principales seront les mêmes mais les Composantes Principales ne seront pas autocorrélées).

On va donc essayer de résoudre le problème de la décomposition d'un champ en composantes décorréliées dont les premières seront plus persistantes que les suivantes. Le but poursuivi est de faire le tri entre la part d'information qui se conserve pendant une certaine durée et celle qui se dégrade rapidement au cours du temps.

Pour mesurer la persistance d'un champ de manière globale, il suffira de considérer la persistance des premières composantes et la part de variance qu'elles occupent dans la décomposition du champ initial.

II. COMPOSANTES PERSISTANTES – DEFINITION ET PROPRIETES

a) Position du problème

Fixons $k \in \mathbb{N}$ intervalle de temps correspondant à l'autocorrélation, $k = 1$ pour chercher la persistance, $k > 1$ pour chercher les durées de retour (par défaut, k vaudra 1).

Considérons une série de N champs de dimension p , Z_n .

On cherche à obtenir des séries x_n^i $n = 1, N$; $i = 1, p$ par combinaison linéaire de Z_n telles que x^1 maximise l'autocorrélation à k jours (ou k unités quelconques de temps) x^2 maximise l'autocorrélation à k jours sous la contrainte d'être non corrélée à x^1 , x^3 maximisant sous la contrainte d'être non corrélée à x^1 et à x^2 etc. Cette démarche est semblable à celle qui mène à la définition des Composantes Principales : on cherche des séries z_n^i $n = 1, N$; $i = 1, p$ par combinaison linéaire de Z_n telles que z^1 maximise la variance, z^2 maximise la variance sous la contrainte d'être non corrélée à z^1 , etc.

Les conditions sur x^i sont indépendantes de la moyenne et de la variance ; pour que le problème ait une solution unique nous poserons des conditions analogues à celles requises pour les Composantes Principales. Celles-ci étant centrées, nous imposerons aux x^i de l'être. Le champ $Z = \{Z_n/n = 1, N\}$ se recompose en $Z_n = \sum_{i=1}^p z_n^i C^i$ où C^i est un vecteur normé de dimension p .

Nous imposerons d'avoir la recombinaison de Z en $Z_n = \sum_{i=1}^p x_n^i \Gamma^i$ où Γ^i est un vecteur normé de dimension p .

b) Résolution du problème

Soit $\bar{Z}$ la moyenne temporelle de Z . Posons

$$V = \frac{1}{N} \sum_{n=1}^N (Z_n - \bar{Z})(Z_n - \bar{Z})'$$

On pose

$$W_k = \frac{1}{N-k} \sum_{n=1}^{n-k} (Z_n - \bar{Z})(Z_{n+k} - \bar{Z})'$$

$\tilde{W}_k = \frac{1}{2} (W_k + W'_k)$ est une matrice symétrique.

Soit B un vecteur de dimension p . Si $x_n = B'(Z_n - \bar{Z})$

L'autocorrélation des x_n est $\frac{B' \tilde{W}_k B}{B' V B} = \frac{B W_k B}{B' V B}$

L'écriture algébrique du problème posé plus haut est :

On cherche B_1 maximisant $\frac{B'_1 \tilde{W}_k B_1}{B'_1 V B_1}$ puis B_2 maximisant $\frac{B'_2 \tilde{W}_k B_2}{B'_2 V B_2}$ sous la contrainte $E(x^1 x^2) = 0$ soit $B'_1 V B_2 = 0$.

La solution de ce problème classique est obtenue à partir des vecteurs propres de $V^{-1} \tilde{W}_k$ et $\tilde{W}_k V^{-1}$. Soient $r_1, r_2, \dots, r_p$ les valeurs propres de $\tilde{W}_k V^{-1}$ ($r_1 \geq r_2 \geq \dots \geq r_p$) respectivement associées aux vecteurs propres $\theta_1, \theta_2, \dots, \theta_p$ orthonormés (pour la métrique V^{-1} :

$$\theta'_i V^{-1} \theta_j = 1 \Rightarrow i = j / \theta'_i V^{-1} \theta_j = 0 \Rightarrow i \neq j).$$

$\alpha_n^i = \theta'_i V^{-1} (Z_n - \bar{Z})$ pour $1 \leq n \leq N$ est la solution du problème de maximisation. La variable α^i est centrée et décorrélée avec les α^j $j \neq i$. On a :

$$Z_n = \bar{Z} + \sum_{i=1}^p \alpha_n^i \theta_i = \bar{Z} + \sum_{i=1}^p (\beta^i \alpha_n^i) \left(\frac{\theta_i}{\beta^i} \right)$$

Nous choisissons les scalaires β^i afin que les vecteurs $\Gamma^i = \frac{\theta_i}{\beta^i}$ soient normés (pour la norme usuelle). Il vient $\beta^i = (\theta'_i \theta_i)^{1/2}$.

La solution du problème peut donc s'écrire sous forme définitive :

$$Z_n = \bar{Z} + \sum_{i=1}^p x_n^i \Gamma^i$$

Où x^i est la i^e Composante Persistante.

$$x_n^i = (\theta'_i \theta_i)^{1/2} \theta'_i V^{-1} (Z_n - \bar{Z})$$

Et Γ^i la i^e Direction Persistante

$$\Gamma^i = (\theta'_i \theta_i)^{-1/2} \theta_i$$

L'autocorrélation de x^i est r_i . Les x^i sont centrés et décorrélés.

La variance de x^i est :

$$(\theta'_i \theta_i) \theta'_i V^{-1} V V^{-1} \theta_i = (\theta'_i \theta_i) (\theta'_i V^{-1} \theta_i) = \theta'_i \theta_i$$

Les θ_i étant V^{-1} orthonormés la somme des variances des x_i est $\text{Tr } V$ comme pour les Composantes Principales. On définira donc la part de variance expliquée par la i^e Composante Persistante : $\theta'_i \theta_i / \text{Tr } V$.

Les m premières Composantes Persistantes portent moins de variance que les m premières Composantes Principales, ce qui est naturel puisqu'elles n'ont pas été construites dans ce dessein.

III. APPLICATION AU GEOPOTENTIEL à 500 mb

a) 10 hivers globaux

Une expérience portant sur une période de 10 ans réalisée en 1981 avec le modèle spectral T 21 (troncature à 21 nombres d'onde zonaux et méridiens soit 484 degrés de liberté) du CEPMMT fournit un important fichier de données globales dont nous n'avons retenu que le géopotentiel à 500 mb en coefficients spectraux tronqués sur une T 13 (troncature à 13 nombres d'onde zonaux et méridiens) afin d'avoir 196 degrés de liberté (sans trop déformer le champ) et de pouvoir traiter des matrices de covariance et d'autocorrélation par des moyens informatiques simples (mémoire centrale du CYBER 175, ordinateur CDC de la Météorologie Nationale). On montre que la résolution du problème est équivalente en points de grille et en coefficients spectraux. Nous avons d'abord étudié l'hiver, ou plus exactement les mois de Janvier et de Février (600 jours). Les résultats sont montrés en projection stéréographique polaire sur l'hémisphère Nord — figure 1 et figure 2. L'analyse en Composantes Principales fournit une première composante de variance expliquée 8,6 % qui correspond à un noyau de baisse (quand elle est positive) sur le Groenland. Son autocorrélation est .96 à 1 jour et .56 à 10 jours. La deuxième composante a une variance expliquée de 6.1 % et correspond à trois noyaux de baisse sur la baie de l'Hudson, la mer Caspienne et le Pacifique Nord et un fort noyau de hausse sur l'Alaska. Son autocorrélation est .94 à 1 jour et .25 à 10 jours. Elle est donc moins persistante. Nous avons fait ensuite une Analyse en Composantes Persistantes avec $k = 1$ sur les 60 premières Composantes Principales (variance expliquée 89 %) de manière à filtrer les phénomènes non significatifs mais qui peuvent être persistants. On obtient des résultats différents. La première composante a une variance expliquée de 6 %, son autocorrélation à 1 jour est de .989 et à 10 jours de .920. Elle correspond grosso modo à la première Composante Principale (noyau de baisse sur le Groenland). La seconde composante a une variance expliquée de 3.6 %, son autocorrélation est .978 à 1 jour et .710 à 10 jours. Son aspect est sensiblement différent de celui de la deuxième Composante Principale. Nous avons testé la sensibilité des deux premières Composantes Persistantes. Lorsqu'on cherche à maximiser l'autocorrélation à 2 jours au lieu de 1 jour, on obtient pratiquement les mêmes résultats ; il en va de même lorsqu'on traite les données désaisonnalisées (en retirant à chaque valeur du jour j la valeur de l'année moyenne correspondant à ce jour).

Toutefois si on ne retient que les 20 premières Composantes Principales au lieu des 60 (variance expliquée 59 %) la deuxième Direction Persistante est plus zonale. Il y a donc suffisamment de persistance dans les Composantes Principales 21 à 60 pour modifier l'aspect de la deuxième Direction Persistante.

b) 10 années globales

Lorsqu'on traite les 10 ans dans leur intégralité, il apparaît une Composante Persistante très autocorrélée et dont la variance expliquée est 60 %, c'est le cycle saisonnier (l'année moyenne se situe à 92 % sur une Direction Principale et la composante correspondante suit une sinusoïde). On enlève donc le cycle annuel par soustraction. Lorsqu'on fait une Analyse en Composantes Principales sur ces données (cf. IV), il apparaît deux composantes (Fig. 7) de 4 % de variance expli-

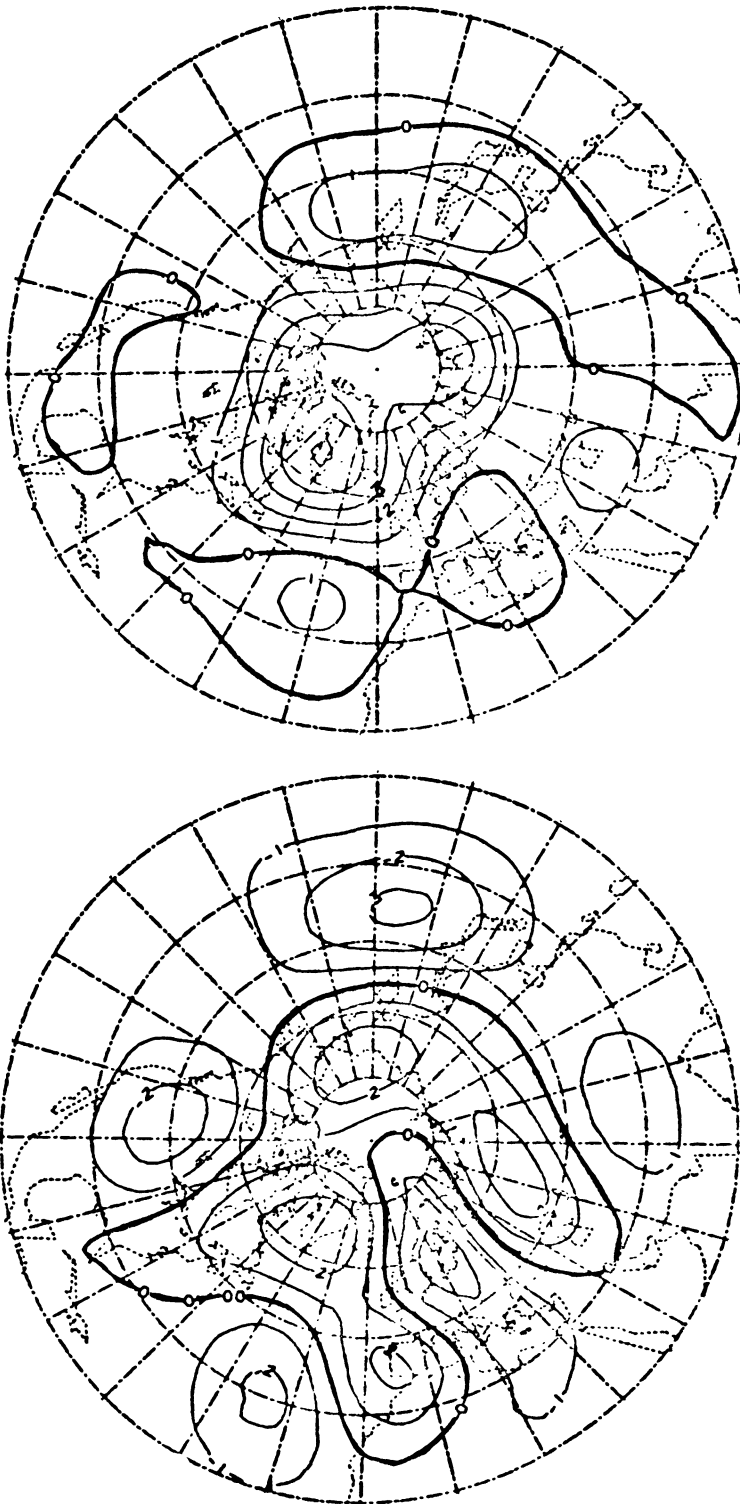


Figure 1. — 1^{re} et 2^e Directions Principales des 10 hivers.

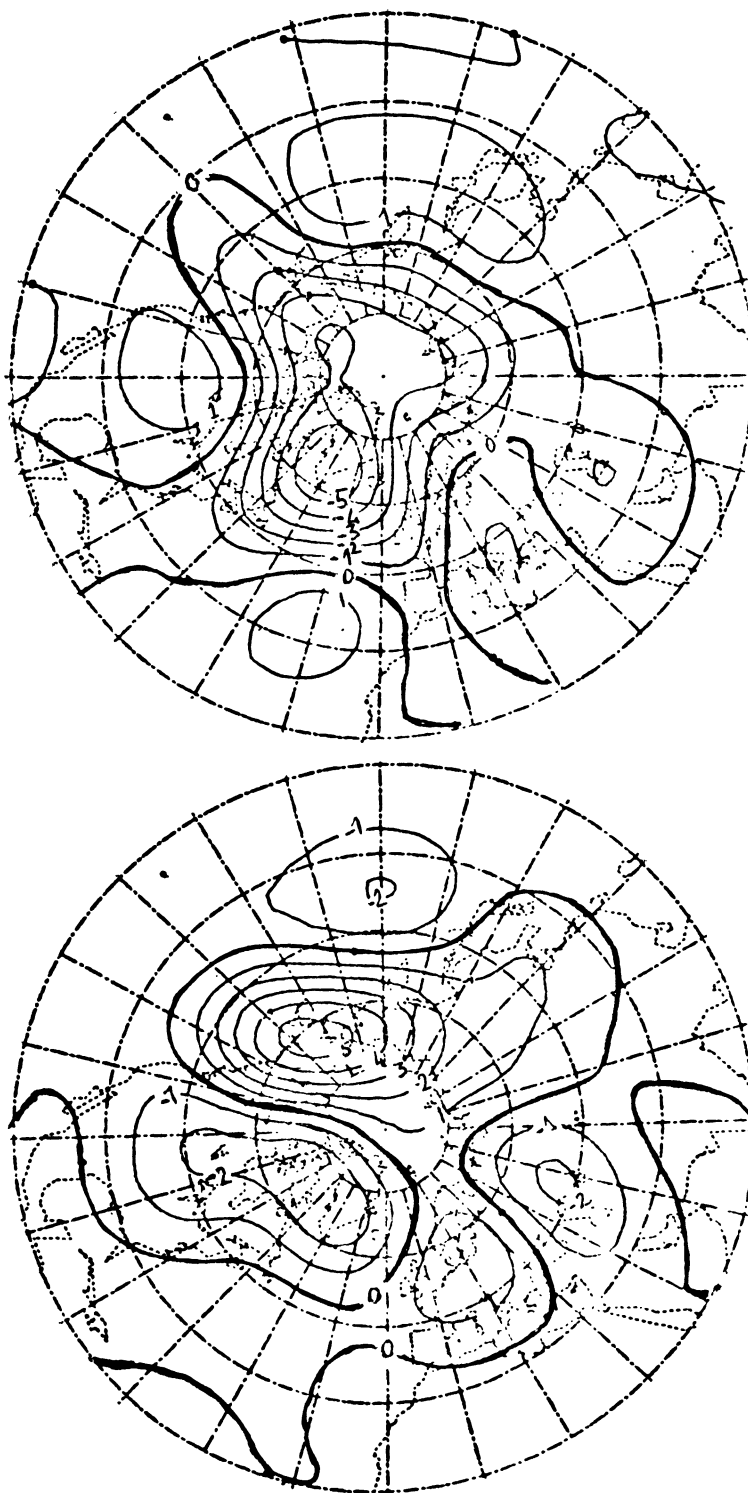


Figure 2. – 1^{re} et 2^e Directions Persistantes des 10 hivers.

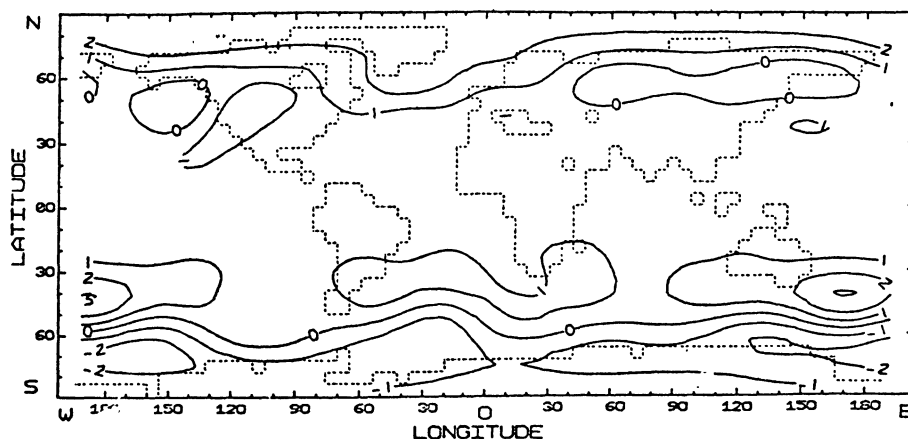


Figure 3. — 1^{re} Direction Persistante des 10 ans désaisonnalisés.

quée dont l'une se situe dans l'hémisphère Nord, l'autre dans l'hémisphère Sud. La première Composante Persistante (Fig. 3) après filtrage par les 20 premières Composantes Principales (variance expliquée 46 %) rétablit le couplage entre les deux pôles. Elle a une autocorrélation plus importante que les Composantes Principales (Fig. 4a) ; à 40 jours elle vaut .55 alors qu'elle est négligeable ($< .02$) pour les premières Composantes Principales, mais sa variance expliquée n'est que de 2.7 %. Il est remarquable d'obtenir une persistance aussi importante à 40 jours (l'autocorrélation demeure supérieure à .55 de 1 à 40 jours) alors que c'est l'autocorrélation à 1 jour que l'on maximise. La carte de la Direction Persistante montre un antagonisme entre le pôle Nord et les anticyclones de l'hémisphère Sud, d'une part et le pôle Sud d'autre part. On observe en effet sur le modèle de circulation générale de l'Etablissement d'Etudes et de Recherches Météorologiques (EERM) (voir ROYER *et al.* 1981, DEQUE *et al.* 1982) des mois de Janvier où la pression moyenne au niveau de la mer est plus faible (que sa valeur moyenne pour tous les mois de Janvier produits par le modèle) au pôle Sud et plus forte au pôle Nord et des mois de janvier où l'inverse se produit. Plus précisément chaque fois qu'un déficit de masse apparaît à un pôle, un excédent apparaît à l'autre. Il s'agit donc d'un accident suffisamment persistant pour qu'il apparaisse au niveau des moyennes mensuelles. Il est difficile de dire, dans l'état actuel des données si ce phénomène se retrouve dans l'atmosphère réelle.

c) 20 années sur un domaine réduit

Un fichier établi par la subdivision prévision à longue échéance de la Météorologie Nationale nous fournit 20 ans de données réelles sur un domaine de 89 points limité à l'Europe et l'Atlantique Nord. Une étude similaire à celle de b) a été effectuée. La première Composante Principale (Fig. 5) représente le phénomène de "Greenland Seesaw" (bascule du Groenland). Quand la composante est positive le flux d'Ouest s'infléchit et suit une trajectoire Sud-Ouest – Nord-Est sur l'Europe du Nord. Quand la composante est négative, l'orientation du flux est Nord-Ouest – Sud-Est. La variance expliquée est 16.3 % et son autocorrélation .91 à un jour, .35 à 5 jours.

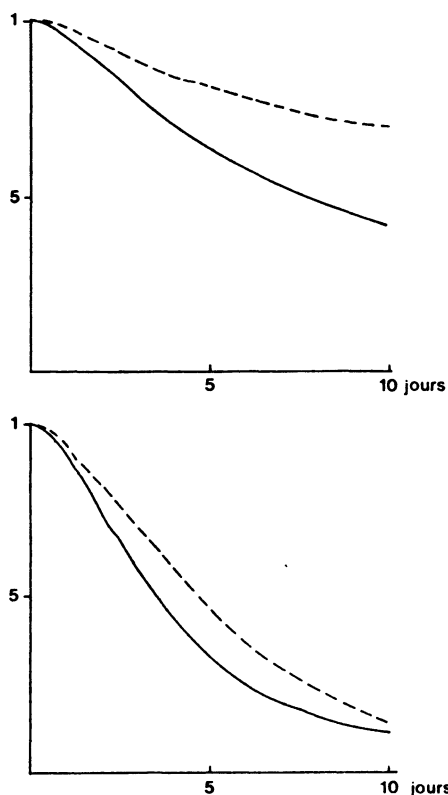
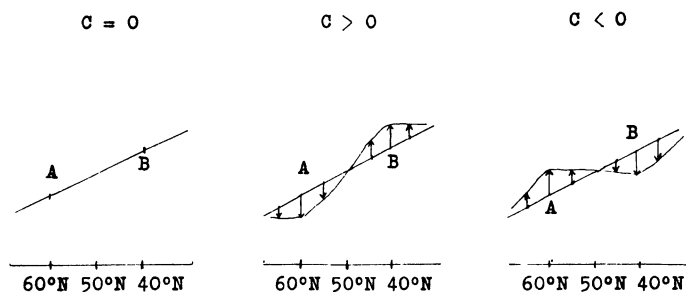


Figure 4. — Diagrammes d'autocorrélation à 10 jours de la première Composante Principale (—) et de la Composante Persistante (---) à partir des données désaisonnalisées
a) sur le domaine global,
b) sur le domaine Europe-Atlantique.

Après filtrage par les 28 premières Composantes Principales (variance expliquée 94 %), on obtient une première Composante Persistante qui ne représente pas le même phénomène. Si on schématise le profil Nord-Sud du géopotentiel en portant A et B les deux centres d'action (Fig. 5) :



On voit que lorsque la composante est positive, le gradient (et donc la circulation) est maximum vers 50° N : il y a accélération du jet. Lorsque la composante est négative le gradient est faible vers 50° N mais il est fort à 30° N et à 70° N : il y a séparation du jet en deux branches, une branche Nord et une branche Sud. C'est le phénomène du blocage. La première composante persistante correspond au blocage sur l'Atlantique Nord. La variance expliquée est de 12.8 % et son auto-

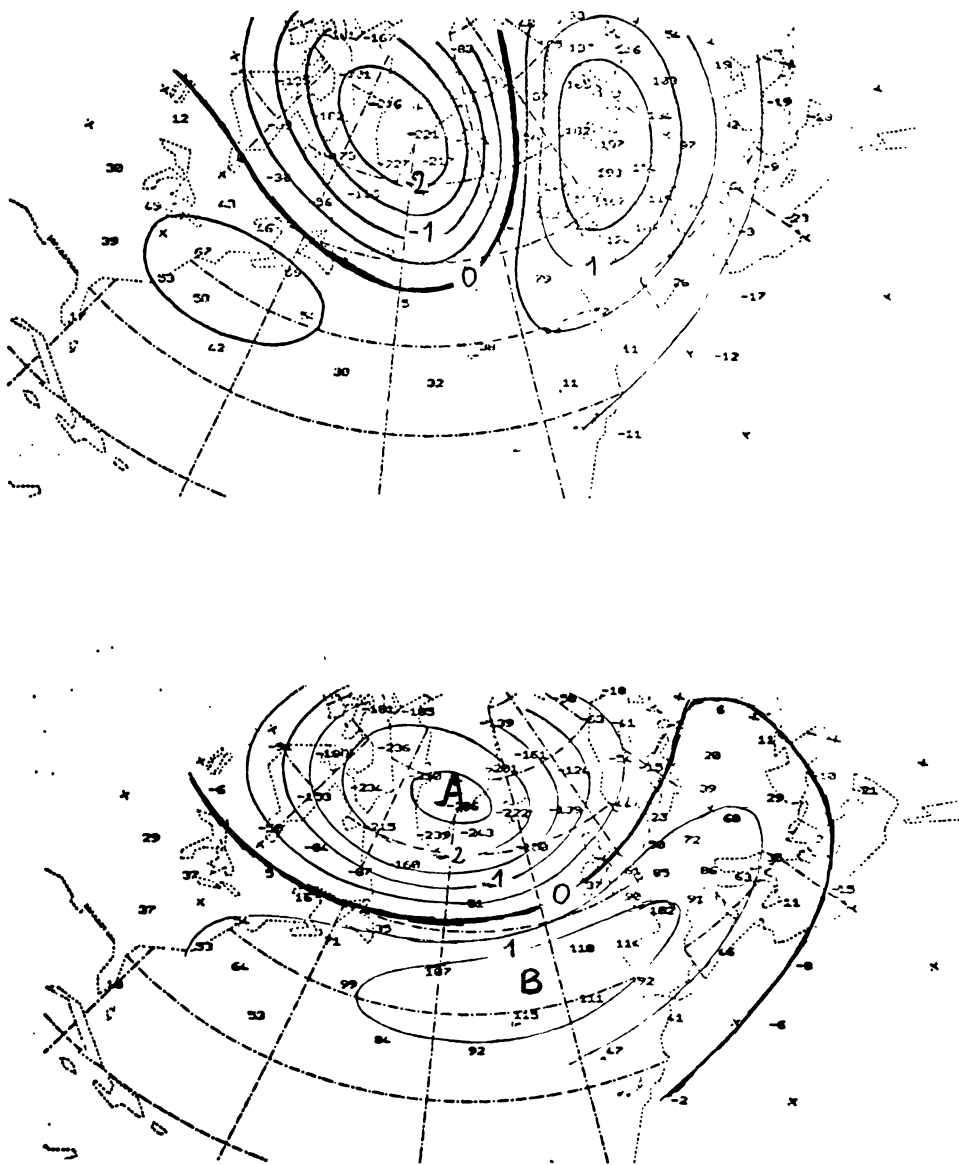


Figure 5. — Première Direction Principale (a) et première Direction Persistante (b) des 20 ans désaisonnalisés.

corrélation est .94 à un jour et .48 à 5 jours. La figure 4b montre que jusqu'à 10 jours (et en fait au-delà) la première Composante Persistante est plus auto-corrélée que la première Composante Principale.

IV. LEVEE DE LA DEGENERESCENCE DANS UNE ANALYSE EN COMPOSANTES PRINCIPALES

Considérons deux Composantes Principales $z_1(t)$ et $z_2(t)$ dégénérées (c'est-à-dire associées à la même valeur propre λ). Elles sont décorrélées, mais toute combinaison :

$$\xi_1 = z_1(t) \cos \theta - z_2(t) \sin \theta$$

$$\xi_2 = z_1(t) \sin \theta + z_2(t) \cos \theta$$

donne un couple (ξ_1, ξ_2) qui satisfait aux mêmes conditions (décorrélation et maximum de variance). Mais si nous supposons que z_1 et z_2 sont indépendantes, il n'en va pas de même pour ξ_1 et ξ_2 . En effet $r_{12}^k = \text{cor}(\xi_1(t), \xi_2(t+k)) = (\rho_1 - \rho_2) \cos \theta \sin \theta$ où ρ_1 et ρ_2 sont les autocorrélations temporelles à k jours de z_1 et z_2 . L'autocorrélation de ξ_1 vaut $\rho_1 \cos^2 \theta + \rho_2 \sin^2 \theta$ et celle de ξ_2 vaut $\rho_1 \sin^2 \theta + \rho_2 \cos^2 \theta$. Elles sont comprises entre ρ_1 et ρ_2 quelque soit θ . Donc dans un sous espace propre de dimension 2 les Composantes Principales qui ont un sens physique (c'est-à-dire celles qui sont indépendantes) sont les Composantes Persistantes. Ce résultat se généralise aisément en dimensions quelconque. Un problème de cette nature s'est posé lors de l'Analyse en Composantes Principales des données désaisonnalisées de l'expérience de 10 ans (Fig. 6 et 7). Les deux premières Composantes Principales expliquent 4.48 % et 4.14 % de la variance. Disposant de 10 ans, cela représente environ 1000 données indépendantes. Si l'on applique le test d'Anderson (MORISSON 1967), on voit que ces deux valeurs ne sont pas significativement différentes (alors qu'on montre que la troisième l'est).

Si l'on considère les deux Directions Principales on voit qu'elles représentent un couplage entre les deux hémisphères (ce qui est étonnant puisque les données sont désaisonnalisées) et que dans l'hémisphère Nord elles ont la même structure et dans l'hémisphère Sud elles ont des structures de signe contraire. On a bien envie de prendre leur somme et leur différence (normées par $1/\sqrt{2}$) mais cela peut sembler artificiel. Considérons les Composantes Principales ξ_1 et ξ_2 que nous donne l'algorithme de diagonalisation. Les coefficients de corrélation croisés r_{12}^k sont peut différents des r_{21}^k , augmentent avec k et valent .10 à 10 jours. Nous ne sommes donc pas en présence de variables indépendantes. On va donc chercher θ pour que $z_1 = \xi_1 \cos \theta + \xi_2 \sin \theta$ et $z_2 = -\xi_1 \sin \theta + \xi_2 \cos \theta$ aient des coefficients de corrélation croisés d'ordre k nuls, ou pour maximiser l'autocorrélation de z_1 il vient :

$$\text{cor}(z_1(t), z_2(t+k)) = (\rho_2 - \rho_1) \sin \theta \cos \theta + r_{12}^k (\cos^2 \theta - \sin^2 \theta)$$

puisque l'on a vu que $r_{12}^k \cong r_{21}^k$. Or il se trouve qu'à toutes les échéances le coefficient d'autocorrélation ρ_1 de ξ_1 est le même que celui de ξ_2 , ρ_2 . On obtient donc $\cos^2 \theta = \sin^2 \theta = 1/2$ et on peut prendre

$$z_1 = \frac{1}{\sqrt{2}} (\xi_1 + \xi_2) \quad z_2 = \frac{1}{\sqrt{2}} (-\xi_1 + \xi_2)$$

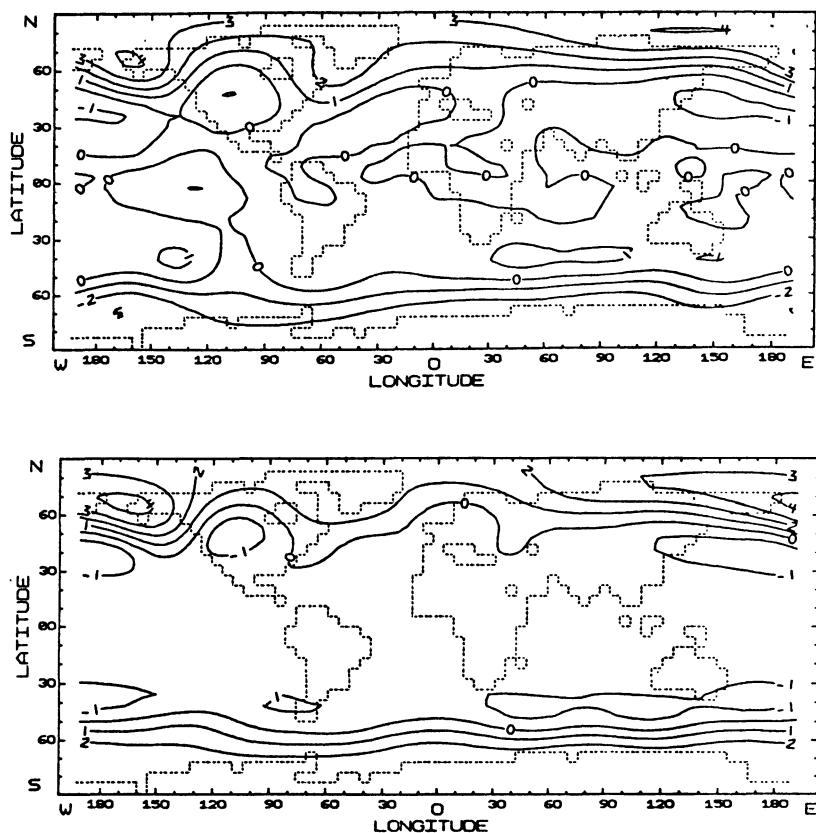


Figure 6. —Directions Principales des données désaisonnalisées C_1 et C_2 .

V. APPLICATION A L'ETUDE DES CYCLES

Nous nous sommes intéressés jusqu'à présent à la maximisation de ρ_1 dont nous avons vérifié qu'elle fournissait des résultats analogues à celle de ρ_k pour les petites valeurs de k . Intéressons-nous maintenant au cas où k devient plus grand (tout en restant petit devant la taille du fichier). Un cas particulier est celui où k vaut une année et où on cherche à mettre en évidence le cycle annuel. Dans le cas de l'expérience de circulation générale de 10 ans (filtrée par les 20 premières Composantes Principales qui expliquent 80 % de la variance totale, on considère ici l'année entière non désaisonnalisée), on trouve une première Composante Persistante (Fig. 8) d'autocorrélation .998 (à un an) et de variance expliquée 60 %. C'est la direction qui porte l'onde annuelle sinusoïdale et que l'on trouve également pour les faibles valeurs de k . La deuxième Composante Persistante possède une autocorrélation de .755 puis on trouve des composantes dont les autocorrélations sont positives et inférieures au seuil de signification estimé plus loin. Le cycle annuel est donc porté par deux directions et les autres composantes sont décorréliées avec celui-ci. Le procédé classique de désaisonnalisation (utilisé au III) se fait en retranchant à une série de valeurs la valeur de l'année moyenne au jour

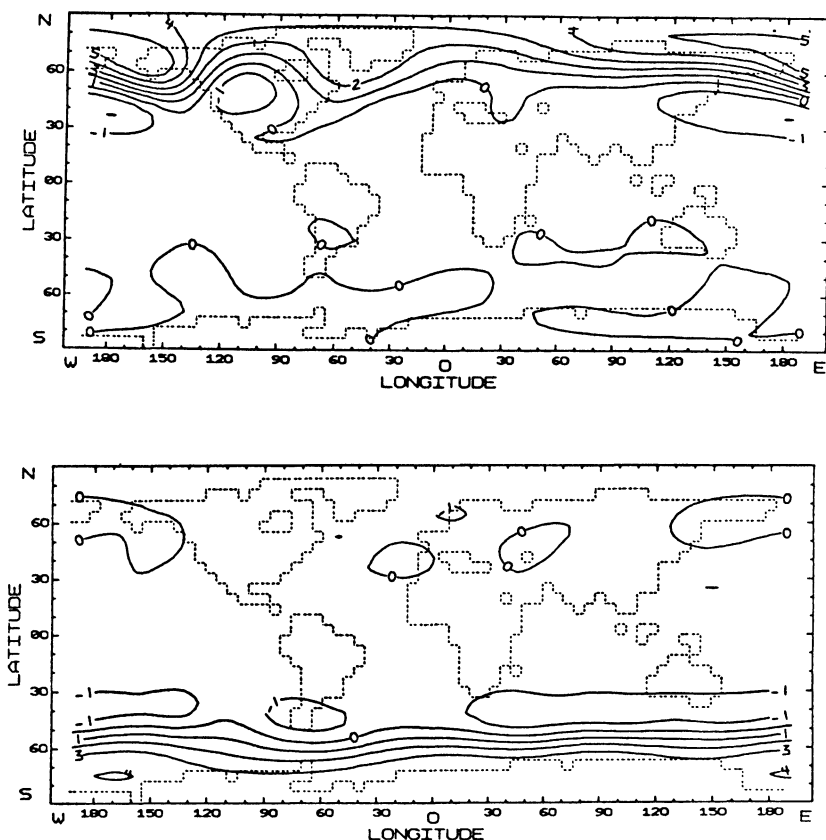


Figure 7. – Directions Principales des données désaisonnalisées

$$\frac{1}{\sqrt{2}} (C_1 + C_2) \text{ et } \frac{1}{\sqrt{2}} (-C_1 + C_2).$$

correspondant $x(\text{jour}, \text{an}) = x(\text{jour}, \text{an}) - x(\text{jour})$. S'il n'y a plus de cycle dans les moyennes de la série, il en reste dans les variances (les variances d'hiver sont plus fortes que celles d'été). Dans le cas d'une variable unidimensionnelle, on peut diviser par l'écart type au jour correspondant, mais dans le cas d'une variable multi-dimensionnelle, c'est plus délicat. De toutes façons il restera un cycle dans les autocorrélations temporelles (ce cycle a été mis en évidence par J.P. LABARTHE). Nous venons de mettre en évidence une procédure pour filtrer le cycle annuel : on décompose le champ en Composantes Persistantes, on annule les deux premières composantes, puis on recompose le champ. Géométriquement cela revient à faire une projection du champ au lieu de faire des translations variables suivant la date. Ce procédé, qui n'a pas encore été testé, nécessite que le cycle annuel d'une variable soit dans une variété de faible dimension.

On peut également regarder ce qui se passe quand k varie de 1 à 100 jours sur les 20 premières Composantes Principales de l'expérience de 10 ans (Fig. 9). Regardons ce que deviennent l'autocorrélation à k jours et la variance expliquée de la première composante Persistante à k jours pour les données brutes et les données désaisonnalisées (par la méthode traditionnelle) ; comme on fait une

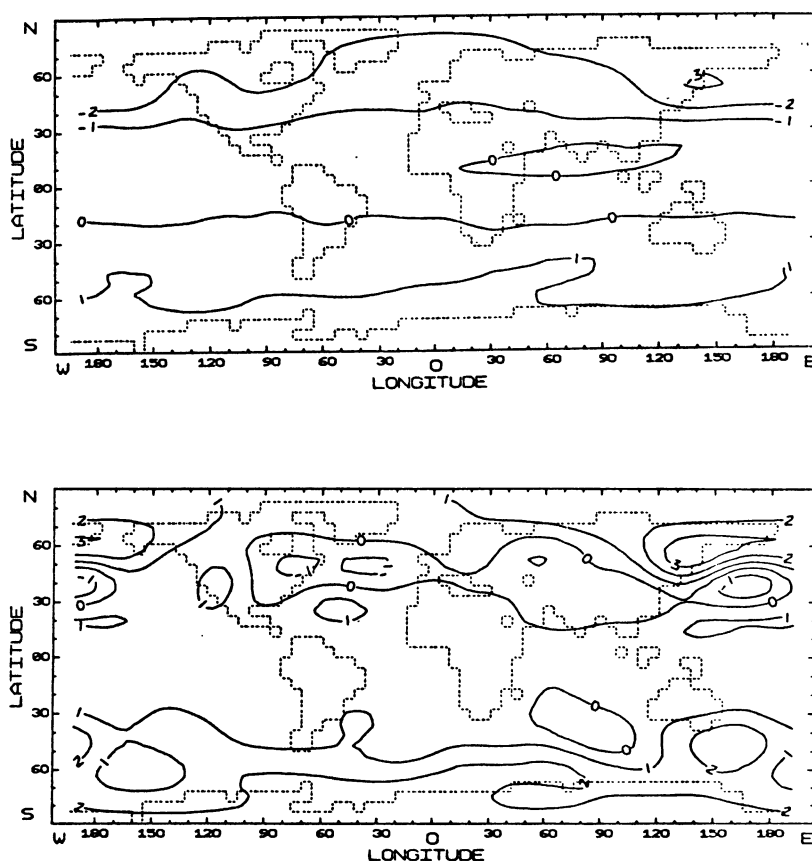


Figure 8. — 1^{re} et 2^e Directions Persistantes du cycle annuel.

Analyse en Composantes Persistantes à chaque échéance la courbe qui joint les autocorrélations est l'enveloppe de tous les diagrammes d'autocorrélation des combinaisons linéaires des 20 Composantes Principales. On observe une décroissance de l'autocorrélation puis une stabilisation à un seuil est de .40 pour les données brutes et .70 pour les données désaisonnalisées (bien que dans les deux cas on ait un fichier 20 x 3 660). La variance expliquée décroît également jusqu'à un seuil légèrement supérieur à la variance expliquée par la 20^e Composantes Principale (par construction il ne peut lui être inférieur). Il n'y a donc pas dans les données de géopotentiel à 500 mb du modèle du CEPMMT de durée de retour privilégiée correspondant à une variance non négligeable (onde à 5 jours, onde à 18 jours, onde à 28 jours (la lune n'est pas prise en compte dans le modèle !) etc.). La méthode que nous venons de décrire peut donc servir pour analyser le spectre temporel d'une série de champs (pour une série de variables unidimensionnelles on dispose de la méthode des coefficients de Fourier). La série des variances expliquées permet de contrôler si l'onde à k jours qui se manifesterait par un pic d'autocorrélation n'est pas de variance négligeable et donc probablement un artefact numérique.

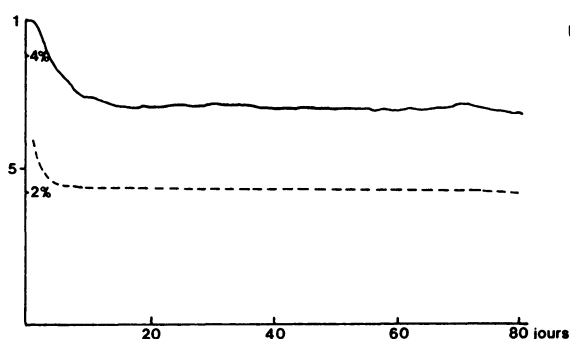
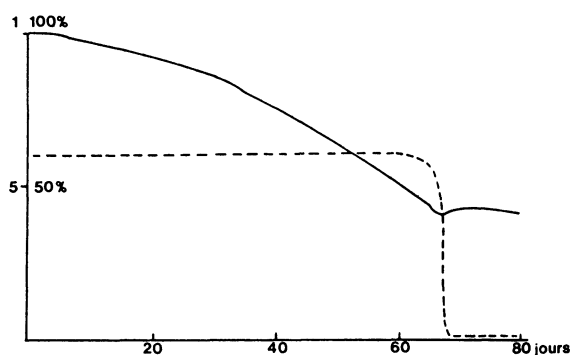


Figure 9. — Autocorrélation (—) et variance expliquée (— —) de la
Première Composante Persistante
a) données brutes,
b) données désaisonnalisées.

VI. COMPOSANTES MIXTES PRINCIPALES PERSISTANTES

Revenons à l'Analyse en Composantes Principales linéaire mais au lieu de considérer les valeurs quotidiennes du géopotential à 500 mb, prenons un fichier typiquement peu persistant : les moyennes mensuelles départementales de la pluie en France. Une première Composante Principale (Fig. 10a) porte 52 % de la variance totale et correspond à la côte atlantique et aux reliefs exposés aux perturbations d'ouest. Les Composantes Principales suivantes expliquent 16 %, 6 %, 5 % mais les autocorrélations sont faibles .02, .15, .11 et .11. Il n'y a pas de correspondance entre l'échelle spatiale du département et l'échelle temporelle du mois. Nous sommes donc conduits à chercher les Composantes Persistantes de ce fichier. On trouve 36 Composantes sur 90 dont l'autocorrélation est supérieure à .50, la première étant .84, mais les variances expliquées sont inférieures à 2 %. On aimerait pourtant bien trouver une projection du champ bien autocorrélée et cependant significative. Une première méthode consiste à ne retenir que les plus grandes Composantes Principales, mais on risque de perdre de l'information car contrairement au géopotential la persistance n'est pas concentrée dans les premières Composantes Principales. Reprenons le passage de la variable réelle au vecteur.

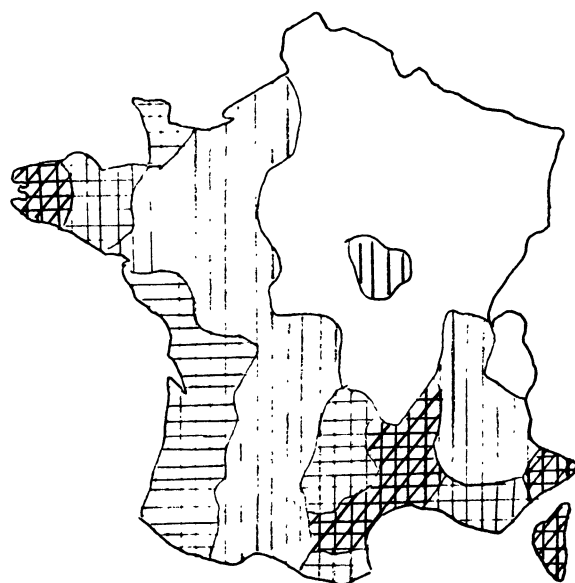
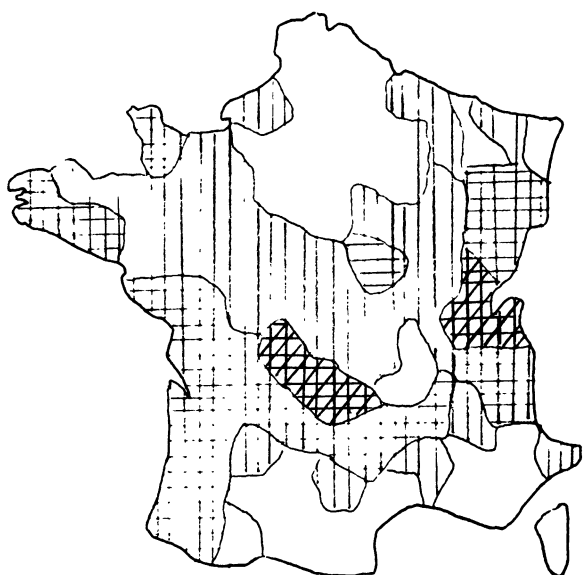


Figure 10. — Précipitations mensuelles départementales,
a) 1^{re} Direction Principale,
b) 1^{re} Direction Mixte Principale Persistante.

- Très faible.
- ▤ Faible.
- ▥ Forte.
- ▧ Très forte.

Dans le premier cas, on dispose d'une mesure de la persistance : l'autocorrélation. Dans le second, on a p coefficients d'autocorrélation dont on peut faire la somme pondérée. Le plus logique est de choisir comme pondération la variance expliquée par chaque coordonnée. Reprenons les notations de II.

L'autocorrélation de la i^e coordonnée z^i est $\rho(z^i) = \frac{W_{ii}}{V_{ii}}$ (nous omettrons le coefficient k fixé une fois pour toutes).

La variance expliquée par z^i est $\frac{V_{ii}}{\sum_j V_{jj}}$

Un coefficient global $\rho(Z)$ est $\sum_{i=1}^p \rho(z^i) \frac{V_{ii}}{\sum_j V_{jj}} = \frac{\text{Tr } W}{\text{Tr } V}$

Cette opération est commode car elle est invariante par rotation des axes. Mais ce coefficient global peut être nul soit parce qu'aucune des coordonnées n'est persistante, soit parce que certaines autocorrélations sont proches de 1 alors que d'autres sont négatives. Nous prendrons donc comme coefficient d'autocorrélation totale

$$\rho(Z) = \sum_{i=1}^p \rho(z^i) \frac{V_{ii}}{\sum_j V_{jj}}$$

Alors

$$\rho(Z) = 1 \Rightarrow \forall i \mid \rho(z^i) \mid = 1$$

$$\rho(Z) = 0 \Rightarrow \forall i \mid \rho(z^i) \mid = 0 \quad \text{aucune persistance dans le champ.}$$

On peut chercher une rotation A maximisant $\rho(AZ)$.

On trouve alors $\text{Max}(\rho(AZ)) = \frac{1}{\text{Tr } V} \text{Max} \left(\sum_i |A_i' \tilde{W} A_i| \right)$ où A_i est la i^e colonne de la matrice A .

Soient D_i les vecteurs propres de $\tilde{W}$ (matrice symétrique) associés aux valeurs propres w_i . En décomposant la base orthonormée (A_i) dans la base des (D_i)

$$\sum_i |A_i' \tilde{W} A_i| = \sum_i \left| \sum_j A_{ij}^2 w_j \right| \leq \sum_j |w_j| \left(\sum_i A_{ij}^2 \right) = \sum_j |w_j| = \sum_j |D_j' \tilde{W} D_j|$$

Donc $\rho(DZ)$ est solution de la maximisation.

Si on ordonne les $|w_i|$ par ordre décroissant on définit une série de composantes

$$x_i^* = \{x_i^{*n} / n = 1, N\}$$

avec

$$x_i^{*n} = D_i'(Z_n - \bar{Z})$$

la première maximise $|\rho(x_i)| \cdot \frac{\text{var}(x_i)}{\text{Tr } V}$ alors que l'analyse en Composantes Principales maximise $\frac{\text{var}(x_i)}{\text{Tr } V}$ et que l'Analyse en Composantes Persistantes maximise $\rho(x_i)$. La seconde n'est pas décorrélée à la première mais on a seulement

$$\text{cov}(x_1^{*n}, x_2^{*n+k}) = 0$$

Si x_i^* est assez persistante, on peut considérer que x_i^* et x_j^* sont décorrélés. Chaque composante x_i^* que nous appellerons Mixte Principale-Persistante maximise l'autocorrélation pondérée (sous la contrainte d'orthogonalité) $|\rho(x_i^*)| \cdot \frac{\text{var}(x_i^*)}{\text{Tr } V}$ et la somme de ces autocorrélations pondérées qui constitue l'autocorrélation totale est maximale et vaut $\rho(x^*) = \frac{\sum_i |w_i|}{\text{Tr } V}$.

Avec la méthode que nous venons de décrire les premières Composantes Mixtes ne peuvent avoir ni une variance expliquée trop faible ni une autocorrélation trop faible et sont sensiblement décorrélées. On dispose donc d'un palliatif lorsque les Composantes Persistantes sont très peu significatives ainsi que d'un coefficient pour mesurer la persistance globale d'un champ. Dans le cas qui nous intéresse (moyennes mensuelles départementales de pluie), l'autocorrélation totale est de .16. On pouvait a priori se douter de la faible persistance de ce champ. La première Composante Mixte (Fig. 10b) possède une autocorrélation pondérée de .05 soit une autocorrélation de .20 et une variance expliquée de 27 %. Elle correspond aux domaines d'influence maritime.

VII. CONCLUSION

La persistance d'un champ peut donc s'étudier en décomposant ce champ suivant des directions privilégiées solutions d'un problème d'optimisation analogue à celui de l'Analyse en Composantes Principales. Les composantes obtenues, baptisées persistantes, sont centrées, décorrélées entre elles et d'autocorrélation temporelle décroissante (la méthode fournit des fonctions linéaires très persistantes mais elle peut se généraliser à des fonctions polynomiales en augmentant la taille du champ par l'adjonction de monômes).

Sur une série de 10 années de champs de géopotential à 500 mb on constate que les Directions Persistantes ne coïncident pas avec les Directions Principales. Cette décomposition permet de lever la dégénérescence dans une Analyse en Composantes Principales. Elle permet de chercher les cycles ou les durées de retour ainsi que de désaisonnaliser une série de champs de manière différente de la méthode de soustraction de la moyenne. Son application permet d'affirmer qu'il n'y a pas de cycle de variance significative dans la série considérée. Toutefois cette méthode ne donne pas des résultats satisfaisants lorsqu'on traite des champs peu persistants. On développe alors la notion de Composante Mixte Principale-Persistante qui permet d'obtenir des composantes autocorrélées et de variance non négligeable. En outre cette méthode fournit un indicateur global de la persistance, l'autocorrélation totale.

REMERCIEMENTS

Je tiens particulièrement à remercier M. Guy DER MEGREDITCHIAN pour son soutien constant et pour ses nombreux conseils.

J'adresse également de vifs remerciements à MM. Jean Pierre VOLMER, Michel JARRAUD et ULRICH CUBASCH qui ont mis en œuvre l'expérience de circulation générale et ont mis les résultats à ma disposition.

Je remercie en outre Jean Pierre LABARTHE, Jean François ROYER et Daniel ROUSSELET pour l'aide occasionnelle qu'ils m'ont apportée.

BIBLIOGRAPHIE

- D. ANDRE (1981). — Analyse d'une expérience de circulation générale effectuée avec le modèle spectral du CEPMMT. *Rapport de stage de l'ENM*.
- M. DEQUE, J.F. ROYER (1982). — Modélisation du cycle annuel à l'échelle globale : étude de janvier et juillet. *Note de Travail de l'EERM n° 35*.
- G. DER MEGREDITCHIAN, C. MANTE (1981). — Autoregression of the 500 mb geopotential field for short and mean range forecasting. *Seminar on Statistical Prediction in Meteorology*, Nice.
- J.P. LABARTHE (1981). — Contribution à la prévision à longue échéance ; étude de la température moyenne mensuelle. *Note technique du SMM n° 9*, Nouvelle série.
- J.P. LABARTHE (1982). — Quelques réflexions sur une tentative de prévision mensuelle. *Note Technique de la DMN* (à paraître).
- N.C. LAU (1981). — A diagnostic study of recurrent meteorological anomalies appearing in a 16-year simulation with a GFDL general circulation model. *Monthly Weather Review n° 109*.
- D.F. MORISSON (1967). — Multivariate statistical methods. Mc Graw Hill Book Company.
- M.F. RILAT, M.A. THEPAUT (1982). — Analyse du cycle hydrologique dans une expérience de .10 ans de simulation du climat. *Note de travail de l'ENM n° 20*.
- J.F. ROYER, M. DEQUE, H.J. CANETTI, M. BOULANGER (1981). — Présentation du modèle spectral de circulation générale à faible résolution. Simulation du climat de janvier. *Note de travail de l'EERM n° 16*.
- J.P. VOLMER, M. DEQUE, M. JARRAUD (1983). — Large scale fluctuations in a long range integration of the ECMWF spectral model. *Tellus n° 35 A*, 173-188.