

REVUE DE STATISTIQUE APPLIQUÉE

K. KOŠMELJ

B. LACHET

Aspect temporel des relations entre les variables hydriques du Haut-Rhône français

Revue de statistique appliquée, tome 31, n° 2 (1983), p. 5-18

http://www.numdam.org/item?id=RSA_1983__31_2_5_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/>

ASPECT TEMPOREL DES RELATIONS ENTRE LES VARIABLES HYDRIQUES DU HAUT RHONE FRANCAIS

K. KOSMELJ^V(*), B. LACHET(**)

RESUME

Dans la région du Haut-Rhône français plusieurs variables hydriques ont été observées pendant les années 1972-1979. L'objectif de notre étude est la mise en œuvre d'une méthodologie statistique qui permette une meilleure compréhension de la dynamique spatio-temporelle de l'eau dans la région étudiée. Cet article est organisé de la façon suivante : dans le premier chapitre le cadre de notre étude est présenté. Les données sont exposées, l'objectif de notre travail est précisé. Le deuxième et le troisième chapitre présentent la méthodologie qui a été utilisée pour comprendre les relations entre les variables et pour effectuer la classification des variables selon plusieurs critères. Les résultats obtenus à partir de ces variables hydriques sont illustrés sur plusieurs graphiques et tableaux. Le dernier chapitre est consacré aux conclusions et ouvre quelques perspectives pour une étude statistique plus globale.

1. PRESENTATION DE L'ETUDE ET DE SON OBJECTIF

Cette étude se situe dans le cadre du "Programme Interdisciplinaire de Recherche en Environnement ASP n° 2 du CNRS", qui a pour but les levés des cartes phytocécologiques qui assureront une couverture complète de la vallée du Rhône entre Lyon et Genève. Plus précisément, notre étude concerne l'écologie végétale dans la zone située entre Groslée et les Avenières. Pour mieux comprendre l'influence et la dynamique de l'eau nous avons entrepris une étude statistique des variables hydriques dans cette zone.

Nous disposons de 12 variables hydriques dans la zone étudiée. Ces variables sont des séries chronologiques ; leurs valeurs ont été relevées chaque semaine pendant 8 ans, de 1972 à 1979 ; la longueur des séries est donc de 412. Les variables sont les suivantes :

- trois variables explicatives : le débit du Rhône, la température et la pluie
- neuf variables expliquées : la hauteur de la nappe phréatique mesurée dans 9 stations.

Nous présentons tour à tour ces variables dont certaines comportent des données manquantes. Les techniques d'analyse des séries chronologiques n'autorisent pas un grand nombre de données manquantes ; pour que les séries soient acceptables, nous avons préalablement remplacé les valeurs manquantes en prenant en compte le plus possible la nature de la série.

(*) Département des Mathématiques Appliquées, Institut Jozef Stefan, Jamova 39, 61001 Ljubljana, Yougoslavie.

(**) Laboratoire de Biologie Végétale, Département de Recherche Fondamentale, Centre d'Etudes Nucléaires de Grenoble, 85 X, 38041 Grenoble Cedex, France.

Mots-clés : Ecologie, Fleuve, Nappe phréatique, Pluie, Température, Débit, Statistique, Série chronologique, Autocorrélation, Corrélation croisée, Modèle de Box-Jenkins, Classification automatique, Distance.

Variables explicatives

Débit du Rhône

Une valeur hebdomadaire est obtenue en calculant la moyenne des valeurs mesurées à la Balme toutes les trois heures. La série du débit est présentée sur la figure 1. Il est évident que le régime du débit a changé plusieurs fois pendant les années observées (1972-1975, 1976 : l'année de sécheresse, 1976-1979).

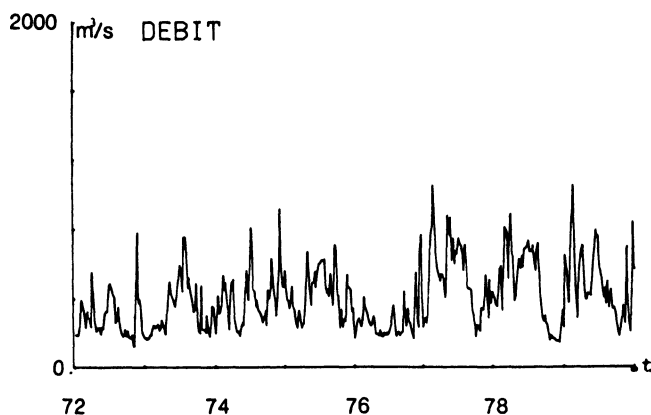


Figure 1. — Débit du Rhône.

Température

Une valeur hebdomadaire représente la moyenne des températures maximales journalières à Ambérieu. La série n'est pas complète, 12 données sont manquantes. Sur la figure 2 on observe une périodicité régulière de la série. Pour cette raison on a remplacé les valeurs manquantes par la moyenne des mesures existantes prises aux mêmes dates, pendant les années observées.

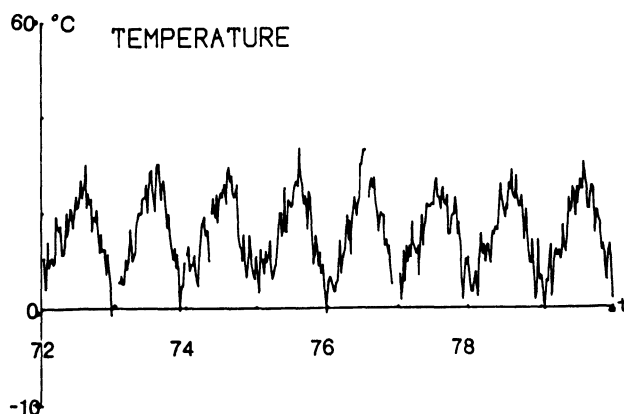


Figure 2. — La température.

Pluie

Une valeur hebdomadaire représente la quantité totale de pluie tombée durant la semaine à Ambérieu. Il n'y a que deux valeurs manquantes. Sur la figure 3

on observe que la série ne possède aucune régularité. Son analyse a montré que cette série, moyenne déduite, est un bruit blanc ; on peut donc remplacer les deux valeurs manquantes par des valeurs tout à fait quelconques.

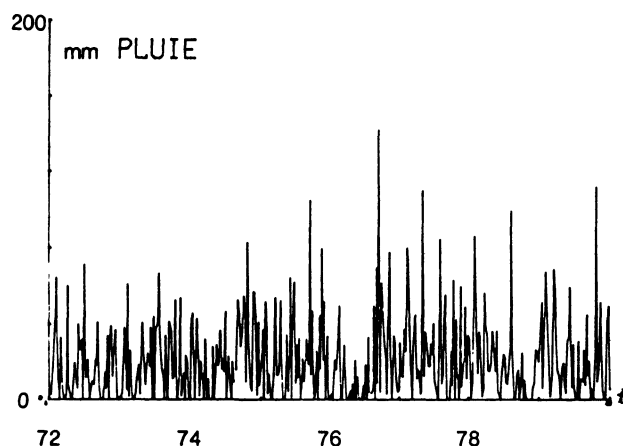


Figure 3. – La pluie.

Variables expliquées

Hauteur de la Nappe phréatique

Elle est mesurée dans 9 stations différentes de la zone étudiée, leur situation est représentée sur les figures 6, 7 et 8. Les valeurs de la hauteur de la nappe (en abrégé "nappe") ont été mesurées tous les lundis. Elles sont exprimées en cm, une valeur positive signifie que le terrain est inondé autour de la station. Les séries sont caractérisées par un grand nombre de données manquantes, les années 1976 et 1977 ne sont souvent représentées que par quelques points isolés. Pour remplacer les données manquantes l'interpolation linéaire entre les parties isolées de la série s'est avérée préférable. Elle constitue une sorte de lissage de la série, tout en conservant la structure du rythme observable mais en éliminant les fluctuations aléatoires. Sur les figures 4 et 5 les variations de deux nappes sont présentées. Les graphiques des séries "lissées" sont représentés juste au-dessous des graphiques des séries originelles.

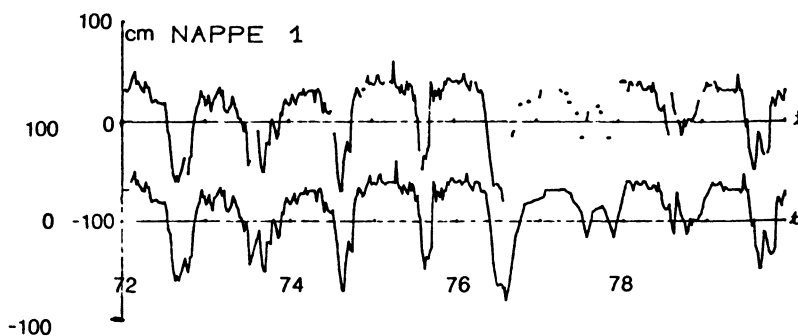


Figure 4. – Hauteur de la nappe phréatique mesurée dans la station 1.

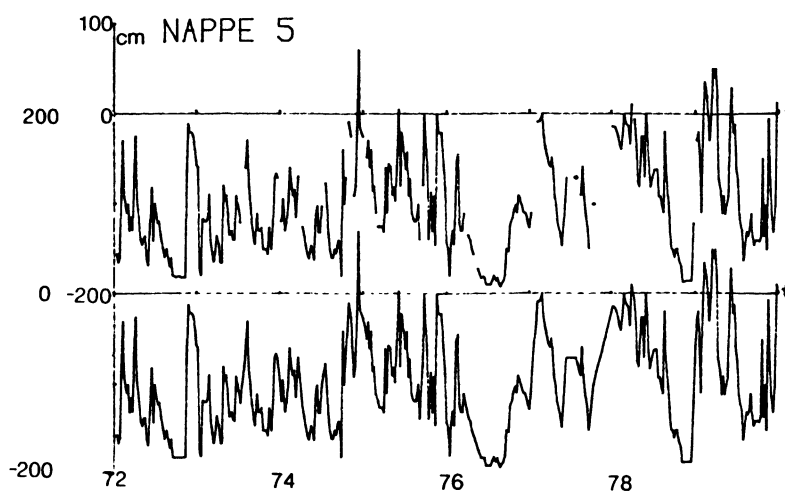


Figure 5. — Hauteur de la nappe phréatique mesurée dans la station 5.

Pour pouvoir précisément définir l'objectif de notre étude, nous introduisons d'abord le terme "écosystème". On suppose que la zone étudiée est assez petite pour que les valeurs des variables explicatives ne changent pas d'une station à l'autre et qu'elles soient bien représentées par les mesures que nous possédons. Sous le terme "écosystème" nous considérons une station décrite par 4 variables : le débit, la température, la pluie et la hauteur de la nappe phréatique mesurée dans cette station.

L'objectif de notre étude est la mise en œuvre d'une méthodologie statistique qui permette :

- A. de comprendre les relations entre les variables étudiées
- B. d'effectuer la classification automatique
 - a) des nappes phréatiques selon la similarité de leurs corrélogrammes ;
 - b) des nappes phréatiques selon les relations qui existent entre elles ;
 - c) des écosystèmes.

2. RELATIONS ENTRE LES VARIABLES ETUDIEES

2.1. Séries chronologiques

Une série chronologique est un ensemble d'observations ordonnées dans le temps. Nous ne considérons que des séries chronologiques discrètes et finies avec des observations prises à des intervalles de temps égaux ; la notation suivante sera utilisée pour une telle série chronologique z ayant N observations $z_1, z_2, \dots, z_N$.

Les valeurs de la série chronologique peuvent être corrélées. Le coefficient d'autocorrélation de la série z avec décalage k , dénoté $\rho_{zz}(k)$, mesure la corrélation entre les observations z_t et z_{t+k} . La suite des coefficients d'autocorrélation cal-

culés pour divers décalages est appelée autocorrélation et est représentée par un corrélogramme.

Soit deux séries chronologiques x et y . Le coefficient de corrélation croisée avec décalage k , dénoté $\rho_{xy}(k)$, mesure la corrélation entre les observations x_t et y_{t+k} .

La suite des coefficients de corrélation croisée pour divers décalages est appelée corrélation croisée ; elle décrit la dépendance entre les deux séries.

Les caractéristiques de l'autocorrélation et de la corrélation croisée sont bien expliquées dans la littérature [2], [5] etc. Les estimateurs $r_{xx}(k)$ de $\rho_{xx}(k)$ et $r_{xy}(k)$ de $\rho_{xy}(k)$ y sont précisément définis.

BARTLETT a montré que la covariance entre $r_{xy}(k)$ et $r_{xy}(1)$, $1 \neq k$, dépend des autocorrélations des deux séries et de leur corrélation croisée ; l'interprétation de $r_{xy}(k)$ est donc impossible. On évite ce problème quand les deux séries sont des bruits blancs, dans ce cas cette covariance est égale à zéro. Donc pour pouvoir interpréter les $r_{xy}(k)$, $k = -K, \dots, K$, il faut d'abord trouver les résidus des séries qui sont des bruits blancs, et calculer ensuite la corrélation croisée entre les résidus des séries originelles [5], [7].

2.2. Présentation des résultats

2.2.1. Modèles de Box-Jenkins pour les variables étudiées.

Pour obtenir des bruits blancs à partir des séries originelles la méthodologie de BOX-JENKINS a été utilisée. Les séries originelles ont été décrites par des modèles ARIMA. Par souci de simplicité nous avons considéré les modèles les moins sophistiqués qui sont toutefois acceptables. Ici on présente pour chaque série chronologique z le modèle retenu, les valeurs estimées des paramètres avec leurs intervalles de confiance à 95 % et le seuil de probabilité p du test de validité du modèle [2].

Débit

$$(1 - B)(1 - B^{52})z_t = (1 - \theta_1 B - \theta_2 B^2)(1 - \theta B^{52})a_t$$

$$\theta_1 = 0.31 [0.21, 0.41] \quad \theta = 0.72 [0.63, 0.81]$$

$$\theta_2 = 0.33 [0.23, 0.43] \quad p = 0.62$$

Température

$$(1 - B^{52})z_t = (1 - \theta_1 B)(1 - \theta B^{52})a_t$$

$$\theta_1 = -0.31 [-0.41, -0.21] \quad \theta = 0.68 [0.60, 0.77]$$

$$p = 0.93$$

Pluie

$$z_t = \theta_0 + a_t$$

$$\theta_0 = 21.27 [19.00, 23.35] \quad p = 0.09$$

2.2.2. Corrélations croisées entre les variables étudiées

Ayant obtenu les résidus des séries chronologiques étudiées, leurs corrélations croisées ont été calculées. L'estimateur proposé par BOX-JENKINS [2] a été

utilisé, au seuil de signification $\alpha = 0.05$ la valeur critique de $r_{xy}(k)$ vaut $2/\sqrt{N} = 2/\sqrt{363} = 0.105$, où N représente la longueur des résidus.

Dans le texte les abréviations suivantes seront utilisées : D = débit, T = température, P = pluie, N_i = hauteur de la nappe phréatique dans la station i .

a) Corrélations croisées entre les variables explicatives : les valeurs significatives suivantes ont été obtenues :

Pluie – Débit

décalage 0 : 0.39

décalage 1 : 0.18

Température – Débit

décalage 1 : 0.16

On a également trouvé quelques valeurs significatives avec des décalages plus grands, par exemple $r_{TD}(18) = 0.14$.

b) Les corrélations croisées entre les variables explicatives et les variables expliquées sont présentées dans le Tableau 1.

TABLEAU 1

Corrélations croisées entre les variables expliquées (Hauteurs des différentes nappes phréatiques) et les 3 variables explicatives (Débit, température, Pluie), avec décalage 0.

Hauteurs	Débit	Température	Pluie
Nappe 1	0.20	– 0.18	0.23
Nappe 2	0.19	– 0.11	0.21
Nappe 3	0.22	– 0.13	0.30
Nappe 4	0.23	– 0.16	0.23
Nappe 5	0.43	– 0.05	0.27
Nappe 6	0.23	– 0.11	0.23
Nappe 7	0.28	– 0.06	0.21
Nappe 8	0.31	– 0.09	0.28
Nappe 9	0.25	– 0.14	0.26

Appelons nappe fluviale une nappe ayant une corrélation plus forte avec le débit qu'avec la pluie, nappe pluviale celle qui a une corrélation plus forte avec la pluie qu'avec le débit et nappe intermédiaire une nappe avec les deux corrélations du même ordre. Selon ces définitions les résultats du Tableau 1 sont présentés sur la figure 6.

On a trouvé deux coefficients de corrélation significatifs avec décalage 1 : $r_{TN_1}(1) = -0.12$ et $r_{TN_6}(1) = -0.13$. Quelques valeurs significatives avec des décalages plus grands, par exemple, $r_{TN_7}(6) = -0.15$, $r_{TN_6}(10) = -0.12$, etc. ont aussi été détectées.

c) Les corrélations croisées entre les variables expliquées avec le décalage 0 sont présentées dans le tableau suivant :

TABLEAU 2

Corrélations croisées entre les hauteurs de nappes phréatiques avec décalage 0.

Décalage 0	Nappe 1	Nappe 2	Nappe 3	Nappe 4	Nappe 5	Nappe 6	Nappe 7	Nappe 8
Nappe 1								
Nappe 2	0.35							
Nappe 3	0.61	0.49						
Nappe 4	0.60	0.43	0.61					
Nappe 5	0.23	0.30	0.35	0.35				
Nappe 6	0.35	0.24	0.39	0.43	0.44			
Nappe 7	0.35	0.36	0.43	0.42	0.41	0.48		
Nappe 8	0.35	0.21	0.37	0.41	0.43	0.55	0.38	
Nappe 9	0.38	0.35	0.50	0.42	0.46	0.44	0.47	0.40

Quelques corrélations croisées entre les variables expliquées avec décalage 1 ont aussi été détectées.

2.3. Discussion des résultats

Les résultats montrent que la pluie augmente le débit du Rhône pendant la semaine en cours et qu'il reste une influence la semaine suivante. Les résultats que nous avons obtenus sont dus à la forme des mesures (les moyennes pour le débit, la quantité totale pour la pluie) qui masquent les variations journalières. On a décelé l'influence de la température sur le débit avec un retard d'une semaine, ce qui est vraisemblablement dû à la fonte des neiges. Le phénomène ne se produit que pendant quelques mois de l'année, donc l'intensité de cette liaison varie considérablement pendant l'année. Les relations entre les variables explicatives et les variables expliquées montrent que toutes les nappes montent avec le débit et avec la pluie, par contre les nappes pluviales et intermédiaires baissent quand la température augmente. La Figure 6 met en évidence que l'influence du débit dépasse l'influence de la pluie près du chenal principal du Rhône, par contre les nappes pluviales et intermédiaires sont plus éloignées du Rhône. Ces résultats montrent aussi que la nappe 5 est fortement corrélée avec le débit : les figures 1 et 5 montrent une similarité des rythmes du débit et de la nappe 5.

Les corrélations croisées les plus fortes entre les nappes existent entre celles qui sont liées par des canaux et entre certaines autres qui sont proches géographiquement.

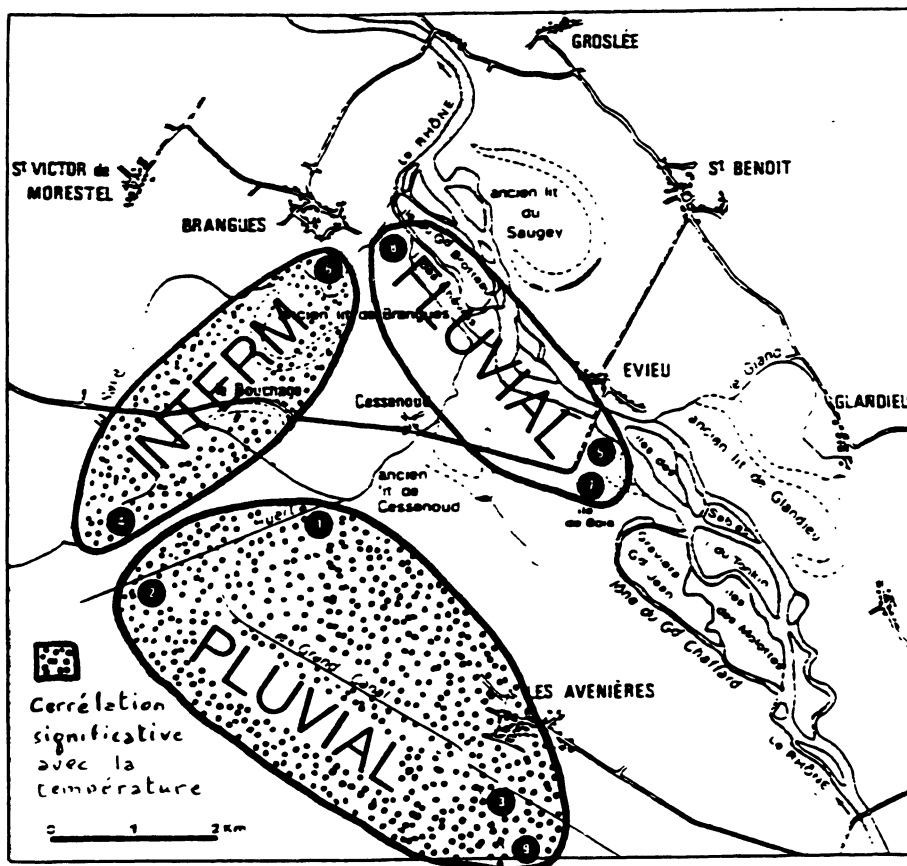


Figure 6. — Représentation graphique des nappes pluviales, intermédiaires et fluviales.

Les corrélations croisées significatives entre les variables étudiées avec des décalages plus grands qu'une semaine nous paraissent douteuses. Elle sont vraisemblablement dues aux fréquences non expliquées par les modèles de BOX-JENKINS.

3. CLASSIFICATION AUTOMATIQUE

3.1. Introduction

Rappelons que le problème de la classification automatique est la partition de n objets X_i^1 , $i = 1, \dots, n$ de l'ensemble E en K groupes X_i , $i = 1, \dots, K$ telle que :

$$X_1 \cup X_2 \cup \dots \cup X_K = E \quad (1)$$

$$X_i \cap X_j = \emptyset, i \neq j \quad (2)$$

Nous estimerons ici que le nombre des groupes k n'est pas fixé *a priori*. Les groupes doivent être les plus différents possibles les uns des autres vis-à-vis d'un critère, cependant chaque groupe doit être le plus homogène possible vis-à-vis de ce même critère. Le critère exprime le but de la classification et doit être prescrit *a priori*. Pour le décrire une application d appelée distance est introduite, elle doit satisfaire aux axiomes suivants :

$$\begin{aligned} d : E \times E &\rightarrow R^+ \cup \{0\} \\ \forall X^i \in E &d(X^i, X^i) = 0 \\ \forall X^i, X^j \in E &d(X^i, X^j) = d(X^j, X^i) \\ X^i = X^j &\Rightarrow d(X^i, X^j) = 0 \end{aligned} \quad (3)$$

La distance mesure la dissimilarité entre les objets par rapport au critère prescrit.

Nous voulons adapter les méthodes de la classification automatique pour classer les séries chronologiques. Il nous paraît important de prendre en compte la dimension spécifique des séries chronologiques : le temps. Pour cette raison nous avons construit des distances entre les séries chronologiques à l'aide des fonctions spécifiques de ces séries : autocorrélation et corrélation croisée [1]. Une fois la matrice des distances calculée, les mêmes algorithmes que dans le cas non temporel sont utilisés.

Nous présentons trois distances conçues pour effectuer la classification automatique des séries chronologiques. Soit x et y deux séries chronologiques, la distance qui les sépare est notée $d(x, y)$.

3.2. Classification des séries chronologiques selon la similarité des corrélogrammes

On cherche une partition de l'ensemble des séries chronologiques telle que dans le même groupe on trouve des séries chronologiques ayant des corrélogrammes semblables, par contre dans des groupes différents des séries chronologiques ayant des corrélogrammes différents.

Il est connu [2] qu'un processus stochastique stationnaire gaussien décrit par une série chronologique est uniquement caractérisé par sa moyenne, sa variance et son autocorrélation. Pour satisfaire le critère prescrit il faut comparer les autocorrélations des séries chronologiques. Nous proposons pour cela la distance suivante entre deux séries stationnaires x et y à distribution gaussienne

$$d(x, y) = \sum_{k=0}^K w(k) \cdot \frac{|r_{xx}(k) - r_{yy}(k)|}{1 + |r_{xx}(k) - r_{yy}(k)|} \quad (4)$$

où $K > 0$ représente la borne supérieure de l'indice courant des décalages et $w(k)$ les poids. Ces deux termes doivent être choisis par l'analyste selon la nature du problème.

On peut vérifier aisément que cette distance satisfait aux axiomes (3). Dans le cas où les autocorrélations des séries x et y sont très différentes la distance est grande ; par contre quand les autocorrélations des séries se ressemblent la distance est petite. Donc on peut conclure que cette distance est acceptable pour effectuer la classification des séries chronologiques selon la similarité des corrélogrammes.

Cette distance ne prend pas en considération les relations entre x et y ni leurs moyennes et variances.

Résultats de la classification des nappes selon la similarité de leurs corrélogrammes.

Pour aboutir aux séries stationnaires des nappes la transformation suivante a été effectuée sur les séries originelles : $(1 - B^{52}) (1 - B)$, où B représente l'opérateur de "recul", $B^k z_t = z_{t-k}$.

Nous avons pris en compte 52 semaines, donc $K = 52$. Nous considérons que tous les décalages ont la même importance, donc nous avons utilisé des poids égaux : $w(k) = 1$; $k = 0, 1, \dots, 52$.

Trois méthodes d'agrégation hiérarchique ont été utilisées : méthode du minimum, méthode de la moyenne et méthode du maximum [4]. A partir des dendrogrammes et en tenant en compte les niveaux d'agrégation des sous-groupes, les partitions suivantes ont été obtenues (les accolades délimitent un groupe).

Minimum

$\{1, 6, 7, 9, 4, 8, 3\}, \{2\}, \{5\}$

Moyenne

$\{1, 6, 7, 9, 4, 8, 2, 3\}, \{5\}$

Maximum

$\{1, 6, 9, 4, \{7, 8\}\}, \{5\}, \{2, 3\}$

Les résultats montrent que la nappe 5 se sépare dans les trois cas, la nappe 2 dans deux cas. On peut donc conclure que la nappe 5 a un caractère particulier par rapport aux autres et que la nappe 2 possède aussi un caractère spécifique.

L'analyse précédente a montré que la nappe 5 est fortement liée au débit et que son rythme suit aussi le rythme du débit. Il nous reste à savoir pourquoi la nappe 2 diffère des autres. Sur le terrain nous avons constaté que la position du piézomètre dans cette station a été changée deux fois pendant les années observées. De plus, la profondeur du piézomètre a été prise comme mesure de cette nappe pendant les semaines où il était à sec faute d'être assez profond.

3.3. Classification des séries chronologiques selon leurs inter-relations.

On cherche une partition de l'ensemble des séries chronologiques telle que dans le même groupe on trouve des séries qui sont liées entre elles, dans des groupes différents des séries chronologiques qui n'ont pas de liaison. Comme indice de dépendance entre des séries chronologiques nous prenons les coefficients de corrélation croisée.

Soit x et y les résidus des séries originelles. Pour respecter le critère ci-dessus, nous proposons la distance suivante :

$$d(x, y) = + \sqrt{\sum_{k=-K_1}^{K_2} w(k) (1 - r_{xy}^2(k))} \quad (5)$$

K_1 , K_2 et $w(k)$ ayant la même signification que dans le cas précédent.

On peut vérifier que (5) satisfait les axiomes (3). La distance est petite dans le cas de la forte dépendance entre x et y par contre elle est grande quand

la dépendance entre x et y est faible. Donc cette distance décrit bien le critère prescrit.

Résultats de la classification des nappes selon leurs inter-relations.

Nous considérons que les relations simultanées entre les nappes sont les plus importantes et que leur importance diminue avec le temps. Pour cette raison, on a utilisé pour les poids la fonction suivante $w(k) = 0.5^{|k|}$; $k = -1, 0, 1$.

La figure 7 présente, avec mention des distances, l'arbre d'expansion minimale⁽¹⁾ [6] qui, dans la théorie des graphes, correspond de façon bijective à la classification pour la méthode du minimum [3]. En coupant les deux distances les plus grandes, trois groupes sont obtenus. Le premier regroupe les nappes 1, 4, 3, 2 et 9, le "centre" du groupe étant la nappe 3 ; le deuxième est constitué des nappes 6, 7 et 8 ; la nappe 5 constitue à elle seule le troisième groupe.

Le premier groupe est constitué de nappes qui sont liées par des canaux. Le deuxième groupe est constitué de trois nappes situées près du Rhône, les nappes

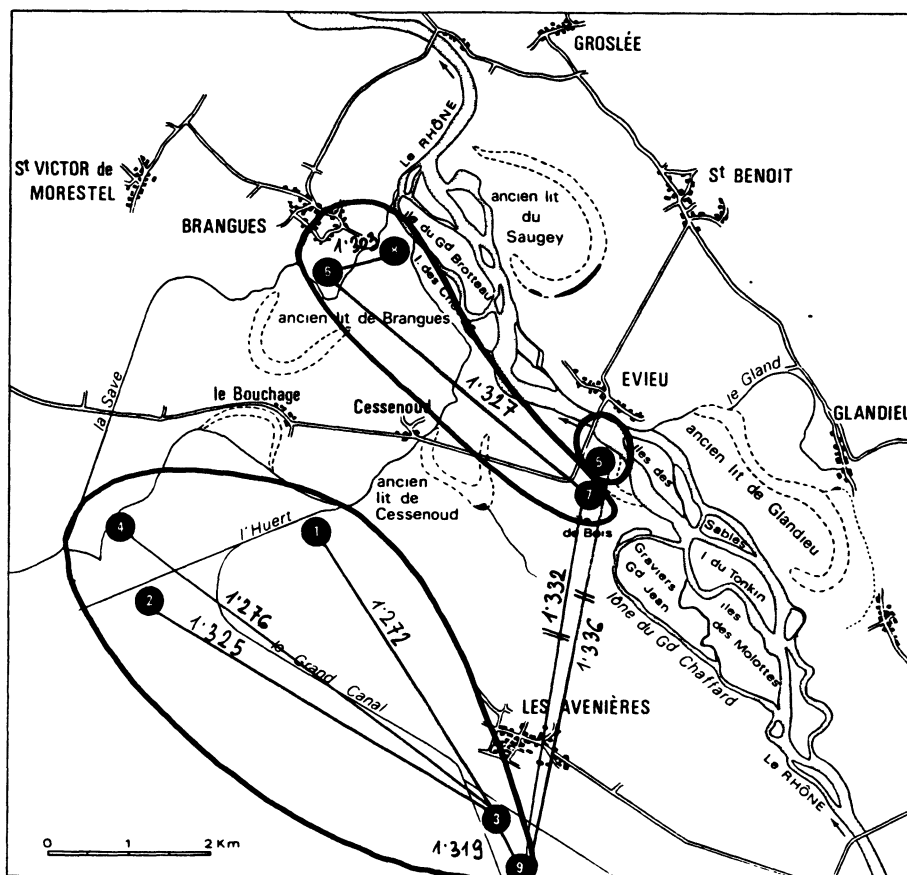


Figure 7. - L'arbre d'expansion minimale : classification des nappes selon leurs relations.

(1) Qui est aussi appelé arbre de longueur minimale.

6 et 8 étant liées par la Save, pourtant des courants d'eau souterrains existent vraisemblablement entre les nappes 6 et 7. La nappe 5 se trouve isolée par rapport aux autres, ce qui suggère l'hypothèse selon laquelle cette nappe est surtout liée au Rhône lui-même.

3.4. Classification des écosystèmes

Rappelons que l'écosystème E_i est défini par trois variables explicatives : le débit, la température et la pluie et une variable expliquée : la nappe. Le but de ce paragraphe est de trouver un indice qui nous permettra de classer ensemble les écosystèmes qui ont des relations semblables entre ces quatre variables.

Les valeurs des variables explicatives ne changeant pas d'un écosystème à l'autre, les relations entre ces trois variables ne nous intéressent pas. Nous supposons que les variables explicatives constituent un espace euclidien à trois dimensions et que les nappes N_i , $i = 1, 9$ y sont représentées par des points N_i dont les coordonnées sont $(r_{DN_i}(0), r_{TN_i}(0), r_{PN_i}(0))$; ces coordonnées dépendent donc des relations entre chaque variable expliquée et les variables explicatives. Comme distance entre deux nappes N_i et N_j , la distance euclidienne peut être considérée. Pour classer les écosystèmes, il nous paraît important de prendre en compte les relations entre les variables explicatives et les variables expliquées ainsi que les relations entre les variables expliquées elles-mêmes. Pour cette raison, nous introduisons au dénominateur de la distance euclidienne le terme $r_{N_i N_j}(0)$.

$$d(E_i, E_j) = + \sqrt{\frac{(r_{DN_i}(0) - r_{DN_j}(0))^2 + (r_{TN_i}(0) - r_{TN_j}(0))^2 + (r_{PN_i}(0) - r_{PN_j}(0))^2}{r_{N_i N_j}^2(0)}} \quad (6)$$

Pour prendre en compte le caractère évolutif des écosystèmes, nous proposons de modifier ainsi la distance (6)

$$d(E_i, E_j) = + \sqrt{\frac{\sum_{k=K_1}^{K_2} (r_{DN_i}(k) - r_{DN_j}(k))^2 + (r_{TN_i}(k) - r_{TN_j}(k))^2 + (r_{PN_i}(k) - r_{PN_j}(k))^2}{\sum_{p=K_3}^{K_4} r_{N_i N_j}^2(p)}} \quad (7)$$

où K_1 , K_2 , K_3 et K_4 sont définis pour chaque cas selon la nature du problème.

Il est facile de vérifier que (7) satisfait les axiomes (3) et que la distance est petite dans le cas d'écosystèmes semblables et grande dans le cas d'écosystèmes différents selon le critère prescrit.

Résultats de la classification des écosystèmes

La formule (6) a été utilisée pour calculer la matrice des distances, les paramètres $k_1 = k_2 = k_3 = k_4 = 1$ ont été pris en compte pour les raisons déjà exposées.

Les trois méthodes d'agrégation donnent les partitions suivantes :

Minimum

$$\{1, 4, 3, 9, 6, 7\}, \{8\}, \{5\}, \{2\}$$

Moyenne

{1, 4, 3, 9}, {6, 7}, {2}, {5, 8}

Maximum

{1, 4, 3, 2}{6, 7, 9}, {5, 8}

La partition obtenue par la méthode de la moyenne est plus intéressante parce qu'elle met en évidence le changement de typologie des écosystèmes au fur et à mesure que l'on s'éloigne du chenal principal du Rhône. Les groupes sont donc caractérisés par des intensités différentes de l'influence du Rhône.

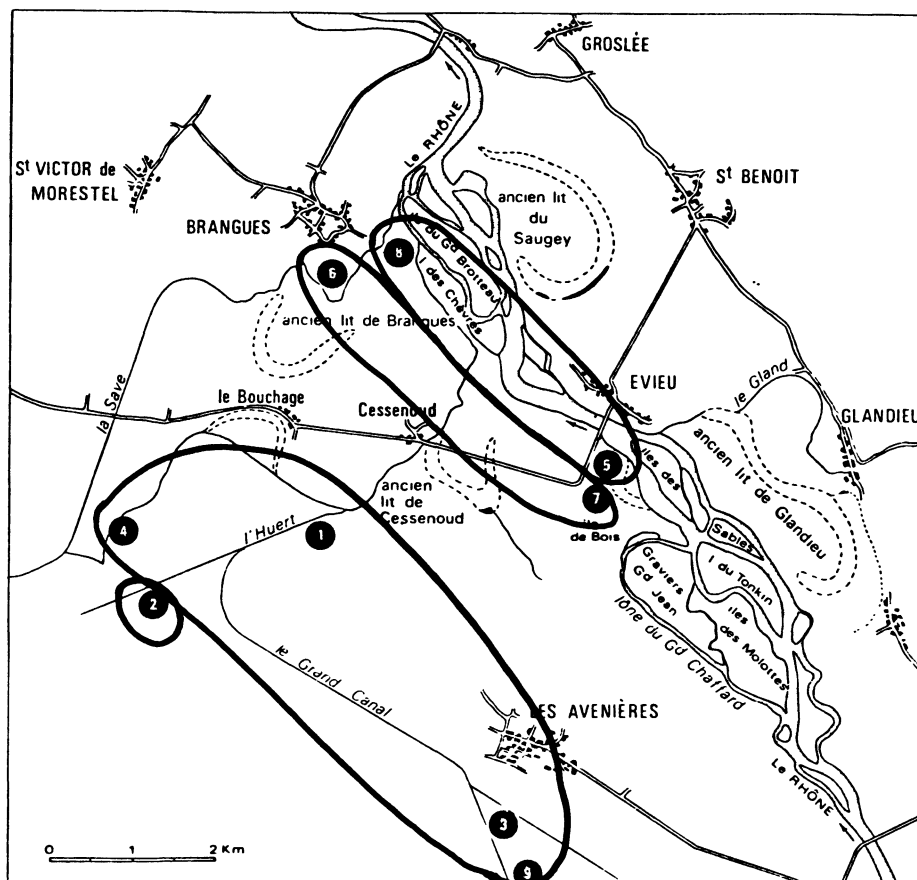


Figure 8. – Classification des écosystèmes : partition obtenue par la méthode de la moyenne.

4. CONCLUSIONS

Le but de notre recherche était de mettre au point une méthodologie statistique qui permette une meilleure compréhension de la dynamique spatio-temporelle de l'eau dans la région du haut Rhône français. Le but a été atteint ; à partir

des données que nous possédons des résultats intéressants ont été obtenus, ce qui peut servir pour une future exploitation à des fins écologiques. Pour approfondir les connaissances sur le fonctionnement de ce biotope, une étude plus globale peut être effectuée à l'aide de cette méthodologie sur des données plus abondantes.

Du point de vue statistique, il existe des sophistications des méthodes que nous avons utilisées, par exemple "analyse d'intervention", modèles de Box-Jenkins à plusieurs entrées : il y aurait lieu d'en faire usage de même que des méthodes de la statistique classique adaptées aux séries chronologiques : analyse spectrale, régression dynamique, méthodes d'analyse multidimensionnelle.

Références

- [1] Z. BOHTE, D. CEPAR, K. KOŠMELJ. — *Clustering of time series*, Compstat 1980, p. 587-593, Edinburg, 1980.
- [2] G.E.P. BOX, G.M. JENKINS. — *Time series analysis, forecasting and control*, Holden Day, San Francisco, 1970.
- [3] J.C. GOWER, G.J.S. ROSS. — Minimum spanning tree and single linkage cluster analysis. *Applied statistics*, 18, p. 54-64, 1969.
- [4] M. JAMBU. — *Classification automatique pour l'analyse des données* ; 1 — méthodes et algorithmes. Dunod ed. p. 145-212, 1978.
- [5] G.M. JENKINS, D.G. WATTS. — *Spectral analysis and its applications*, Holden Day, San Francisco, 1968.
- [6] G.J.S. ROSS. — Minimum spanning tree, algorithm AS 13 (Algol 60). *Applied statistics*, 18, p. 103-104, 1969.
- [7] A. RAMACHANDRA RAO, R.L. KASHYAP. — Causality in hydrologic systems, *IFAC, Water and related land resource systems*, Cleveland, Ohio, 1980.
- [8] P.H.A. SNEATH, R.R. SOKAL. — *Numerical Taxonomy*, W.H. Freeman and Co., San Francisco, 1973.