

**Introduction à l'ACP sur intervalles**  
**Visualiser la distribution des quantiles coûts spécifiques de production du lait**  
**de vache dans l'Union européenne**

*par*

**Dominique DESBOIS**

**Session ingénieurs**

**JRSS 2018**

**Nantes, 13 décembre 2018**

**12<sup>èmes</sup> JOURNÉES DE RECHERCHES EN SCIENCES SOCIALES**  
**JRSS**

**Nantes**



**Organisateurs :**

**Lieu de la manifestation : Oniris**

Problématique :

**Comment conserver le maximum d'informations pour l'analyse de distributions de paramètres estimés?**

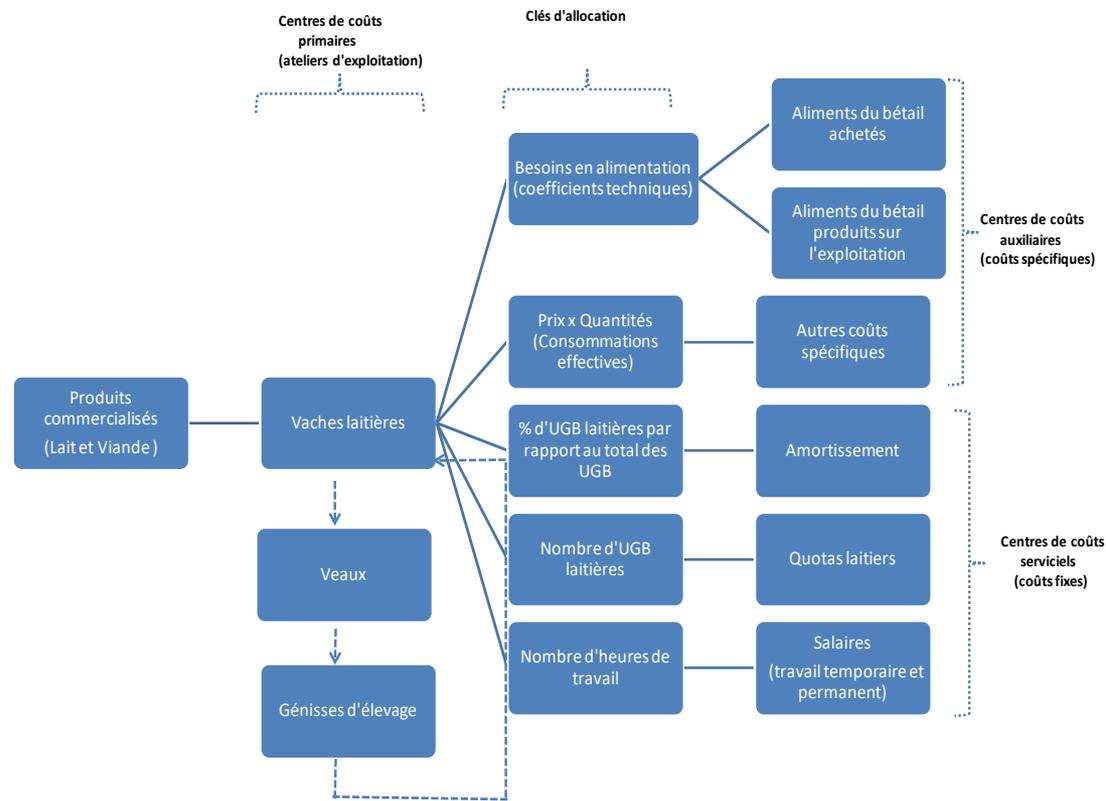
## Plan de la présentation :

- 1. L'estimation par intervalles de quantiles conditionnels**
- 2. Analyse en composantes principales sur intervalles**
- 3. Visualiser les distributions de quantiles conditionnels estimés**

# 1 Evaluation des coûts de production agricole

## 1. Comptabilité technique des coûts de production

### Choix d'une méthode de Comptabilité technico-économique



## 2. Modélisation économétrique des coûts de production

### Choix d'un 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	$\sum Y_{1h}$	...	$\sum Y_{kh}$	...	$\sum Y_{Kh}$	$\sum_k Y_{kh} = \sum_l X_{lh}$



## 1.2 Estimations quantitatives des coûts de production agricoles

**Problème :** face à l'hétérogénéité des structures de production agricole et des comportements de production en Europe, comment retenir le maximum d'information utile lors de l'estimation des coûts de production ?

**Modèle conceptuel :** le tableau des entrées-sorties

**Méthode :**

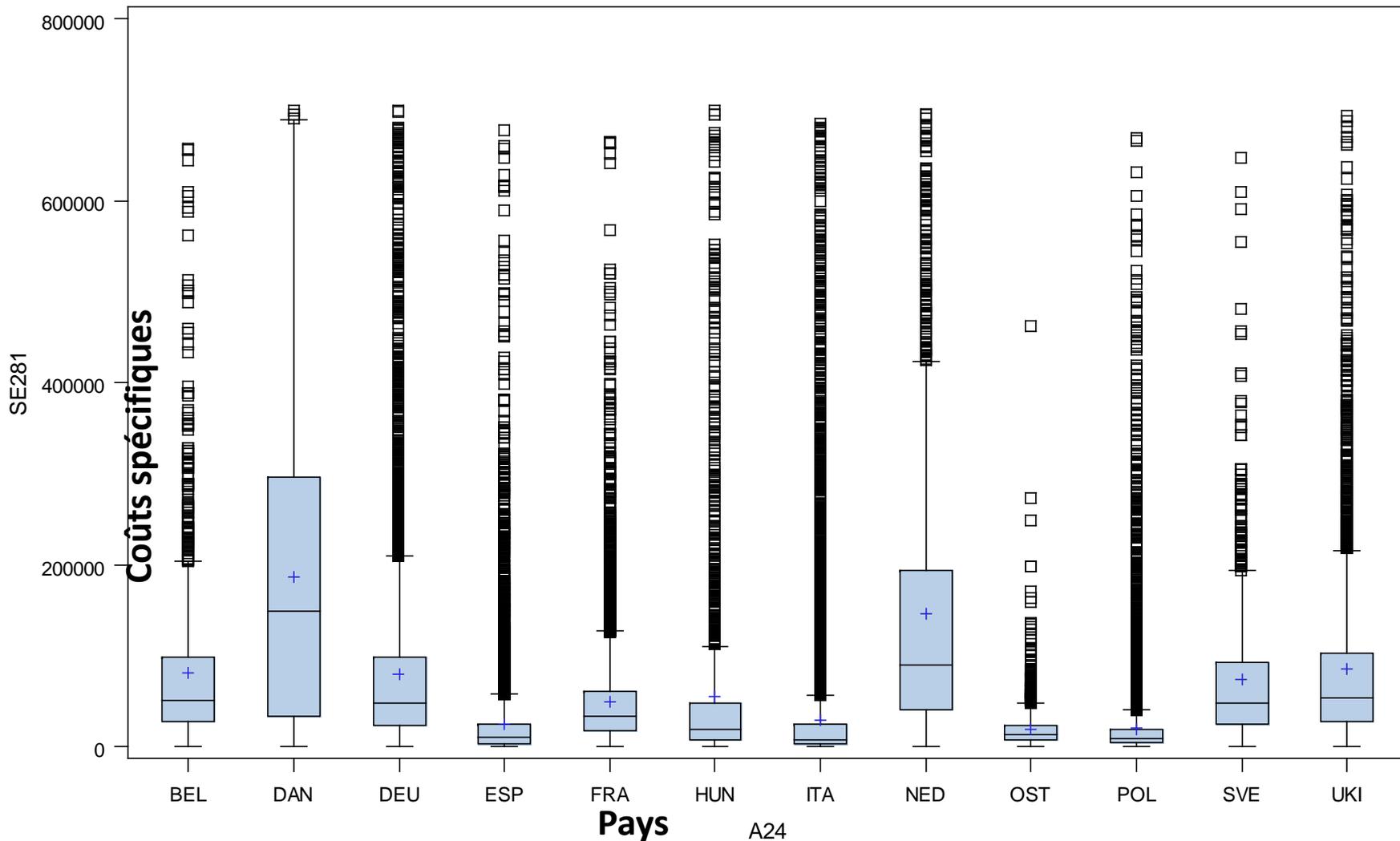
- i) la régression en norme L1 permet, en estimant les quantiles conditionnels, de générer une distribution de coûts de production spécifiques et des marges brutes ;
- ii) l'analyse de données symboliques permet une visualisation optimale de l'échelle et de la forme des distributions d'estimations par intervalle.

**Données :** le Réseau d'information comptable agricole (RICA) en France et en Europe

**Application :** Estimation des coûts spécifiques pour déduire la marge brute

## I.3 Distribution des charges spécifique de production agricole dans l'UE

Distribution des charges spécifiques (<750 000 €) par pays, RICA 2006, UE12



## I.4 Estimation du quantile conditionnel.

L'estimation des quantiles conditionnels est obtenue en résolvant un problème de minimisation de la fonction de perte suivante pour chaque quantile  $q$ , selon le critère de l'écart absolu en moyenne (MAD) :

$$\sum_{i: y_i \geq x_i' \beta} q |x_i - y_i' \beta| + \sum_{i: y_i < x_i' \beta} (1 - q) |x_i - y_i' \beta|$$

Pour un processus générateur de données

$$X = Y' \beta + u \quad \text{tel que } u = Y' \alpha * \varepsilon$$

suivant un model linéaire avec hétéroscédasticité multiplicative

(i.e.  $u = Y' \alpha * \varepsilon$ ), sous l'hypothèse  $Y' \alpha > 0$

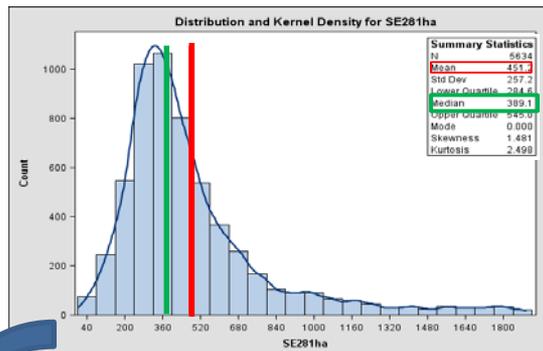
le  $q^e$  quantile conditionnel du coût  $X$  conditionnellement à  $Y$  est égal à :

$$\mu_{q(Y, \beta, \alpha)} = Y' [\beta + \alpha * F_{\varepsilon}^{-1}(q)]$$

donc linéaire en  $Y$ .

# I.5 Estimations quantitatives des coûts de production agricole

## 1. Charges phytosanitaires

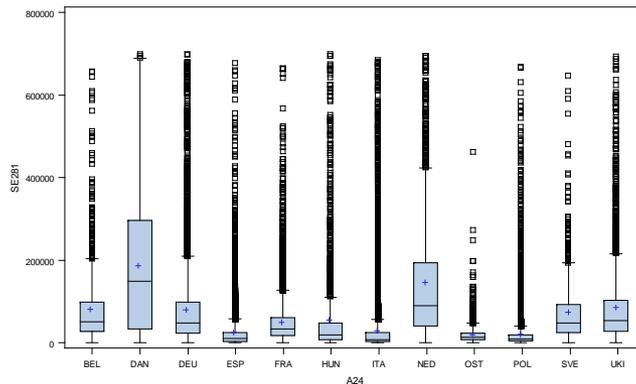


*Méthode : régression quantile*

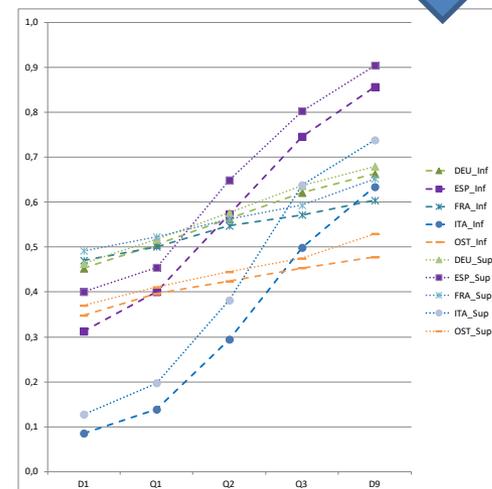
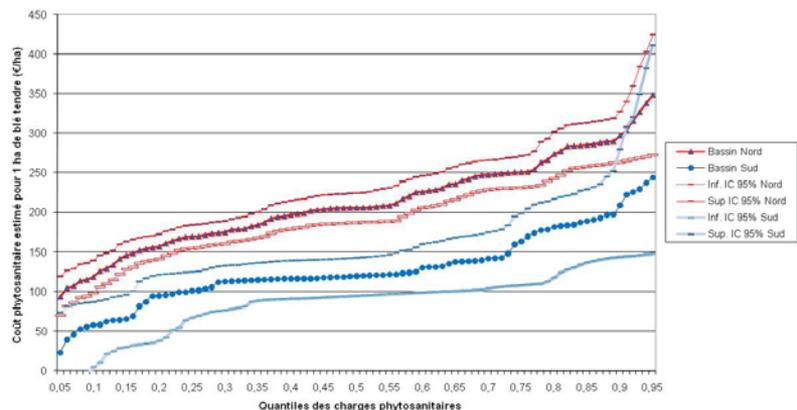
$$\text{Min}_{\beta} \left\{ \sum_{\vec{i} x_i \geq y_i \beta} q |x_i - y_i \beta| + \sum_{\vec{i} x_i \leq y_i \beta} (1 - q) |x_i - y_i \beta| \right\}$$

*Estimation: quantiles conditionnels*

## 2. Coûts spécifiques



**Blé tendre : processus quantile du coût phytosanitaire, bassins Nord et Sud (RICA 2006, grandes cultures, estimations basées sur les surfaces)**



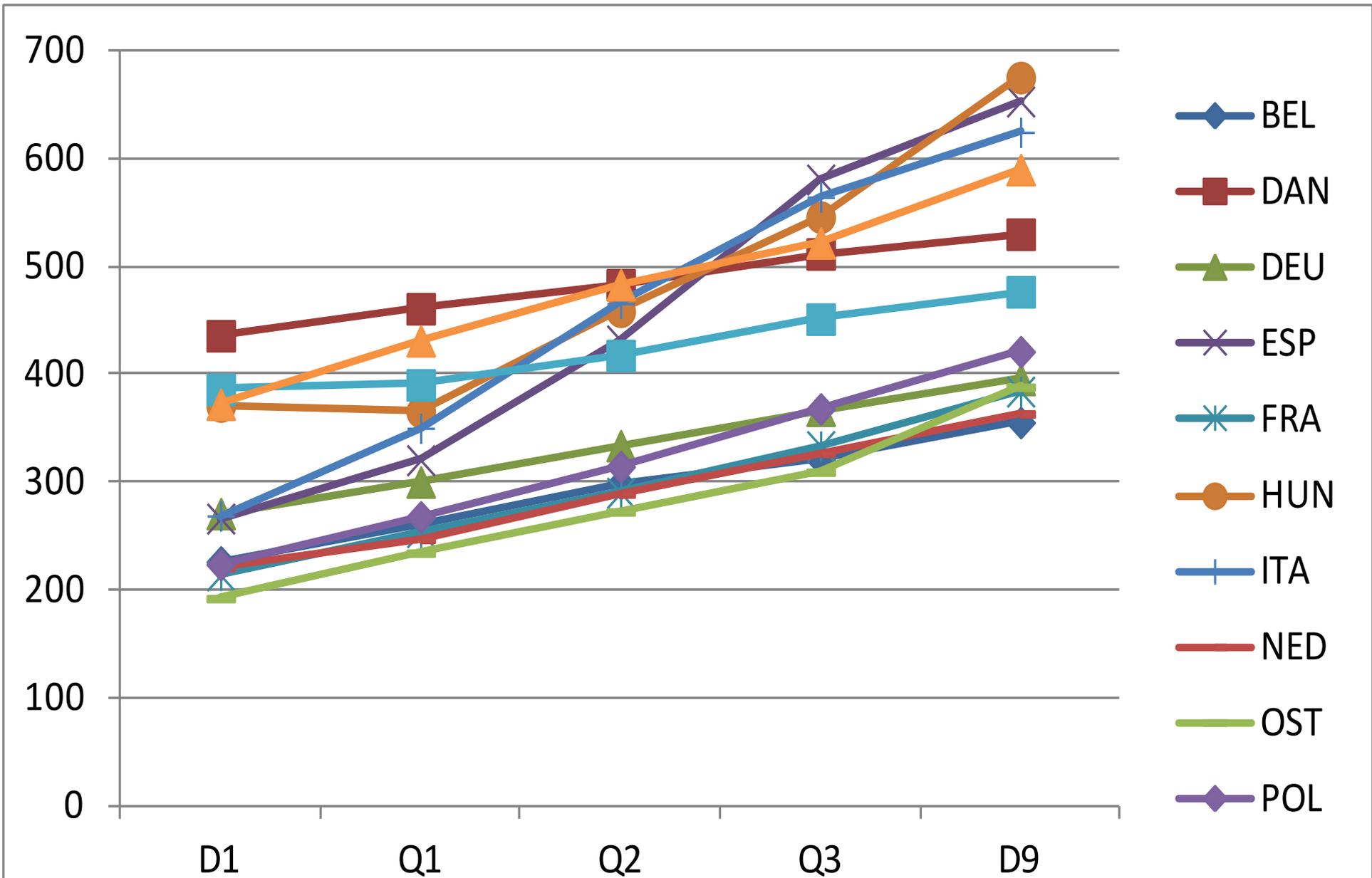
## II.1 Distribution des coûts spécifiques de production du lait de vache dans l'UE : Estimations ponctuelles des coûts spécifiques pour 1000 € de Produit brut

<b>Coûts spécifiques (€)</b>						
<b>Pays</b>	<b>D1</b>	<b>Q1</b>	<b>Q2</b>	<b>Q3</b>	<b>D9</b>	<b>MCO</b>
Allemagne	277,0	301,6	336,0	370,8	403,8	380,4
Autriche	194,7	235,5	267,2	310,2	389,1	307,6
Belgique	224,8	259,7	299,3	318,2	362,4	297,9
Danemark	437,5	466,9	489,7	518,6	531,8	498,2
Espagne	270,9	330,5	437,3	583,9	658,9	475,3
France	216,7	251,1	288,4	329,6	381,9	304,4
Hongrie	391,5	393,5	523,3	638,9	755,7	733,2
Italie	266,3	346,0	459,3	562,5	609,0	413,4
Pays-Bas	220,5	251,9	284,4	327,3	360,2	287,7
Pologne	224,7	266,7	312,4	364,8	418,8	333,0
Royaume-Uni	375,0	433,2	481,1	523,3	583,5	489,6
Suède	384,0	387,6	415,8	449,2	470,6	427,3



## II.1.1 Estimations quantiles des coûts de production en agriculture

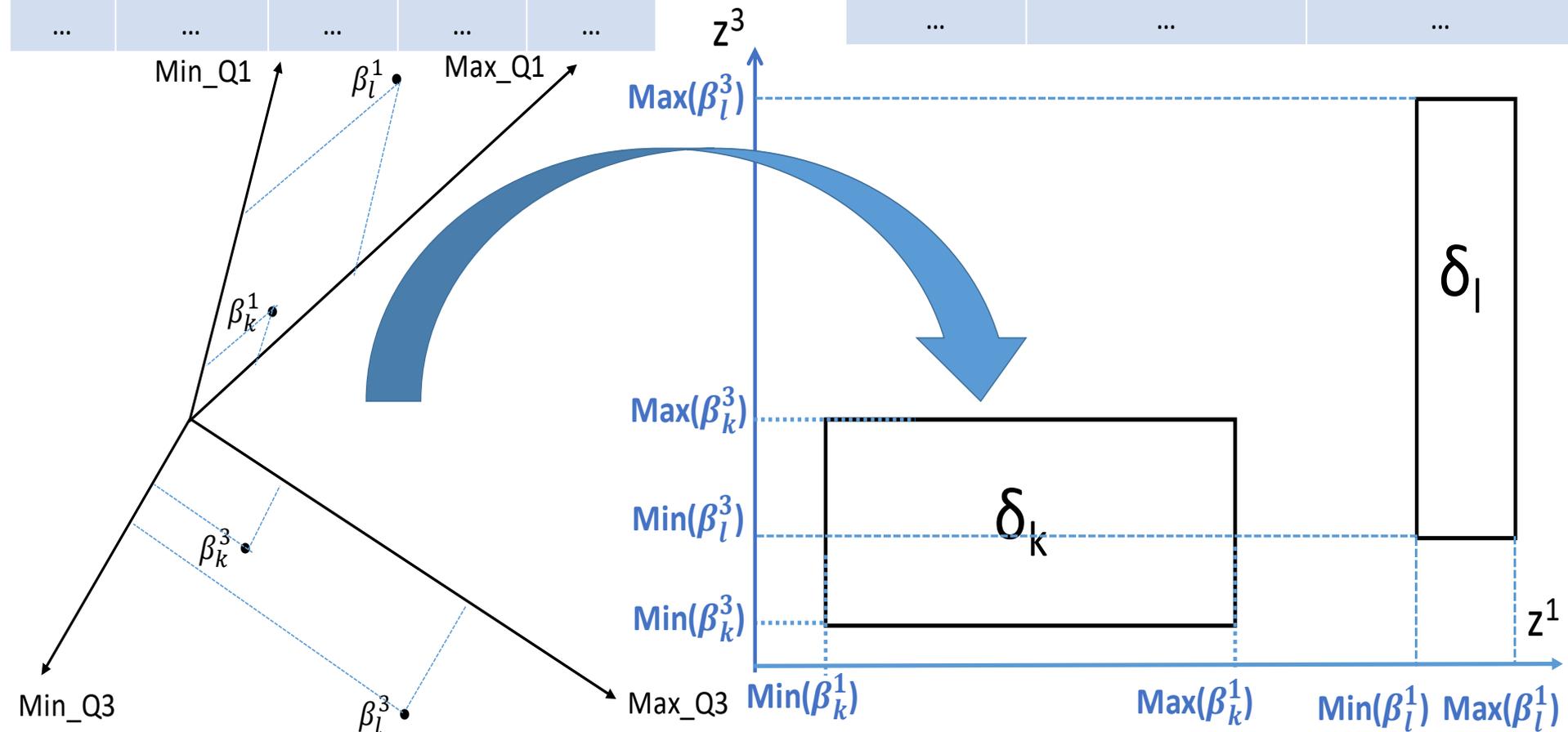
### Lait de vache



## 2.2 distribution des intervalles d'estimation : représentation graphique en 2 dimensions

	Min_Q1	Max_Q1	Min_Q3	Max_Q3
...	...	...	...	...
$\delta_k$	$\text{Min}(\beta_k^1)$	$\text{Max}(\beta_k^1)$	$\text{Min}(\beta_k^3)$	$\text{Max}(\beta_k^3)$
...	...	...	...	...
$\delta_l$	$\text{Min}(\beta_l^1)$	$\text{Max}(\beta_l^1)$	$\text{Min}(\beta_l^3)$	$\text{Max}(\beta_l^3)$
...	...	...	...	...

	$Z^1$	$Z^3$
...	...	...
$\delta_k$	$[\text{Min}(\beta_k^1); \text{Max}(\beta_k^1)]$	$[\text{Min}(\beta_k^3); \text{Max}(\beta_k^3)]$
...	...	...
$\delta_l$	$[\text{Min}(\beta_k^1); \text{Max}(\beta_k^1)]$	$[\text{Min}(\beta_l^3); \text{Max}(\beta_l^3)]$
...	...	...



## II.3 ACP des extrema d'intervalles d'estimation (ACP-S)

Comme en ACP classique, le sous-espace propre (optimal pour la représentation duale) est structuré par des axes orthonormés dans  $\mathbb{R}^K$

$$\frac{1}{N} Z' Z v_m = \lambda_m v_m$$

L'analyse duale dans  $\mathbb{R}^p$  conduit à une équation similaire

$$\frac{1}{N} Z Z' w_m = \lambda_m w_m$$

Relation entre les vecteurs propres :

$$v_m = \lambda_m^{-1/2} Z' w_m$$

L'interprétation des axes de l'ACP-S s'effectue en fonction des quantiles conditionnels (variables de l'ACP-S) présentant les contributions les plus fortes

$$CTA_j^m = \frac{(\lambda_m^{1/2} v_j^m)^2}{\lambda_m} = (v_j^m)^2$$

Les coordonnées des projections des extrémités des intervalles d'estimation des quantiles conditionnels (sommets  $s(k)$ )

de la distribution empirique  $\delta_k$  des coûts spécifiques (objet symbolique)

sur les axes factoriels principaux sont fournies par la relation :

$$\psi_k^m = Z_k v_m$$

Retenir parmi les représentations celles dont les sommets sont les mieux représentés en utilisant comme critère la contribution relative (CTR) définie en termes de cosinus carré, soit pour une extrémité d'intervalle d'estimation  $s(k)$  de la distribution quantile  $k$  :

$$CRT_{s(k)}^m = \frac{\sum_{j=1}^p (z_{s(k)}^m v_m)^2}{\sum_{j=1}^p z_{s(k)}^j{}^2}$$

## II.4 ACP symboliques d'intervalles d'estimation (ACP-OS)

Soit  $A$ , la matrice booléenne indicatrice de l'appartenance des  $N$  extrema d'intervalle d'estimation aux  $K$  distributions empiriques. L'expression de la variance entre les objets symboliques est donnée par :

$$\frac{1}{N} Z' A (A' A)^{-1} A' Z$$

Soit, le projecteur orthogonal  $P_A$  associé à la matrice  $A$  sur le sous-espace de référence codant l'appartenance des extrema aux différentes distributions :

$$P_A = A (A' A)^{-1} A'$$

Dans l'espace  $\mathbb{R}^p$ , les axes principaux d'inertie sont obtenus comme solution de l'équation suivante aux valeurs propres :

$$\frac{1}{N} Z' P_A Z \tilde{v}_m = \tilde{\lambda}_m \tilde{v}_m$$

Les coordonnées de l'hyper-rectangle associé à la distribution empirique sont alors calculées comme suit :

$$\tilde{\psi}_{i,m} = Z_i \tilde{v}_m$$

L'analyse dans  $\mathbb{R}^p$  revient à résoudre l'équation suivante aux valeurs propres :

$$(A' A)^{-1/2} (A' Z Z' A) (A' A)^{-1/2} \tilde{w}_m = \tilde{\lambda}_m \tilde{w}_m$$

où  $\tilde{w}_m = (A' A)^{-1/2} A' Z \tilde{v}_m$

Les contributions relatives des variables (CRT) sont définies de la même façon que pour l'ACP-S

$$CRT_{s(k)}^m = \frac{\sum_{j=1}^p (z_{s(k)}^m v_m)^2}{\sum_{j=1}^p z_{s(k)}^j{}^2}$$

## II.5 ACP des étendues d'intervalles d'estimation

Lauro et Palumbo (2000) ont proposé une ACP partielle basée sur la translation à l'origine des données d'intervalle  $e_k^j = \overline{x_k^j} - \underline{x_k^j}$  qui retient l'étendue comme information utile pour décrire la taille et la forme des objets symboliques.

L'hyper-pavé est associé à la description de la  $k^e$  distribution empirique, domaine défini par le produit cartésien des étendues  $E_k = e_i^1 \times \dots \times e_k^j \dots e_k^p$  des  $p$  paramètres associés à la description de l'objet symbolique  $\delta_k$ ,  
où  $\tilde{e}_k^j = e_k^j / s_j$

$$\pi(\delta_k) = \prod_{j=1}^p \tilde{e}_k^j$$

on utilise la mesure linéaire du potentiel de description :

$$\sigma(\delta_k) = \sum_{j=1}^p \tilde{e}_k^j$$

Soit  $X$ , la matrice de dimension  $K \times p$  et de terme général L'ACP des étendues (ACP-E) d'intervalle d'estimation (*Range Transformation Principal Component Analysis*) décompose le critère  $LDP = tr(X'X) = tr(XX') = \sum_{k=1}^K \sigma(\delta_k)$  selon les équations caractéristiques suivantes :

$$x_k^j = \sqrt{\tilde{e}_k^j} :$$

$$X'X t_m = \mu_m t_m$$

$$XX' u_m = \mu_m u_m$$

$$\phi_m = X t_m$$

Les coordonnées factorielles de la représentation des distributions de coûts spécifiques dans le sous-espace optimal sont données par

La contribution relative (CTR) mesure la qualité de représentation de la distribution empirique dans le sous-espace factoriel de représentation choisi :

$$CTR_k^m = \frac{\sum_{m=1}^{M^*} (\phi_k^m)^2}{\sum_{j=1}^p (x_k^j)^2}$$

## II.6 ACP des intervalles d'estimation en stratégie mixte (ACP-OS)

La stratégie mixte de l'analyse en composantes principales des objets symboliques (ACP-OS) combine l'ACP des sommets (ACP-S) et l'ACP des étendues (ACP-E) dans une approche en trois étapes pour prendre en compte les différences d'échelle et de forme entre les distributions empiriques de coûts spécifiques :

i) ACP des étendues afin d'extraire les axes principaux qui restituent le mieux les échelles et les formes des distributions empiriques de quantiles conditionnels ;

ii) Projection de  $Z$  en  $\hat{Z} = P_A Z$  afin de prendre en compte les relations entre les différents extrema, compte-tenu des relations d'ordre entre les différents quantiles conditionnels de la distribution des coûts spécifiques ;

iii) ACP des projections des lignes  $\hat{Z}_i$  sur le sous-espace de représentation optimale  $\Phi = \{\phi_1, \dots, \phi_m, \dots, \phi_{M^*}\}$  pour la représentation des échelles et des formes des distributions, avec l'opérateur de projection

$$\hat{Z}' P_{\Phi} \hat{Z} = Z' A (A' A)^{-1/2} P_{\Phi} (A' A)^{-1/2} A' Z S_m = \rho_m S_m$$

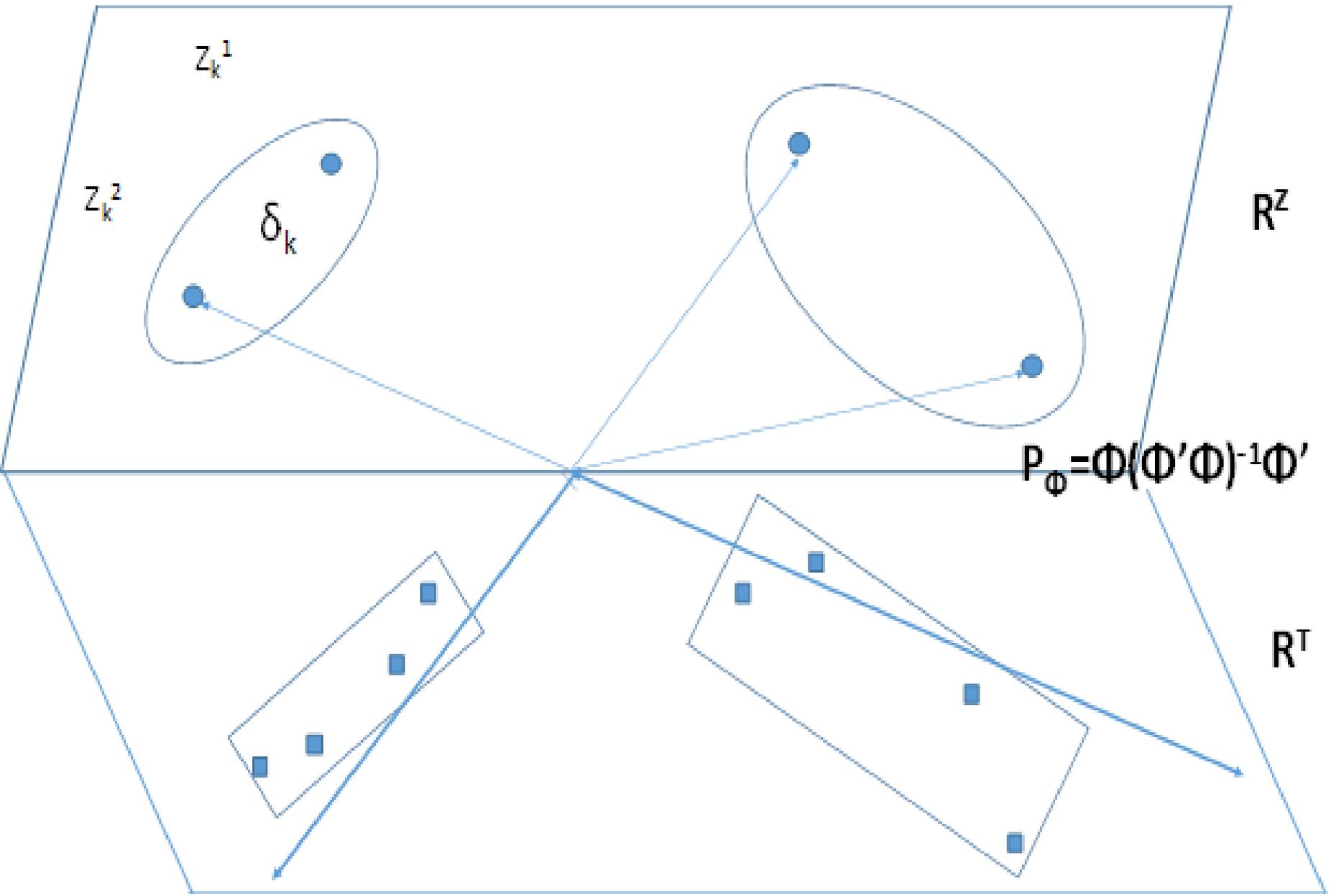
La stratégie mixte d'analyse est donc basée sur la solution de l'équation caractéristique suivante :

$$\hat{Z}' P_{\Phi} \hat{Z} = Z' A (A' A)^{-1/2} P_{\Phi} (A' A)^{-1/2} A' Z S_m = \rho_m S_m$$

L'interprétation des résultats de l'analyse dépend du choix de l'opérateur de projection  $P_{\Phi}$ , dont le terme diagonal, interprétable comme un poids normalisé, est égal à :

$$\phi_k (\phi_k' \phi_k)^{-1} \phi_k' = \sum_{m=1}^{M^*} (\phi_k^m)^2 / \mu_m$$

## II.7 ACP symbolique : projection sur le sous-espace de référence



### III.1 Distribution des coûts spécifiques de production agricole dans l'UE12 :

#### Lait : coûts spécifiques pour 1000 € de produit brut, 2006

##### Coûts spécifiques (€)

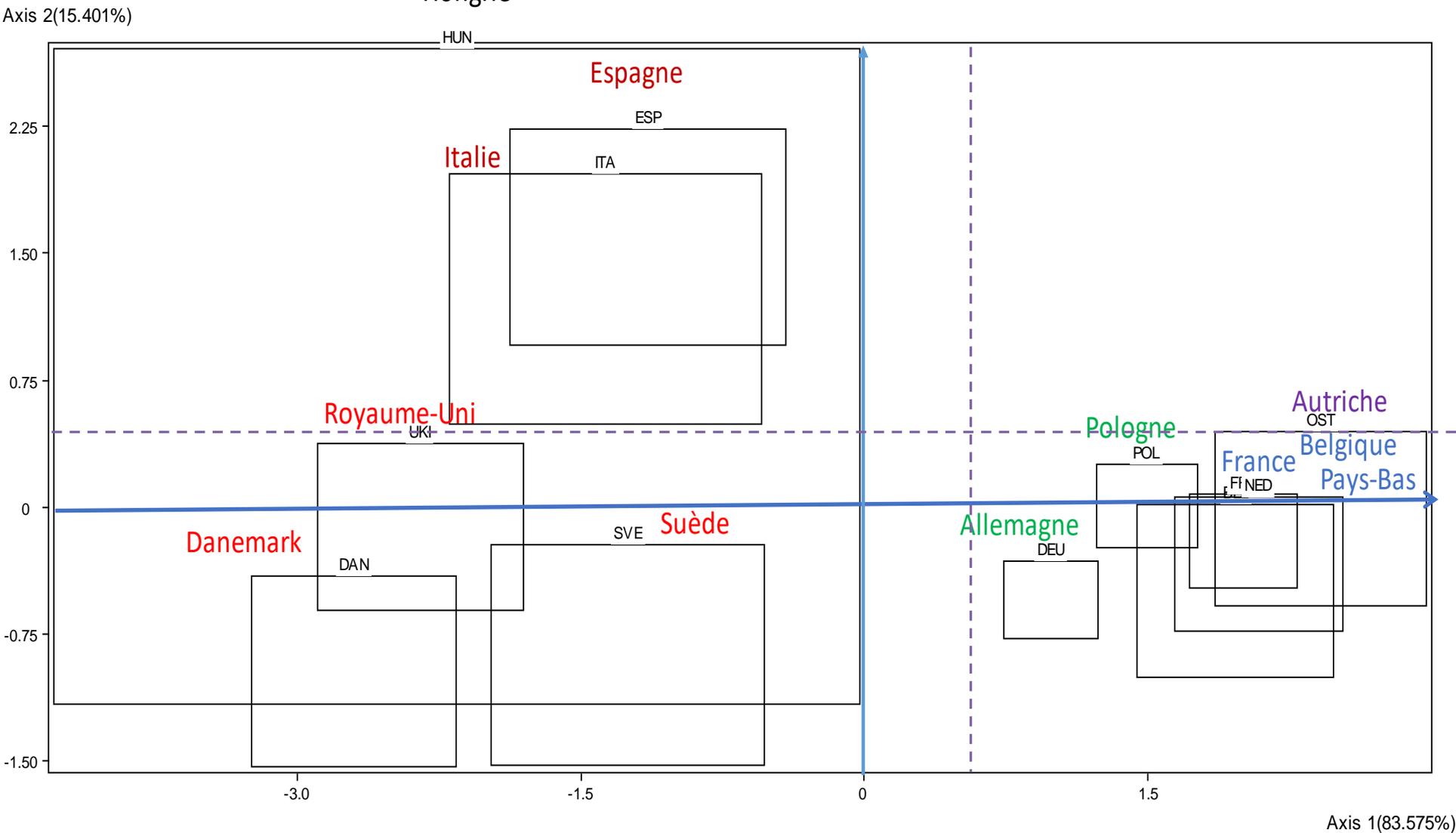
Pays	D1 [min ; max]	Q1 [min ; max]	Q2 [min ; max]	Q3 [min ; max]	D9 [min ; max]	MCO [min ; max]
<i>Allemagne</i>	[270,8 ; 283,2]	[296,5 ; 306,7]	[331,7 ; 340,3]	[366,1 ; 375,5]	[396,1 ; 411,5]	[376,7 ; 384,1]
<i>Autriche</i>	[181,3 ; 208,1]	[226,7 ; 244,3]	[257,4 ; 277,0]	[298,3 ; 322,1]	[371,6 ; 406,6]	[300,8 ; 314,4]
<i>Belgique</i>	[214,2 ; 235,4]	[240,8 ; 278,6]	[288,4 ; 310,2]	[302,5 ; 333,9]	[326,4 ; 398,4]	[284,1 ; 311,6]
<i>Danemark</i>	[429,1 ; 445,9]	[457,9 ; 475,9]	[482,2 ; 497,2]	[503,7 ; 533,5]	[505,9 ; 557,7]	[489,7 ; 506,8]
<i>Espagne</i>	[254,5 ; 287,3]	[314,6 ; 346,4]	[420,7 ; 453,9]	[569,8 ; 598,0]	[643,4 ; 674,4]	[466,0 ; 484,6]
<i>France</i>	[210,3 ; 223,1]	[245,5 ; 256,7]	[282,9 ; 293,9]	[322,7 ; 336,5]	[371,8 ; 392,0]	[299,8 ; 309,1]
<i>Hongrie</i>	[353,6 ; 429,4]	[354,0 ; 433,0]	[466,0 ; 580,6]	[578,7 ; 699,1]	[712,7 ; 798,7]	[719,1 ; 747,3]
<i>Italie</i>	[244,4 ; 288,2]	[330,7 ; 361,3]	[443,4 ; 475,2]	[549,8 ; 575,2]	[589,1 ; 628,9]	[408,2 ; 418,6]
<i>Pays-Bas</i>	[207,8 ; 233,2]	[242,4 ; 261,4]	[277,8 ; 291,0]	[320,7 ; 333,9]	[350,1 ; 370,3]	[265,5 ; 309,9]
<i>Pologne</i>	[218,6 ; 230,8]	[261,3 ; 272,1]	[307,4 ; 317,4]	[357,6 ; 372,0]	[409,3 ; 428,3]	[327,7 ; 338,3]
<i>Royaume-Uni</i>	[363,0 ; 387,0]	[422,9 ; 443,5]	[470,7 ; 491,5]	[515,2 ; 531,4]	[571,9 ; 595,1]	[483,3 ; 495,9]
<i>Suède</i>	[368,6 ; 399,4]	[376,7 ; 398,5]	[399,3 ; 432,3]	[436,9 ; 461,5]	[447,5 ; 493,7]	[416,5 ; 438,2]



# III.2 Estimations quantiles des coûts de production en agriculture

## Lait de vache, ACP symbolique sur les intervalles d'estimation

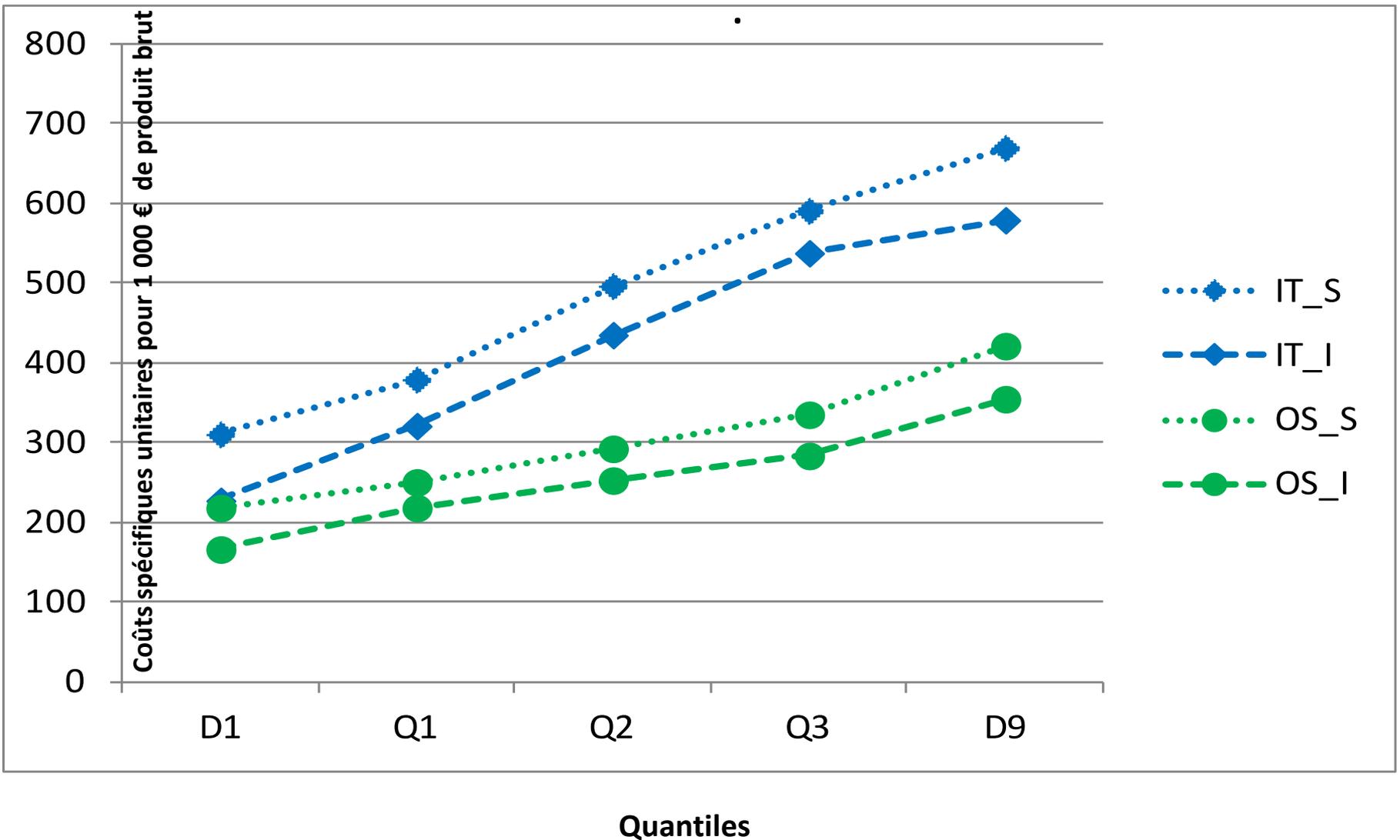
SPCA - SOs Interval Coordinates





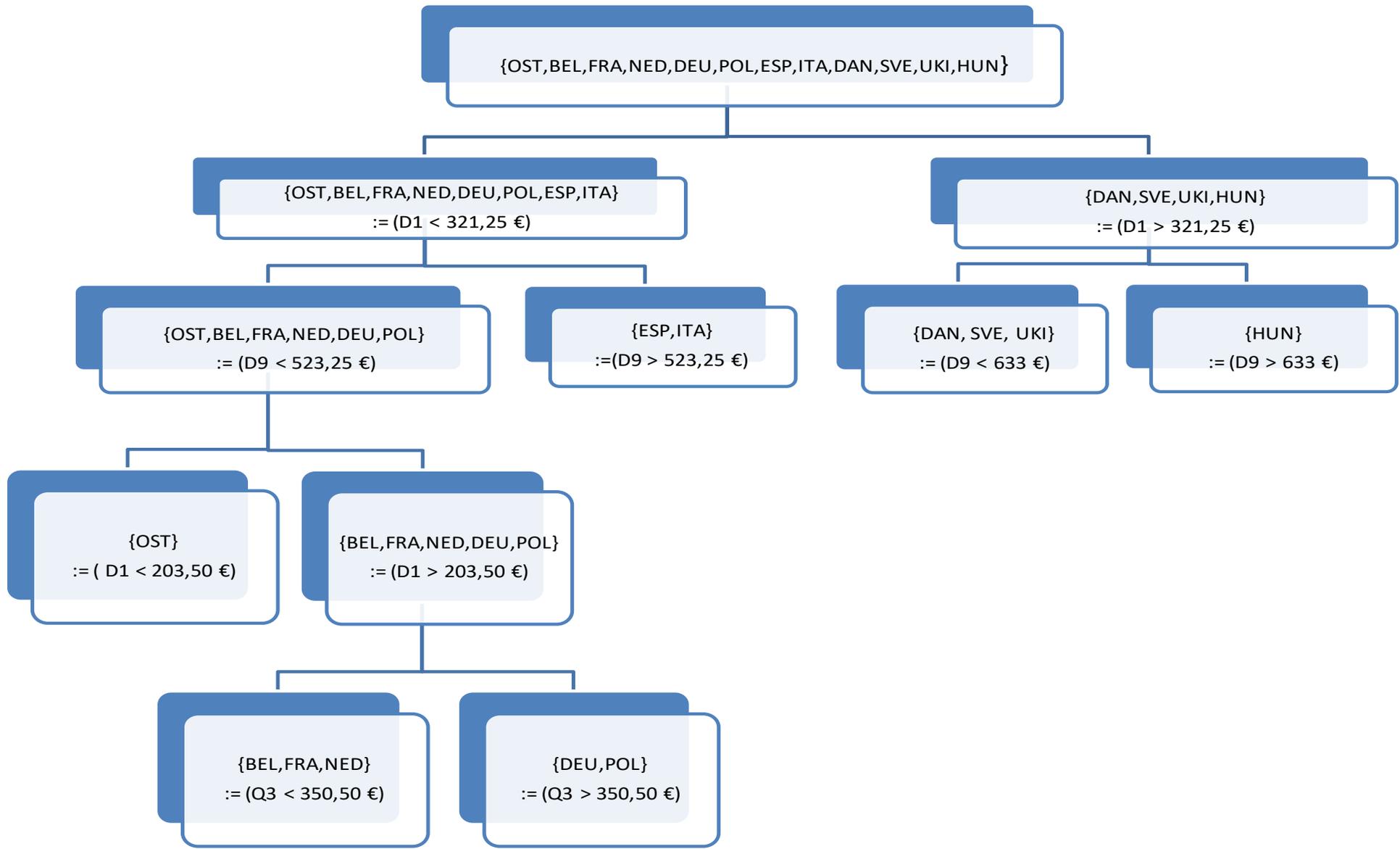
### III.3 Distribution des coûts spécifiques de production agricole dans l'UE12 : Autriche versus Italie

Lait : coûts spécifiques pour 1000 € de produit brut, 2006



### III.4 Estimations quantiles des coûts de production en agriculture : lait

#### Classification hiérarchique divisive sur les intervalles d'estimation





## Conclusions

**Coûts spécifiques :** dans l' UE12, pour le lait de vache

Le contexte national est facteur significatif d'hétérogénéité;

L'influence de la dimension économique est significative ;

L'analyse selon l'orientation technico-économique ne révèle pas de différences significatives.

.

# Références

- Desbois D. (2017) Displaying empirical distributions of conditional quantile estimates: an application of symbolic data analysis to the cost allocation problem in agriculture », *Proceedings of the 17th Conference of the Applied Stochastic Models and Data Analysis International Society*, Londres, pp. 189-202.
- Desbois D., Butault JP., Surry Y. (2017) Distribution des coûts spécifiques de production dans l'agriculture de l'Union européenne : une approche reposant sur la régression quantile, *Économie rurale*, vol. 361, n°5, pp. 3-22.
- Desbois D. (2015) Estimation des coûts de production agricoles : approches économétriques. *PhD Thesis*, ABIES-AgroParisTech, dirigé par J.C. Bureau et Y.Surry, 249 p.
- Desbois D., Butault J.-P., et Surry Y. (2013) Estimation des coûts de production en phytosanitaires pour les grandes cultures. Une approche par la régression quantile. *Economie Rurale*, n° 333, 27-49.
- Surry Y., Desbois D., et Butault J.-P. (2012) Quantile Estimation of Specific Costs of Production. *FACEPA*, D8.2, 49 p.

International Workshop:  
ADVANCES IN DATA SCIENCE FOR BIG AND COMPLEX DATA  
From data to classes and classes as new statistical units  
UNIVERSITY PARIS-DAUPHINE \*  
January 10-11, 2019

Call for contributions:

- Academics with mathematical, **statistical** and computer solutions to the problems **raised** by complex and standard or massive data as well as **industrials** wishing to present open problems concerning this type of data, **are welcome to participate**. Propositions must be sent as a one-page summary including references before **December 15th, 2018**, to: [datasciencejanv2019@gmail.com](mailto:datasciencejanv2019@gmail.com)
- **Free of charge for** registered participants, the number of places being limited. **Registration:** [datascience.sda2019@oniris-nantes.fr](mailto:datascience.sda2019@oniris-nantes.fr) (Name of the participant, e-mail address, Institution, Department and Country are needed).