

Méthodes de couplage de deux K-tableaux et collections de graphiques

J. Thioulouse^a, A. Siberchicot^a, A. Julien-Laferrière^a, A.B. Dufour^a, S. Dray^a

^aUMR CNRS 5558 - LBBE "Biométrie et Biologie évolutive"
UCB Lyon 1 - Bât. Grégor Mendel
43 bd du 11 novembre 1918
69622 VILLEURBANNE cedex
jean.thioulouse@univ-lyon1.fr

Mots clefs : Analyse de données, Écologie, ade4.

Le nouveau package `adegraphics` [1] est une ré-écriture des fonctions graphiques du package d'analyse de données multivariées `ade4`, utilisant des classes `S4`, et basée sur le package `lattice` [2]. Parmi les nombreux avantages de ce nouveau package [3], figure en particulier la possibilité, grâce à `lattice`, de gérer des collections de graphiques, par juxtaposition, insertion et superposition. Cette possibilité s'inspire directement des anciennes versions du logiciel **ADE4**, mais elle avait été laissée de côté lors de la ré-écriture sous forme de package **R**.

D'autre part, les méthodes d'analyse de données dites "multitableaux", et en particulier les méthodes de couplage de deux K-tableaux, conduisent naturellement à produire des collections de graphiques. On peut ainsi collectionner les graphiques des différents tableaux, et des différentes variables ou des différents individus d'un ou de plusieurs tableaux. On peut également collectionner les biplots en superposant les individus et les variables de chaque tableau. Dans le cas des méthodes de couplage de deux K-tableaux, on peut aussi vouloir juxtaposer et/ou superposer les graphiques correspondant aux deux tableaux de chaque paire.

Dans le domaine de l'analyse de données écologiques, les méthodes de couplage de deux K-tableaux sont particulièrement intéressantes car elles permettent par exemple d'analyser la stabilité des relations espèces-environnement. Dans ce cas, un des K-tableaux est constitué des mesures de variables environnementales répétées au cours du temps, et le second par des mesures d'abondance floreo-faunistique, également répétées.

Après un bref rappel sur l'utilisation des méthodes K-tableaux et 2K-tableaux dans le package `ade4`, nous présentons plusieurs exemples de mise en oeuvre des fonctions du package `adegraphics` pour réaliser des collections de graphiques coordonnées et adaptées aux objectifs des méthodes d'analyse de K couples de tableaux (méthodes 2K-tableaux). Trois méthodes sont plus précisément détaillées: `BGCOIA`, `STATICO`, et `COSTATIS`. Elles sont toutes les trois disponibles dans le package `ade4`, et un article récent ([4]) propose une comparaison des trois, ainsi qu'une mise en oeuvre interactive par le biais d'un site Web permettant de reproduire les calculs et les représentations graphiques à l'aide des anciennes fonctions graphiques du package `ade4`: <http://pbil.univ-lyon1.fr/SAOASOPET/>.

- La `BGCOIA` (Between Group Co-Inertia Analysis, [5]) est une analyse de coinertie inter-groupes. Chaque tableau est un groupe, les moyennes des variables par tableau sont calculées et arrangées en deux tableaux qui sont ensuite soumis à une analyse de coinertie.
- `STATICO` (`STATIS` et Coinertie, [6]) est une Analyse Triadique Partielle portant sur la série de tableaux de covariance croisées de chaque couple de tableau.
- `COSTATIS` (Coinertie et `STATIS`, [4]) est une analyse de coinertie des deux compromis obtenus par l'Analyse Triadique Partielle de chaque K-tableau pris séparément.

Les trois étapes des méthodes K-tableaux (interstructure, compromis, intrastructure) fournissent des coordonnées factorielles qui peuvent être utilisées pour faire diverses représentations graphiques. Dans le cas de l'intrastructure, on projette en général les lignes et les colonnes des tableaux initiaux dans l'analyse du compromis, ce qui fournit de nombreuses autres possibilités de représentations graphiques.

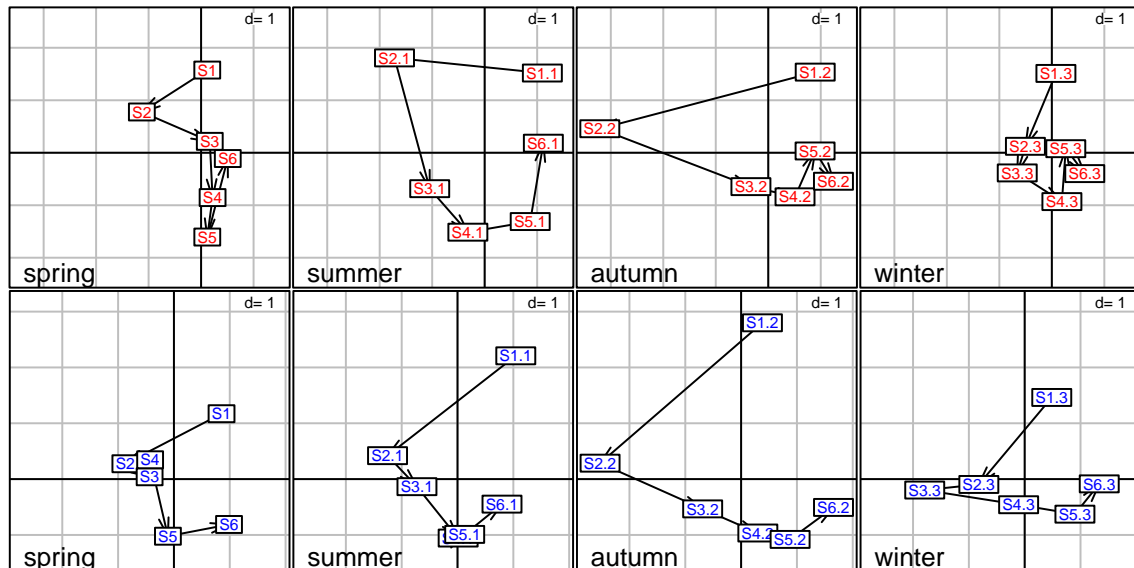


Figure 1: Intrastructure des stations pour les variables environnementales (en haut) et pour les espèces (en bas). Méthode STATICO, jeu de données `meau`, package `ade4`.

Par exemple, dans le cas de STATICO (Figure 1), on peut représenter la collection de graphiques obtenue par juxtaposition des graphiques des projections des individus par tableau dans l'analyse du compromis. L'utilisation du package `adegraphics` facilite alors grandement la réalisation de ces graphiques, en automatisant le processus de collection des graphiques par tableau, et en autorisant une grande souplesse dans leur positionnement et l'ajustement des paramètres secondaires (couleurs, type de points, labels, etc).

Références

- [1] Julien-Laferrière, A., and Dray, S. (2012). Visualisation de données multivariées: réimplémentation des fonctionnalités graphiques de la librairie `ade4`. In *Premières Rencontres R*, Bordeaux, France.
- [2] Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*. Springer, New York.
- [3] Siberchicot, A., Julien-Laferrière, A., Thioulouse, J., and Dray, S. (2013) `adegraphics` : un package pour la représentation et l'analyse de données multivariées. In *Deuxièmes Rencontres R*, Lyon, France.
- [4] Thioulouse, J. (2011). Simultaneous analysis of a sequence of paired ecological tables: a comparison of several methods. *Annals of Applied Statistics*, **5**, 2300-2325.
- [5] Franquet, E., Doledec S., and Chessel D. (1995) Using multivariate analyses for separating spatial and temporal effects within species-environment relationships. *Hydrobiologia*, **300**, 425-431.
- [6] Thioulouse J., Simier M. and Chessel D. (2004). Simultaneous analysis of a sequence of paired ecological tables. *Ecology*, **85**, 272-283.