



HAL
open science

Application de l'ACP et de la classification par intervalles aux estimations de quantiles des distributions empiriques du coût des engrais pour les cultures annuelles dans les pays de l'UE

Dominique Desbois

► **To cite this version:**

Dominique Desbois. Application de l'ACP et de la classification par intervalles aux estimations de quantiles des distributions empiriques du coût des engrais pour les cultures annuelles dans les pays de l'UE. 14ièmes Journées de Recherches en Sciences sociales, SFER, Apr 2021, Clermont-Ferrand, France. hal-03193404

HAL Id: hal-03193404

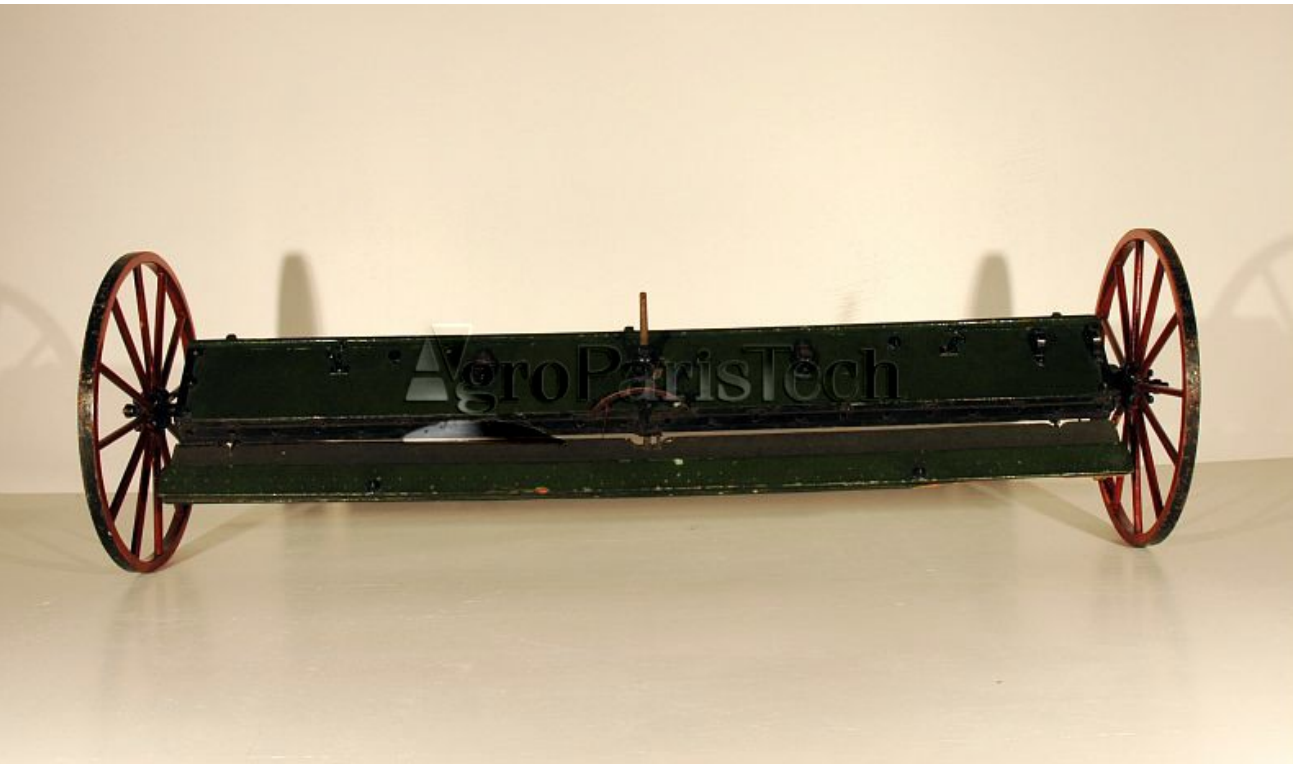
<https://hal.inrae.fr/hal-03193404>

Submitted on 8 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright



**I Estimations des quantiles
conditionnels des coûts de fertilisation**

**II ACP symboliques des intervalles
d'estimation des distributions de coûts**

**III Classification symbolique des
intervalles d'estimation des distributions de
coûts**

Distributeur d'engrais (maquette, vers 1850), Musée du vivant

Dominique DESBOIS, UMR Economie publique, INRAE-AgroParisTech

Journées de Recherches en Sciences sociales, 7 avril 2021

Gestion de l'exploitation agricole et pratiques agro-écologiques

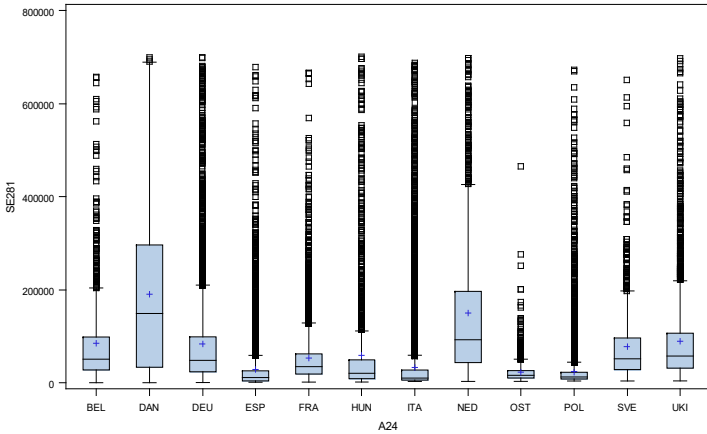
I.1 Estimation du coût des engrais : le modèle input-output

Modélisation économétrique des coûts de production agricoles: *le modèle à coefficients constants*

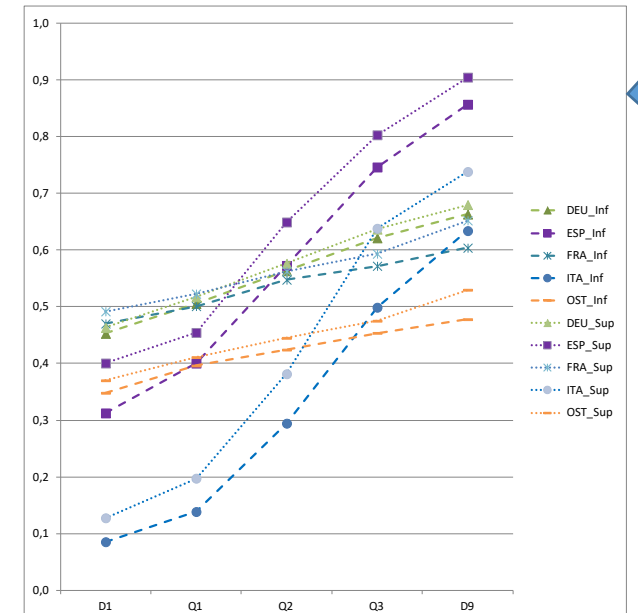
$$X_{ih} = \sum_{k=1}^K \alpha_{ih}^k Y_{kh} + \varepsilon_{ih} \text{ with } \varepsilon_{ih} \text{ i.i.d.}$$

CHARGES	PRODUITS					TOTAL CHARGE
	Y_{1h}	...	Y_{kh}	...	Y_{Kh}	
X_{1h}	a_{1h}^1	...	a_{1h}^k	...	a_{1h}^K	$\sum X_{1h}$
\vdots	\vdots		\vdots		\vdots	\vdots
X_{ih}	a_{ih}^1	...	a_{ih}^k	...	a_{ih}^K	$\sum X_{ih}$
\vdots	\vdots		\vdots		\vdots	\vdots
X_{Ih}	a_{Ih}^1	...	a_{Ih}^k	...	a_{Ih}^K	$\sum X_{Ih}$
TOTAL PRODUIT	$\sum Y_{1h}$...	$\sum Y_{kh}$...	$\sum Y_{Kh}$	$\sum_k Y_{kh} = \sum_i X_{ih}$

1.2 Estimation quantile des coûts de production agricoles



$$\hat{\beta}_\omega(q) = \underset{\beta \in \mathcal{R}^{P+1}}{\operatorname{argmin}} \left\{ \sum_{i \in \{i/x_i \geq [\frac{\partial_i}{Y_i}]' \beta\}} w_i q \left| x_i - [\frac{\partial_i}{Y_i}]' \beta \right| + \sum_{i \in \{i/x_i < [\frac{\partial_i}{Y_i}]' \beta\}} w_i (1 - q) \left| x_i - [\frac{\partial_i}{Y_i}]' \beta \right| \right\}$$



id	D1I	D1S	Q1I	Q1S	Q2I	Q2S	Q3I	Q3S	D9I	D9S
Bel	0,009	0,019	0,023	0,030	0,038	0,047	0,056	0,080	0,082	0,110
Dan	0,018	0,024	0,035	0,035	0,056	0,056	0,094	0,094	0,140	0,140
Deu	0,004	0,009	0,025	0,033	0,082	0,082	0,140	0,140	0,181	0,181
Esp	0,013	0,017	0,025	0,033	0,058	0,058	0,103	0,103	0,169	0,169
Fra	0,023	0,028	0,053	0,065	0,125	0,125	0,182	0,182	0,232	0,232
Hun	0,020	0,038	0,056	0,071	0,093	0,110	0,138	0,164	0,197	0,197
Ita	0,007	0,011	0,019	0,022	0,041	0,041	0,078	0,078	0,121	0,121
Ned	0,001	0,004	0,004	0,006	0,009	0,012	0,017	0,022	0,026	0,029
Ost	0,000	0,029	0,043	0,057	0,068	0,086	0,106	0,127	0,155	0,179
Pol	0,024	0,032	0,052	0,059	0,088	0,099	0,146	0,165	0,215	0,228
Sve	-0,007	0,016	0,003	0,038	0,100	0,100	0,215	0,215	0,293	0,293
Uki	0,006	0,029	0,036	0,047	0,088	0,088	0,137	0,137	0,171	0,171

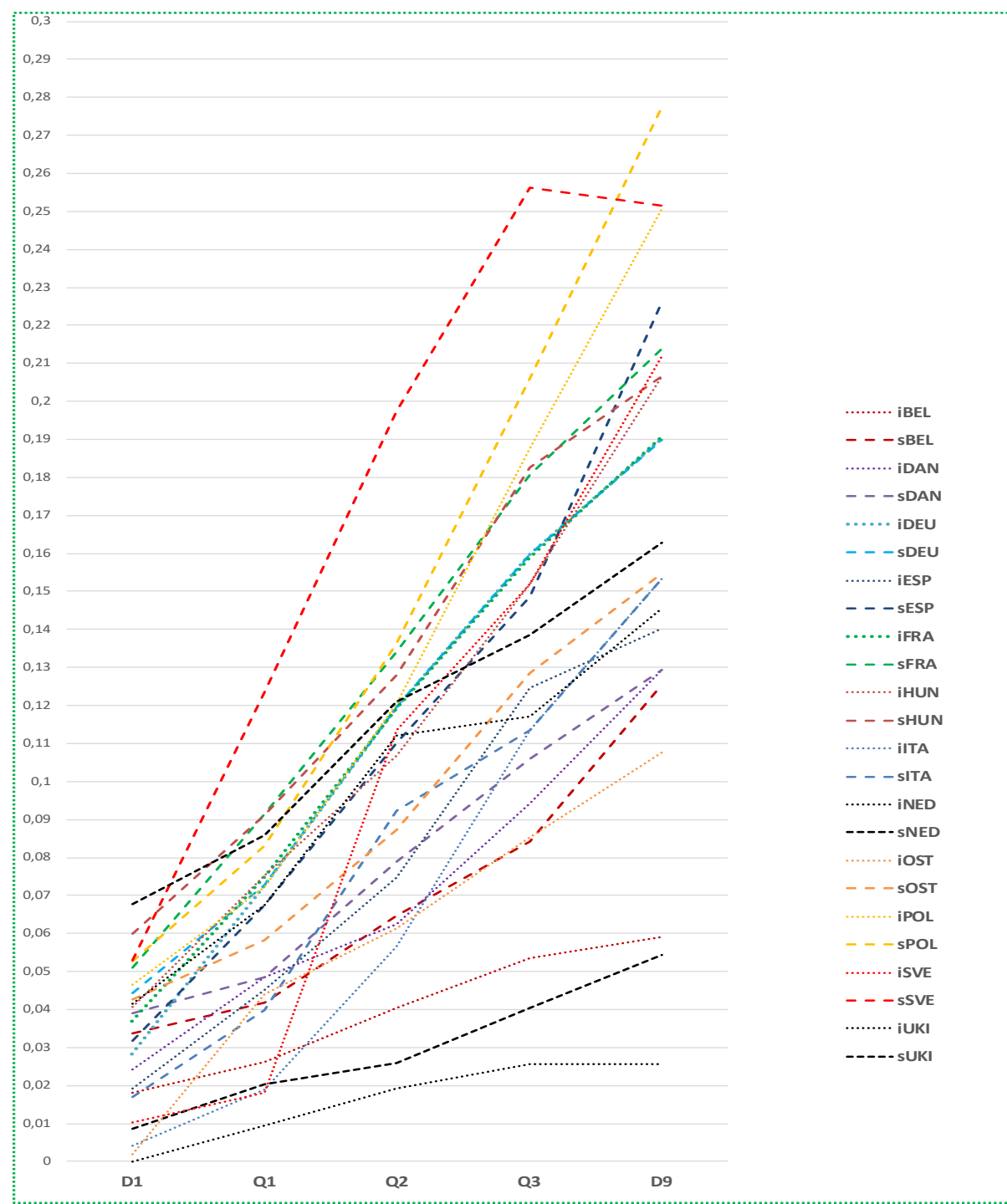
(D.Desbois , FADN-UE 2006)

Belgique:

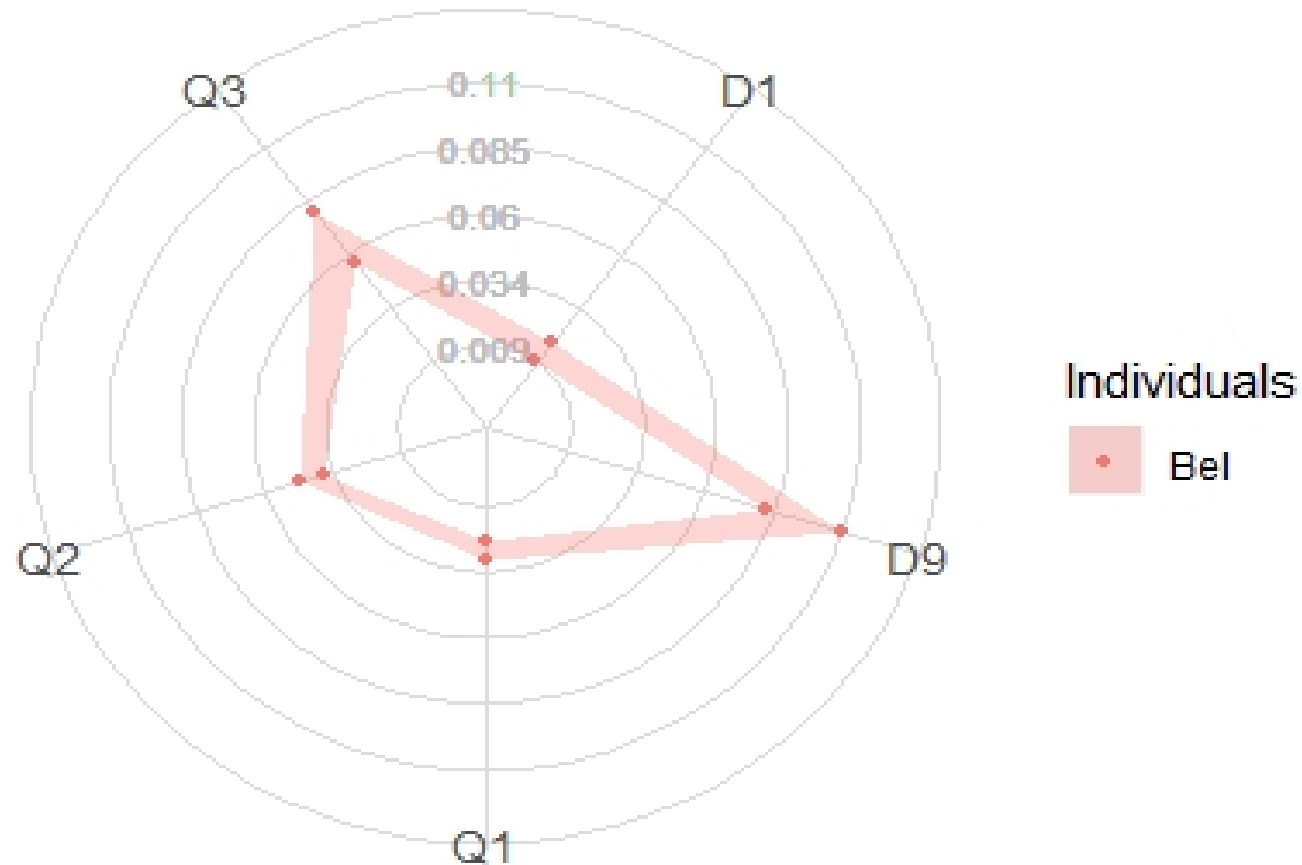
entre 9 to 19 € d'engrais sur 1 000 € de produit brut en grandes cultures pour le premier décile de dépenses en engrais

```
syecrop2<-read.sym.table("~/FERTI/syecrop2b.txt",header=TRUE,sep='\t',dec=',',row.names=1)
print.data.frame(syecrop2)
```

I.4 Diagramme des quantiles conditionnels des distributions de coûts de fertilisation en grandes cultures, EU12



Fertilization Cost

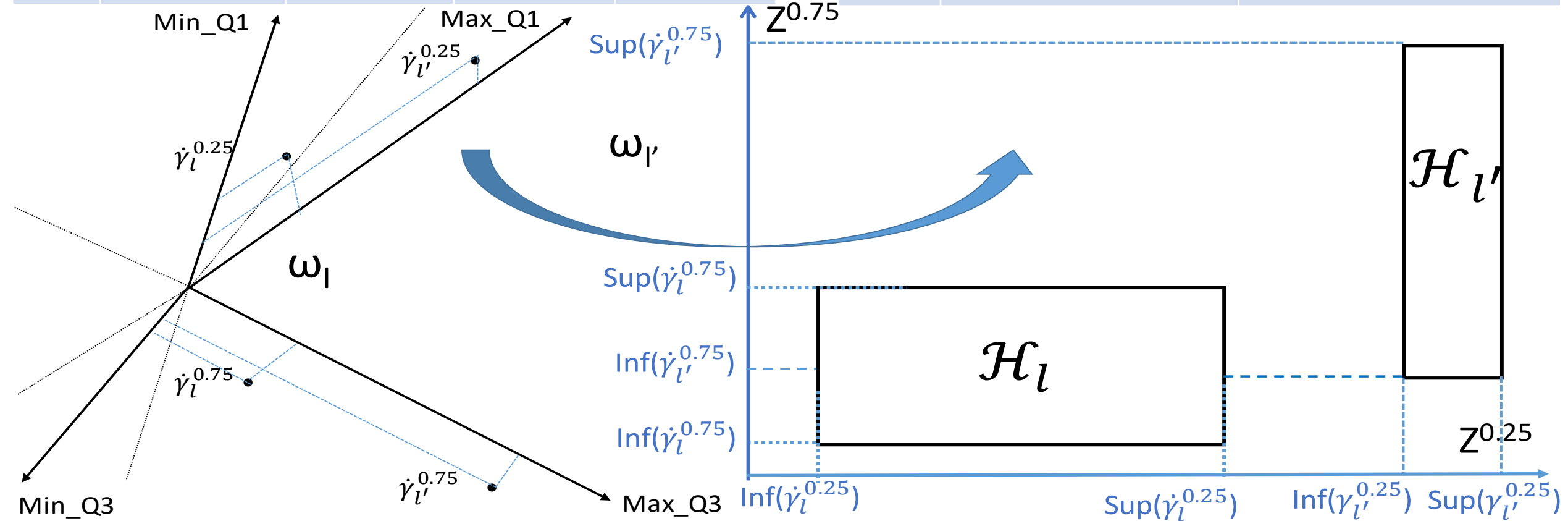


```
library(ggplot2)
```

```
sym.radar.plot(syearcrop2[1:1,2:6])
```

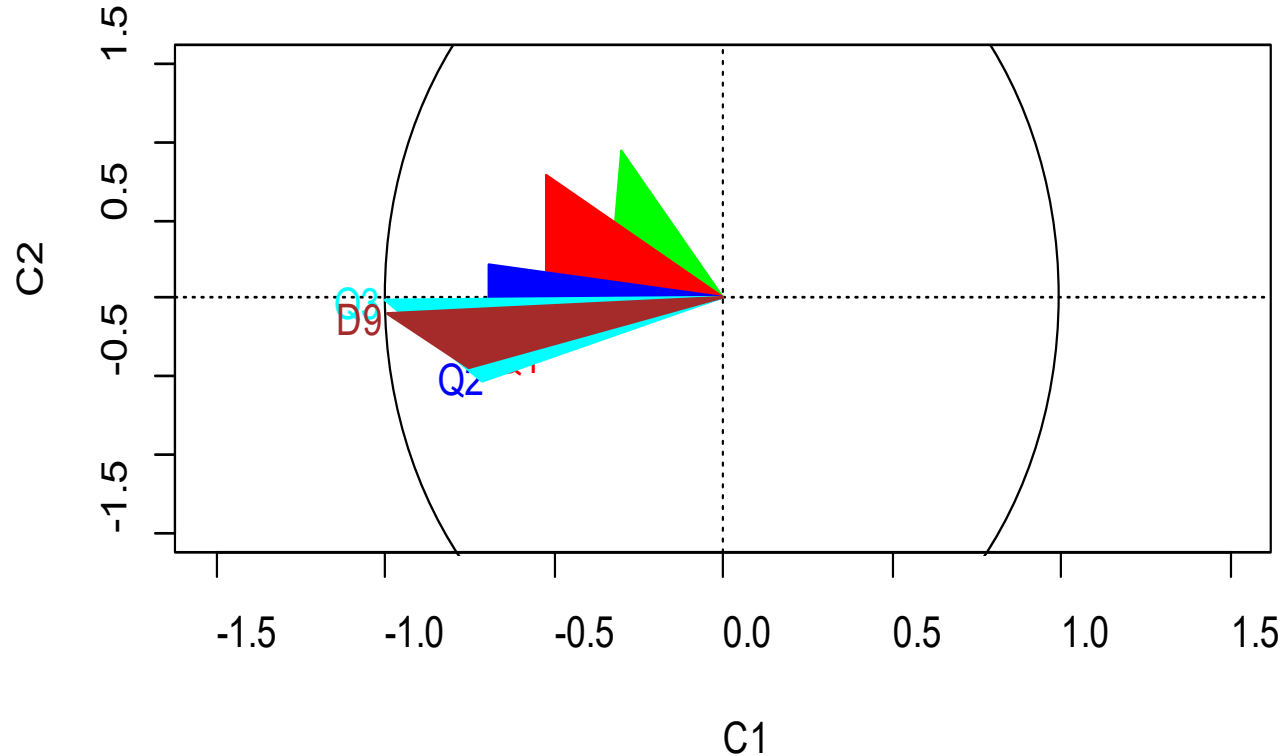
II.2 ACP symbolique : représentation des intervalles d'estimation quantiles

	Min_Q1	Max_Q1	Min_Q3	Max_Q3		$z^{0.25}$	$z^{0.75}$
...
ω_l	$\text{Inf}(\dot{\gamma}_l^{0.25})$	$\text{Sup}(\dot{\gamma}_l^{0.25})$	$\text{Inf}(\dot{\gamma}_l^{0.75})$	$\text{Sup}(\dot{\gamma}_l^{0.75})$	ω_l	$[\text{Inf}(\dot{\gamma}_l^{0.25}); \text{Sup}(\dot{\gamma}_l^{0.25})]$	$[\text{Inf}(\dot{\gamma}_l^{0.75}); \text{Sup}(\dot{\gamma}_l^{0.75})]$
...
$\omega_{l'}$	$\text{Inf}(\dot{\gamma}_{l'}^{0.25})$	$\text{Sup}(\dot{\gamma}_{l'}^{0.25})$	$\text{Inf}(\dot{\gamma}_{l'}^{0.75})$	$\text{Sup}(\dot{\gamma}_{l'}^{0.75})$	$\omega_{l'}$	$[\text{Inf}(\dot{\gamma}_{l'}^{0.25}); \text{Sup}(\dot{\gamma}_{l'}^{0.25})]$	$[\text{Inf}(\dot{\gamma}_{l'}^{0.75}); \text{Sup}(\dot{\gamma}_{l'}^{0.75})]$
...



II.3 ACP symbolique, option « centres » cercle des corrélations

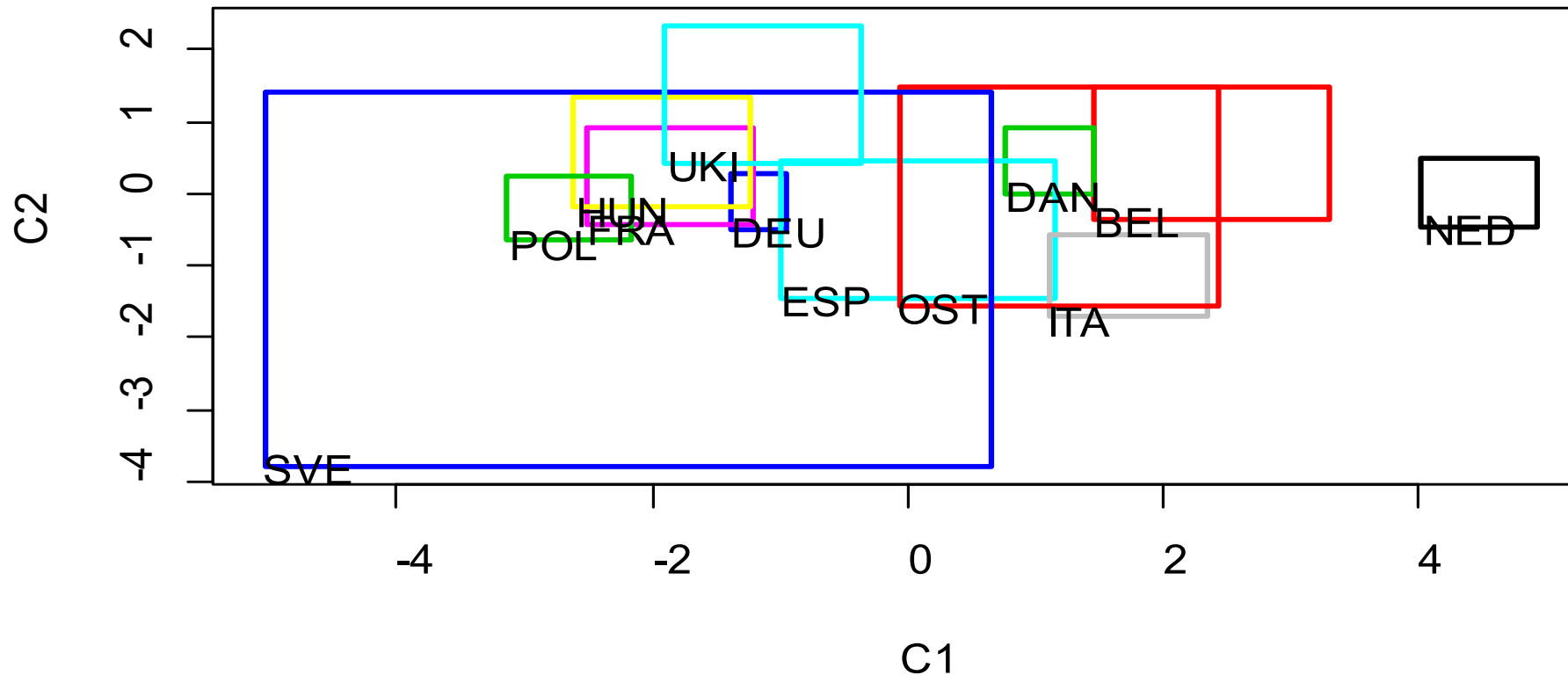
Correlation Circle



```
res <- sym.pca(syearcrop2[,2:6],'centers')
plot(res, choix = "var")
```

L'option "centers" de la fonction `sym.pca` du logiciel RSDA correspond à Cazes, Chouakria, Diday and Schektman (1997)

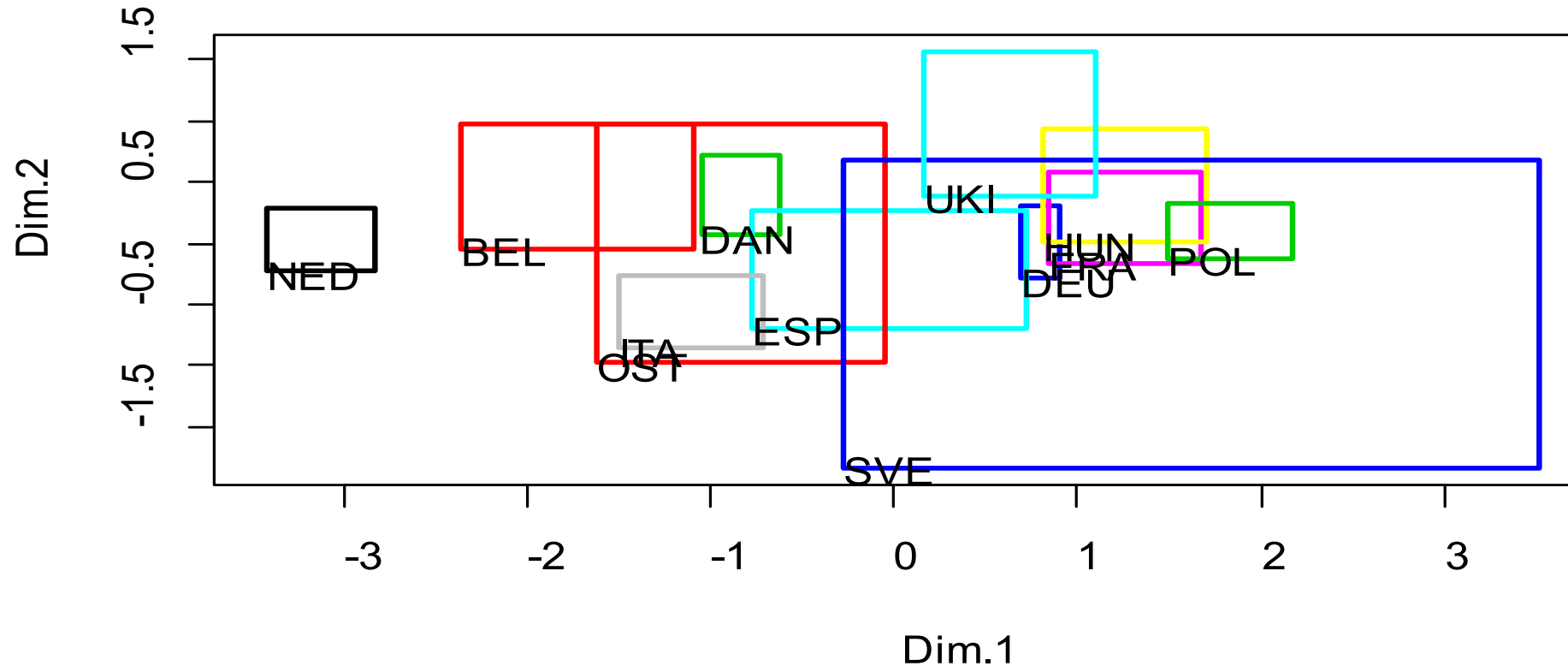
II.4 ACP symbolique, option « centres » plan factoriel des pays



```
res <- sym.pca(syearcrop2[,2:6],'centers')
plot(res, choix = "ind")
```

L'option "centers" de la fonction `sym.pca` correspond à Cazes, Chouakria, Diday and Schektman (1997)

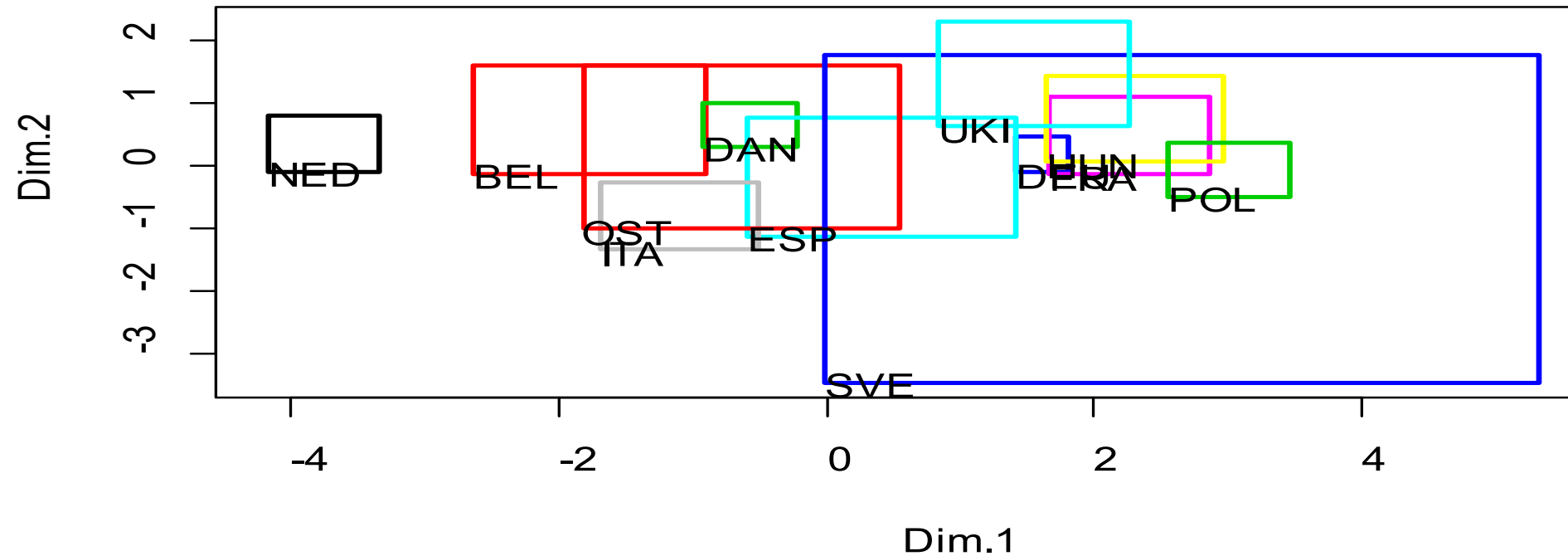
II.6 ACP intervalles, option « distance optimisée » : premier plan factoriel des pays



```
res <- sym.pca(syearcrop2[,2:6], 'optimized.distance')
plot(res, choix = "ind")
```

Les options de la fonction `sym.pca` sont décrites dans "Optimized Dimensionality Reduction Methods for Interval-Valued Variables and Their Application to Facial Recognition", Jorge Arce Garro et Oldemar Rodríguez Rojas, *Entropy* 2019, 21(10), 1016

II.7 ACP symbolique, option « variance optimisée » : le premier plan factoriel des pays

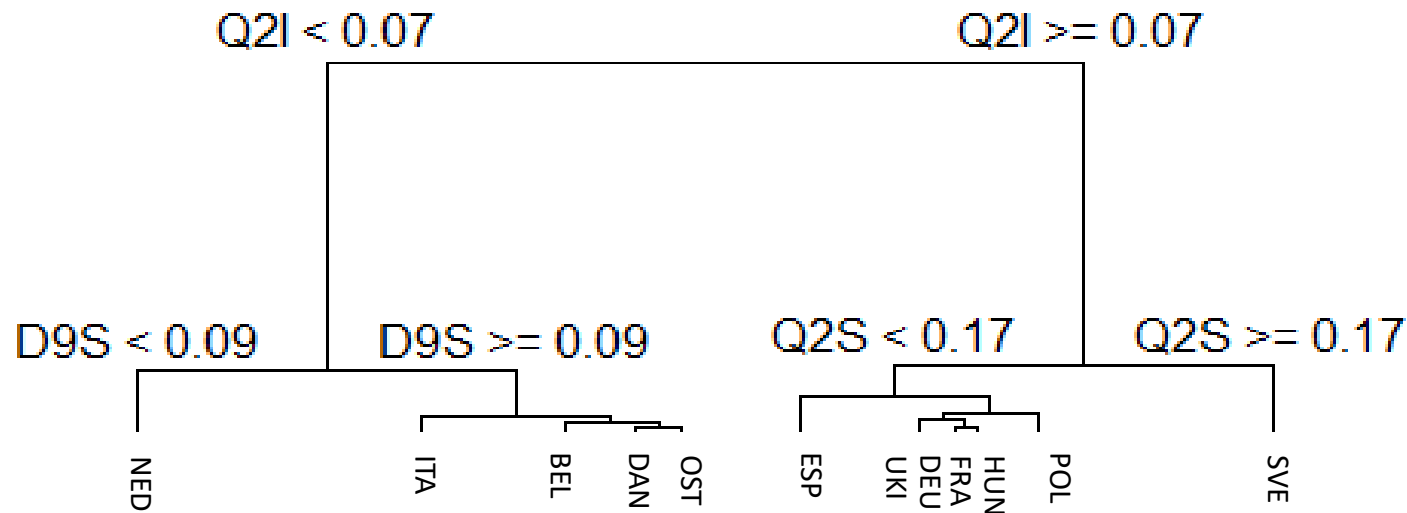


```
res <- sym.pca(syearcrop2[,2:6], 'optimized.variance')
plot(res, choix = "ind")
```

Les options optimisées de la fonction `sym.pca` sont décrites dans "Optimized Dimensionality Reduction Methods for Interval-Valued Variables and Their Application to Facial Recognition" by Jorge Arce Garro and Oldemar Rodríguez Rojas, *Entropy* 2019, 21(10), 1016

III.1 Classification descendante symbolique : la partition en 4 classes

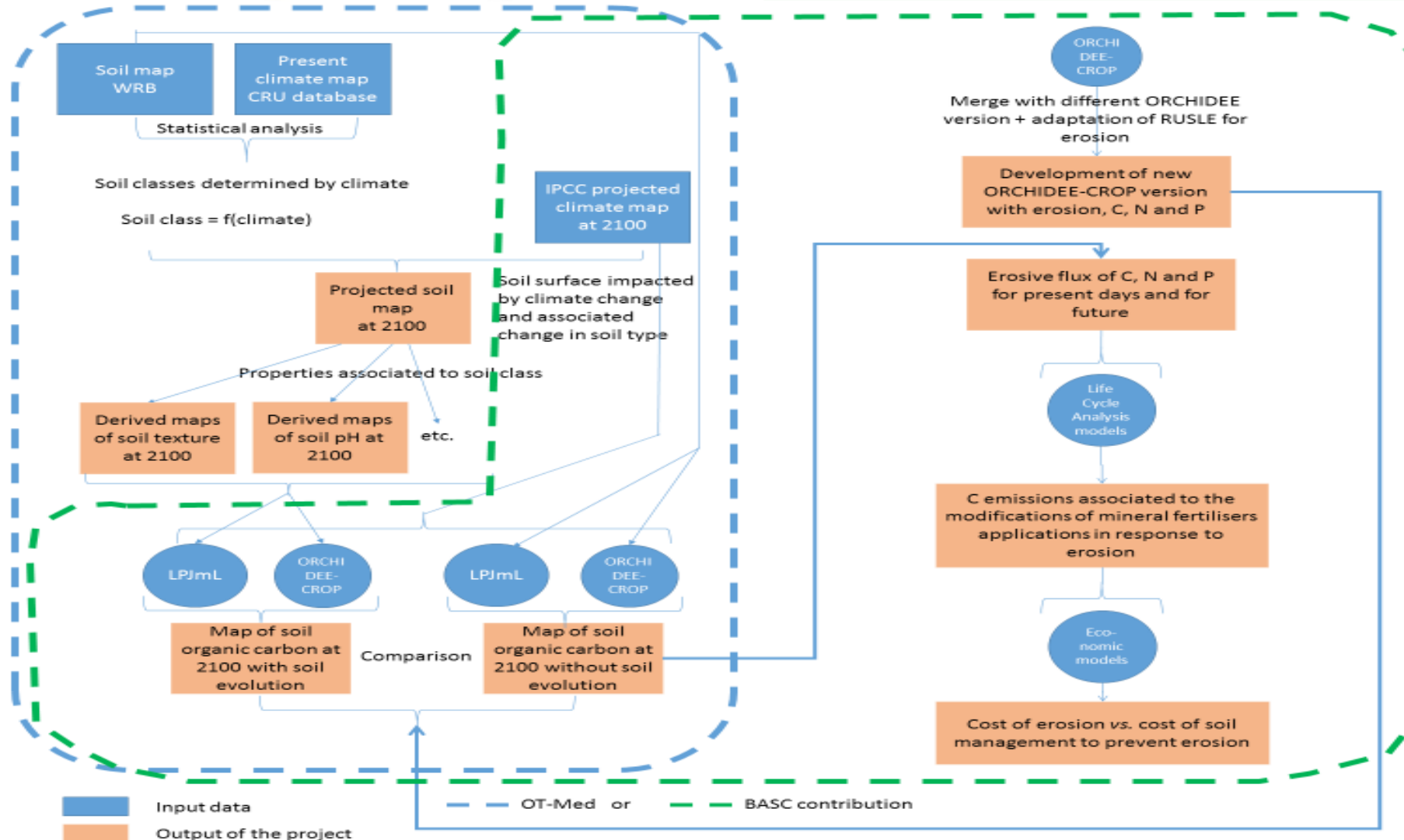
$$W(P_{K+1}) = W(P_K) - h(C_K^*)$$



```

syear2crop<-read.delim2("~/FERTI/syear2crop.txt", row.names=1, stringsAsFactors=FALSE)
tree <- divclust(syear2crop[,4:13])
plot(tree)
  
```

La fonction symbolique de classification descendante est adaptée de la librairie R `divclust`, basée sur Chavent M., Lechevalier Y., Briant O. (2007) DIVCLUS-T: A monothetic divisive hierarchical clustering method. *Computational Statistics & Data Analysis*, 52, 2, 687-701.



- Afonso F., Diday E. and Toque C. (2018) *Data science par analyse des données symboliques*, Technip, Paris, 444 p.
- Billard L., Diday E. (2006) *Symbolic Data Analysis: Conceptual Statistics and Data Mining*, 321 p.
- Cazes P., Chouakria A., Diday E., Schektman Y. (1997) Extensions de l'analyse en composantes principales à des données de type intervalle. *Revue de Statistique Appliquée*, n°24, pp. 5-24.
- Carvalho F., Souza R., Chavent M., and Lechevallier Y. (2006) Adaptive Hausdorff distances and dynamic clustering of symbolic interval data. *Pattern Recognition Letters*, Volume 27, Issue 3, pp. 167-179.
- Chavent M., Lechevalier Y., Briant O. (2007) DIVCLUS-T: A monothetic divisive hierarchical clustering method. *Computational Statistics & Data Analysis*, 52, 2, 687-701.
- Desbois D. (2015) *Estimation des coûts de production agricoles : approches économétriques*. PhD dissertation directed by J.C. Bureau and Y. Surry, ABIES-AgroParisTech, Paris, 2015.
- Desbois D., Butault J.-P., Surry Y. (2013) Estimation des coûts de production en phytosanitaires pour les grandes cultures. Une approche par la régression quantile, *Economie Rurale*, n° 333. pp.27 49.
- Desbois, D., Butault J.-P. and Surry Y. (2017). Distribution des coûts spécifiques de production dans l'agriculture de l'Union européenne : une approche reposant sur la méthode de régression quantile, *Economie rurale*, 361, 3-22.
- Garro J.A., Rodrigues Rojas O. (2019) Optimized Dimensionality Reduction Methods for Interval-Valued Variables and Their Application to Facial Recognitions, *Entropy* 2019, 21(10), 1016.
- Halkidi M., Batistakis Y., and Vazirgiannis Mi. On clustering validation techniques. *J. Intell. Inf. Syst.*, 17(2-3):107{145, 2001.
- Ichino M., Yaguchi, H. (1994): Generalized Minkowski metrics for mixed feature type data analysis. *IEEE Transactions on Systems, Man and Cybernetics*, 24 (4), 698–708.
- Koenker R. and Bassett G. (1978) Regression quantiles. *Econometrica*, 46, 3350, 1978.
- Lauro C.N. and Palumbo F. (2000) Principal component analysis of interval data: a symbolic data analysis approach. *Computational Statistics*, 15, 1, 73-87.
- Rodrigues Rojas O. (2019) *R to Symbolic Data Analysis: Package 'RSDA'*, Version 3.0, October 21, 2019