

**Arbeitsberichte der Schweizerischen Meteorologischen Zentralanstalt  
Rapports de travail de l'Institut Suisse de Météorologie  
Rapporti di lavoro dell'Istituto Svizzero di Meteorologia  
Working Reports of the Swiss Meteorological Institute**

**Zürich**

No 122

L'ANALYSE EN COMPOSANTES PRINCIPALES ET  
SON APPLICATION EN METEOROLOGIE

par

J. Quiby, Zurich

Avril 1984

Statistique  
Météorologie

31  
551.5

Résumé

Pour la réalisation du projet DIAGNO 2, qui consiste à déterminer par des méthodes statistiques la précipitation instantanée (en pratique pendant un intervalle de temps court) aux stations du réseau d'observation automatique ANETZ, nous disposons d'une quantité de prédicteurs telle que nous devons les comprimer en perdant le minimum d'information. Nous avons choisi, pour effectuer cette compression, l'analyse en composantes principales, méthode que nous avons tenté de présenter dans cette note de façon simple, compacte et bien structurée en n'oubliant pas les cas particuliers.

Zusammenfassung

Für die Durchführung des Projektes DIAGNO 2, welches aufgrund statistischer Methoden die momentanen Niederschläge (praktisch während eines kurzen Zeitintervalls) an den Stationen des automatischen Beobachtungsnetzes ANETZ bestimmen soll, verfügen wir über eine so grosse Menge von Prädiktoren, dass sie mit einem minimalen Verlust an Information verdichtet werden muss. Um diese Bedingung zu erfüllen, wurde die Hauptkomponentenanalyse gewählt. Man hat in dieser Arbeit versucht, diese Methode in einer einfachen, kompakten und gut strukturierten Weise darzustellen, ohne die Sonderfälle zu vergessen.

## Riassunto

Per il progetto DIAGNO 2, il cui scopo è di determinare con metodi statistici la precipitazione momentanea (in pratica in un intervallo di tempo molto corto) alle stazioni della rete automatica ANETZ, si hanno a disposizione una grande quantità di predittori. Per la realizzazione del programma è però necessario comprimere il numero dei predittori con la minima perdita possibile di informazioni. A questo scopo è stata scelta l'analisi delle componenti principali. Nel presente scritto si è cercato di illustrarne il metodo in maniera semplice, compatta e ben strutturata ma senza tralasciare i casi particolari.

## Summary.

For the realization of the DIAGNO 2 project, whose aim is the determination by statistical methods of the instantaneous precipitation (in fact for a short period of time) at the stations of the automatic observation network ANETZ, we have at our disposal so large a number of predictors that we have to compress them with a minimum loss of information. We have chosen to this end the principal component analysis. In this note, we have tried to present this method in an easy, compact and well structured manner that also includes the special cases.

## Table des matières

	Page
I. <u>Introduction</u>	
A. L'augmentation permanente du volume des données	3
B. La nécessité de réduire le volume des données	3
C. Raisons de l'étude de la transformation en composantes principales	4
D. Raisons pour la rédaction de ce rapport	4
E. Plan de ce rapport	5
II. <u>Principe et propriétés de la méthode</u>	
A. Le principe	6
B. Les propriétés	7
III. <u>Dérivation mathématique de la méthode</u>	
A. Exposition du principe	9
B. Démonstration des propriétés	15
C. Cas particuliers	21
IV. <u>Une application pratique de la méthode</u>	23
V. <u>Références bibliographiques</u>	25

## Chapitre I

### Introduction

---

#### A. L'augmentation permanente du volume des données

La transformation technologique que subit la météorologie depuis environ deux décades et qui est causée principalement par la révolution électronique et l'utilisation des satellites a déjà eu pour conséquence un accroissement de plusieurs ordres de grandeur du volume des données reçues dans les centres météorologiques. Ce n'est pas seulement le volume de l'information observée qui a augmenté (surtout à cause des stations automatiques et des satellites), mais aussi l'information en prévision fournie par les modèles de simulation numérique.

#### B. La nécessité de réduire le volume des données

Il n'est dès lors pas étonnant que l'on assiste à un regain d'intérêt des milieux météorologiques pour les méthodes mathématiques susceptibles de réduire cette énorme masse de données. Cependant une réduction du volume des données doit être faite à bon escient, c'est-à-dire avec une perte minimum de l'information qu'elles contiennent. Il s'agit de conserver les traits ou modes principaux et de n'éliminer que le bruit.

Il y a en météorologie deux domaines d'application pour les méthodes de réduction ou de compression des données.

#### Compression des données à archiver

Afin de conserver aux banques de données une taille raisonnable, il est nécessaire de réduire le nombre des données. Cette réduction est d'autant plus justifiée que les données sont anciennes, car, dans ce cas, l'on peut se dispenser de conserver des données fortement corrélées à d'autres déjà archivées et l'on peut éliminer le bruit qui n'est plus d'une quelconque utilité.

#### Réduction du nombre des prédictes

Pour la détermination par des méthodes statistiques d'une grandeur météorologique en prévision, l'on dispose aujourd'hui d'un nombre tel de prédictes possibles qu'il est absolument nécessaire de les comprimer pour n'en retenir qu'un petit nombre.

### C. Raisons de l'étude de la transformation en composantes principales

Pour la réalisation du projet DIAGNO 2, nous devons trouver une méthode pour comprimer de l'information. Le projet DIAGNO 2 consiste à prévoir par des méthodes statistiques la précipitation instantanée - en pratique pendant un intervalle de temps court - aux stations du réseau automatique suisse (Réseau "ANETZ") en s'appuyant sur les nombreux champs que les modèles de prévision numérique nous livrent. Comme chacun des paramètres en chacun des points de grilles de ces modèles peut être un prédicteur, l'on en aurait un nombre énorme si on les prenait tous. Il faut donc restreindre très fortement leur nombre tout en gardant le maximum de l'information qu'ils contiennent.

De toutes les méthodes pouvant être utilisées pour comprimer de l'information, la transformation en composantes principales est la plus universellement employée car elle possède des propriétés très puissantes comme nous allons le voir au prochain chapitre.

### D. Raisons pour la rédaction de ce rapport

L'on peut s'étonner de la publication d'un tel rapport, car la transformation en composantes principales est une méthode très classique que l'on trouve plus ou moins décrite dans presque tous les ouvrages de statistique supérieure.

Or, lorsque nous avons étudié cette méthode, nous n'avons pas été enthousiasmés par ce que nous avons trouvé dans les livres. D'abord, la plus grande partie des auteurs recourent pour la démonstration de la méthode aux multiplicateurs de Lagrange. Nous estimons que l'emploi exclusif de l'algèbre linéaire est plus élégant car il donne à la méthode plus d'unité. De plus, à part quelques exceptions, les auteurs n'envisagent pas toutes les possibilités, à savoir ce qu'il advient de la méthode lorsque, par exemple, les valeurs propres ne sont pas uniques. Souvent la présentation du sujet n'est pas compacte, l'auteur s'appuyant sur des résultats démontrés dans les chapitres précédents.

Nous avons donc tenté dans ce rapport de décrire de façon simple, compacte et bien structurée la méthode de réduction en composantes principales.

### E. Plan de ce rapport

Le chapitre II contient la description sans mathématiques de la méthode et de ses propriétés. Les justifications mathématiques sont fournies au chapitre III. Pour la lecture de ce chapitre, de bonnes connaissances en algèbre linéaire sont requises du lecteur. On notera aussi que les démonstrations des propriétés I à V présentent des redondances. Elles ne sont pas involontaires car nous avons voulu, pour garder une structure simple, démontrer chaque propriété séparément. Le chapitre IV décrit une application pratique de la méthode.

## Chapitre II

### Principe et propriétés de la méthode

---

#### A. Le principe

Supposons  $n$  observations, chaque observation consistant en  $p$  paramètres différents. Par exemple, prenons pendant une année les observations d'une station automatique à 12 z. Nous aurons dans ce cas  $n=365$  et  $p=15$  si cette station mesure 15 paramètres différents.

Chacune de ces  $n$  observations peut être représentée par un vecteur dans un espace vectoriel orthonormé à 15 dimensions où chaque axe du référentiel de base correspond à une grandeur météorologique observée (voir Fig. 1).

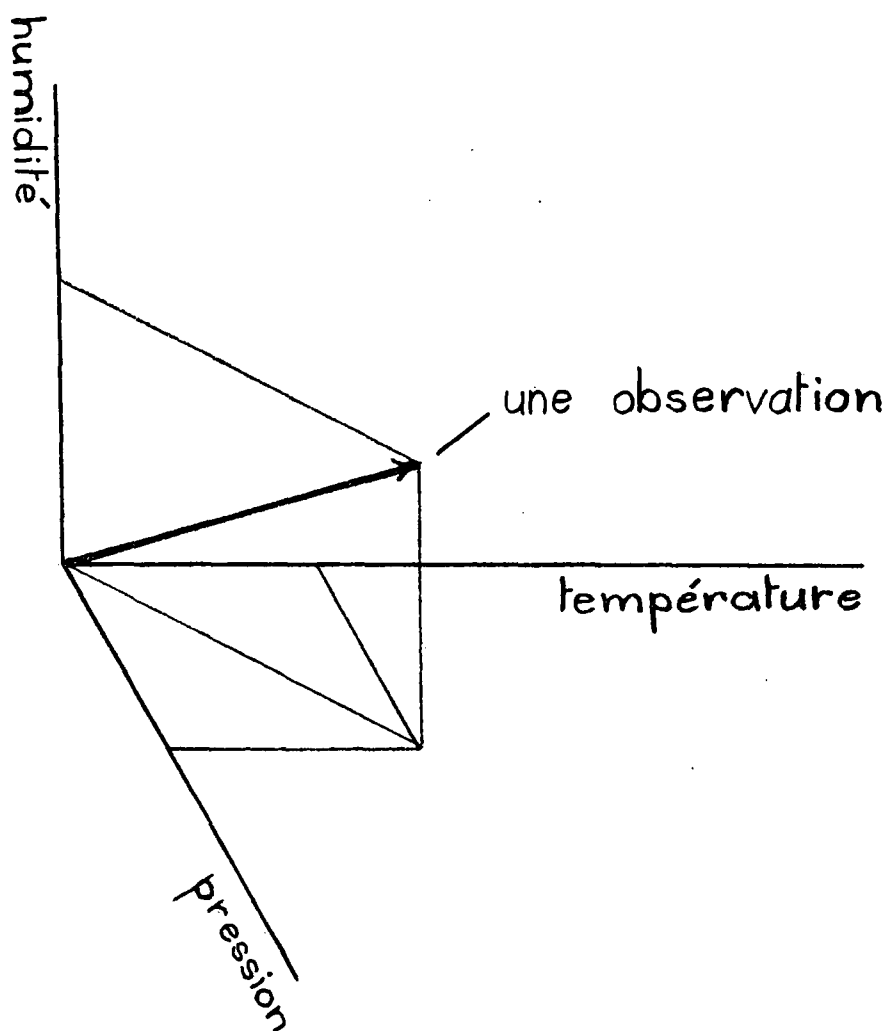


Fig. 1: Représentation des vecteurs observations.



Afin que l'information apportée par les variables météorologiques ne soit pas fonction des unités employées, les  $p$  classes de  $n$  variables auront été au préalable centrées et normées, c'est-à-dire que de chaque variable l'on aura soustrait la moyenne de sa classe et que cette différence aura été divisée par l'écart type de la classe.

La méthode consiste à chercher la droite passant par l'origine et sur laquelle la somme des projections au carré des vecteurs observations centrés et normés sera maximale. Cette droite sera le premier axe d'une nouvelle base orthonormée. Le deuxième axe, qui doit être perpendiculaire au premier, sera celui sur lequel la somme des projections sera la plus grande. Ainsi de suite chaque nouvel axe, qui doit être perpendiculaire aux axes déjà définis, sera celui sur lequel la somme des projections au carré des vecteurs observations centrés et normés aura la plus grande valeur. Les axes de cette nouvelle base s'appellent les axes principaux.

La détermination du premier axe principal nous oblige à considérer la matrice dite de variance-covariance appelée aussi matrice de dispersion des variables. Cette matrice est au coeur de la méthode.

Une fois les axes principaux déterminés, l'on exprime les vecteurs observations dans la base que ces axes définissent. Les composantes  $y$  relatives s'appellent les composantes principales.

Nous avons plus haut utilisé le mot "information". Par information, nous entendons la variance totale de toutes les observations centrées et normées, c'est-à-dire la somme des variances calculées sur chaque axe du référentiel de base.

## B Les propriétés

- la variance d'une composante principale est égale à la valeur propre associée à cette composante.
- deux composantes principales quelconques ne sont pas corrélées.
- la variance totale calculée avec les composantes principales est égale à celle calculée dans le référentiel de base. Elle est égale à la somme des valeurs propres.
- le premier axe principal (l'axe sur lequel la somme des projections au carré des vecteurs observations est maximale) est déterminé

par la direction du vecteur propre associé à la plus grande valeur propre de la matrice de dispersion.

- similairement, la direction du deuxième axe principal (l'axe perpendiculaire au premier et sur lequel la somme des projections au carré des vecteurs observations est maximale) est donnée par la direction du vecteur propre correspondant à la deuxième valeur propre. La relation d'ordre impliquée par le mot "deuxième" est donnée par l'arrangement des valeurs propres en ordre décroissant. Plus généralement, la direction du  $j^{\text{ième}}$  axe principal (l'axe perpendiculaire aux  $j-1$  premiers axes et sur lequel la somme des projections au carré des vecteurs observations est maximale) est donnée par la direction du vecteur propre associé à la  $j^{\text{ième}}$  valeur propre de la matrice de dispersion.

Comme indiqué dans l'introduction, le but de la méthode est de pouvoir fortement réduire le nombre  $n$   $p$  des données de base (dans notre cas  $365 \times 15 = 5475$  données) en perdant le minimum d'information. Vu que la variance totale est égale à la somme des valeurs propres et que la variance d'une composante principale est égale à la valeur propre correspondante, l'on peut fixer le taux de variance total que l'on veut retenir (par exemple 90%) et calculer le nombre de valeurs propres (en commençant par la première et en respectant l'ordre décroissant) pour lequel le taux choisi est satisfait. Supposons que, dans notre cas, le 90% de la variance totale soit obtenu par les trois premières valeurs propres. L'on aura ainsi au minimum le 90% de la variance totale en ne gardant que  $365 \times 3 = 1095$  valeurs au lieu de 5475, soit une réduction de la masse des données de 80%.

### Chapitre III

#### Dérivation mathématique de la méthode

---

##### A. Exposition du principe

Soient  $\underline{x}_i$ ,  $i = 1, 2, \dots, n$   $n$  observations.

Chaque observation comprend  $p$  variables:

$$\underline{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$$

Les  $n$  valeurs correspondant à chaque variable sont centrées et normées:

$$x_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma(x_j)} \quad \forall i, j \quad i = 1, \dots, n \quad j = 1, \dots, p$$

$$\text{où } \bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad \text{et} \quad \sigma(x_j) = \left\{ \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \right\}^{1/2}$$

Nous avons:

$$\bar{x}_j = 0 \quad \text{et} \quad \sigma(x_j) = 1 \quad \forall j, \quad j = 1, 2, \dots, p$$

Preuve:

$$\begin{aligned} \bar{x}_j &= \frac{1}{n} \sum_{i=1}^n \frac{x_{ij} - \bar{x}_j}{\sigma(x_j)} = \frac{1}{n \sigma(x_j)} \sum_{i=1}^n x_{ij} - \frac{n \bar{x}_j}{n \sigma(x_j)} \\ &= \frac{\bar{x}_j}{\sigma(x_j)} - \frac{\bar{x}_j}{\sigma(x_j)} = 0 \end{aligned}$$

$$\begin{aligned} \sigma(x_j) &= \left\{ \frac{1}{n} \sum_{i=1}^n (x_{ij})^2 \right\}^{1/2} \\ &= \left\{ \frac{1}{n} \frac{1}{\sigma^2(x_j)} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \right\}^{1/2} = \frac{1}{\sigma(x_j)} \sigma(x_j) \\ &= 1 \end{aligned}$$

Considérons  $S^2$  la somme des projections au carré des vecteurs  $\tilde{x}_i$  sur la droite définie par  $\underline{u}$  (voir Fig. 2).

$$S^2 = \sum_{i=1}^n (\underline{u} \cdot \tilde{x}_i)(\underline{u} \cdot \tilde{x}_i) \quad (1)$$

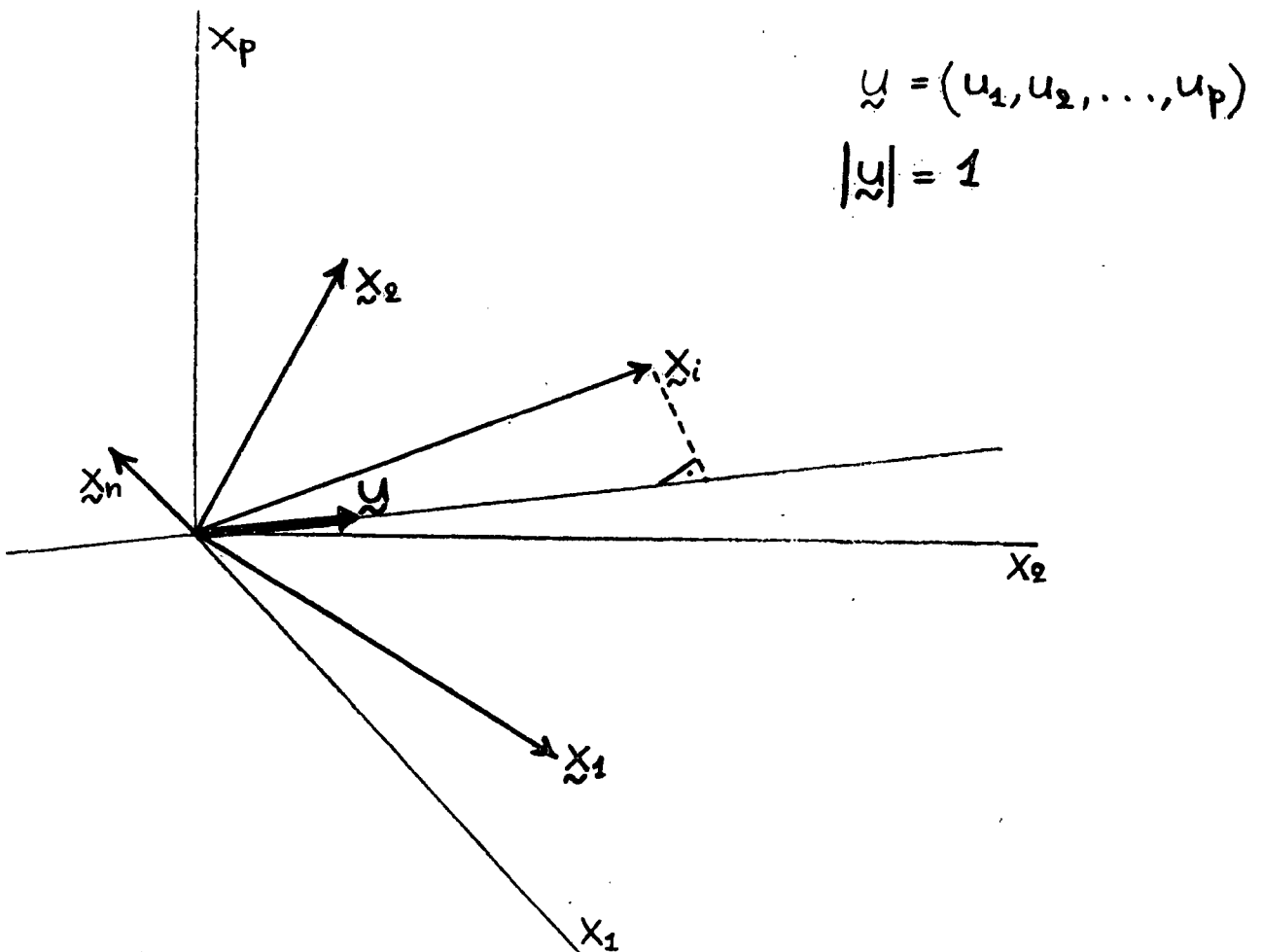


Fig. 2:  $x_1, x_2, \dots, x_p$  le référentiel de base  
 $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n$  les vecteurs observations

$$S^2 = \sum_{i=1}^n \left\{ \left( \sum_{l=1}^p u_l x_{il} \right) \left( \sum_{k=1}^p u_k x_{ik} \right) \right\}$$

Ecrivons

$$\sum_{i=1}^n x_{il} x_{ik} = a_{lk}$$

On remarque que

$$a_{lk} = a_{kl}$$

Il vient

$$S^2 = \sum_{l=1}^p \sum_{k=1}^p u_l u_k a_{lk}$$

que nous préférons écrire

$$S^2 = \sum_{l=1}^p u_l \left( \sum_{k=1}^p a_{lk} u_k \right)$$

ou

$$S^2 = \sum_{k=1}^p \left( \sum_{l=1}^p u_l a_{lk} \right) u_k$$

Ces deux dernières expressions peuvent être représentées sous forme matricielle.

Si l'on pose  $A = (a_{lk})$ , une  $(p,p)$  - matrice,

nous avons respectivement

$$S^2 = \underline{u}^T (A \underline{u})$$

ou

$$S^2 = (\underline{u}^T A) \underline{u}$$

ce que nous écrirons

$$S^2 = \underline{u}^T A \underline{u} \quad (2)$$

$\underline{u}$  est, suivant le contexte, soit un vecteur, soit une  $(p,1)$  - matrice.  $A$  est la matrice de dispersion (appelée aussi matrice des variances-covariances).

On voit que la matrice A a pour diagonale les variances (de façon stricte les variances multipliées par n) des différentes variables et hors de la diagonale les covariances de toutes les variables prises deux à deux (et multipliées par n). Comme les variables sont centrées et normées, les termes de la diagonale ont tous la valeur n.

On remarque aussi que la matrice est symétrique ( $a_{lk} = a_{kl}$ ). Elle est même définie positive si les vecteurs observations ne se trouvent pas tous dans un sous-espace orthogonal à  $\underline{u}$ .

Supposons la matrice A définie positive

Nous avons dans ce cas:

$$\underline{u}^T A \underline{u} = \beta^2 > 0 \quad \forall \underline{u} \neq 0$$

Considérons la relation

$$A \underline{u} = \beta^2 \underline{u} \quad (3)$$

Par définition,  $\beta^2$  est une valeur propre de A et  $\underline{u}$  un vecteur propre associé à cette valeur propre.

Multiplions la relation (3) à gauche par  $\underline{u}^T$ .

$$\underline{u}^T (A \underline{u}) = \underline{u}^T (\beta^2 \underline{u})$$

$$\underline{u}^T A \underline{u} = \beta^2 \quad \text{Rappelons que } |\underline{u}| = 1$$

On retrouve la relation (2).

Soient  $\beta_1^2, \beta_2^2, \beta_3^2, \dots, \beta_p^2$  les p valeurs propres de la matrice A. Comme la matrice A est supposée définie positive (ce qui est toujours le cas lorsque la dimension de l'espace des vecteurs observations est égale à p), toutes ses valeurs propres sont positives.

Supposons toutes les valeurs propres de la matrice A de multiplicité 1

Dans ce cas, les valeurs propres peuvent être strictement ordonnées de façon décroissante

$$S_1^2 > S_2^2 > \dots > S_j^2 > \dots > S_p^2 > 0$$

et la détermination du vecteur propre associé à chacune d'elle

$$\underline{u}_1, \underline{u}_2, \dots, \underline{u}_j, \dots, \underline{u}_p \quad \text{où} \quad |\underline{u}_j| = 1 \quad \forall j, j = 1, \dots, p$$

est univoque (car pour une matrice symétrique, la dimension de l'espace des vecteurs propres associés à une valeur propre est égale à la multiplicité de cette valeur propre).

Nous savons aussi que tous les  $\underline{u}_j$  sont orthogonaux entre eux (propriété des matrices symétriques).

$$\underline{u}_j \cdot \underline{u}_k = \begin{cases} 0 & j \neq k \\ 1 & j = k \end{cases}$$

Les vecteurs  $\underline{u}_j$  forment une base orthonormée.

Détermination des composantes des vecteurs observations dans cette nouvelle base

Ces composantes sont, par définition, les composantes principales. La théorie de l'algèbre linéaire nous dit que la matrice A est diagonalisable (c'est-à-dire qu'elle est semblable à une matrice diagonale) et que l'on a les deux propriétés suivantes:

- Soit D la (p,p)-matrice diagonale. Les éléments de sa diagonale sont les valeurs propres de A.

$$D = D(S_1^2, \dots, S_j^2, \dots, S_p^2)$$

- Soit  $K$  la  $(p,p)$ - matrice auxiliaire. Ses vecteurs colonnes sont les vecteurs propres de la matrice  $A$ .

$$K = (\underline{u}_1 | \underline{u}_2 | \dots | \underline{u}_j | \dots | \underline{u}_p)$$

Nous avons:

$$K^{-1}AK = D \quad (4)$$

par définition des matrices semblables.

Appliquons la matrice  $K$  aux vecteurs<sup>o</sup> observations centrés et normés exprimés dans le référentiel de base. Nous avons que les vecteurs définis par

$$\underline{x}'_i = K \underline{x}_i \quad i = 1, \dots, n$$

ont les mêmes composantes, lorsqu'on les exprime dans le référentiel des vecteurs propres, que les vecteurs observations centrés et normés exprimés dans le référentiel de base.

$$x'_{ij} = x_{ij} \quad \forall i, j \quad i = 1, \dots, n \quad j = 1, \dots, p$$

Mais nous ne voulons pas déplacer les vecteurs observations. Nous voulons seulement les exprimer dans le référentiel des vecteurs propres, c'est-à-dire les exprimer avec leurs composantes principales. Il s'en suit:

$$\underline{x}_i = K^{-1} \underline{x}'_i \quad i = 1, \dots, n$$

Comme la matrice  $K$  est une matrice orthogonale, l'on peut écrire:

$$\underline{x}_i = K^T \underline{x}'_i \quad i = 1, \dots, n \quad (5)$$

La matrice  $K$  est orthogonale car ses vecteurs colonnes sont orthogonaux entre eux et leur norme est égale à 1.



B. Démonstration des propriétés

I. Montrons que la variance (de façon stricte la variance multipliée par n) d'une composante principale est égale à la valeur propre associée à cette composante

La variance de la composante principale k est égale à

$$V_k^2 = \frac{1}{n} \sum_{i=1}^n (\chi_{ik} - \bar{\chi}_{ik})^2$$

Montrons que la moyenne d'une composante principale est égale à zéro:

$$\frac{1}{n} \sum_{i=1}^n \chi_{ij} = \frac{1}{n} \sum_{i=1}^n (u_j \cdot \tilde{x}_i) = \frac{1}{n} u_j \cdot \underbrace{\sum_{i=1}^n \tilde{x}_i}_{\mathbf{0}}$$

ce qui permet d'écrire

$$V_k^2 = \frac{1}{n} \sum_{i=1}^n \chi_{ik}^2$$

Nous avons vu à la relation (5) que

$$\tilde{\chi}_i = K^T \tilde{x}_i$$

ce qui donne

$$\chi_{ij} = u_j \cdot \tilde{x}_i$$

En remplaçant

$$nV_k^2 = \sum_{i=1}^n (u_k \cdot \tilde{x}_i)(u_k \cdot \tilde{x}_i)$$

En suivant la même procédure que celle qui a conduit à la relation (2), l'on peut écrire:

$$nV_k^2 = \underline{u}_k^T A \underline{u}_k$$

où A est la même matrice que dans la relation (2) et où  $\underline{u}_k$  est un vecteur propre.

Il s'en suit que

$$\begin{aligned} nV_k^2 &= \underline{u}_k^T (\underline{B}_k^2 \underline{u}_k) \\ &= \underline{B}_k^2 (\underbrace{\underline{u}_k^T \cdot \underline{u}_k}_1) \end{aligned}$$

II. Montrons que les composantes principales ne sont pas corrélées

Soient  $\chi_{ik}$  et  $\chi_{il}$ ,  $i = 1, 2, \dots, n$

Montrons que  $\text{cov}(\chi_{ik}, \chi_{il}) = 0$  si  $k \neq l$

Nous avons:

$$\text{cov}(\chi_{ik}, \chi_{il}) = \frac{1}{n} \sum_{i=1}^n \chi_{ik} \chi_{il}$$

En vertu de la relation (5):

$$\text{cov}(\chi_{ik}, \chi_{il}) = \frac{1}{n} \sum_{i=1}^n (\underline{u}_k \cdot \underline{x}_i)(\underline{u}_l \cdot \underline{x}_i)$$

En suivant la même procédure que celle qui a conduit à la relation (2), nous obtenons:

$$\text{cov}(\chi_{ik}, \chi_{il}) = \frac{1}{n} (\underline{u}_k^T A \underline{u}_l)$$

où A est la même matrice que dans la relation (2) et où  $\underline{u}_k$  et  $\underline{u}_l$  sont deux vecteurs propres.

Nous pouvons dès lors écrire

$$\begin{aligned} \text{cov}(\chi_{ik}, \chi_{il}) &= \frac{1}{n} \{ \underline{u}_k^T (\underline{B}_l^2 \underline{u}_l) \} \\ &= \frac{1}{n} \underline{B}_l^2 (\underline{u}_k^T \underline{u}_l) = 0 \text{ si } k \neq l \end{aligned}$$

III. Montrons que la variance totale des composantes principales est égale à celle calculée dans le référentiel de base

Variance totale des composantes principales:

$$V_x^2 = \frac{1}{n \cdot p} \sum_{j=1}^p \sum_{i=1}^n \chi_{ij}^2$$

Nous avons que (en vertu de la Propriété I)

$$\frac{1}{n} \sum_{i=1}^n \chi_{ij}^2 = \frac{1}{n} S_j^2$$

Il vient

$$V_x^2 = \frac{1}{p} \sum_{j=1}^p \frac{1}{n} S_j^2 = \frac{1}{n \cdot p} \sum_{j=1}^p S_j^2$$

Variance totale des observations centrées et normées:

$$V_x^2 = \frac{1}{n \cdot p} \sum_{j=1}^p \sum_{i=1}^n X_{ij}^2$$

Nous voulons montrer que

$$\frac{1}{n \cdot p} \sum_{j=1}^p S_j^2 = \frac{1}{n \cdot p} \sum_{j=1}^p \sum_{i=1}^n X_{ij}^2$$

Comme les variables  $X_{ij}$  sont centrées et normées, nous savons que

$$\frac{1}{n \cdot p} \sum_{j=1}^p \sum_{i=1}^n X_{ij}^2 = 1$$

Nous avons, d'après la relation (4), que les matrices A et D sont semblables. Leurs traces sont donc égales.

$$\sum_{j=1}^p S_j^2 = \underbrace{\sum_{j=1}^p \left( \sum_{i=1}^n X_{ij}^2 \right)}_{n \cdot p}$$

Il s'en suit:

$$\frac{1}{n \cdot p} (n \cdot p) = 1$$

IV. Montrons que le maximum possible de variance se trouve sur le vecteur propre correspondant à la plus grande valeur propre

On peut formuler cela différemment:

Montrons que la plus grande valeur propre de A est la plus grande valeur possible pour  $\mathcal{B}^2$  (cf. relation (1)).

Supposons  $\mathcal{B}^2$  maximum sur  $\underline{u}^*$  avec  $|\underline{u}^*| = 1$

$$\exists c_1, \dots, c_p \quad \circ \quad \underline{u}^* = \sum_{j=1}^p c_j \underline{u}_j \quad \text{avec} \quad \sum_{j=1}^p c_j^2 = 1$$

Nous avons à nouveau en vertu de la relation (2):

$$\mathcal{B}^2 = (\underline{u}^*)^T A \underline{u}^*$$

En remplaçant:

$$\mathcal{B}^2 = \left( \sum_{j=1}^p c_j \underline{u}_j \right)^T A \left( \sum_{j=1}^p c_j \underline{u}_j \right)$$

$$\mathcal{B}^2 = \left( \sum_{j=1}^p c_j \underline{u}_j^T \right) A \left( \sum_{j=1}^p c_j \underline{u}_j \right)$$

Considérons

$$A \left( \sum_{j=1}^p c_j \underline{u}_j \right) = \sum_{j=1}^p c_j (A \underline{u}_j) = \sum_{j=1}^p c_j (\mathcal{B}_j^2 \underline{u}_j)$$

Il vient

$$\begin{aligned} \mathcal{B}^2 &= \left( \sum_{j=1}^p c_j \underline{u}_j^T \right) \left( \sum_{\ell=1}^p c_\ell \mathcal{B}_\ell^2 \underline{u}_\ell \right) \\ &= \sum_{j=1}^p c_j^2 \mathcal{B}_j^2 \end{aligned}$$

Supposons au moins deux  $c_j$  non égaux à zéro.

Dans ce cas nous pouvons écrire:

$$\mathfrak{S}^2 < \underbrace{\mathfrak{S}_1^2 \sum_{j=1}^p c_j^2}_1$$

Or  $\mathfrak{S}^2$  doit au moins être égal à la plus grande valeur propre.

Il faut donc supposer qu'un seul  $c_j$  n'est pas égal à zéro.

Sa valeur doit être 1.

Si  $c_j=1$ , nous avons:

$$\mathfrak{S}^2 = \mathfrak{S}_j^2$$

$\mathfrak{S}^2$  doit être égal à une valeur propre.

$\mathfrak{S}^2$  est maximum sur  $U_1$ .

V. Montrons que de tous les vecteurs perpendiculaires aux k premiers axes principaux, c'est sur le (k+1)<sup>ème</sup> axe principal que l'on a la plus grande variance (k = 1, 2, ..., p-1).

Supposons la (k+1)<sup>ème</sup> variance  $\mathfrak{S}_*^2$  maximale sur  $U^*$ .

$$U^* = \sum_{j=1}^p c_j U_j \quad \text{avec} \quad \sum_{j=1}^p c_j^2 = 1$$

Comme  $U^*$  est perpendiculaire aux k premiers axes principaux, nous avons:

$$c_1 = c_2 = \dots = c_k = 0$$

Il s'en suit:

$$\sum_{j=k+1}^p c_j^2 = 1$$

$$U^* = \sum_{j=k+1}^p c_j U_j$$

En vertu de la relation (1), nous pouvons écrire:

$$B_*^2 = (\underline{u}^*)^T A \underline{u}^*$$

En remplaçant:

$$B_*^2 = (\underline{u}^*)^T A \left( \sum_{j=k+1}^p c_j \underline{u}_j \right)$$

Nous savons que les  $\underline{u}_j$  sont des vecteurs propres.

$$B_*^2 = (\underline{u}^*)^T \sum_{j=k+1}^p c_j B_j^2 \underline{u}_j$$

De même pour  $(\underline{u}^*)^T$

$$B_*^2 = \left( \sum_{j=k+1}^p c_j \underline{u}_j^T \right) \left( \sum_{j=k+1}^p c_j B_j^2 \underline{u}_j \right)$$

Nous obtenons:

$$B_*^2 = \sum_{j=k+1}^p c_j^2 B_j^2$$

Supposons au moins 2  $c_j$  différents de zéro. Il vient:

$$B_*^2 < B_{k+1}^2 \sum_{j=k+1}^p c_j^2$$

Or  $B_*^2$  doit au moins être égal à  $B_{k+1}^2$ . Il faut donc supposer qu'un seul  $c_j$  est différent de zéro.

On voit donc qu'au maximum  $B_*^2$  peut avoir  $B_{k+1}^2$  pour valeur.

### C. Cas particuliers

#### a). Supposons la matrice de dispersion définie semi-positive

La matrice A est définie semi-positive lorsqu'il existe au moins un tel qu'il est défini dans la relation (1), perpendiculaire à tous les vecteurs observations ou, formulé plus généralement, lorsque les vecteurs observations forment un sous-espace de dimension inférieure à p. De par sa définition,  $\mathfrak{S}^2$  ne peut pas être négatif mais peut être nul lorsque la condition décrite ci-dessus est remplie.

De par la théorie de l'algèbre linéaire, nous savons que les valeurs propres d'une matrice définie semi-positive sont positives ou nulles. Le fait que, si elle est définie semi-positive, la matrice A n'est plus de rang maximum ne joue dans ce problème pas de rôle: elle reste diagonalisable.

#### Supposons toutes les valeurs propres de multiplicité 1

Comme dans ce cas les valeurs propres peuvent être strictement ordonnées de façon décroissante, seul  $\mathfrak{S}_p^2$  sera égal à zéro car toute matrice définie semi-positive possède au moins une valeur propre nulle.

#### Règle

Si la matrice de dispersion est définie semi-positive mais que toutes les valeurs propres restent de multiplicité égale à un, la théorie que nous avons développée précédemment ne subit aucun changement. En particulier, les cinq propriétés de la transformation en composantes principales que nous avons démontrées restent valables. Si les valeurs propres sont ordonnées en ordre décroissant, la p<sup>ième</sup> composante principale de chaque observation est nulle.

#### b). Supposons que les valeurs propres ne soient pas toutes de multiplicité 1

#### La matrice de dispersion est définie positive ou semi-positive

Supposons que nous ayons comme valeurs propres de la matrice de dispersion

$$\mathfrak{S}_1^2 > \mathfrak{S}_2^2 > \dots > \mathfrak{S}_{j+1}^2 = \mathfrak{S}_{j+2}^2 = \dots = \mathfrak{S}_{j+t}^2 > \dots > \mathfrak{S}_p^2 \geq 0$$

De par la théorie de l'algèbre linéaire, nous pouvons comme précédemment trouver  $p$  vecteurs propres  $\underline{y}$  définissant une base orthonormée, mais nous n'avons plus de relation bijective avec les valeurs propres. En d'autres termes, pour la suite des vecteurs propres donnée plus haut, il existe  $t!$  matrices  $K$  (voir relation (4)) se différenciant par l'ordre des vecteurs colonnes.

Règle

Si, pour une transformation en composantes principales, l'on utilise toujours la même matrice  $K$  et l'on "oublie" les  $(t!-1)$  autres, les propriétés I à V démontrées précédemment restent valables.



## Chapitre IV

### Une application pratique de la méthode

---

Concernant l'application de l'analyse en composantes principales pour le projet DIAGNO 2 décrit au Point C de l'Introduction, le lecteur est prié de se référer au travail de Monsieur Jacques AMBUHL intitulé "Interprétation statistique des prévisions numériques" et publié dans la même série que la présente note. Le travail de M. AMBUHL décrit la stratégie statistique de tout le projet DIAGNO 2.

Nous allons dans ce chapitre nous borner à présenter un petit exemple pratique d'une utilisation possible de l'analyse en composantes principales.

Reprenons l'exemple donné au début du Chapitre II: une station automatique mesure 15 paramètres différents chaque jour à 12 z. Nous aimerions en déduire un 16<sup>ième</sup> paramètre que l'on connaît par un autre moyen. Par exemple la nébulosité observée. La marche à suivre peut se décomposer en cinq pas principaux.

Premier pas: création de l'échantillon de développement.

Comme toujours en statistique, le premier travail sera de créer un échantillon de développement. Pour fixer les idées, supposons que nous ayons rassemblé les données de 12 z du 1er novembre au 28 février pour les années 1971 - 1980. Cela fait 1200 vecteurs observations à 15 composantes chacun.

Deuxième pas: les données sont centrées et normées.

L'on calcule la moyenne et l'écart type (appelé aussi déviation standard) pour chacun des 15 paramètres: pression, température du thermomètre sec, vitesse du vent, etc, et l'on centre et norme chaque mesure avec la moyenne et l'écart type de la classe à laquelle elle appartient.

Troisième pas: calcul de la matrice des vecteurs propres.

Après avoir calculé toutes les variances et covariances pour former la matrice de dispersion (une matrice carrée symétrique de 15 x 15), l'on en détermine les 15 valeurs propres et, pour chacune de ces valeurs propres, un vecteur propre de norme 1. Ces 15 vecteurs propres formeront les vecteurs colonnes de la matrice des vecteurs propres.

Quatrième pas: transformation en composantes principales.

Chacun des 1200 vecteurs observations est multiplié par la transposée de la matrice des vecteurs propres pour être exprimé en composantes principales. Puis on les tronque en fonction du taux de variance total que l'on veut retenir. Supposons que l'on ne garde que les 4 premières composantes. Notre ensemble de 1200 vecteurs à 15 composantes est ainsi réduit à un ensemble de 1200 vecteurs à 4 composantes.

Cinquième pas: détermination de la fonction statistique.

Ces 4 composantes principales sont maintenant les 4 prédicteurs que l'on utilisera pour déterminer la relation avec la nébulosité. Pour l'établissement de cette relation, l'on utilisera une procédure habituelle telle que la régression multiple ou l'analyse discriminante. Supposons maintenant que, au cours de l'été 1980, l'observation de la nébulosité ait cessé d'être disponible (par exemple par suite du retrait de l'observateur). Sa détermination pourra néanmoins se faire tous les jours de la façon suivante: L'observation de 12 z - un vecteur de 15 composantes - est centrée et normée en utilisant les moyennes et les déviations standards de l'échantillon de développement. Ce vecteur est alors multiplié par la transposée de la matrice des vecteurs propres de l'échantillon de développement et le vecteur résultant est tronqué (l'on ne garde que ses 4 premières composantes). Ces 4 prédicteurs sont alors insérés dans la relation statistique choisie et la nébulosité est déterminée.

Adresse de l'auteur:

Jean Quiby  
Institut suisse de météorologie  
Krähbühlstrasse 58  
CH-8044 Zurich

## CHAPITRE V

### Références bibliographiques

Nous allons citer dans ce chapitre un certain nombre d'ouvrages de statistique où l'analyse en composantes principales est traitée.

- ANDERSON T.W.            An Introduction to Multivariate Statistical Analysis  
John Wiley and Sons, Inc., 1958  
Voir Chapitre 11, pp 272-279
- CHAKRAVARTI I.M.,  
LAHA R.G., ROY J.        Handbook of Methods of Applied Statistics  
Volume 1  
John Wiley and Sons, Inc., 1967  
Voir Point 9.6, pp 435-436
- HOPE K.                    Methods of Multivariate Analysis  
Gordon and Breach, Science Publishers, 1969  
Voir Chapitre 4, pp 40-67
- KSHIRSAGAR A.M.         Multivariate Analysis  
Marcel Dekker, Inc., 1972  
Voir Chapitre 11, pp 424-439
- MORRISON D.F.            Multivariate Statistical Methods  
Second Edition, 1976  
McGraw-Hill Book Company  
Voir Chapitre 8, pp 266-289
- MOSTELLER F.,  
TUKEY J.W.                Data Analysis and Regression  
Addison-Wesley Publishing Company, 1977  
Voir Point 15F, pp 397-401
- PRESS S.J.                Applied Multivariate Analysis  
Holt, Rinehart and Winston, Inc., 1972  
Voir Chapitre 9, pp 283-302
- RAO C.R.                  Linear Statistical Inference and Its  
Applications  
John Wiley and Sons, Inc., 1965  
Voir Point 8g.2, pp 501-504
- SEAL H.L.                 Multivariate Statistical Analysis for  
Biologists  
Methuen and Co Ltd, 1964  
Voir Chapitre 6, pp 101-122
- WILKS S.S.                Mathematical Statistics  
John Wiley and Sons, Inc., 1962  
Voir Point 18.7, pp 564-568

