

La méthode *ClustOfVar* pour mesurer les conditions de vie à l'échelle communale

Le complexe littoral/estuaire de la Gironde

Vanessa Kuentz-Simonet ¹, Amaury Labenne ^{1,2} & Tina Rambonilaza ¹

¹ IRSTEA, UR ETBX, 33612 Cestas Cedex, France.

² INRIA, CQFD, F-33400 Talence, France.



PLAN

- 1 Introduction
- 2 La méthode ClustOfVar
- 3 Application de la méthode ClustOfVar

INTRODUCTION I

Problématique et objectifs

- Effectuer un diagnostic socio-économique des territoires à l'échelle des communes à partir d'un ensemble de variables.
- Rendre compte de la multi-dimensionnalité du concept de qualité de vie.
- Problème de réduction de dimension bien connu en Statistique.
- Création d'indices synthétiques qui résument au mieux l'information apportée par les variables initiales.
- Intégrer des variables environnementales.

INTRODUCTION II

Quelques approches existantes

- “Une approche de la qualité de vie dans les territoires”. Reynard et Vialette, INSEE Première, 2014.
 - ▶ Échelle du territoire de vie sur toute la France métropolitaine,
 - ▶ ACP puis CAH sur composantes principales.
- IDH régional
 - ▶ Échelle régionale,
 - ▶ Moyenne pondérée de trois indicateurs.

L'approche par classification de variables de ClustOfVar

- Réduction de dimension.
- Possibilité de mixité des données (quantitatives/qualitatives).
- Création de variables synthétiques comme combinaison linéaire d'un nombre restreint de variables initiales.
- Facilité d'interprétation.

LES DONNÉES ET LA ZONE D'ÉTUDE 1/2

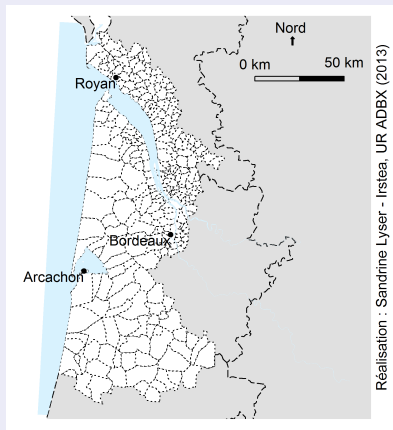
Les données :

- Variables concernant l'année 2009.
- $n = 303$ communes décrites par $p = 47$ variables (quantitatives et qualitatives). Les variables proviennent principalement de l'INSEE ainsi que de l'IGN et de Corine Land Cover pour les variables environnementales.
- Les 47 variables ont été extraites selon les 13 domaines de la qualité de vie dans le Système Européen des indicateurs Sociaux [Noll, 2012]. Seulement 8 thèmes ont été retenus, le tableau ci dessous montre les variables utilisées ainsi que leur appartenance aux différents thèmes.

Les 8 domaines de la qualité de vie retenus							
Conditions de logement	Marché de l'emploi et conditions de travail	Niveau de vie et inégalités économiques	Accès aux services	Accès à l'éducation	Accès aux soins	Interactions sociales et modes de vie	Environnement
RPTypApp_09	Emploi1524_09	RNI Moy_09	Boul_09	EcolElem_09	Pharmacie_09	Densite_09	Bati
RPTypMai_09	Emploi2554_09	ArtComCE_09	BanqueCE_09	EcoleMat_09	MedOmni_09	Menage_Fseule_09	EAU
RPHLMLoc_09	Emploi5564_09	OuvEmp_09	Supermarche_09	College_09	ChiDentiste_09	Menage_famprinc_cplesansenfants_09	Agri
RPOccProp_09	ActOqp_09	Profinter_09	Restaurant_09			Menage_famprinc_monoparentale_09	Vegetation
RPLogTot_09	EmploiDpt_09	SAP_09	Veterinaire_09			Menage_famprinc_cpleaveceenfants_09	
	EmploiComm_09	AgrExpl_09	BouChar_09			Menage_Hseul_09	
		Ndip_ns_09	Poste_09			Ret_09	
		Cadint_09	Superette_09				
		BacBP_ns_09	Epicerie_09				
			GardPrescol_09				

LES DONNÉES ET DE LA ZONE D'ÉTUDE 2/2

Le complexe littoral/estuaire de la Gironde :



- 303 communes pour une surface de 1062000 hectares.
- 3 départements (Gironde, Landes, Charente-Maritime).
- “Territoires de l'eau” délimités par les frontières hydrographiques.
- Délimitation de 5 SAGE (Schéma d'Aménagement et de Gestion des Eaux).
- Complexe littoral/estuaire de la Gironde considéré comme un socio-écosystème.

LA MÉTHODE CLUSTOFVAR

- La méthode *ClustOfVar* [Chavent, Kuentz, Liquet et Saracco, 2012] est une méthode de classification ascendante hiérarchique de variables. Elle maximise un **critère d'homogénéité** basé sur la notion de **corrélation** pour les variables **quantitatives** et de **rapport de corrélation** pour les variables **qualitatives**.
- L'homogénéité $H(C_k)$ de la classe C_k est une mesure d'adéquation entre les variables de la classes et la **variable synthétique quantitative** de la classe notée y_k . Elle est définie par :

$$H(C_k) = \sum_{x_j \in C_k} r^2_{x_j, y_k} + \sum_{z_j \in C_k} \eta^2_{y_k | z_j} .$$

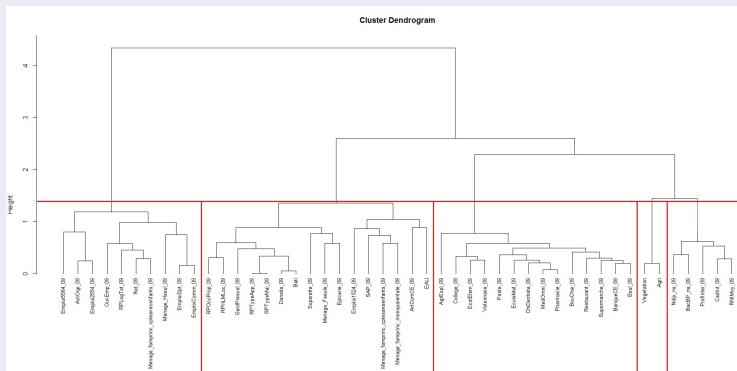
Où r^2 désigne la corrélation de Pearson au carré entre y_k et la **variable quantitative** x_j et η^2 désigne le rapport de corrélation entre y_k et la **variable qualitative** z_j .

- La variable synthétique quantitative y_k est la variable “la plus liée” aux variables de la classe au sens du critère H qu'elle maximise. C'est la première composante principale issue de PCAmix (ACP pour données mixtes) [Chavent, Kuentz, Liquet et Saracco, 2012].

APPLICATION DE CLUSTOFVAR 1/5

Afin de choisir un nombre de clusters pertinent on utilise le dendrogramme issu de la classification de variables avec ClustOfVar.

Le dendrogramme issu de ClustOfVar



La “découpe la plus naturelle” est en 4 clusters de variables, cependant on choisit le niveau inférieur, plus fin, en 5 clusters afin de conserver un cluster relatif à l’environnement contenant les variables “Agriculture” et “Végétation”.

APPLICATION DE CLUSTOFVAR 2/5

Suite à l'application de *ClustOfVar* sur notre zone d'étude, nous avons retenu la partition en $k = 5$ **clusters de variables**, le tableau ci dessous montre le lien entre les variables de chaque cluster et la variable synthétique associée.

Les clusters de variables

Variable synthétique 1		
Variables	Squared Loadings	Corrélation
Pharmacie_09	0.89	NA
MedOmn_09	0.82	NA
Boul_09	0.80	NA
BanqueCE_09	0.76	NA
EcoleElem_09	0.73	NA
EcoleMat_09	0.73	NA
ChiDentiste_09	0.73	NA
Supermarche_09	0.72	NA
Restaurant_09	0.67	NA
Veterinaire_09	0.62	NA
BouChar_09	0.60	NA
College_09	0.59	NA
Poste_09	0.59	NA
AggExpl_09	0.24	-0.49

Variable synthétique 2		
Variables	Squared Loadings	Corrélation
RPTypApp_09	0.91	-0.96
RPTypMai_09	0.91	0.95
Bati	0.75	-0.87
Densite_09	0.74	-0.86
RPHMLoc_09	0.61	-0.78
RPOccProp_09	0.60	0.77
GardPrescol_09	0.50	NA
Epicerie_09	0.35	NA
Menage_Fseul_09	0.34	-0.58
Menage_famprinc_cplesansenfants_09	0.27	0.52
SAP_09	0.17	-0.41
Superette_09	0.14	NA
Menage_famprinc_monoparentale_09	0.11	-0.34
Emploi1524_09	0.06	0.24
ArtComCE_09	0.04	0.20
EAU	0.02	-0.13

Variable synthétique 3		
Variables	Squared Loadings	Corrélation
RNI Moy_09	0.75	0.87
Ndlip_ns_09	0.73	-0.85
Cadint_09	0.71	0.84
BacBP_ns_09	0.53	0.72
Profinter_09	0.49	0.70

Variable synthétique 4		
Variables	Squared Loadings	Corrélation
Menage_famprinc_cpleavecevenfants_09	0.72	-0.85
Ret_09	0.60	0.78
EmploiDpt_09	0.58	-0.76
EmploiComm_09	0.57	0.75
RPLogTot_09	0.57	-0.75
Emploi2554_09	0.47	-0.69
ActDap_09	0.38	-0.62
OuvEmp_09	0.27	-0.52
Menage_Hseul_09	0.22	0.47
Emploi5564_09	0.18	-0.42

Variable synthétique 5		
Variables	Squared Loadings	Corrélation
Agri	0.9	-0.95
Vegetation	0.9	0.95

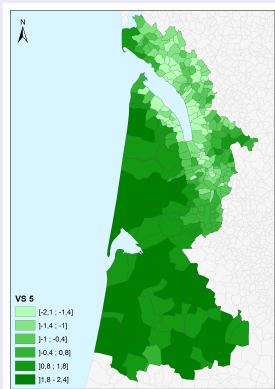
Interprétation des variables synthétiques

- VS 1 : Accès aux services.
- VS 2 : Conditions de logement.
- VS 3 : Niveaux de diplôme, Catégories socio-professionnelles et revenus.
- VS 4 : Conditions familiales et accès à l'emploi.
- VS 5 : Environnement.

APPLICATION DE CLUSTOFVAR 3/5

Les VS obtenues avec ClustOfVar sont des indices de conditions de vie résumant l'information apportées par plusieurs variables. On peut donc représenter sur une carte les valeurs de ces indices.

Carte de la VS 5



Interprétation de la VS 5

- Au Nord de la Zone et le long de l'estuaire, on remarque des faibles valeurs de la VS5, celle-ci étant corrélée négativement avec la variable "Agriculture", on en déduit que ces territoires sont majoritairement agricoles.
- Inversement, au Sud de la zone et sur certaines communes littorales, on observe des fortes valeurs de la VS5 qui est corrélée négativement avec la variable "Végétation", ces territoires sont donc des territoires végétalisés.

APPLICATION DE CLUSTOFVAR 4/5

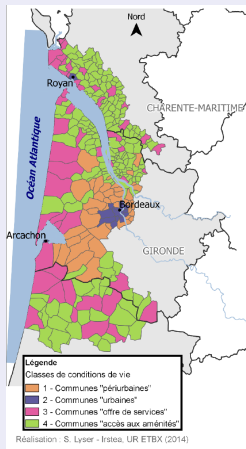
Afin de mieux identifier les spécificités des différentes communes, nous effectuons une **typologie des communes en 4 classes** à l'aide d'une CAH sur les 5 variables synthétiques obtenues à l'aide de ClustOfVar. Le tableau ci dessous montre quelles sont les variables les plus caractéristiques de chaque classe. Ceci va nous aider à mieux caractériser les classes.

Classes de communes	Variables synthétiques	V.test	Distribution intra-classe		Distribution globale
			Moyenne	Ecart type	Ecart type
Classe 1	VS3	9.95	2.48	1.74	1.79
	VS1	8.68	3.73	1.60	3.08
	VS5	3.21	0.60	1.15	1.34
	VS2	-2.35	-0.84	1.79	2.55
	VS4	-7.50	-2.23	1.01	2.14
Classe 2	VS1	6.59	5.52	0.11	3.08
	VS3	3.56	1.73	1.88	1.79
	VS2	-13.92	-9.65	2.62	2.55
Classe 3	VS4	9.28	2.56	1.45	2.14
	VS1	8.37	3.34	1.67	3.08
	VS5	3.02	0.52	1.25	1.34
	VS2	-3.20	-1.06	1.32	2.55
Classe 4	VS2	10.12	1.10	0.79	2.55
	VS4	-2.23	-0.20	1.74	2.14
	VS5	-5.26	-0.30	1.35	1.34
	VS3	-8.51	-0.65	1.24	1.79
	VS1	-15.70	-2.06	1.08	3.08

Rq: Par construction les VS sont centrées

APPLICATION DE CLUSTOFVAR 5/5

Carte de la typologie des communes



Description des classes

- **Les communes "périurbaines"** : proches de Bordeaux / peu denses / accès aux services très important ;
- **Les communes "urbaines"** : Bordeaux et voisins immédiats / forte densité / large offre de services ;
- **Les communes "offre de services"** : majorité de la façade littorale / proportion de retraités importante / offre d'emploi à l'échelle de la commune / offre de services importante ;
- **Les communes "accès aux aménités"** : conditions de vie proches des communes du rural isolé / maisons individuelles / espaces agricoles / très faible accès aux services.

UTILISATION DE CLUSTOFVAR DANS

```
#Load package and data
library(ClustOfVar)
data(QLV)
dat.quantif <- QLV[,c(1:35)]
dat.qualif <- QLV[,c(36:47)]

#Perform ClustOfVar
tree <- hclustvar(X.quantif=dat.quantif, X.qualif=dat.qualif)

#Plot of the dendrogram (slide 8)
plot(tree)


#Cut the dendrogram in 4 groups of variables
treecut <- cutreevar(tree, k=4)

#List of matrices of squared loadings (slide 9).
treecut$var

#Vector indicating the cluster to which each variable is allocated.
treecut$cluster

#Dataframe with n rows and k columns containing synthetic variables
treecut$scores
```

CONCLUSION

- La méthode ClustOfVar apparait comme une alternative intéressante aux méthodes classique de réduction de dimension (ACP, ACM).
- Des VS plus facile à interpréter que les composantes principales d'une ACP par exemple.
- Parallèlement à cette étude nous avons mené une enquête afin de mieux comprendre les perceptions des individus vis à vis de leur qualité de vie. L'analyse des résultats de l'enquête est en cours.
- Le package  **ClustOfVar** est disponible sur le CRAN.

BIBLIOGRAPHIE

- 1 Chavent M., Kuentz V., Liquet B., Saracco J., ClustOfVar: An R Package for the Clustering of Variables, Journal of Statistical Software. Vol. 50, pp. 1-16.
- 2 Chavent M, Kuentz-Simonet V, Saracco J, (2012), Orthogonal rotation in PCAMIX, *Advances in Data Analysis and Classification*, 6 : 131-146.
- 3 Noll Heinz-Herbert , Towards a European System of Social Indicators: Theoretical Framework and System Architecture, Social Indicators Research, 2002, vol. 58, issue 1, pages 47-87
- 4 Reynard R., Vialette P., “Une approche de la qualité de vie dans les territoires”, Insee Première n 1519, octobre 2014.

MERCI DE VOTRE ATTENTION...