

Approches pour l'analyse quantitative des **trajectoires biographiques** en socio-démographie

Séminaire "Temporalité, perceptions et
traitements" (PIREH-SAMM)

21 février 2014

Nicolas Robette

Laboratoire Printemps (CNRS-UVSQ)



Émergence de l'analyse des parcours de vie

- Elder (1974): « *Children of the Great Depression* »
- Fin 70's-80's: perspective de recherche importante
- Parallèlement au développement :
 - des **données longitudinales micro-individuelles** (panels, enquêtes rétrospectives, biographiques)
 - des **méthodes d'analyse statistique**

Les données longitudinales

des données individuelles

des données qui relatent pour chaque personne son histoire dans plus ou moins de détails

elles peuvent être incomplètes, fragmentaires, tronquées, interrompues, censurées...

→ Un **corpus complexe de récits tronqués**

Données prospectives

Panels, suivis de cohorte

Registres de population, EDP, systèmes de suivis démographiques

+ : fiabilité des données

- : coût, délais, attrition, « ménage »

Les enquêtes biographiques

+ : moins cher, immédiat

Fiabilité ?

la question de la remémoration

des outils pour la faciliter:

mode d'interview

fiche *ageven*

La fiche *ageven*

Une grille âge-événement, chronologique

3 domaines traditionnels: familial, résidentiel et professionnel; mais aussi nvx domaines, perceptions

Unités de d'observation complexifiées

Une approche « atomiste »

Unité d'analyse	événement
Objectif	modélisation de probabilités de transitions/durée
Genèse du parcours	processus stochastique (paramétrique)
Démarche	explicative, causale
Méthodes	modèles de Markov, analyse de survie, modèle de Cox, multilevel-multiprocess...

Degenne, 2001: la sortie de chômage

Variables à expliquer : Log de la durée de la première période de chômage

Variables	Df	Coefficient β	Erreur standard	χ^2	Pr> χ^2	Rapport des chances
Constante	1	2.18022407	0.010477	43303.73	0.0001	
Hommes		(réf.)				
Femmes	1	0.1933963	0.004204	2116.703	0.0001	0.817
16-17	1	-0.5682302	0.018342	959.6933	0.0001	1.812
18-19	1	-0.4984859	0.010471	2266.477	0.0001	1.685
20-21	1	-0.4169744	0.009211	2049.43	0.0001	1.547
22-23	1	-0.3197611	0.009019	1256.934	0.0001	1.397
24-25	1	-0.1951142	0.009224	447.4156	0.0001	1.227
26-27	1	-0.0971934	0.009617	102.1322	0.0001	1.107
28-29	1	-0.0441713	0.009987	19.5626	0.0001	1.047
30-31		(réf.)				
32-33	1	0.03883473	0.010859	12.78948	0.0003	0.960
34-35	1	0.06609906	0.011275	34.37005	0.0001	0.933
36-37	1	0.09581869	0.011571	68.56992	0.0001	0.905
38-39	1	0.10776655	0.011923	81.70199	0.0001	0.893
40-41	1	0.12774174	0.012211	109.4429	0.0001	0.875
42-43	1	0.19531385	0.012506	243.9212	0.0001	0.815
44-45	1	0.20358319	0.012809	252.6135	0.0001	0.808
Lic. économique	1	0.22074852	0.006661	1098.279	0.0001	0.794
Lic. autre	1	0.21452524	0.007211	885.1399	0.0001	0.799
Démission	1	-0.1362495	0.008442	260.4591	0.0001	1.153
Fin contrat		(réf.)				
Fin intérim	1	-0.0306424	0.010568	8.407707	0.0037	1.033
Autres	1	0.09540754	0.006697	202.9704	0.0001	0.905

Blossfeld & Mills: fécondité et mariage

Partial Likelihood Estimates of the Transition
from Consensual Union to Marriage (Final Model)
for West and East Germany, Canada, Latvia and the Netherlands

Covariates	Final Model Results by Country				
	West Germany	East Germany	Canada	Latvia	Netherlands
Pregnancy/Birth Process					
Time-dependent (1)					
[time before pregnancy]	-1.2595	-0.6179	-1.0768	-1.3918	-1.0909
month of pregnancy	0.1131	0.1729	-0.1157	0.3822	-0.2217
1 month since pregnancy	0.4783	0.2715	0.7107	0.2009	0.3769
2 months since pregnancy	0.8837 *	0.4225	1.0851 *	1.0109 *	0.9374 *
3 months since pregnancy	1.026 *	0.7723 *	0.5849	1.2959 *	1.3229 *
4 months since pregnancy	0.8578 *	1.3903 *	0.6563	1.0817 *	1.5587 *
5 months since pregnancy	0.9905 *	0.7938 *	0.248	0.9328 *	1.0743 *
6 months since pregnancy	0.8701 *	0.151	-0.8948	0.7525 *	0.0227
7 months since pregnancy	0.8158 *	-0.5166	-0.0365	0.4793	0.1028
8 months since pregnancy	-0.8121 *	-2.5449 *	-0.5693	-0.4727	-0.235
Month of birth	-1.4709	-0.6254	-0.1115	-1.6669	-1.2711
1-3 months after birth	-0.7513	0.2875	0.0096	-0.0136	-0.4595
4-6 months after birth	-0.7638	0.1351	0.0363	-1.3576 *	-0.4404
More than 7 months after birth	-0.9877 *	-0.0921	-0.5263 *	-1.2336 *	-1.6771 *
Birth cohort (2)					
1965-69	-0.3094	-0.6001	-0.4341 *	-1.3096 *	-2.2829 *
1960-64	-0.17	-0.0536	-0.3589 *	-0.8563 *	-1.4258 *
1955-59	-0.1486	0.092	-0.4324 *	-0.6154	-0.8228 *
[1950-54]	0	0	0	0	0
Historical period					
[Before 1974]	0	0	0	0	0
1974-83	0.0882	0.3521	-0.3027	0.001	-0.2488
After 1983	-0.1554	0.0363	-0.2905	-0.3164	-1.7642 *
Highest education level					
Low	0.1722 *	-0.0189	0.1563	-0.0164	0.249 *
[Medium]	0	0	0	0	0
High	-0.0354	0.0941	-0.1092	-0.0763	-0.1962 *
Educational enrollment					
Time-dependent					
In school	-0.3575 *	0.0061	-0.3187	0.27	-0.1856
[Out of school]	0	0	0	0	0

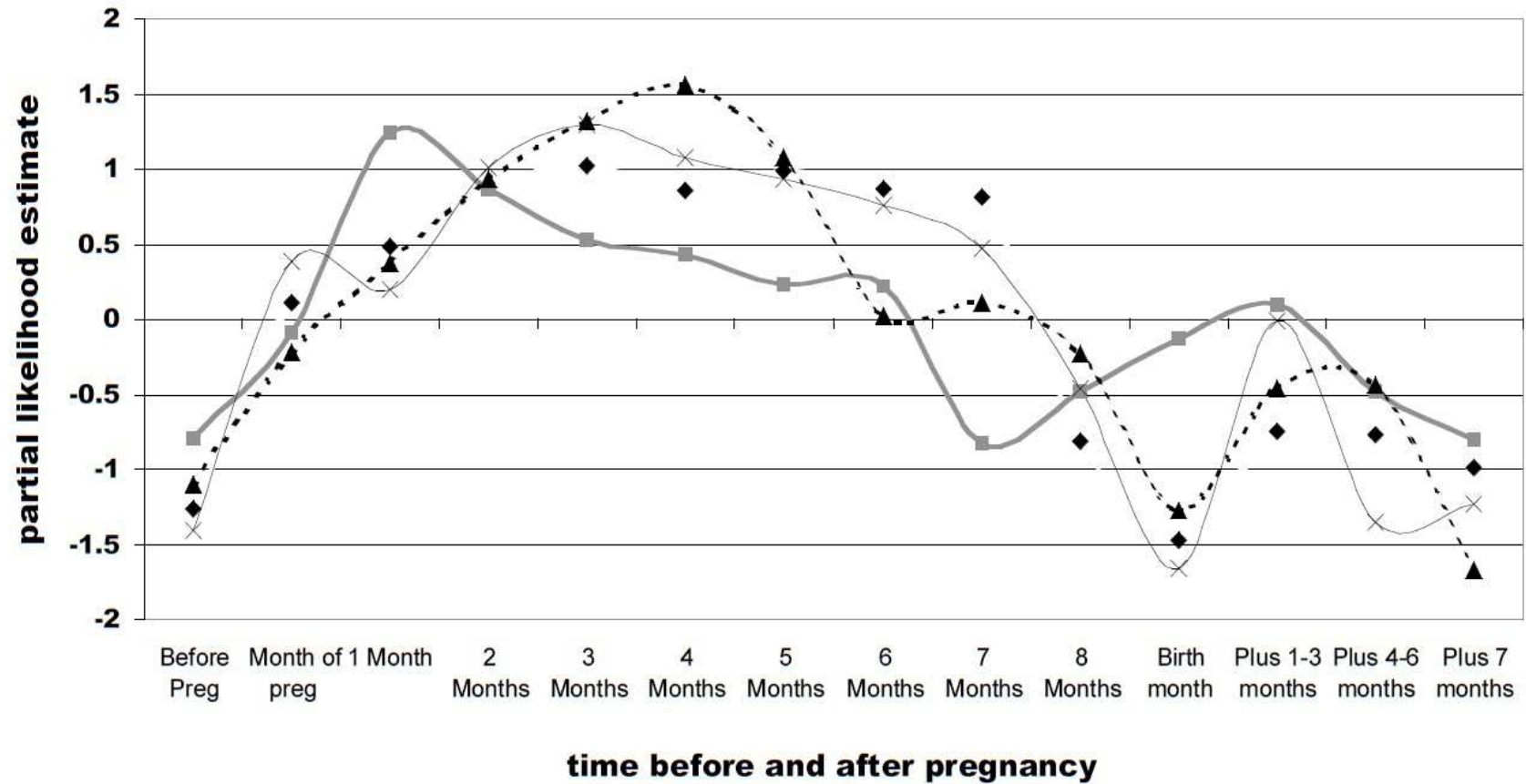
* = Significant at the 0.05 level. Results are shown for the final model 5.

Notes: (1) First covariate coded as centered effects, all others as comered effects. Reference groups denoted by brackets.

(2) Birth cohorts for West and East Germany are represented by 1968-72, 1963-67, 1958-62 and 1953=57.

Source: Blossfeld et al. (1999) for West and East Germany and Mills and Trovato (forthcoming 2001) for Canada, Latvia and the Netherlands.

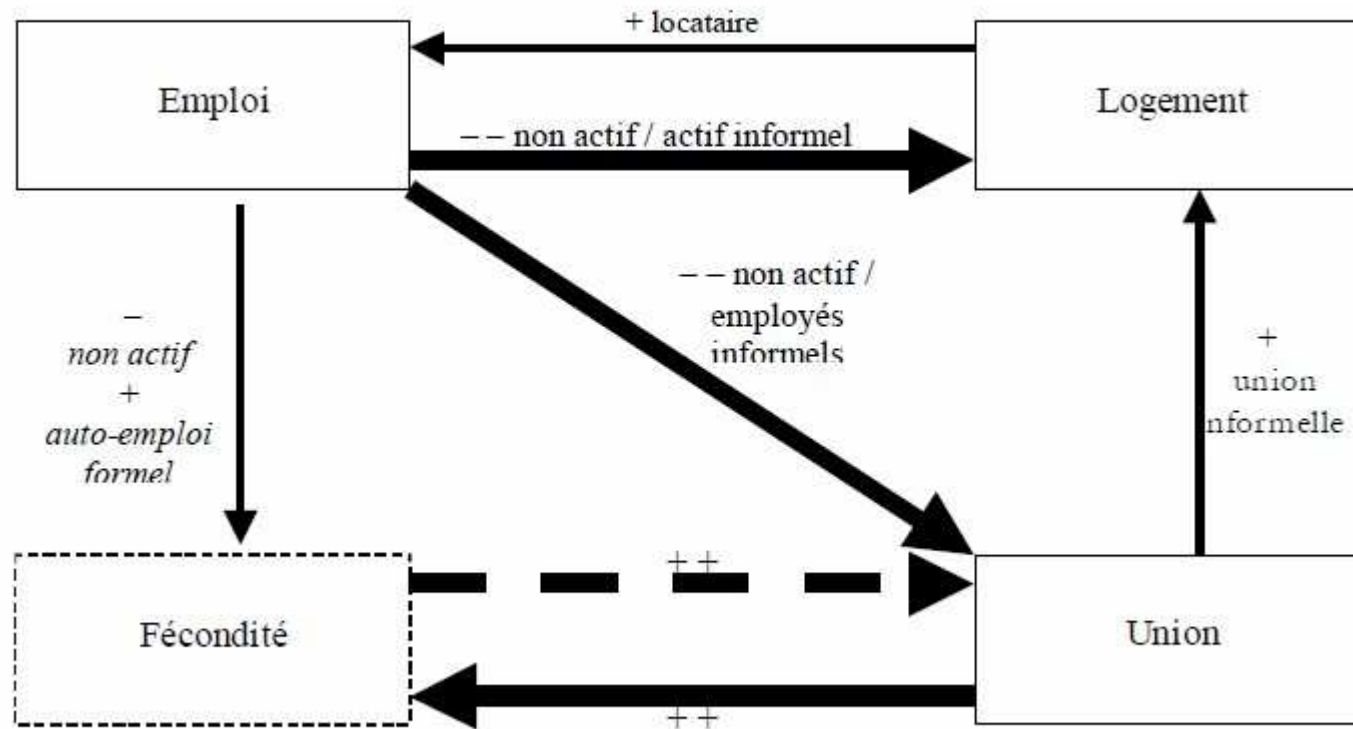
Blossfeld & Mills: fécondité et mariage



◆ West Germany —■— Canada - -▲- -Netherlands —×— Latvia

Agwanda *et al*, 2004

Figure 3 : Influences réciproques des principaux événements marquant l'entrée dans la vie adulte à Nairobi (Hommes)⁷



Source : Agwanda *et al.*, 2004

Une approche « holiste »

Unité d'analyse	parcours
Objectif	recherche des différences/régularités
Genèse du parcours	inconnue (non paramétrique)
Démarche	descriptive, exploratoire, culture "algorithmique"
Méthodes	mesures de similarité (analyse factorielle, de séquences...), classifications (CAH, arbres de décision...)

Démarche typologique

Similarité entre l'ensemble des parcours

→ Matrice de distance

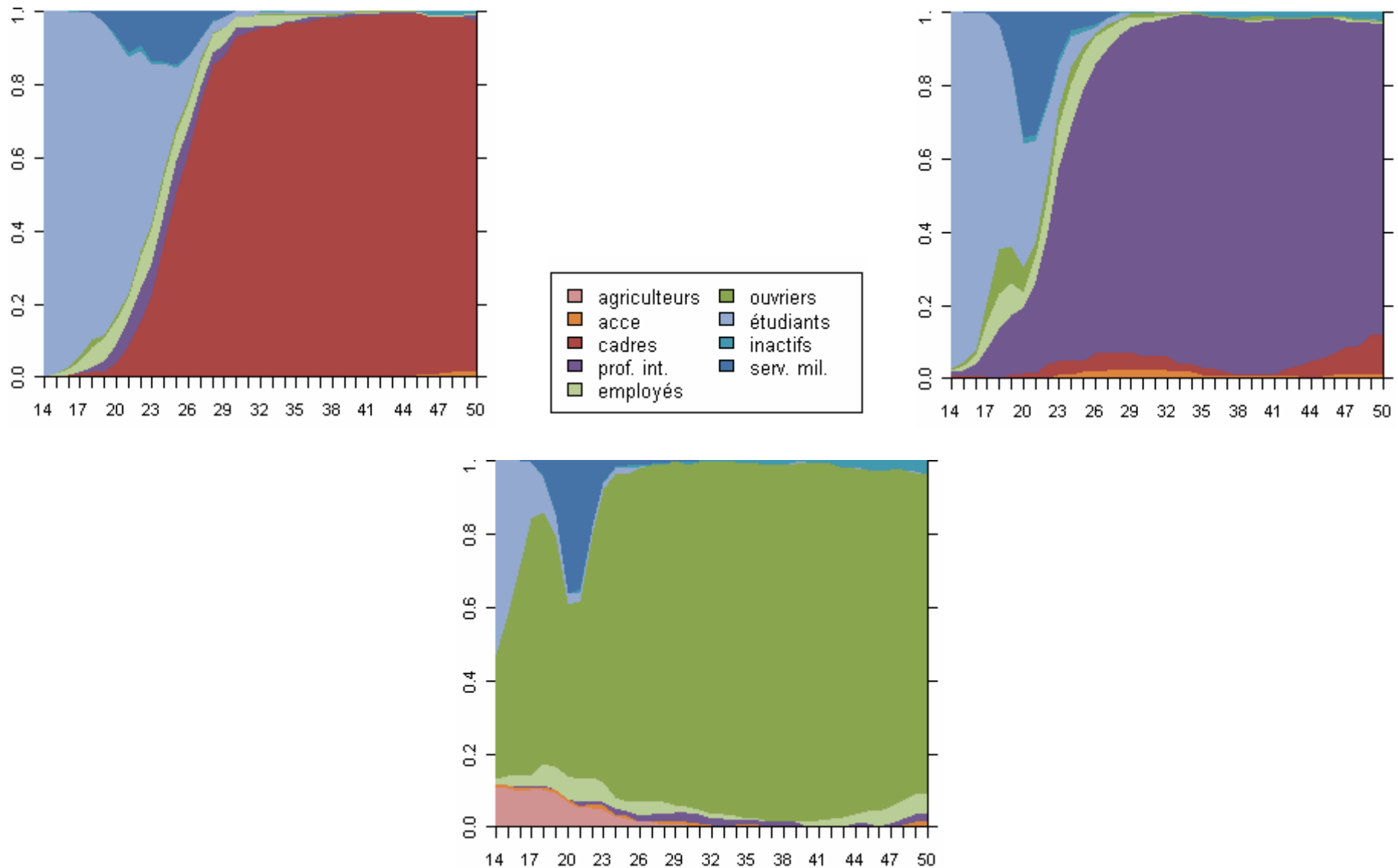
→ Classification (CAH, ...)

→ **Typologie** de parcours

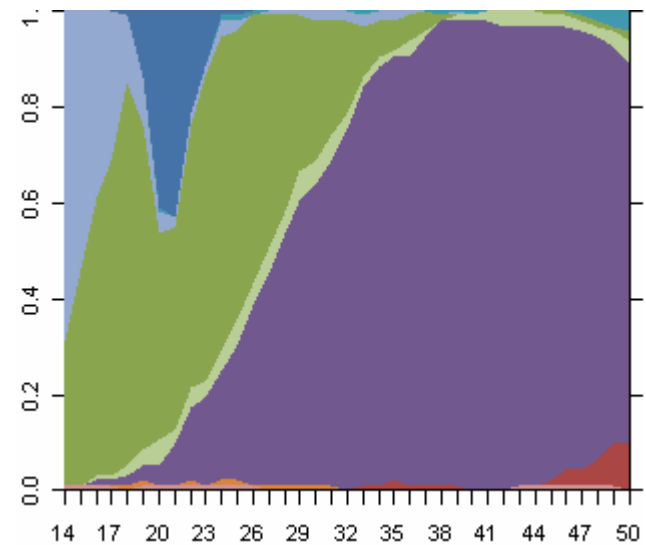
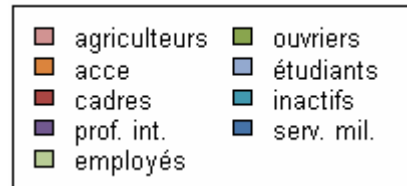
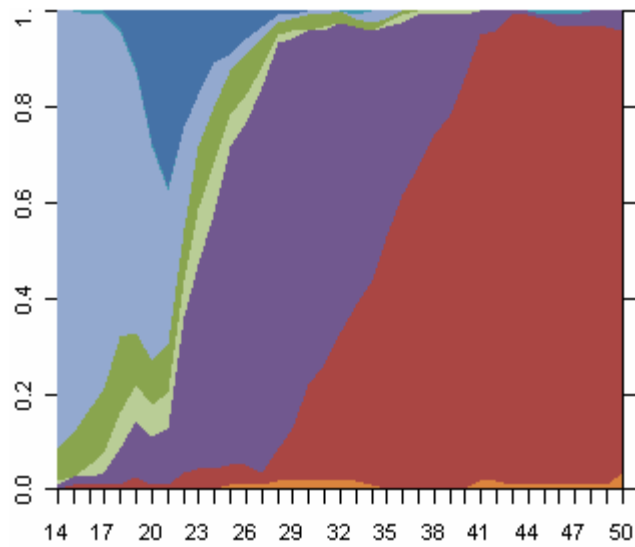
Exemple de typologie (1)

- Enquête *Biographies et entourage* (INED, 2001)
- **Carrières professionnelles** de 1421 hommes
- 37 années, de 14 à 50 ans
- 9 états:
 - 6 catégories socioprofessionnelles,
 - étudiant,
 - service militaire,
 - autre inactivité

Exemple de typologie (2)



Exemple de typologie (3)



Exemple de trajectoire

Trajectoire d'insertion professionnelle:

E = étudiant

C = chômeur

S = salarié

18	19	20	21	22	23	24	25
E	E	E	C	S	S	S	S

Les méthodes factorielles (1)

18	19	20	21	22	23	24	25
E	E	E	C	S	S	S	S

- **Codage disjonctif**

18E	18C	18S	...	25E	25C	25S
1	0	0	...	0	0	1

avec ACP -> distance euclidienne

avec AFC -> distance du χ^2

→ prise en compte de la **durée** et du **moment**

Les méthodes factorielles (2)

18	19	20	21	22	23	24	25
E	E	E	C	S	S	S	S

Calendrier simplifié (Analyse Harmonique Qualitative)

18-20 E	18-20 C	18-20 S	21-25 E	21-25 C	21-25 S
1	0	0	0	0,2	0,8

avec AFC -> distance du χ^2

→ prise en compte de la **durée** et du **moment**

(moment moins précis, mais moins sensible au « décalage »)

→ permet de « pondérer » des sous-périodes

Les méthodes factorielles (3)

18	19	20	21	22	23	24	25
E	E	E	C	S	S	S	S

- **Indicateurs:**

- Durées

E	C	S
3	1	4

- Transitions

EE	EC	ES	CE	CC	CS	SE	SC	SS
2	1	0	0	0	1	0	0	3

- Nombre d'épisodes, temps d'accès à un état...

Voir (Grelet, 2002)

Les méthodes séquentielles

- Les parcours individuels sont construits comme des **séquences d'états**
- Mesure de la **similarité** entre paires de séquences
- **Algorithmes** = **optimal matching analysis**, ...

Optimal Matching Analysis (1)

- Méthode utilisée en bio-informatique (ADN)
- Introduction dans les sciences sociales par Andrew **Abbott** (années 80)
- **Principe:** mesurer la dissimilarité entre paires de séquences en évaluant le **coût** représenté par la **transformation** de l'une des séquences en l'autre

Voir par exemple (Macindoe & Abbott, 2004)

Optimal Matching Analysis (2)

- 3 opérations élémentaires:
 - insertion
 - suppression
 - substitution

Optimal Matching Analysis (2)

ASTROLOGIE

Optimal Matching Analysis (2)

option 1:

~~A~~STROLOGIE

→ 1 suppression

Optimal Matching Analysis (2)

option 1:

~~A~~SOTROLOGIE

→ 1 suppression, 1 insertion

Optimal Matching Analysis (2)

option 1:

~~A~~SOLOGIE

→ 1 suppression, 1 insertion, 2 substitutions

Optimal Matching Analysis (2)

ASTROLOGIE

Optimal Matching Analysis (2)

option 2 :

SSTROLOGIE

Optimal Matching Analysis (2)

option 2 :

SOTROLOGIE

Optimal Matching Analysis (2)

option 2 :

SOCROLOGIE

Optimal Matching Analysis (2)

option 2 :

SOCIOLOGIE

→ 4 substitutions

Optimal Matching Analysis (2)

- 3 opérations élémentaires:
 - insertion
 - suppression
 - substitution
- à chaque opération est associé un **coût**
- la **distance** entre deux séquences est équivalente au **coût minimal** de transformation de l'une des séquences en l'autre

Le choix des coûts (1)

Question centrale de l'OMA:

- **Substitution:**

conserve la structure temporelle (**moment**)
mais altère les événements (ordre)

- **Insertion/suppression:**

déforme le temps
mais conserve l'**ordre** des événements

Le choix des coûts (2)

- Matrice des **coûts de substitution**:
 - selon **hypothèses théoriques**: hiérarchie des états...
 - à partir des **données**: en fonction des probabilités de transition...
- Coûts d'**insertion/suppression** (*indel*) :
 - **ordre** privilégié → *indel* faible /substitution
 - **moment** privilégié → *indel* élevé /substitution

Les métriques d'Elzinga (2003;2008)

- **Critique** : OMA prend mal en compte l'ordre (substituer A à B ou B à A est équivalent)
- **Plusieurs propositions** :
 - Plus long préfixe commun
 - Plus longue sous-séquence commune
 - Nombre de sous-séquences communes
 - ...

Dynamic Hamming (*Lesnard*)

- **Critique:** Les probabilités de transition dépendent du temps
- **Principe:**
 - Pas d'insertion/suppression
 - Coûts de substitution calculés à chaque période
- Applications à des emplois du temps

Quelles différences ? (1)

- Un grand nombre de méthodes, chacune appréhendant de manière spécifique:
moment + durée + ordre
- Nécessité de comparaisons systématiques
- Premiers résultats → **robustesse**

Quelles différences ? (2)

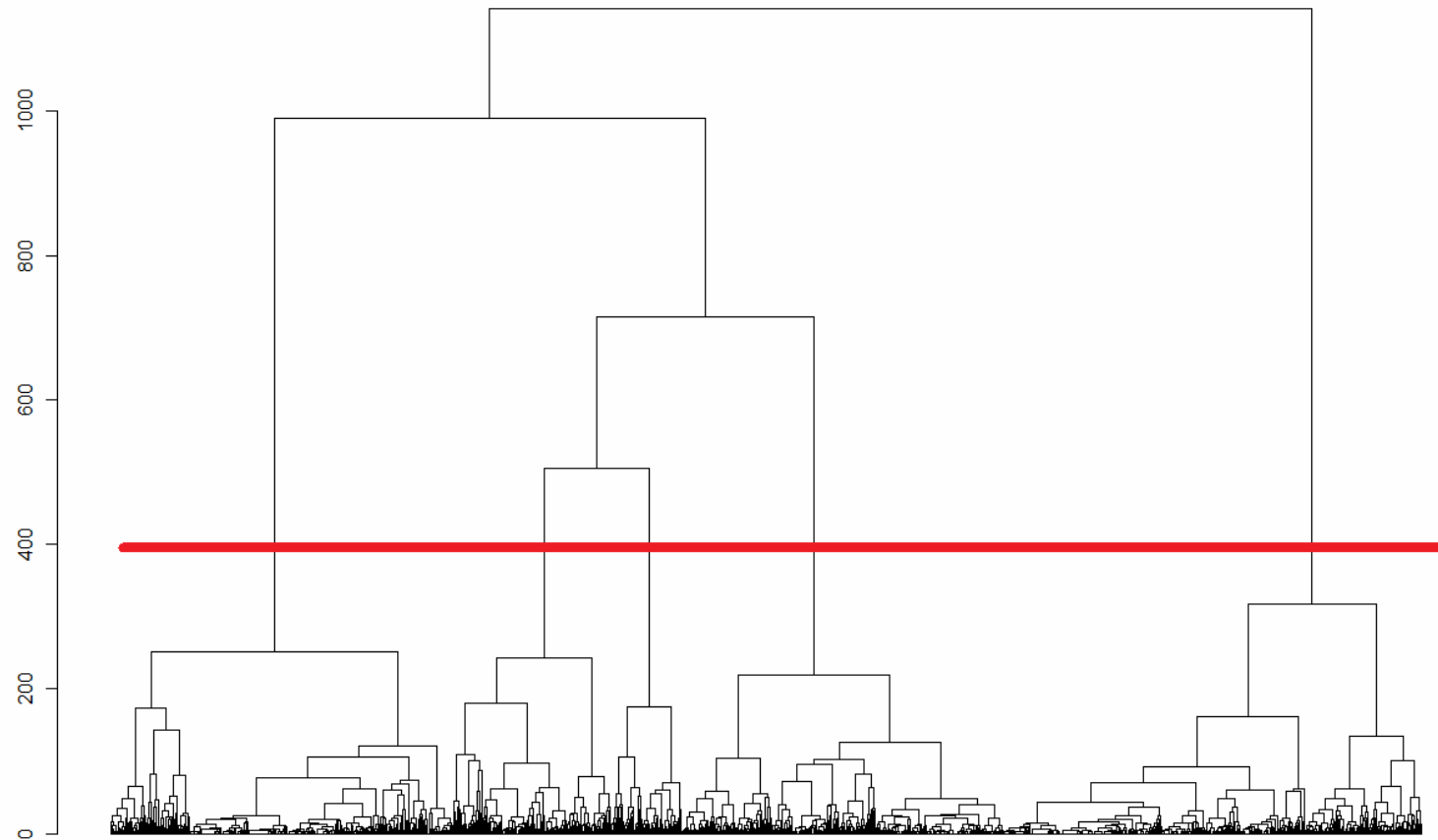
- 3 sous-ensembles =
OMA *vs* AF *vs* Elzinga
- Principales **différences** :
 - Transition *vs* durée (i.e. mobilité *vs* stabilité)
 - AF : parfois petites classes avec *états rares*
 - Parcours très *chaotiques* et/ou totalement distincts

Choix du nombre de classes

- **Selon des indicateurs statistiques:**
 - Inertie, variance inter/intra, critère de Calinski...
- **Empiriquement :**
 - arbitrage entre nb de classes interprétable et prise en compte de la diversité
 - Jusqu'à observer un processus donné, jusqu'à obtenir classes homogènes...
 - Éventuellement, observer à plusieurs niveaux de la classification

Arbre de classification

(=*dendrogramme*)



Typologie en 5 classes

cadres	32,6%
PI	27,0%
ouvriers	26,2%
employés	8,7%
--> acce	5,4%

Comment caractériser les classes ?

- **Indicateurs:**
 - ✓ durées, transitions, nb d'épisodes, temps d'accès...
 - ✓ homogénéité (distance intra, entropie...)
- **Parangon** : trajectoire « idéal-typique »
- **Représentations graphiques:**
 - ✓ chronogrammes
 - ✓ tapis (= *index plots*)

Indicateurs

		cadres	PI	ouvriers	empl	acce
Effectif		437	362	352	117	73
%		32,6	27,0	26,2	8,7	5,4
Entropie		0,310	0,357	0,351	0,434	0,471
Dintra		20,8	23,2	21,9	28,0	31,8
nb états		3,5	3,7	3,0	3,6	4,2
nb transitions		3,0	3,5	2,9	3,5	3,9
durée moyenne	agri	0,0	0,1	0,8	0,8	0,1
	acce	0,2	0,2	1,7	0,4	19,9
	cadre	22,1	0,7	0,0	0,4	2,2
	pi	3,8	23,8	1,8	0,9	4,3
	empl	1,1	2,9	1,0	23,5	2,1
	ouvr	0,7	3,0	28,3	5,5	2,7
	sm	8,0	4,7	1,8	3,2	4,1
	inact	0,1	0,3	0,3	1,2	0,4
	etu	1,1	1,3	1,2	1,1	1,2
nombre d'épisodes	agri	0,0	0,0	0,1	0,1	0,0
	acce	0,1	0,1	0,2	0,1	1,2
	cadre	1,2	0,2	0,0	0,1	0,4
	pi	0,6	1,3	0,2	0,3	0,7
	empl	0,2	0,5	0,2	1,4	0,4
	ouvr	0,2	0,7	1,8	1,0	0,7
	sm	1,1	0,9	0,6	0,7	0,9
	inact	0,1	0,1	0,1	0,2	0,1
	etu	0,6	0,7	0,6	0,6	0,7

Parangons

CADRES: 26 ans → professeur de maths

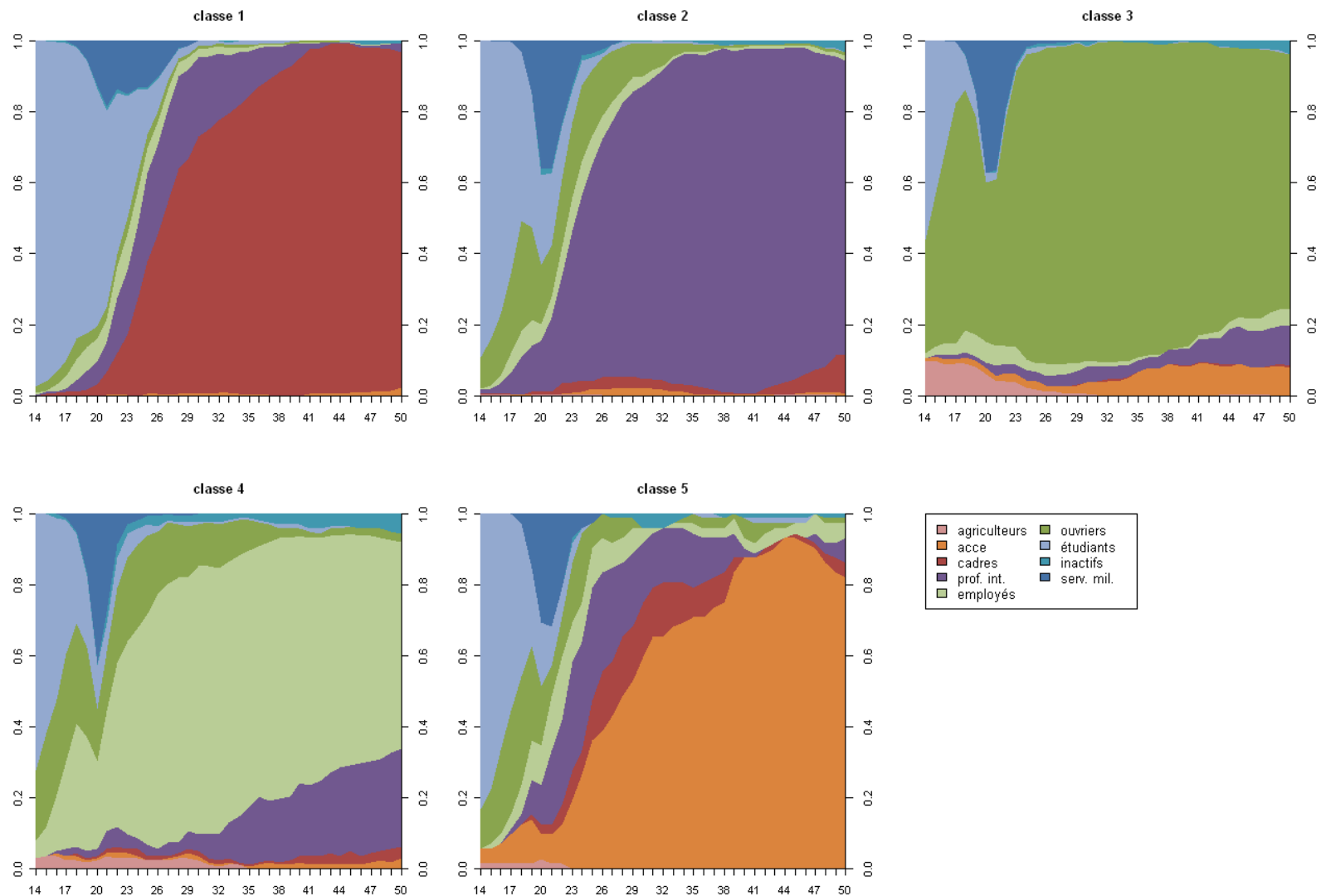
PI: 19 ans → ouvrier (ouvr)
23 ans → chef d'atelier
27 ans → professeur d'atelier

OUVRIERS: 16 ans → manutentionnaire, monteur de cloisons,
chauffeur magasinier

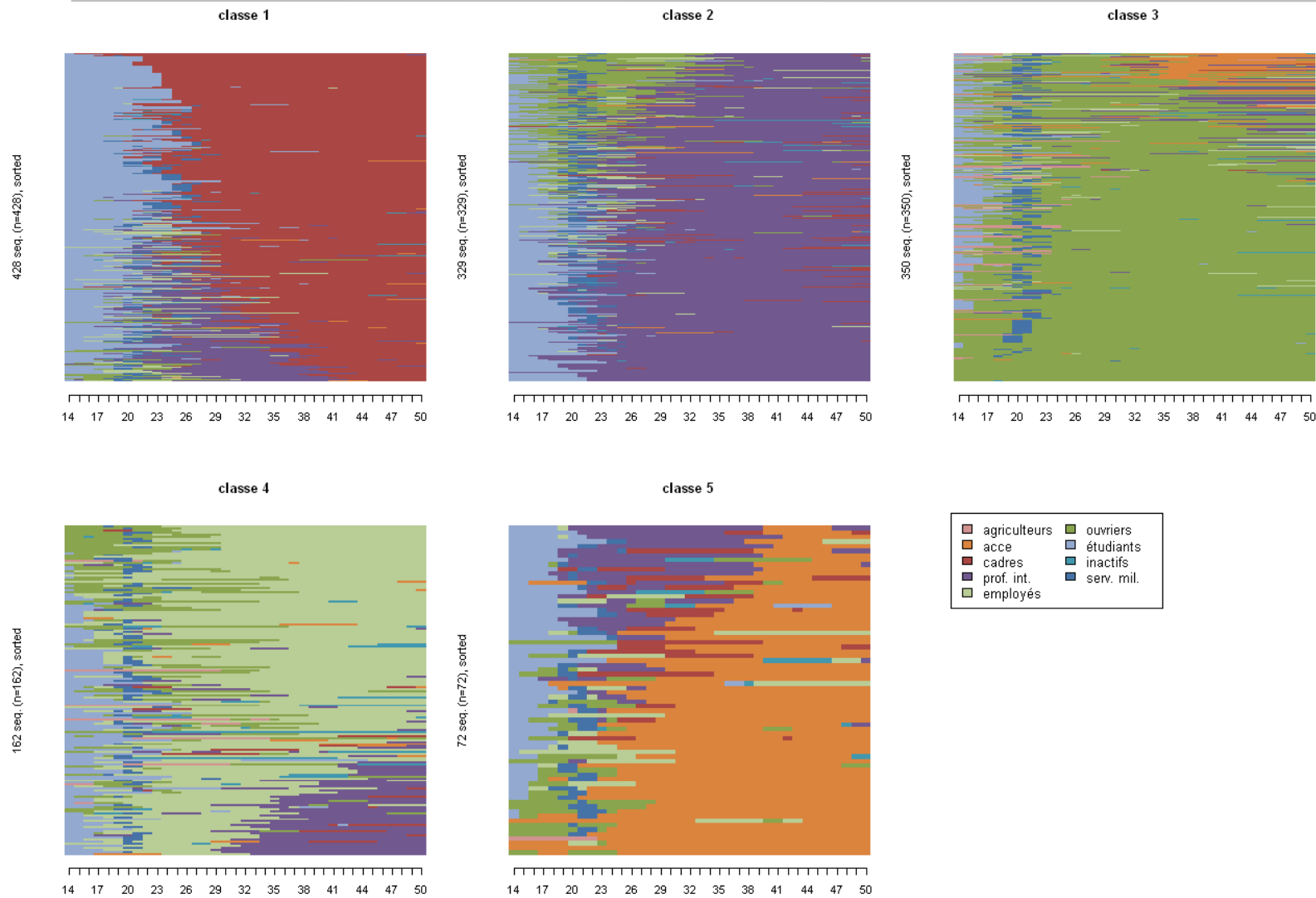
EMPLOYES: 17 ans → maroquinier (ouvr)
24 ans → employé de presse

ACCE: 21 ans → représentant dans l'entreprise familiale (PI)
28 ans → directeur

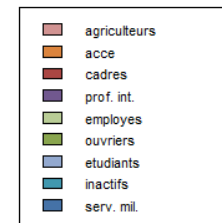
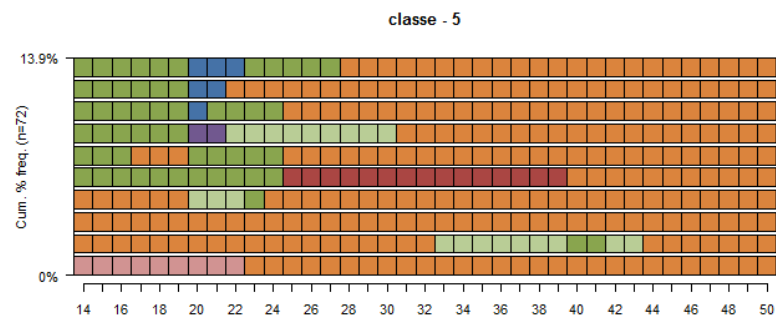
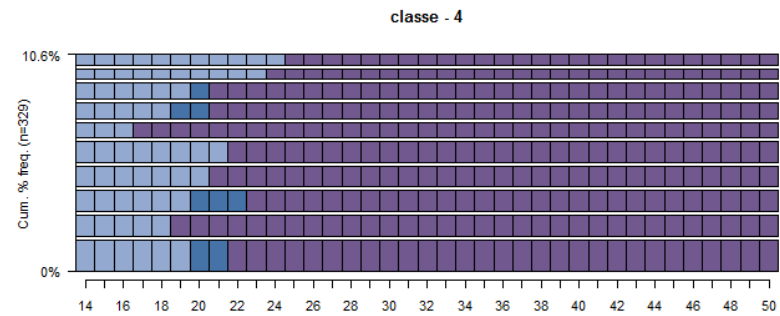
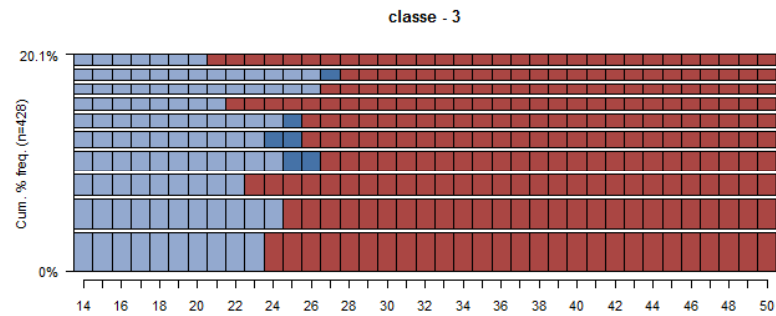
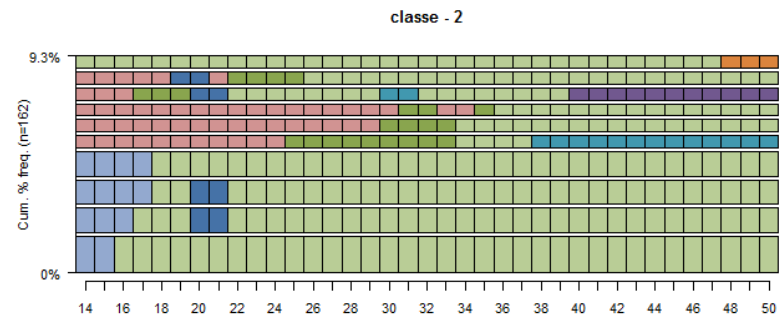
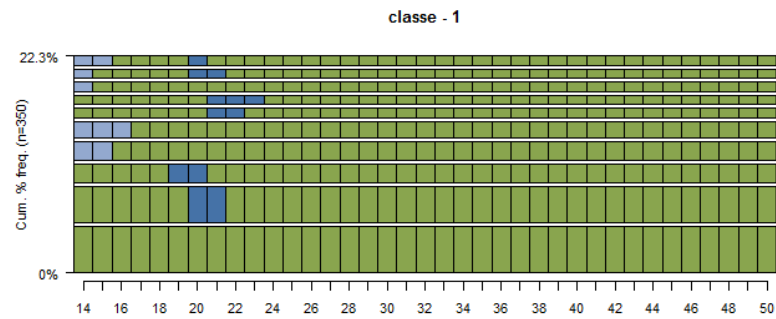
Chronogrammes



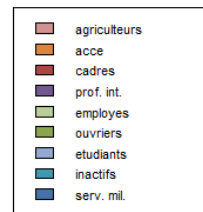
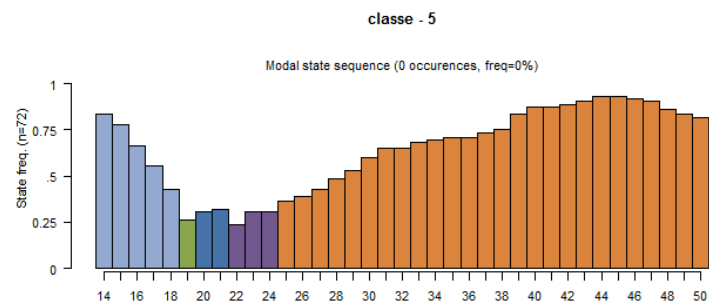
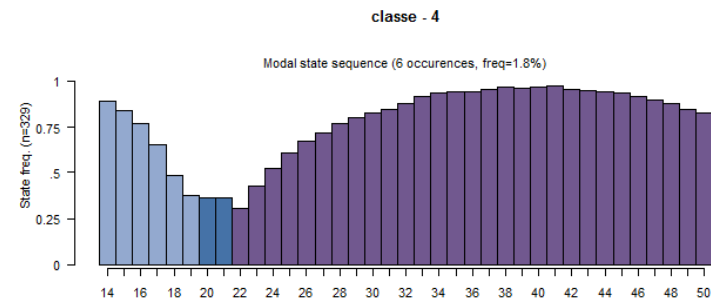
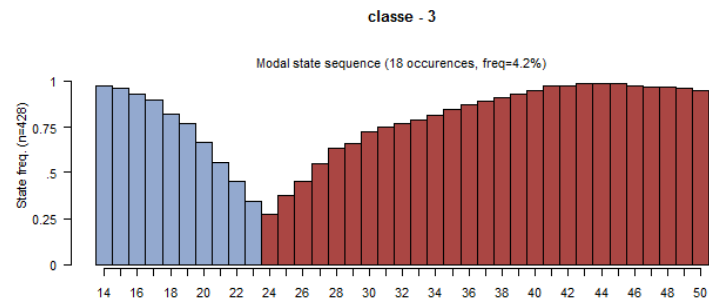
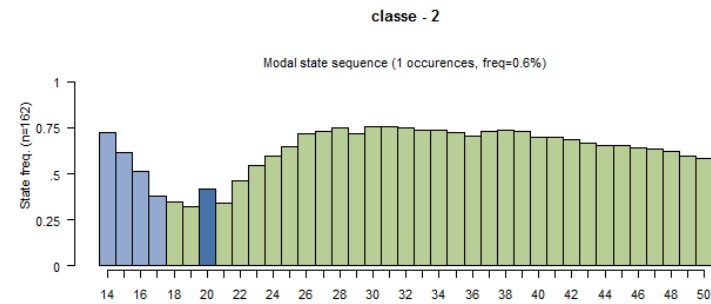
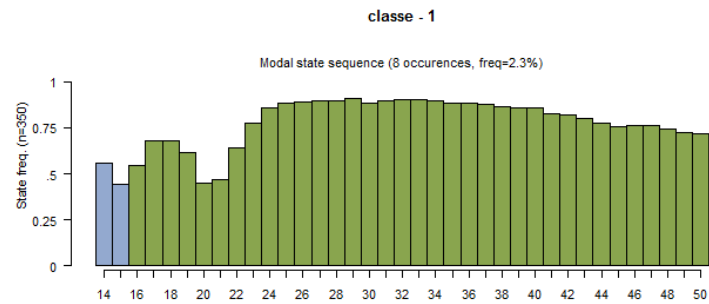
Tapis (ou index plots)



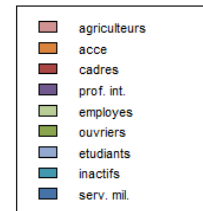
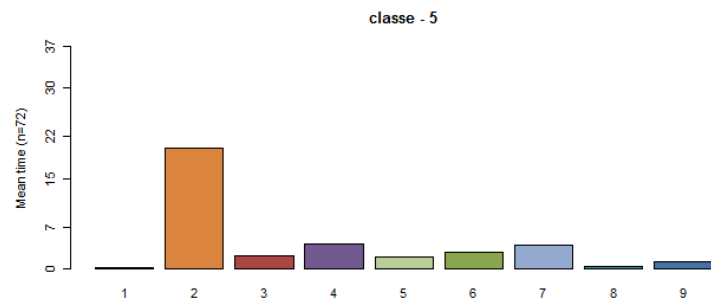
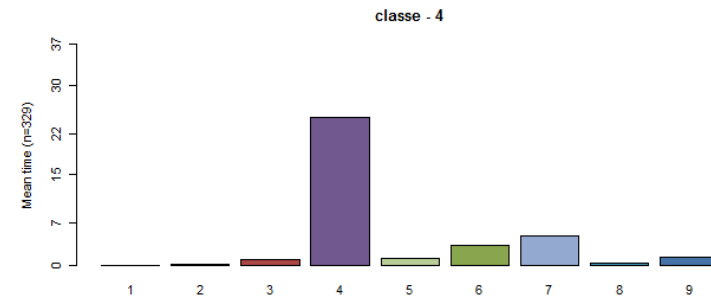
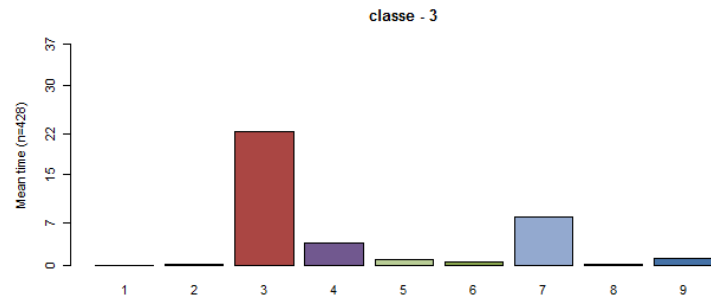
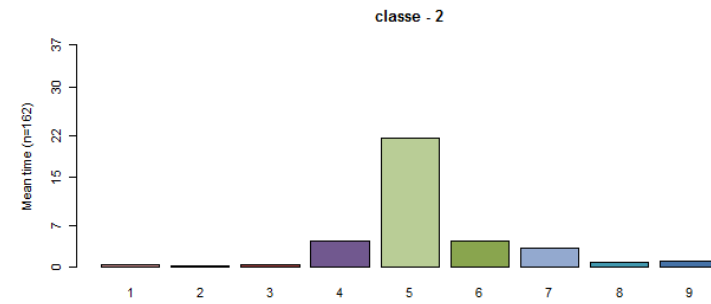
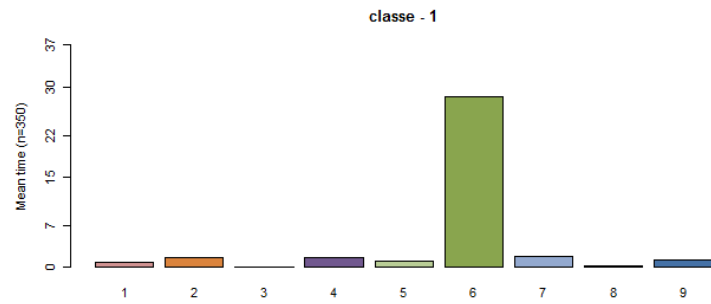
Frequency plots



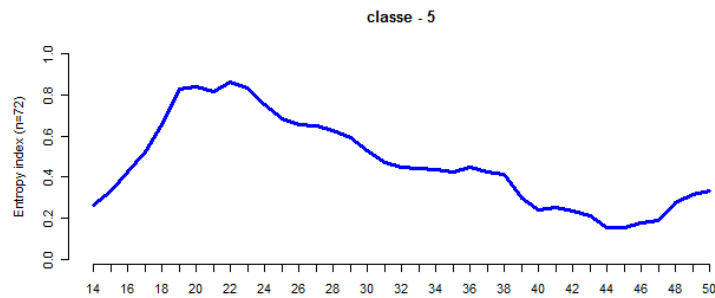
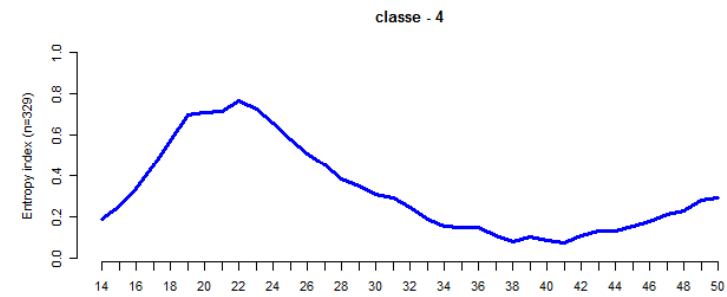
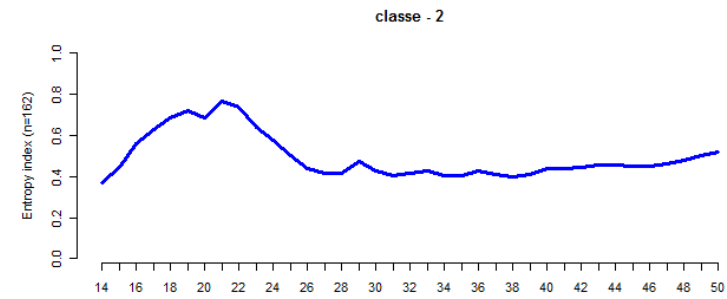
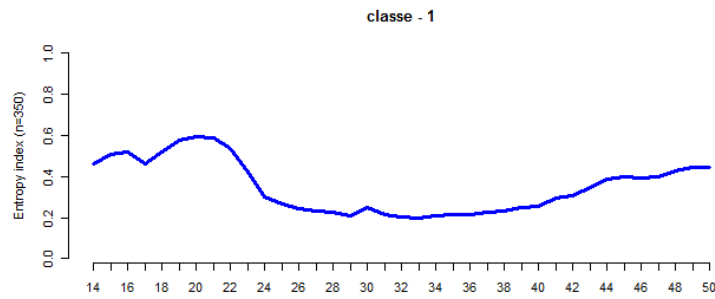
Modal state plots



Mean time plots



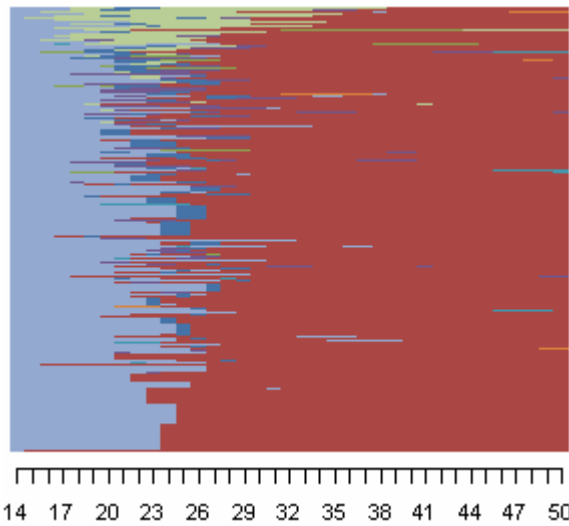
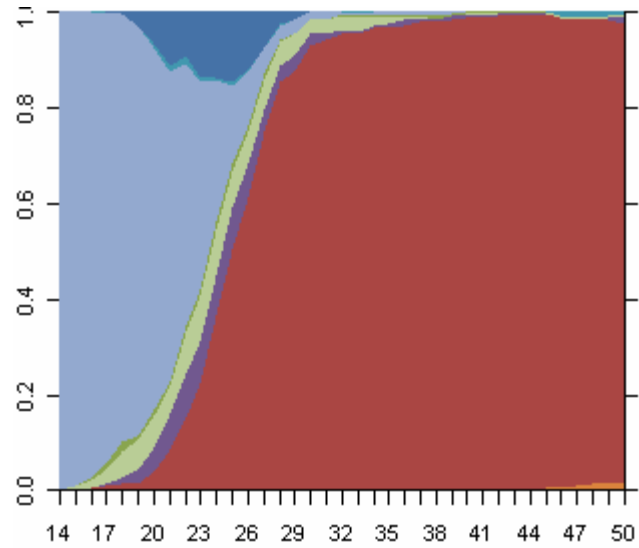
Transversal entropy plots



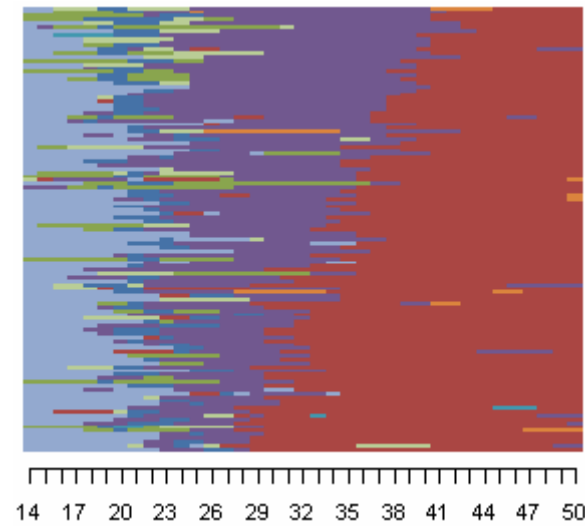
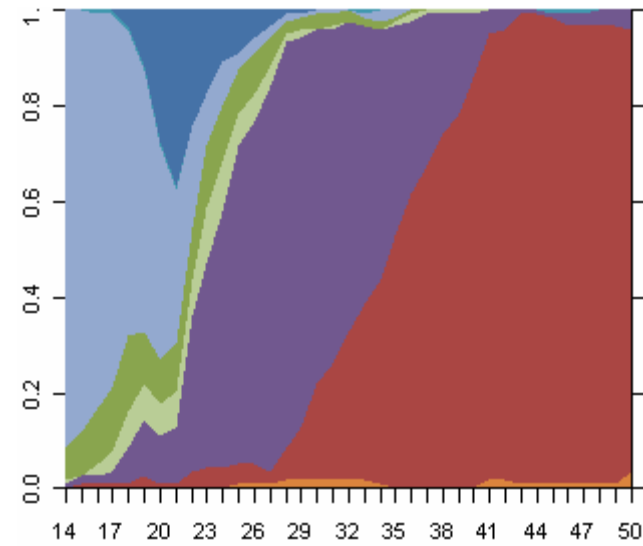
Typologies en 5 et 10 classes

Typologie en 5 classes		Typologie en 10 classes	
cadres	32,6%	cadres	26,2%
		PI --> cadres	6,4%
PI	27,0%	PI	15,5%
		ouvriers --> PI	6,3%
		employés --> PI	5,2%
ouvriers	26,2%	ouvriers	19,2%
		ouvriers --> PI/acce	7,0%
employés	8,7%	employés	6,5%
		ouvriers --> employés	2,2%
* --> acce	5,4%	* --> acce	5,4%

Cadres

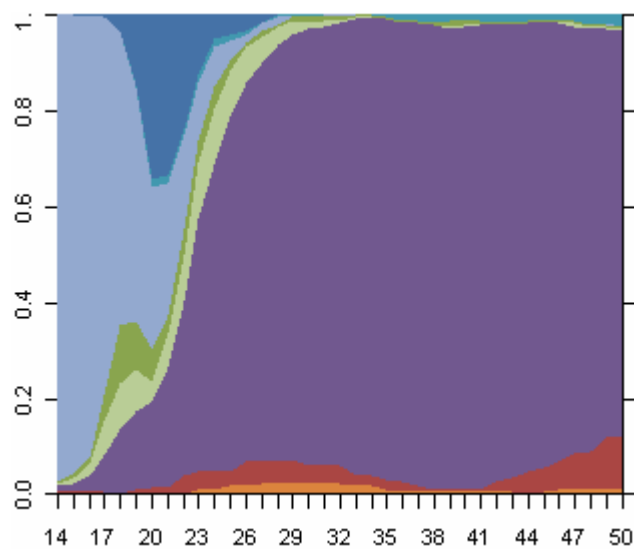


PI → Cadres

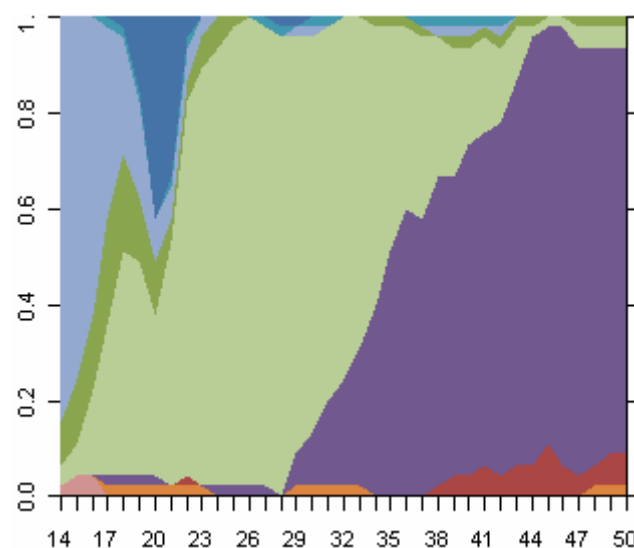
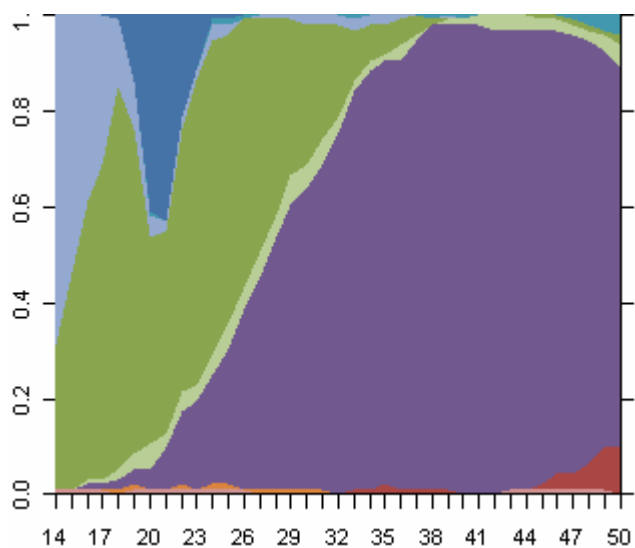


Professions intermédiaires

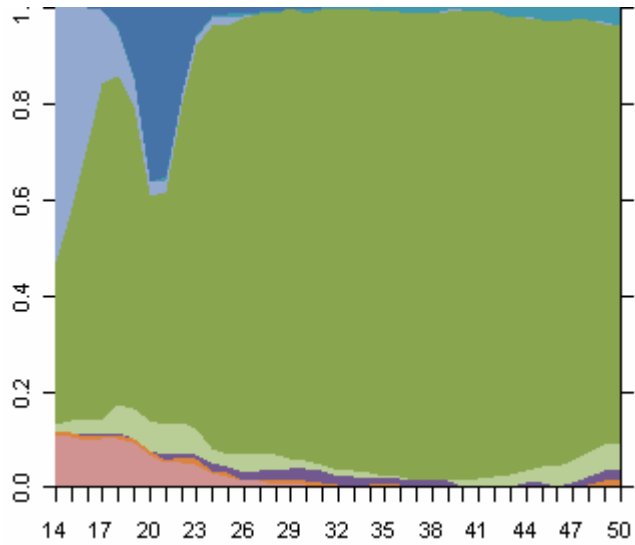
Ouvr \rightarrow PI



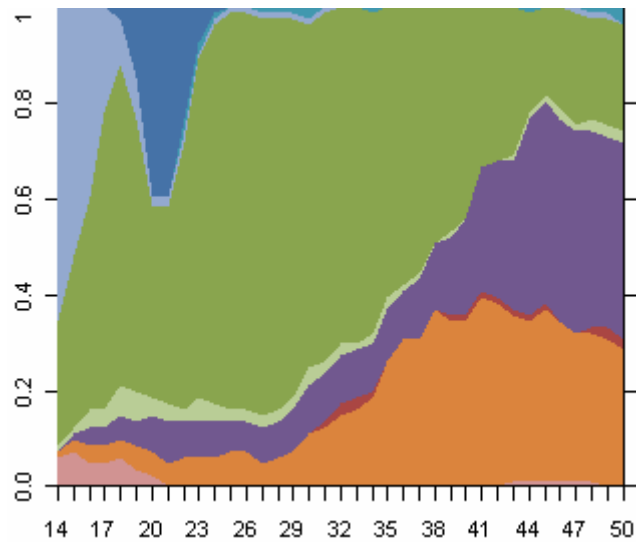
Empl \rightarrow PI



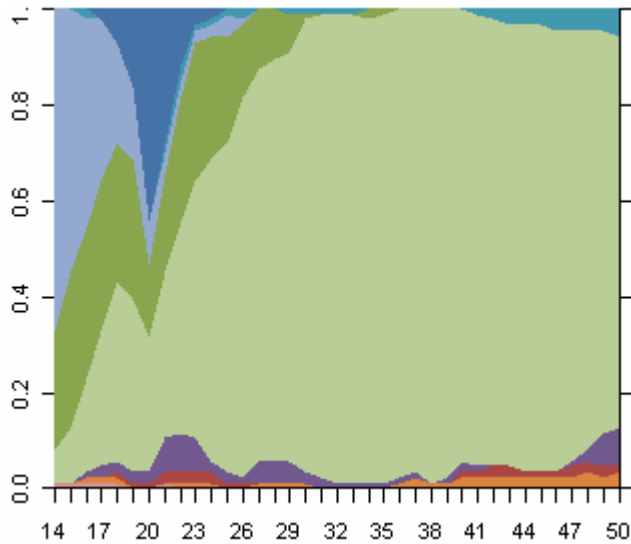
Ouvriers



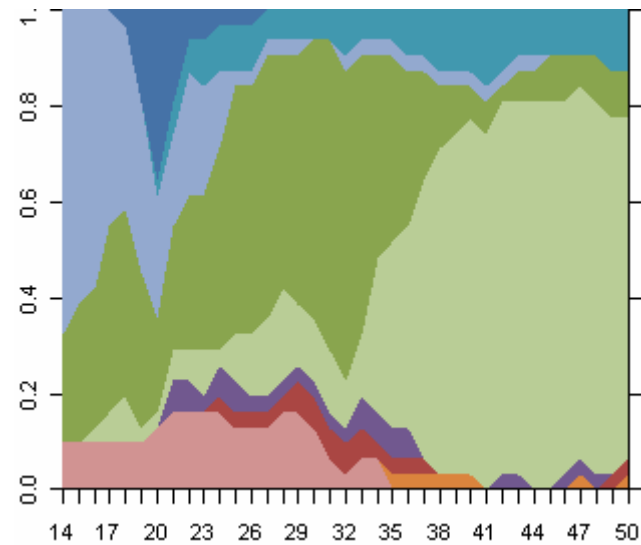
Ouvr \rightarrow PI/acce



Employés



Ouvr \rightarrow empl



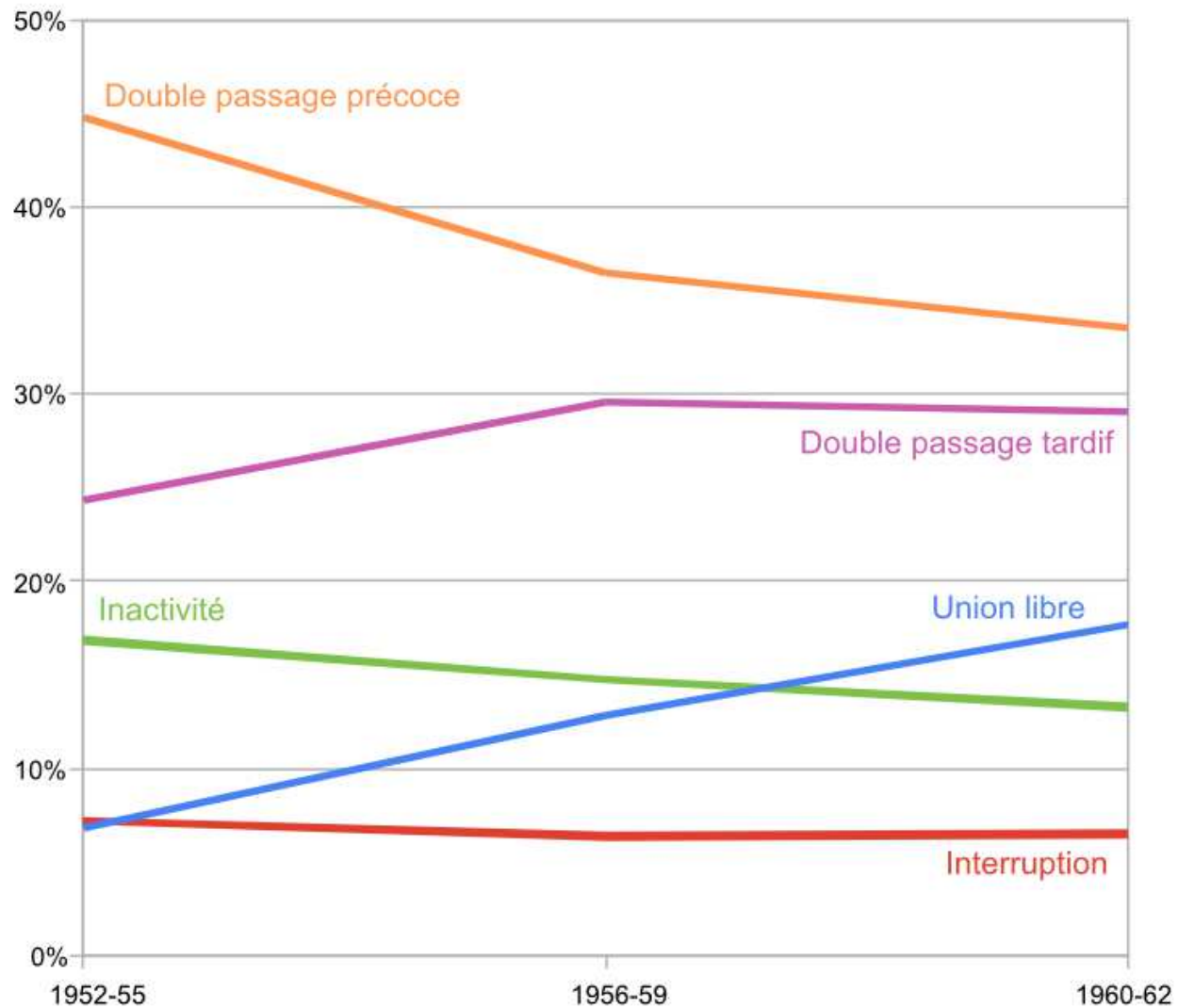
Poursuivre l'exploration

- **Caractérisation** statistique et graphique, à différents niveaux de partition (nb de classes)

Poursuivre l'exploration

- **Caractérisation** statistique et graphique, à différents niveaux de partition (nb de classes)
- Quels sont les **déterminants** de l'appartenance à un type de trajectoire ? *génération, origine sociale...*
 - classe = variable à expliquer

Évolution des trajectoires biographiques des jeunes femmes



Certification scolaire et trajectoire biographique des jeunes femmes

différences de probabilité de « destinée »
(... vs cap/bep, en %)

	interruption	inactivité	dbl passage précoce	dbl passage tardif	union libre
>bac+2	-7	-7	-21	37	-1
bac+2	-7	-6	-5	21	-3
bac	-4	-6	-1	12	-2
<i>cap/bep</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
bepc	1	1	-2	0	1
aucun	2	20	-16	-3	-3

Certification scolaire et trajectoire biographique des jeunes femmes

Rapports de chances de « destinée »

(aucun diplôme vs >bac+2)

	inactivité
1952-1955	5,9
1956-1959	15,6
1960-1962	28,4

(>bac+2 vs aucun diplôme)

	dbl passage tardif
1952-1955	5,9
1956-1959	6,3
1960-1962	6,4

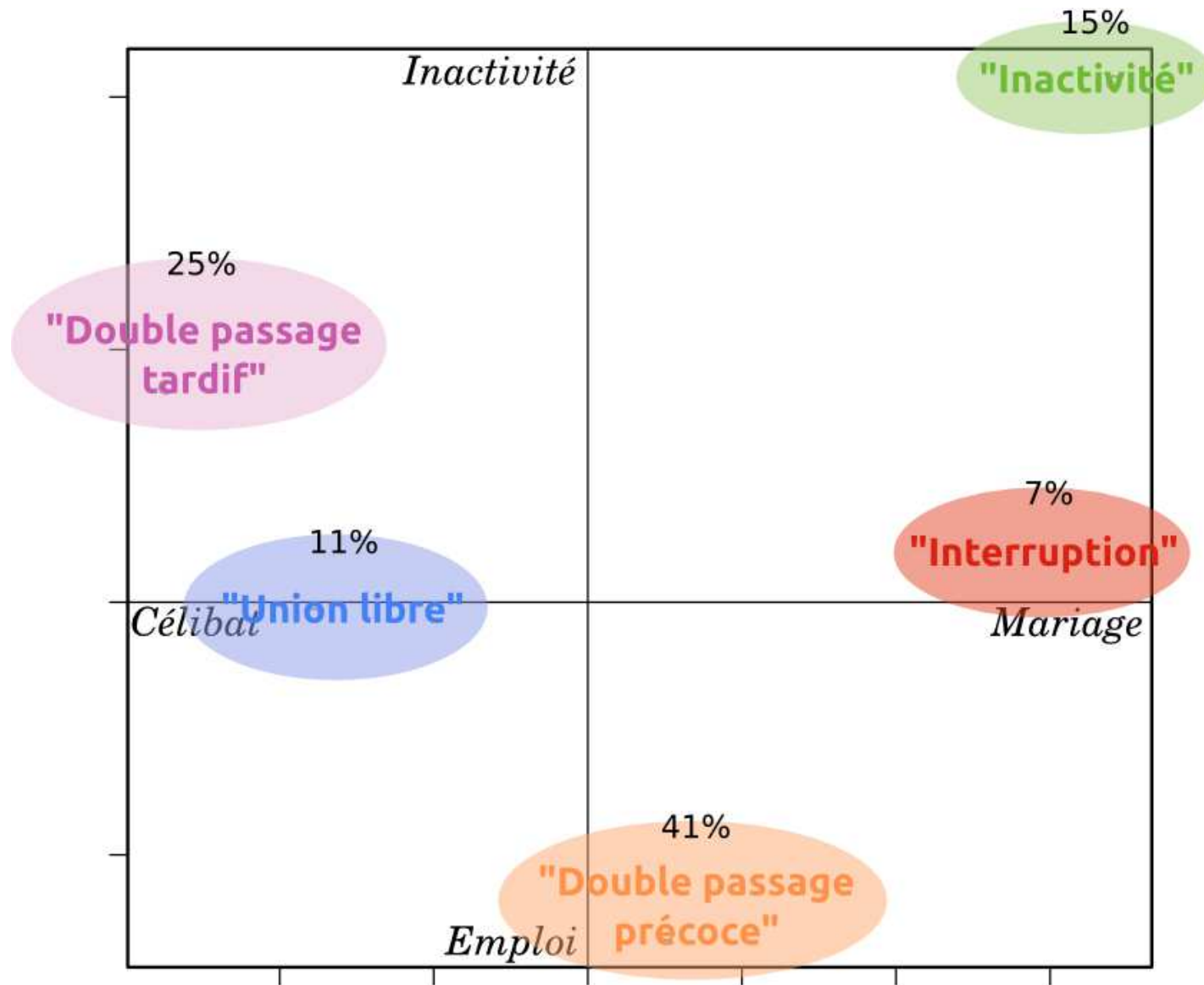
Poursuivre l'exploration

- **Caractérisation** statistique et graphique, à différents niveaux de partition (nb de classes)
- Quels sont les **déterminants** de l'appartenance à un type de trajectoire ? *génération, origine sociale...*
 - classe = variable à expliquer
- La trajectoire a-t-elle des **conséquences** sur le parcours ultérieur ? *passage à la retraite...*
 - classe = variable explicative

Démarches « non-typologiques »

- Projection sur plan factoriel (MDS)

Espace des trajectoires biographiques des jeunes femmes



Abbott & DeViney, 1992

- séquences d'adoption des « programmes » d'assurance sociale
- 5 programmes:
 - ✓ *accident*
 - ✓ *santé, maternité*
 - ✓ *vieillesse, invalidité, décès*
 - ✓ *famille*
 - ✓ *chômage*
- 18 pays développés

Abbott & DeViney, 1992

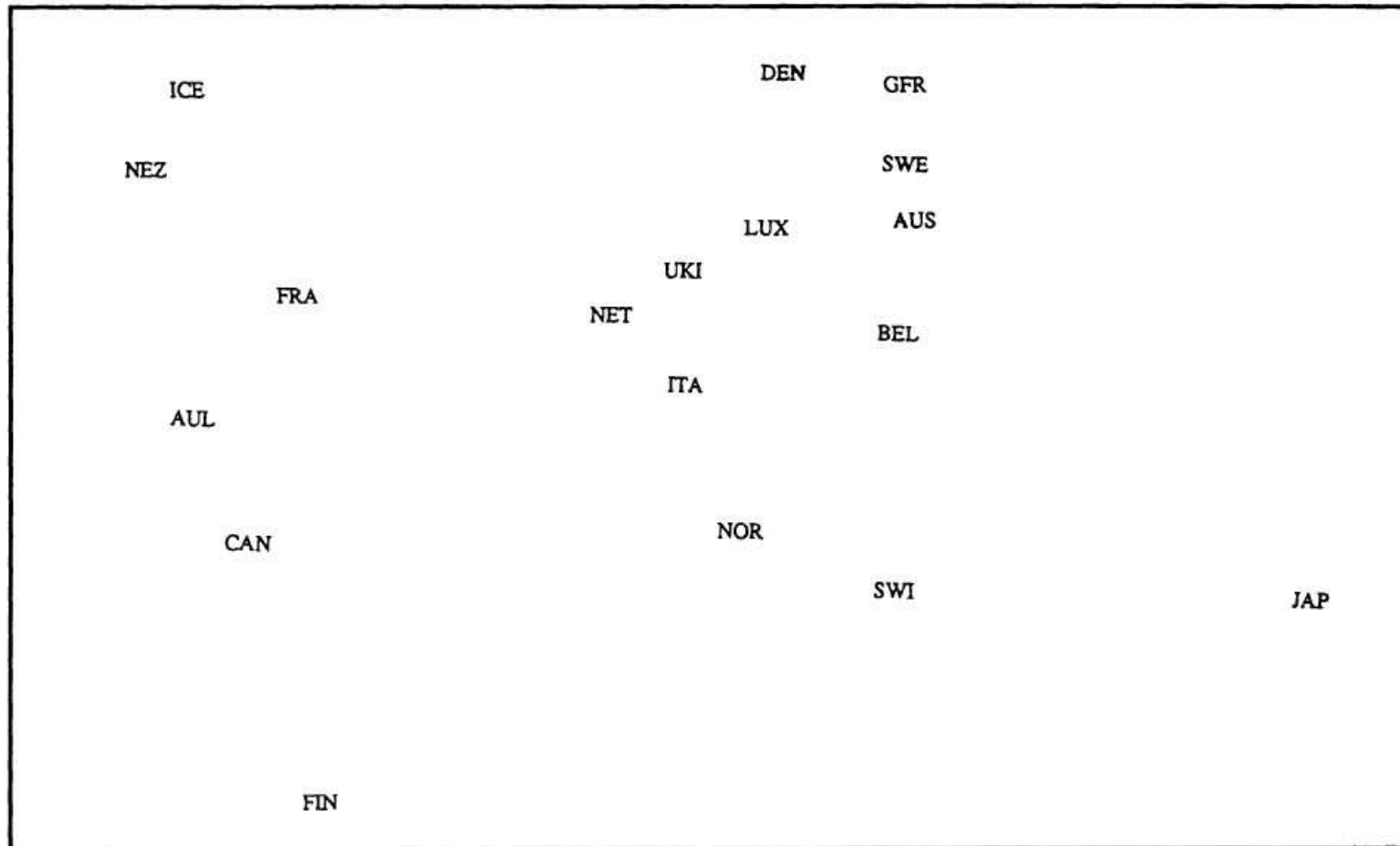


Figure 2 Scaling of distances between welfare sequences

Abbott & DeViney, 1992

- **axe 1 : position de l'assurance santé**
 - % socialistes au parlement (-)
 - part des services dans l'économie (-)

- **axe 2 : position de l'assurance vieillesse**
 - recettes de l'état (+)
 - corporatisme (-)

Démarches « non-typologiques »

- Projection sur plan factoriel (MDS)
- Distance à un parcours de référence, modal, « normal »...

Kogan, 2003

Employment Careers of Immigrants in Germany

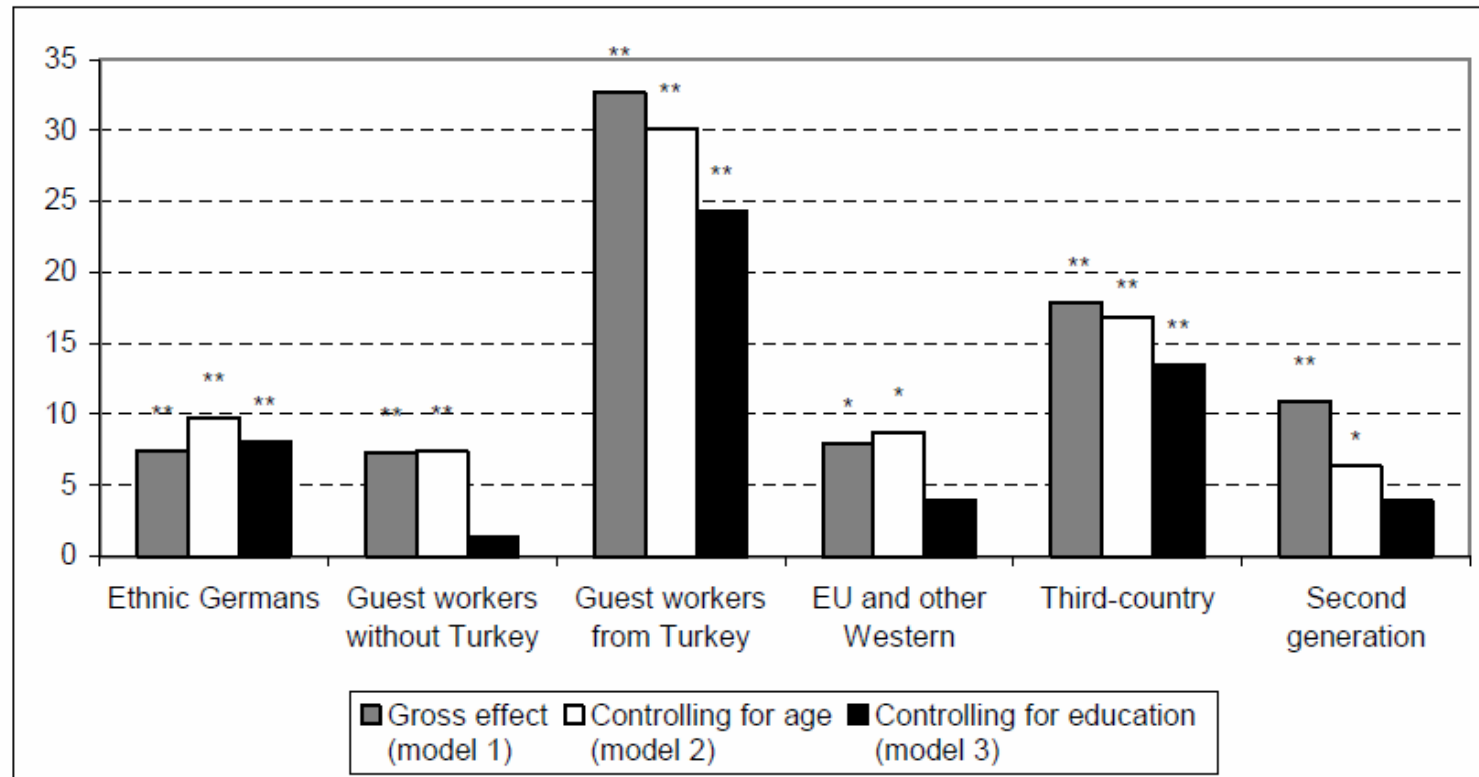


Figure 2: Unstandardized coefficients of the dummy-coded variables pertaining to immigrant groups in the OLS regression analysis of the distance to continuous employment

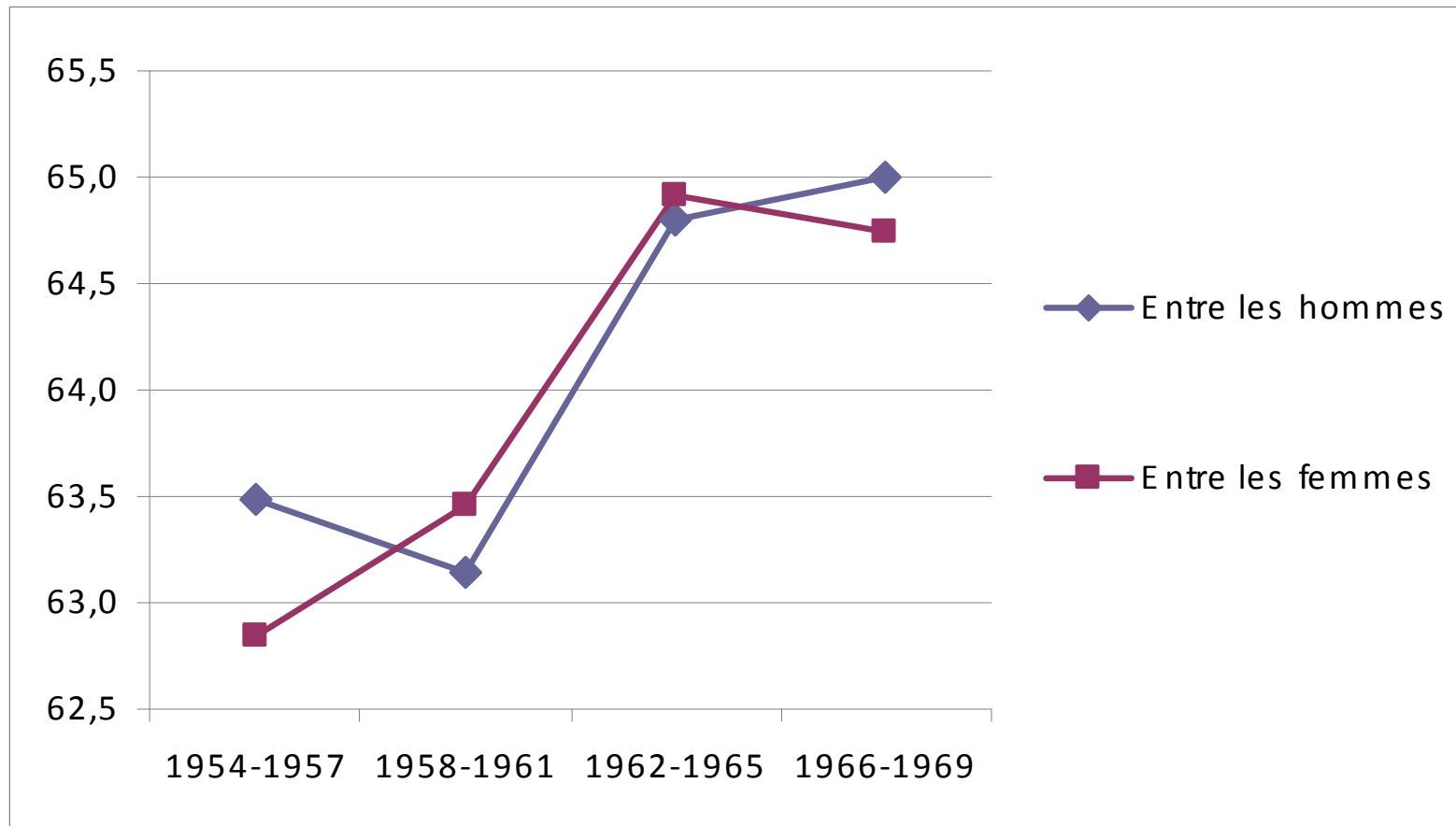
Source: GSOEP, Waves L-Q (1995-2000), monthly unweighted data

Notes: * $p < 0.05$, ** $p < 0.01$

Démarches « non-typologiques »

- Projection sur **plan factoriel** (MDS)
- Distance à un **parcours de référence**, modal, « normal »...
- Similarité moyenne par **groupes** d'individus

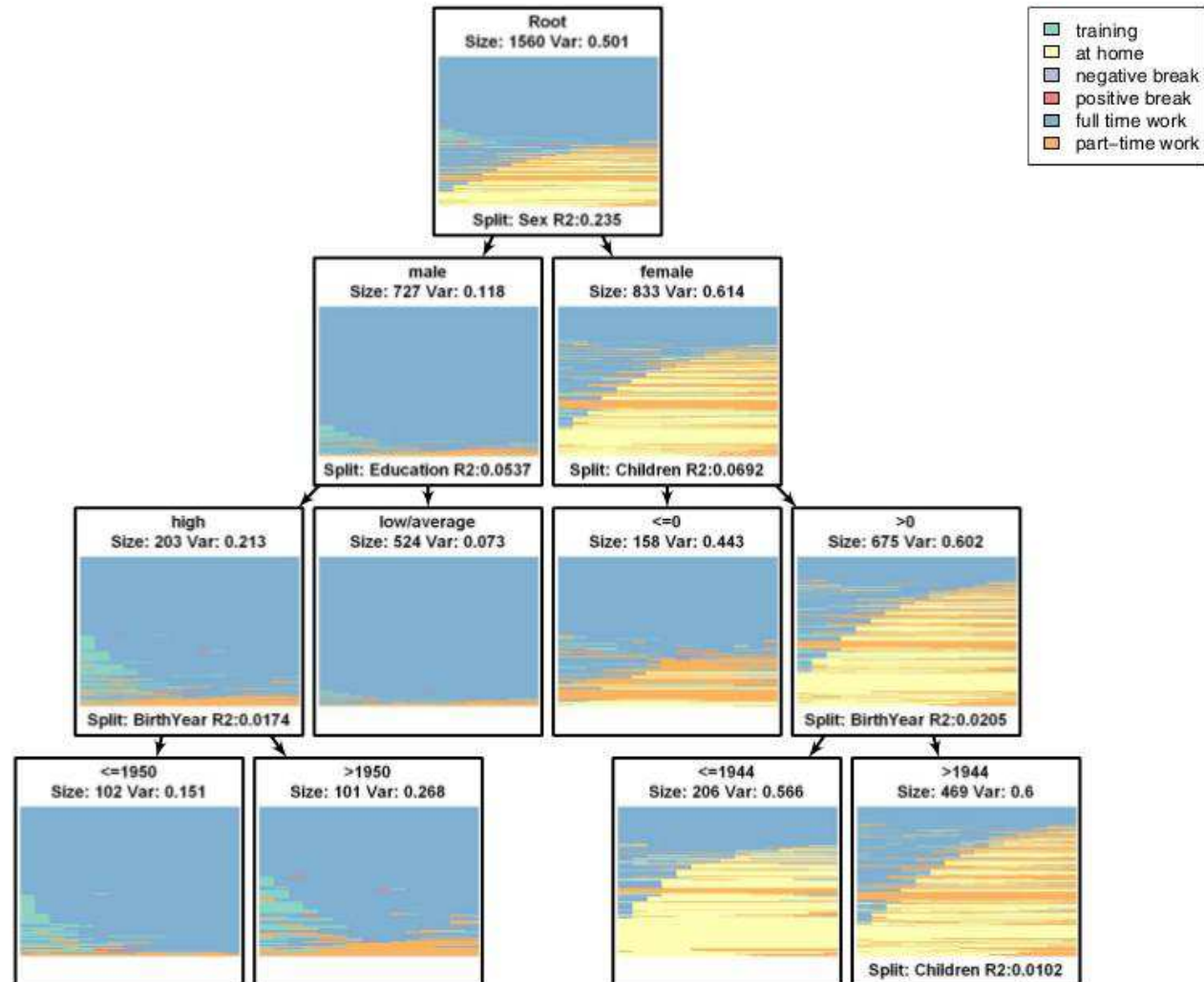
Diversification des trajectoires biographiques



Démarches « non-typologiques »

- Projection sur **plan factoriel** (MDS)
- Distance à un **parcours de référence**, modal, « normal »...
- Similarité moyenne par **groupes** d'individus
- pseudo-ANOVA et arbres d'induction

Studer et al, 2009



Quelles applications ?

- Carrières professionnelles...
 - ... ou autres trajectoires individuelles...
 - ... ou tous types de séquences
- Différents éléments des parcours de vie (Elder):
 - dans le temps et l'espace
 - parcours « multidimensionnels »
 - parcours « liés »

Séquences multi-dimensionnelles


Strategy	Multidimensionality	Parsimony	Interdependence
Combining states (1)	No	Yes	Local
Combining costs (2)	Yes	Yes	Local
Combining distance matrices (3)	Yes	Yes	No
Combining typologies (4)	Yes	No	Global
GIMSA (5)	Yes	Yes	Global

Robette et al, forthcoming

Logiciels pour l'analyse de séquences

- package R « *TraMiner* »
- TDA (téléchargeable gratuitement)
- Module STATA (*SQ*)
- Logiciels spécifiques: Optimize, Distance...

Conclusion

- Démarche **exploratoire** 
 - souple
 - robuste
 - puissante
- La **description** comme outil d'analyse

Bibliographie

- **ma page perso:** <http://nicolas.robette.free.fr/Publis.htm>
- Robette N., 2011, *Explorer et décrire les parcours de vie : les typologies de trajectoires*, Paris : Ceped (série « les clefs pour »), 86 p.
- Robette N., 2010, « The diversity of pathways to adulthood in France: evidence from a holistic approach », *Advances in Life Course Research*
- Robette N., Thibault N., 2008, « L'analyse exploratoire de trajectoires professionnelles: analyse harmonique qualitative ou appariement optimal? », *Population-F*, 64(3), p.621-646.

- Billari F., 2005, « Life course analysis: two (complementary) cultures? Some reflections with examples from the analysis of the transition to adulthood », in R. Levy, P. Ghisletta, J.-M. Le Goff, D. Spini and E. Widmer (Eds.), *Towards an Interdisciplinary Perspective on the Life Course*, Oxford: Elsevier, pp. 261–282.
- Elzinga C., 2008, « Sequence analysis: Metric representations of categorical time series », *Sociological Methods and Research*, in revision.
- Grelet Y., 2002, « Des typologies de parcours. Méthodes et usages », *Document Génération* 92, (20), 47 p.
- Lesnard L., Saint-Pol T. (de), 2004, « Introduction aux méthodes d'appariement optimal (Optimal Matching Analysis) », *Document de travail du Crest*, (15), 30 p.
- Macindoe H., Abbott A., 2004, « Sequence analysis and optimal matching techniques for social science data », in Hardy Melissa, Bryman Alan, *Handbook of Data Analysis*, London, Sage, p. 387-406.