

OPTIMISATION DE LA REPRESENTATION DES SERIES CHRONOLOGIQUES PLANCTONIQUES MULTIVARIABLES *

F. IBANEZ

Groupe d'Ecologie Numérique - Station Zoologique 06236 Villefranche-sur-mer.

ABSTRACT

A smoothing technique for chronological series is presented in order to obtain an ordination of the observations by principal component analysis, which takes into account the spatio-temporal contiguity constraint.

Les séries chronologiques planctoniques (ou spatiales en continu) correspondent à des fluctuations cycliques et/ou aléatoires autour d'une tendance générale. L'approximation de la tendance par un modèle linéaire ou polynomial reste arbitraire. De même, le choix de l'intervalle de lissage pour les moyennes mobiles est indéterminé. De plus si les séries sont courtes, on perd un nombre trop important d'observations aux extrémités des séries et si au contraire celles-ci sont longues on introduit des périodicités artificielles avec un lissage très étendu.

Je propose donc ici une technique empruntée à l'économétrie, la méthode des points médians (LIORZOU 1966), qui consiste à définir la tendance comme le lieu des points situés à une distance moyenne de l'enveloppe de la courbe originale (figure 1).

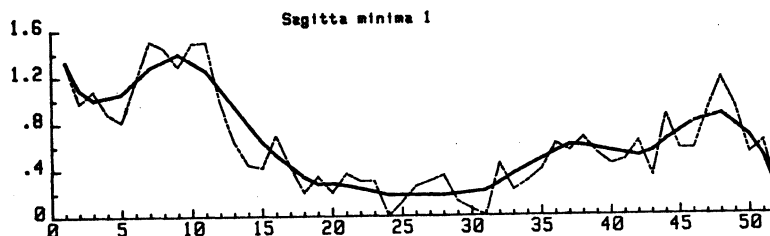


Figure 1 : Cycle annuel du chaetognathe *Sagitta minima* stade 1 dans la rade de Villefranche-sur-mer. Données hebdomadaires. La tendance générale est en trait plein.

Propriétés du lissage

Les phénomènes de basse fréquence sont inchangés par le lissage. La figure 2 montre les cohérences entre les séries lissées et originales (280 prélèvements par pompage sur une radiale en continu), pour trois catégories planctoniques : *Nauplii*, Copépodites, Copépodes adultes. On constate que les cohérences sont toutes significatives pour des fréquences inférieures à 0,2 cycles/minutes : le lissage élimine les oscillations dont la période est inférieure à 5 observations (soit 5 minutes ou 300 mètres parcourus).

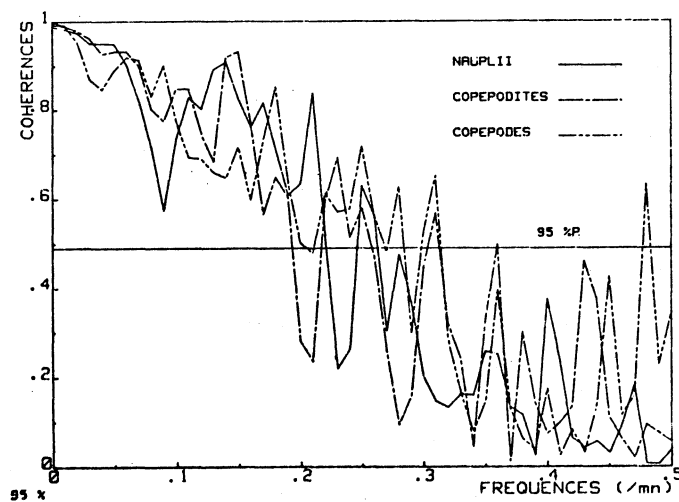


Figure 2 : Fonction de cohérence entre les données lissées et originales pour trois catégories planctoniques.

Le lissage des points médians a les propriétés suivantes :

- conservation intégrale du nombre d'observations
- élimination des phénomènes de haute fréquence supérieurs à 0,2 cycles/unité de temps
- pas d'introduction de périodicités longues artificielles
- identité du processus de lissage pour tous les descripteurs.

Justification de l'analyse en composantes principales sur les données lissées

L'analyse en composantes principales (ACP) ou l'analyse factorielle des correspondances (AFC), ont été déjà utilisées pour décrire les séries chronologiques planctoniques (Binet et al 1972, Cassie 1967, Ibanez et Dallot 1969, Ibanez et Seguin 1972). Théoriquement, ces méthodes d'inertie devraient être employées uniquement à partir de bruits blancs, c'est-à-dire d'observations successives sans corrélation. De ce fait, la représentation de ces observations dans l'espace factoriel réduit ne restitue pas rigoureusement la dépendance spatio-temporelle, et on peut visualiser des groupes constitués d'échantillons fort éloignés dans le temps et/ou l'espace les uns des autres.

L'application de l'ACP sur des données brutes et lissées met en évidence tout d'abord les propriétés numériques suivantes :

- 1 - forte diminution de l'hétérogénéité des variances
- 2 - corrélations amplifiées entre les descripteurs
- 3 - augmentation des pourcentages d'inertie afférents aux premiers axes

Exemple : radiale en continu de la campagne Prétrophos (52 stations, 12 espèces)

	Données originales	Données lissées
Axe 1	28,43	34,43
Axe 2	47,56	56,93
Axe 3	59,63	73,06

Tableau 1 : pourcentages cumulés de variance expliquée par les trois premiers axes.

- 4 - corrélations amplifiées entre les descripteurs et les axes principaux qui n'altèrent pas sensiblement la reconnaissance des groupes.

Mais la particularité la plus remarquable de ce lissage c'est la qualité de l'ordination des observations dans l'espace factoriel réduit. Comme l'autocorrélation entre les observations successives est renforcée par le lissage, on obtient une ordination qui conserve la connexité spatio-temporelle. La succession des stations définit un tracé régulier, assez monotone, marqué par des inflexions relativement peu nombreuses. La figure 3 met en évidence la clarté de cette représentation par rapport à celle que l'on peut obtenir sur des données originales (après transformation logarithmique).

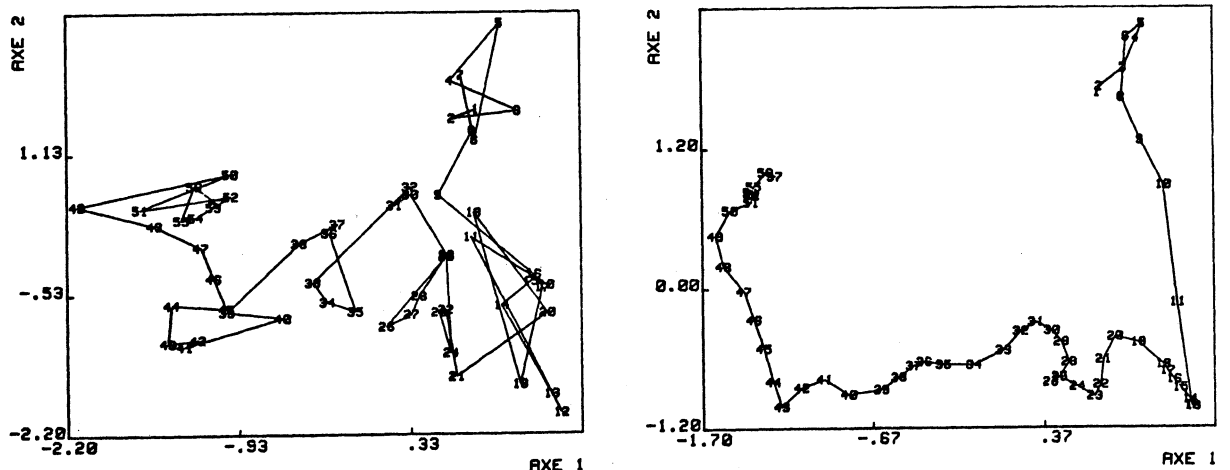


Figure 3 : Ordination des observations par l'ACP dans le premier plan propre. Données originales à gauche, lissées à droite.

Les points d'inflexion de la courbe continue qui ressortent dans l'ordination sur les tendances, permettent un découpage de la série chronologique multivariable. Ils signalent les changements de gradients de population dont l'identification peut se faire en examinant les corrélations entre les axes et les descripteurs.

Le plus souvent, seules les trois premières composantes sont conservées et pour trouver les points d'inflexion, on peut effectuer une représentation triangulaire, chaque côté du triangle correspondant à une composante. Les coordonnées des observations sur chaque côté seront égales aux contributions relatives (en pourcentage) de chaque axe vis-à-vis de chaque observation. La figure 4 donne les représentations triangulaires de l'exemple précédent, à partir d'une ACP sur les tendances ou les données originales.

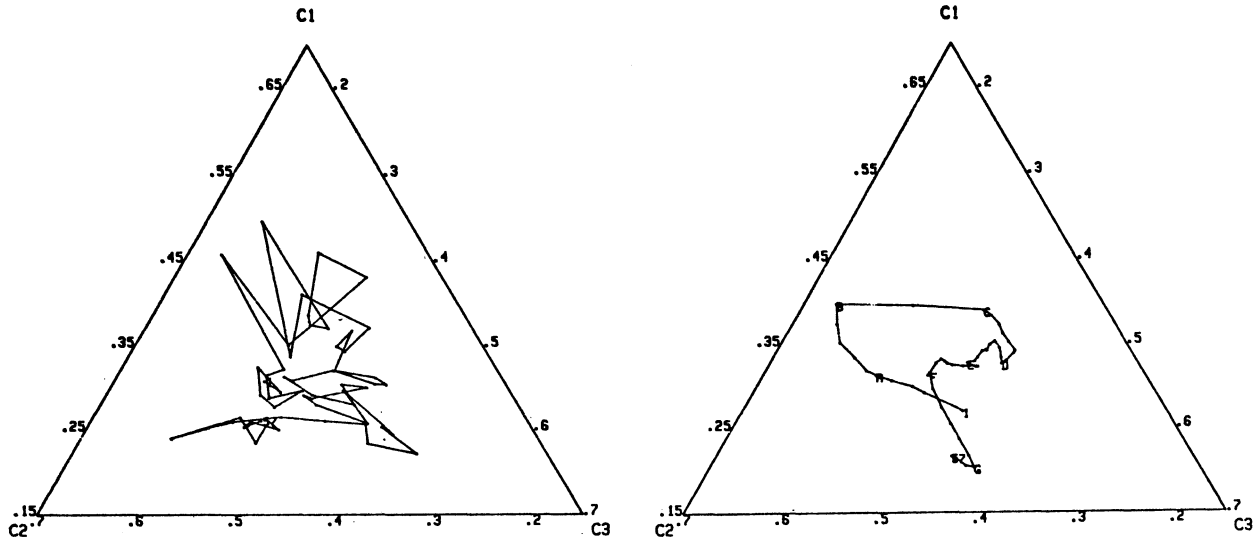


Figure 4 : Ordination en coordonnées triangulaires (trois premiers axes principaux). Données originales à gauche, lissées à droite.

Conclusion

Le principe du lissage des points médians est simple et son application est aisément programmable. Quelle que soit l'hétérogénéité des séries il donne une courbe figurant l'évolution temporelle générale d'une série même si celle-ci est très brève. La description par l'ACP de plusieurs séries devient très efficace car les variances sont très homogènes et l'ordination des observations respecte la connexité spatio-temporelle. La relative stabilité des variances permet d'utiliser les données originales si une transformation des données n'est pas justifiée par une hypothèse écologique. La visualisation des observations dans l'espace factoriel peut servir à découper les séries de départ en périodes où les gradients de population gardent chacun le même sens.

Notons également qu'après avoir soustrait la tendance des données originales, on obtient des séries résiduelles qui peuvent être analysées par les techniques propres aux processus aléatoires.

CASSIE R.M. 1967. Principal component analysis of the zooplankton of lake Maggiore. Mem. Inst. Ital. Idrobiol., 21 : 129 - 144.

DESSIER A. et A. LAUREC. 1978. Le cycle annuel du zooplancton à Pointe-Noire (RP Congo). Description mathématique. Oceanol. Acta 1 (3) : 285 - 304.

IBANEZ F. et S. DALLOT. 1969. Etude du cycle annuel du zooplancton d'Abidjan. Comparaison de plusieurs méthodes d'analyses multivariées. Inv. Pesq. 36 (1) : 81 - 108.

LIZORZOU A. 1966. Initiation pratique à la statistique. Eyrolles Ed. PARIS. 310.

* Ce travail a été réalisé dans le cadre de l'ATP 9-82-65.

