

Principe de l'analyse en composantes principales

L'analyse en composantes principales (ACP) sert à mettre en évidence des similarités ou des oppositions entre variables et à repérer les variables les plus corrélées entre elles (Figure 1).

Ces variables sont par exemple des concentrations hebdomadaires ou bihebdomadaires mesurées pendant plusieurs périodes successives sur un ensemble de tubes. A ces concentrations peuvent s'ajouter des variables auxiliaires connues, migrées ou interpolées en ces mêmes tubes.

Dans l'exemple de la figure 1, les variables introduites dans l'ACP sont quatorze concentrations hebdomadaires d'ozone, la concentration moyenne qui en résulte, les coordonnées de l'espace x et y, l'altitude et la densité de population. Les valeurs de ces variables sont connues en au moins 25 des 32 points d'échantillonnage. En raison de leur hétérogénéité, ces variables sont préalablement centrées et réduites (cette opération est effectué automatiquement par Isatis).

Principe de l'ACP

L'ACP consiste à remplacer une famille de variables par de nouvelles variables de variance maximale, non corrélées deux à deux et qui sont des combinaisons linéaires des variables d'origine. Ces nouvelles variables, appelées *composantes principales*, définissent des plans factoriels qui servent de base à une représentation graphique plane des variables initiales. L'interprétation des résultats se restreint généralement aux deux premiers plans factoriels, sous réserve que ceux-ci expliquent la majeure partie de la variance du nuage des variables initiales.

Interprétation des résultats de l'ACP

En ACP normée, les variables projetées sur chaque plan factoriel se trouvent à l'intérieur d'un cercle de rayon unité. Plus une variable est projetée vers le bord du cercle, mieux elle est représentée. Par ailleurs, deux variables bien représentées et proches l'une de l'autre sont corrélées positivement tandis que deux variables qui s'opposent sont corrélées négativement. Une orthogonalité entre deux variables traduit l'absence de corrélation linéaire.

Notons que l'ACP ne mesure que des liens linéaires entre variables. Avant de conclure sur l'existence ou l'absence de relations entre variables, il est donc utile d'examiner l'allure de leurs nuages de corrélation.

Le diagramme en haut à droite de la figure 1 indique le pourcentage de l'inertie totale du nuage des variables porté par chaque axe factoriel. Ce pourcentage est respectivement de 66%, 10% et 5% pour les trois premiers axes qui sont retenus pour la représentation graphique. La première composante principale sépare les variables de concentration (concentration moyenne, minimale et maximale et concentrations des quatorze semaines) de la densité de population. Compte tenu des différences de concentration entre ville et campagne, cette opposition était prévisible. On observe de plus une corrélation positive entre la concentration d'ozone et l'altitude. Aucune corrélation n'apparaît avec les coordonnées de l'espace.

Les autres composantes sont essentiellement liées à la position géographique des tubes passifs, représentée par les coordonnées X et Y, et à la densité de population. Cela se vérifie aisément en examinant le plan F2-F3 dans lequel toutes les variables de concentration sont projetées vers le centre du cercle.

Notons que le nuage des variables de concentration associées aux quatorze semaines de mesure est dispersé autour de la concentration moyenne. Quelques périodes se démarquent, telle la semaine 8 qui correspond à une chute des concentrations liée à une diminution brutale de température.

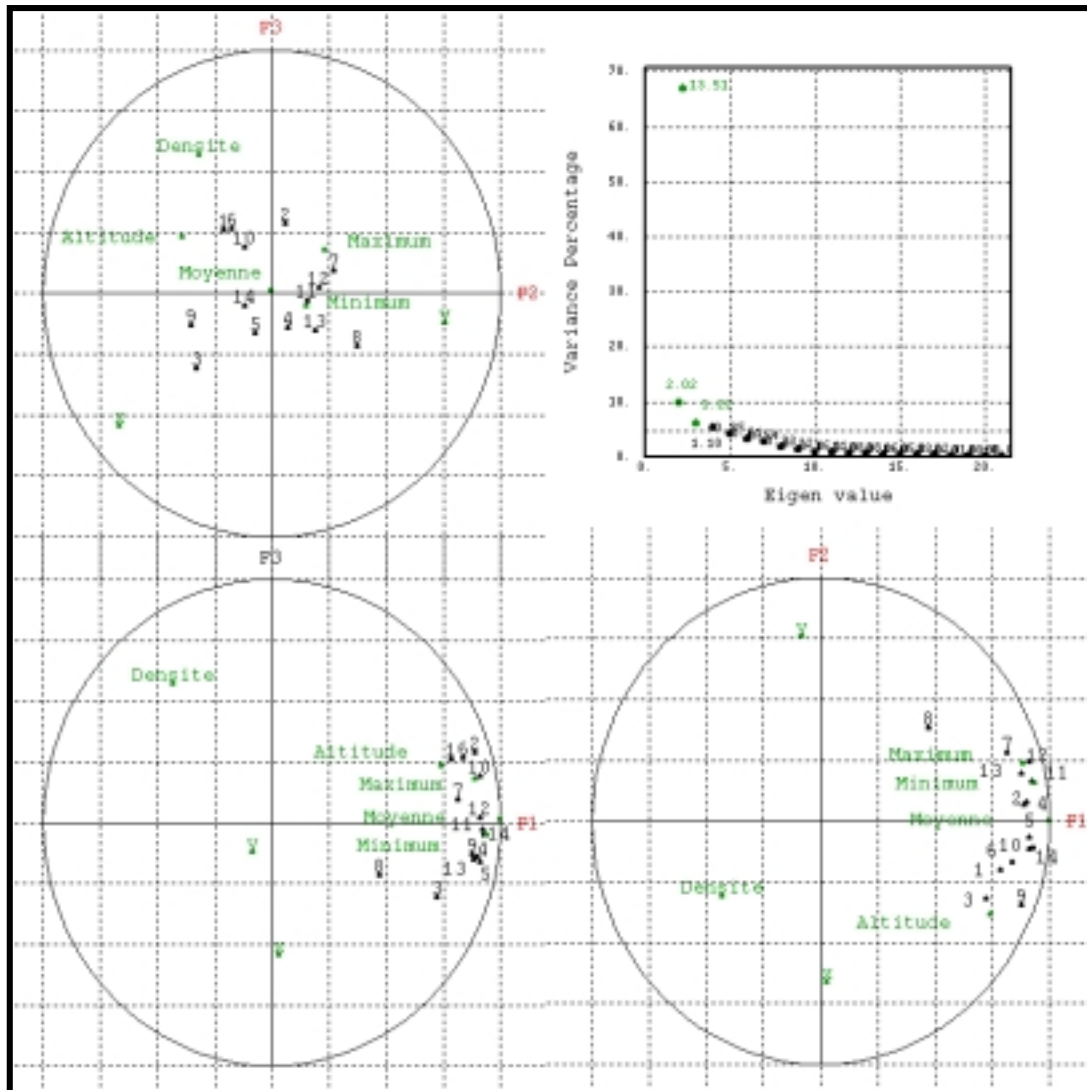


Figure 1 : Exemple d'ACP. Données d'ATMO Auvergne : 14 semaines consécutives de mesure de l'ozone (été 2001) et variables auxiliaires. Projection du nuage des variables sur les deux premiers plans factoriels