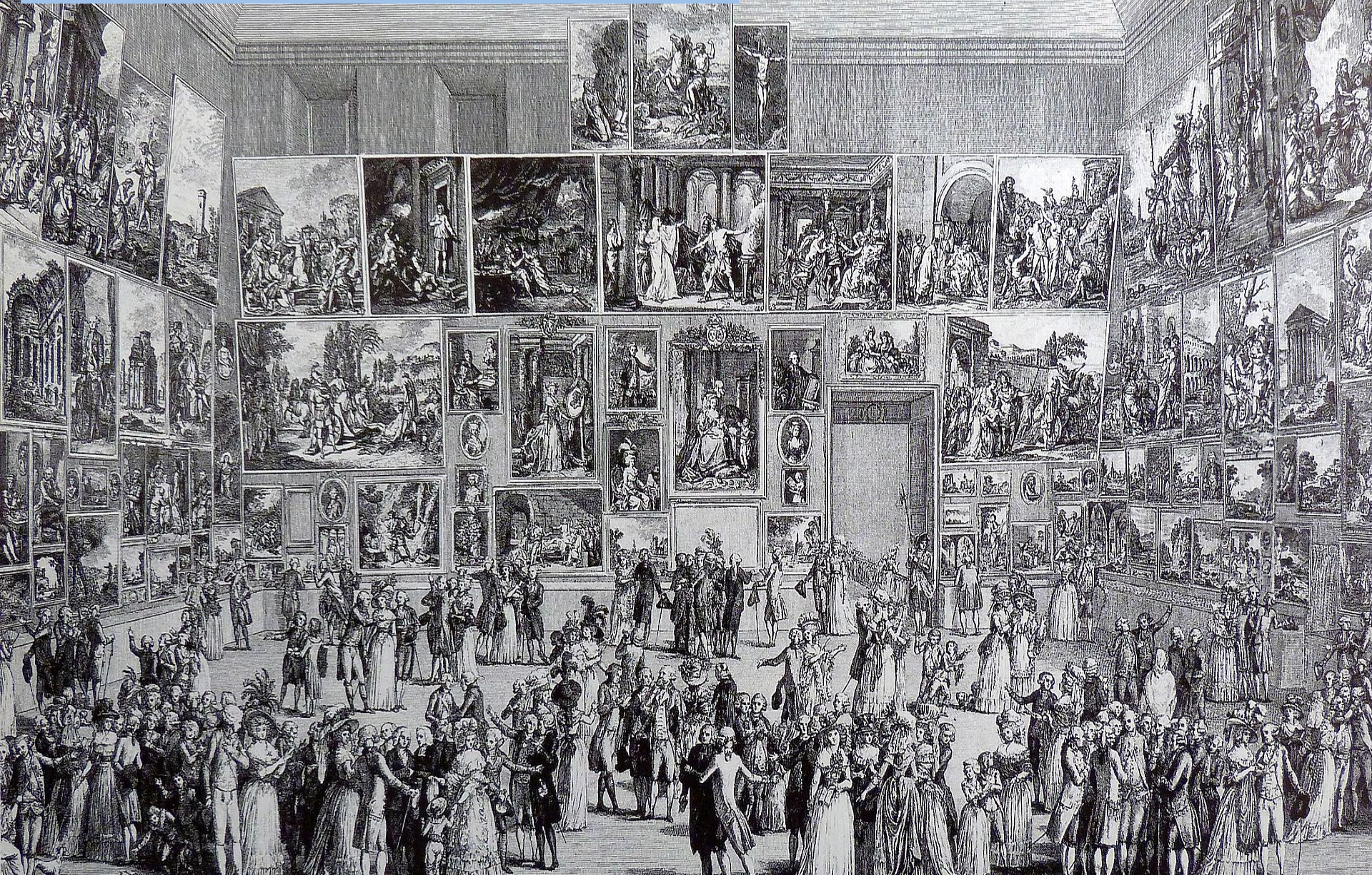


*Ordinations multi-tableaux dans
Emmanuel Castella, Institut Foré C.I.S.E*



avertissement

ce programme contient des scènes
susceptibles de choquer certaines
sensibilités mathématiques. c'est
pourquoi il est accompagné de ce sigle:



LIII. On Lines and Planes of Closest Fit to Systems of Points in Space. By KARL PEARSON, F.R.S., University College, London*.

(1) IN many physical, statistical, and biological investigations it is desirable to represent a system of points in plane, three, or higher dimensioned space by the "best-fitting" straight line or plane. Analytically this consists in taking

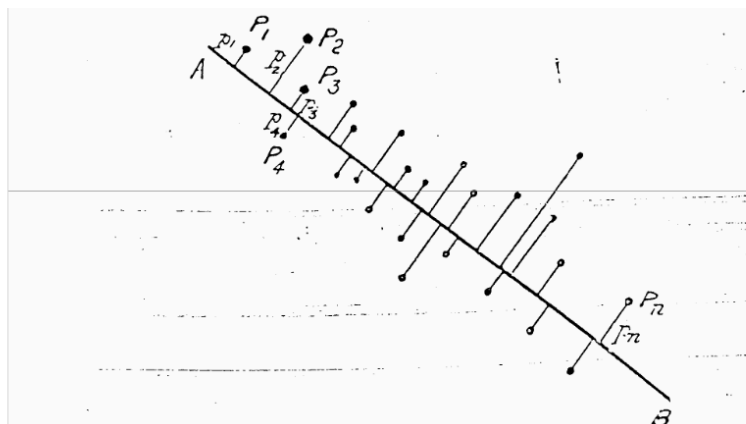
$$y = a_0 + a_1x, \text{ or } z = a_0 + a_1x + b_1y, \\ \text{or } z = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \dots + a_nx_n,$$

where $y, z, x_1, x_2, \dots, x_n$ are variables, and determining the "best" values for the constants $a_0, a_1, b_1, a_2, a_3, \dots, a_n$ in relation to the observed corresponding values of the variables. In nearly all the cases dealt with in the text-books of least squares, the variables on the right of our equations are treated as the independent, those on the left as the dependent variables. The result of this treatment is that we get one straight line or plane if we treat some one variable as independent, and a quite different one if we treat another variable as the independent variable. There is no paradox about this; it is, in fact, an easily understood and most important feature of the theory of a system of correlated variables. The most probable value of y for a given value of x , say, is not given by the same relation as the most probable value of x for a given value of y . Or, to take a concrete example, the most probable stature of a man with a given length of leg l being s , the most probable length of leg for a man of stature s will not be l . The "best-fitting" lines and planes for the cases of z up to n variables for a correlated system are given in my memoir on regression†. They depend upon a determination of the means, standard-deviations, and correlation-coefficients of the system. In such cases the values of the independent variables are supposed to be accurately known, and the probable value of the dependent variable is ascertained.

(2) In many cases of physics and biology, however, the "independent" variable is subject to just as much deviation or error as the "dependent" variable. We do not, for example, know x accurately and then proceed to find y , but both x and y are found by experiment or observation. We observe x and y and seek for a unique functional relation between them. Men of given stature may have a variety

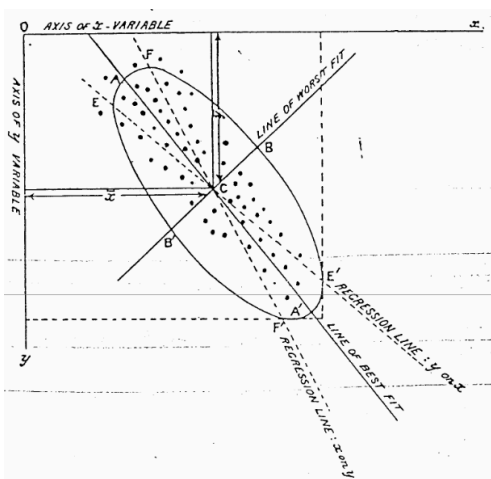
* Communicated by the Author.

† Phil. Trans. vol. clxxxvii. A, pp. 301 et seq.



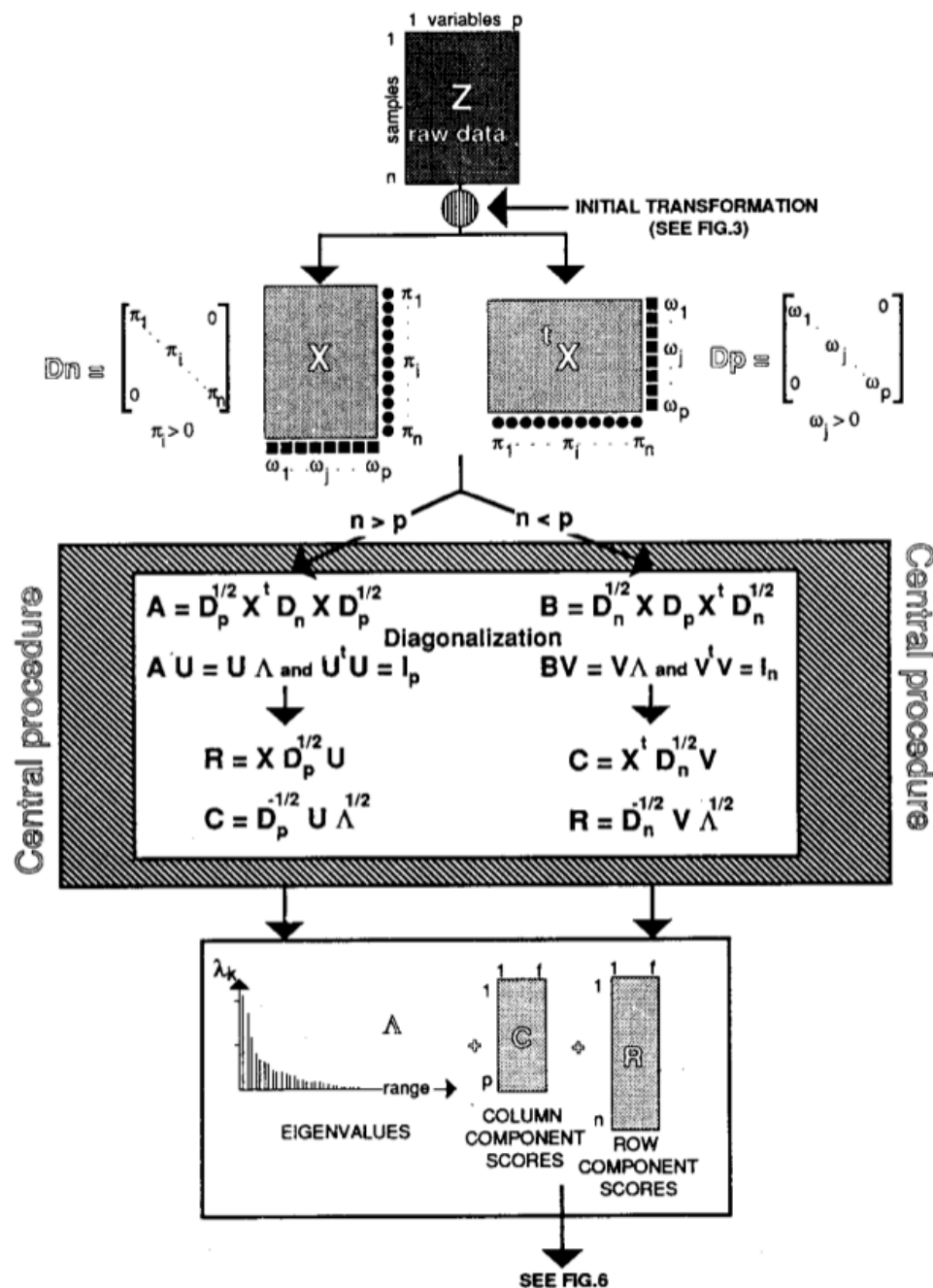
Now clearly $U = S(p^2)$ is the moment of momentum, the second moment of the system of points, supposed equally loaded, about the line AB. But the second moment of a system about a series of parallel lines is always least for the

Pearson, K. 1901. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* 2:559-572. <http://pbi.univ-lyon1.fr/R/pearson1901.pdf>



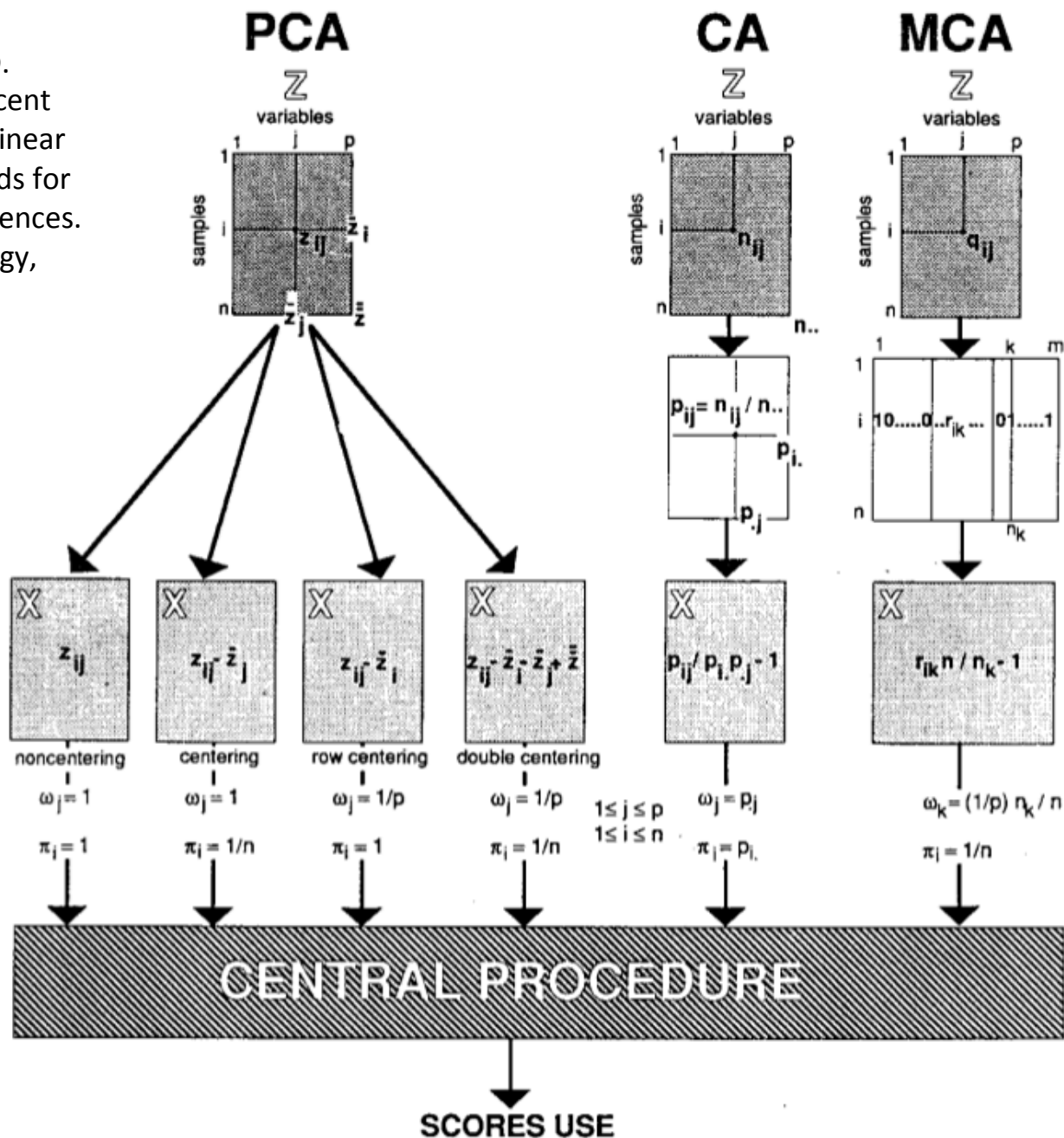
Physically the axes of the correlation type-ellipse are the directions of independent or uncorrelated variation. Hence the line of best fit is a direction of uncorrelated variation.





Dolédec, S., and D. Chessel. 1991. Recent developments in linear ordination methods for environmental sciences. *Advances in Ecology, India* 1:133-155.

Dolédec, S., and D. Chessel. 1991. Recent developments in linear ordination methods for environmental sciences. *Advances in Ecology, India* 1:133-155.

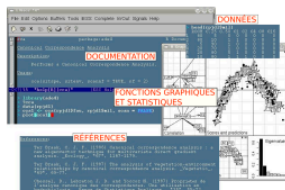


Accueil ade4

ade4 est un logiciel développé au laboratoire de Biométrie et Biologie Évolutive (UMR 5558) de l'Université Lyon 1. Il contient des fonctions d'Analyse de Données destinée d'abord à la manipulation des données Écologiques et Environnementales avec des procédures Exploratoires d'essence Euclidienne, d'où la dénomination **ade4**.

ade4 se caractérise par :

- l'implémentation de **fonctions statistiques et graphiques**
- la mise à disposition de **données numériques**
- la rédaction d'une **documentation** technique et thématique
- l'inclusion de **références bibliographiques**



fonctions, données, documentation et références dans ade4 (OS : Debian avec Emacs et ESS)

Depuis 2002, **ade4** est un package du logiciel R, disponible sur les systèmes Windows, Mac OS, Linux et Unix.

Les méthodes dédiées à l'analyse d'un tableau sont décrites dans Chessel et al. (2004) ([pdf](#), [html](#)).

Les méthodes pour l'analyse de 2 et K tableaux sont décrites dans Dray et al. (2007) ([pdf](#), [html](#)).

La théorie du schéma de dualité et son implémentation dans **ade4** sont discutés dans Dray et Dufour (2007) ([pdf](#)).

D'autres packages liés à **ade4** sont décrits sur la page de [téléchargements](#), en particulier le [package ade4TkGUI](#) qui propose une interface utilisateur graphique pour **ade4**.

L'ancienne version du logiciel (ADE-4, en majuscules) présenté dans Thioulouse et al. (1997) ([pdf](#), [html](#)) est disponible [ici](#) mais n'est plus maintenue depuis 2004.

Les documents du stage de formation ade4 de Sept 2008 à Lyon sont disponibles [ici](#)

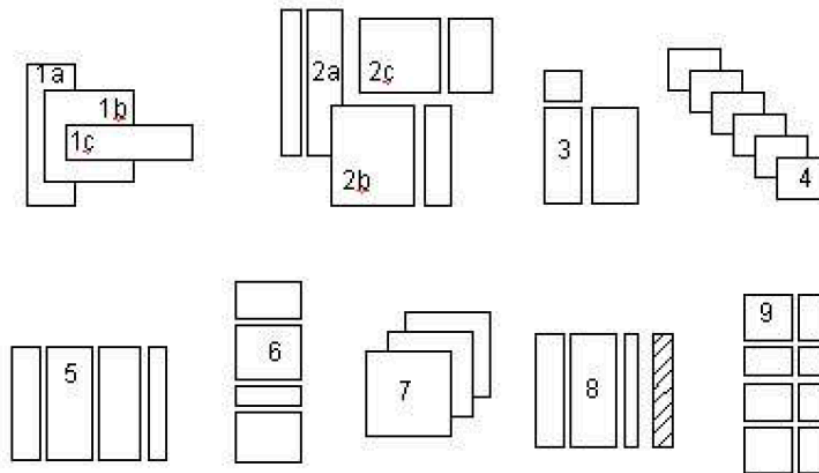
Une liste des articles citant ade4/ADE-4 est disponible [ici](#).

<http://pbil.univ-lyon1.fr/ade4/>

Des données originales

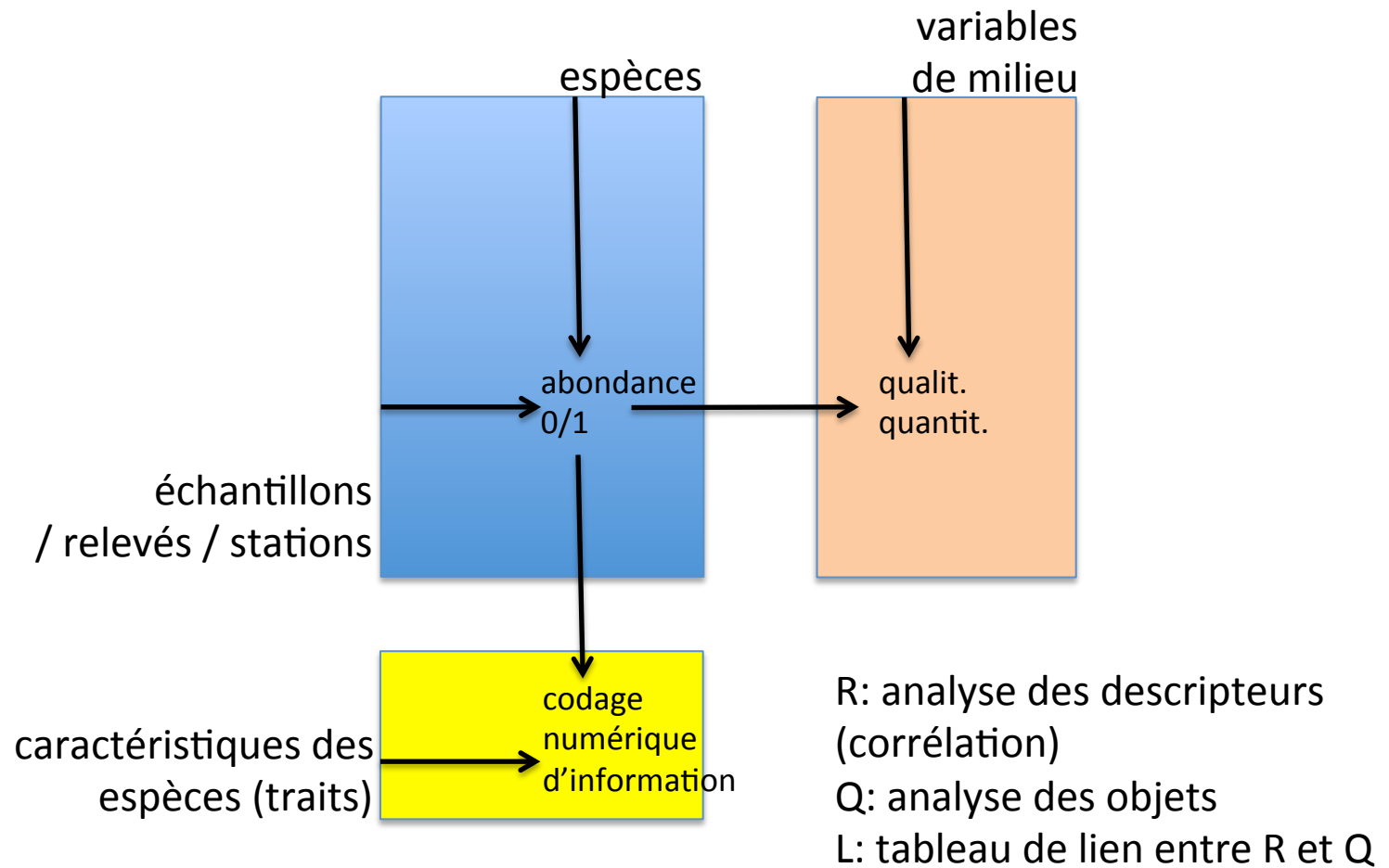
Ce qui retiendra le statisticien en écologie des communautés, c'est l'extraordinaire diversité de structures de données, de conditions numériques et de questions plus ou moins clairement formulées susceptibles d'orienter les analyses. L'acquisition d'observations basées sur la notion de *relevés* conduit naturellement à des questions portant sur :

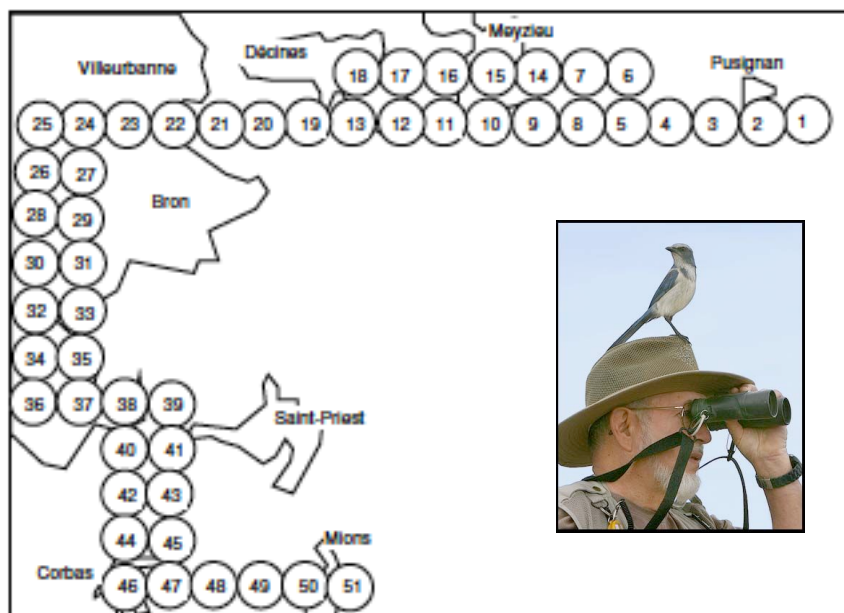
- 1 - les analyses à un tableau relevés-espèces et les méthodes d'ordination
- 2 - les couples de tableaux dans des conditions d'analyse canonique (a), de variables instrumentales (b), de co-inertie (c)
- 3 - les analyses à trois tableaux (biologique-écologique-environnemental)
- 4 - les cubes de données (stations-espèces-dates)
- 5 - les K-tableaux par blocs de variables (groupes faunistiques)
- 6 - les K-tableaux par blocs de stations (expertises régionales)
- 7 - les K-matrices de distances (spatiale, génétique, écologique, environnementale)
- 8 - les K+1-tableaux (groupes taxonomiques et variables de milieu)
- 9 - les K-couples de tableaux (faune-milieu par dates, régions, ...)



Structures de données fréquentes en écologie des communautés.

D. Chessel, 2000



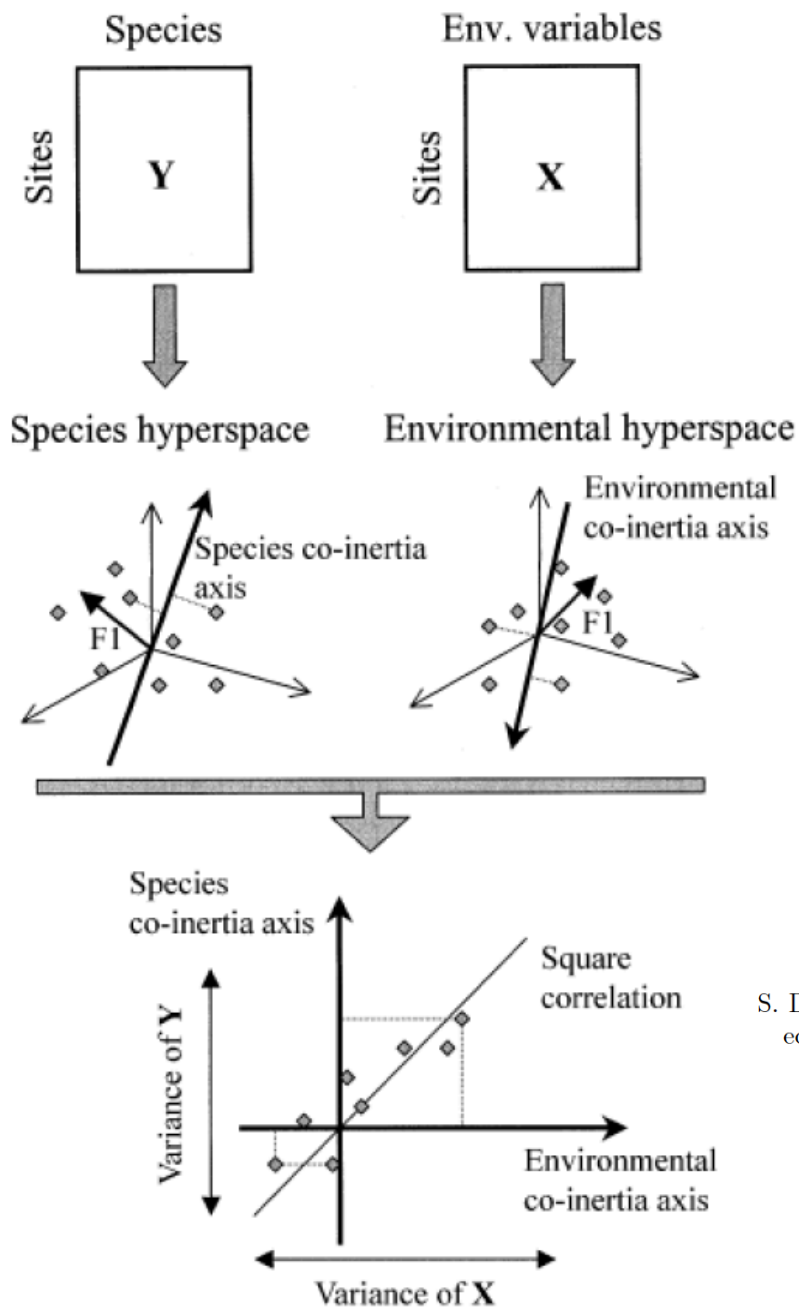


- 1) 51 secteurs d'écoute x 40 sp d'oiseaux (notes d'abondance 0 - 4)
- 2) 51 secteurs d'écoute x 11 variables décrivant l'environnement (catégories)
- 3) 40 sp d'oiseaux x 3 variables biologiques (traits; catégories)

Y-a-t'il dans la variation des traits biologiques des espèces des correspondances avec la variation des caractéristiques de l'environnement ?

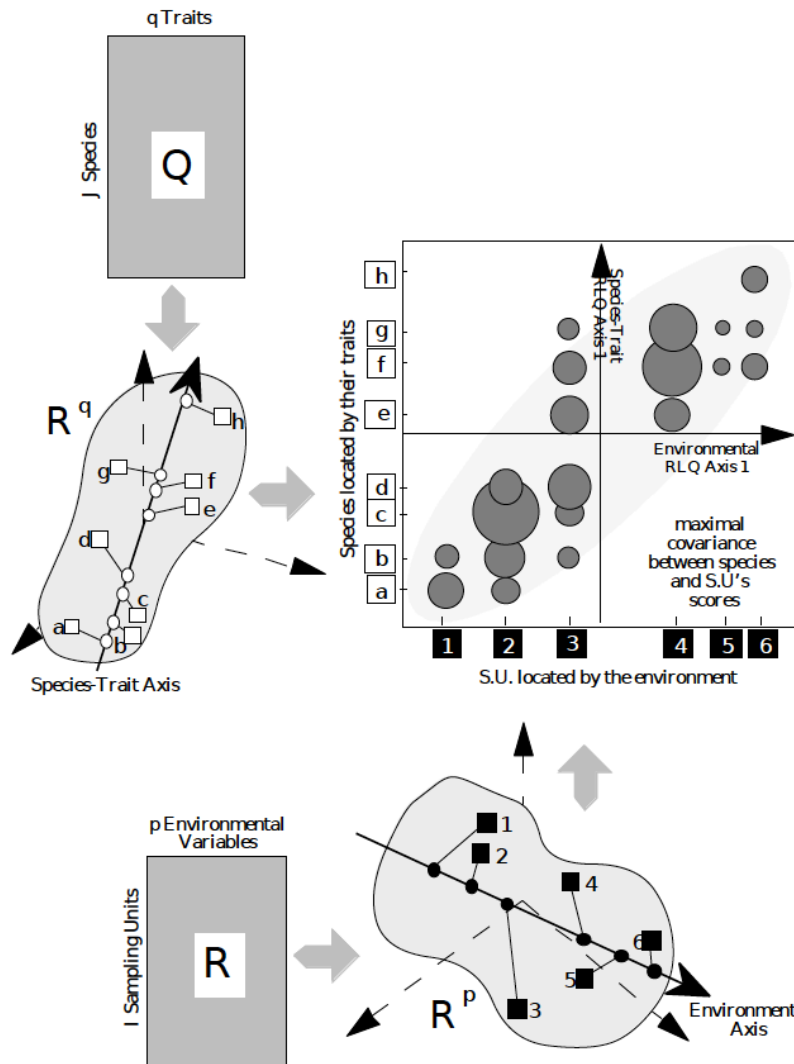
Tatibouet F. 1981. Approche écologique d'un établissement humain (environnement et structure). Exemple de la communauté urbaine de Lyon. Thèse . Université Lyon1.

S. Dolédec, D. Chessel, C.J.F. Ter Braak, and S. Champely. 1996. Matching species traits to environmental variables : a new three-table ordination method. *Environmental and Ecological Statistics* , 3 : 143-166.



Analyse de Co-inertie

S. Dray, D. Chessel, and J. Thioulouse. Co-inertia analysis and the linking of ecological tables. *Ecology*, 84(11) :3078–3089, 2003.



Analyse RLQ