REVUE DE STATISTIQUE APPLIQUÉE

P. D. ALEVIZOS
A. MORINEAU

Analyse factorielle conditionnelle : une application sur des données concernant le masticage de vitraux

Revue de statistique appliquée, tome 41, nº 2 (1993), p. 5-22
<http://www.numdam.org/item?id=RSA_1993__41_2_5_0>

© Société française de statistique, 1993, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (http://www.sfds.asso.fr/publicat/rsa.htm) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM
Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

Rev. Statistique Appliquée, 1993, **XLI** (2), 5-22

ANALYSE FACTORIELLE CONDITIONNELLE : UNE APPLICATION SUR DES DONNÉES CONCERNANT LE MASTICAGE DE VITRAUX

P.D. Alevizos(1), A. Morineau(2)

(1) Université de Patras, Département de Mathématiques, Patras, Grèce
(2) Centre International de Statistique et d'Informatique Appliquées, Saint-Mandé, France

RÉSUMÉ

On dispose d'un tableau numérique relatif à une expérimentation sur le masticage des vitraux. Les données sont constituées, pour chacun des 56 vitraux, de 30 couples de valeurs représentant à 30 instants successifs la pression du vent exercée sur le vitrail et la flèche résultant, c'est-à-dire la déformation du panneau sous l'effet de la pression. On obtient donc pour chaque vitrail une trajectoire dessinée avec 30 points. On a déjà étudié la dispersion des localisations des trajectoires dans le plan «flèche-pression» (P. Alevizos, A. Morineau, 1992). Dans le présent article on décrit la dispersion des courbes autour de leur point moyen. On compare les formes des trajectoires des courbes indépendamment de leur localisation ou, en d'autres termes, on effectue l'analyse des formes conditionnée par la localisation. L'article conclut sur la possibilité de synthétiser les 60 paramètres décrivant une épreuve par seulement deux paramètres.

Mots-clés : *Analyse en composantes principales. Analyse factorielle conditionnelle. Comparaison d'analyses factorielles.*

ABSTRACT

A numerical table concerns 56 stained glass windows in which different adhesive materials have been used for their joints.They have undergone an artificial ageing test under variable conditions. The collected data for each window make up 30 couples of successive values, which give the wind pressure and the deformation caused under the effect of the pressure. For each window we draw a curve with 30 points. In a preceding paper we studied the global location of points in the plane "deformation-pressure" (P. Alevizos, A. Morineau, 1992). In the present paper we describe the dispersion of the curves around their average point. We compare shape of curves without taking into account their location (analysis of shape of curves conditional to location). Paper ends with a proposal to synthetize the whole experiment on any window with only two special parameters.

Key-words : *Principal components analysis. Conditional principal analysis. Comparison of principal analyses.*

1. Introduction

On dispose d'un tableau numérique concernant 56 vitraux mastiqués au moyen de différents matériaux et soumis à des épreuves de vieillissement sous des conditions variées. Les données recueillies au cours d'un essai constituent 30 couples de valeurs successives : d'une part la pression du vent exercée sur le vitrail et d'autre part la flèche, c'est-à-dire le déplacement sous l'effet de la pression. On obtient donc pour chaque vitrail une trajectoire dessinée avec 30 points, résultat de quatre mois d'observations (correspondant à environ 300 années de vieillissement naturel).

Dans un premier papier on a étudié la localisation des courbes dans le plan «flèche-pression» (P.Alevizos, A.Morineau, 1992). On étudie ici la dispersion des courbes autour de leur point moyen et on compare les «déformations» de chacun des vitraux au cours du temps.

Ces déformations sont consignées dans les trajectoires tracées avec 30 points chacunes. Pour étudier les formes indépendamment de la «localisation», les courbes peuvent être au préalable ramenées à leur centre de gravité propre. On procède alors à l'analyse des formes des trajectoires conditionnée par la localisation. On s'interrogera ensuite sur l'indépendance entre la forme et la localisation. On utilisera des méthodes de simulation afin de valider les conclusions.

2. Typologie des formes de trajectoires

2.1 Première étude des formes de trajectoires

La caractérisation des formes des courbes est un exercice difficile. Il faut tenir compte de la longueur de la courbe, de sa plus ou moins forte linéarité, de l'orientation entre les points successifs, etc. Pour avoir un aperçu de la nature des courbes, on consultera la figure 1 où 4 des 56 courbes sont représentées.

Pour effectuer l'étude des formes des courbes, nous choisissons, après avoir ramené chaque courbe à son centre de gravité, les paramètres suivants :

– La moyenne des 5 premières et des 5 dernières observations des variables flèche et pression (après centrage). On aura ainsi une indication «robuste» de la longueur des courbes et de leur orientation générale.

– L'écart-type des 30 mesures de flèche et celui des 30 mesures de pression : un écart-type petit indique que le facteur correspondant varie peu au cours de l'épreuve (une «courbe idéale» correspondrait aux deux écarts-types très petits).

– Le coefficient de corrélation linéaire entre les mesures de flèche et de pression : on verra que ce coefficient peut être un indicateur de la linéarité de la trajectoire et de sa direction.

Le choix des moyennes des 5 observations extrêmes, assez intuitif, est en fait suggéré par un résultat de la seconde analyse décrite ci-dessous. Les libellés des paramètres retenus pour caractériser les formes sont :

F*S	l'écart-type de flèche,
P*S	l'écart-type de pression,
CORR	le coefficient de corrélation entre flèche et pression,
FD5C	la moyenne centrée des 5 premières flèches
FF5C	la moyenne centrée des 5 dernières flèches
PD5C	la moyenne centrée des 5 premières pressions
PF5C	la moyenne centrée des 5 dernières pressions

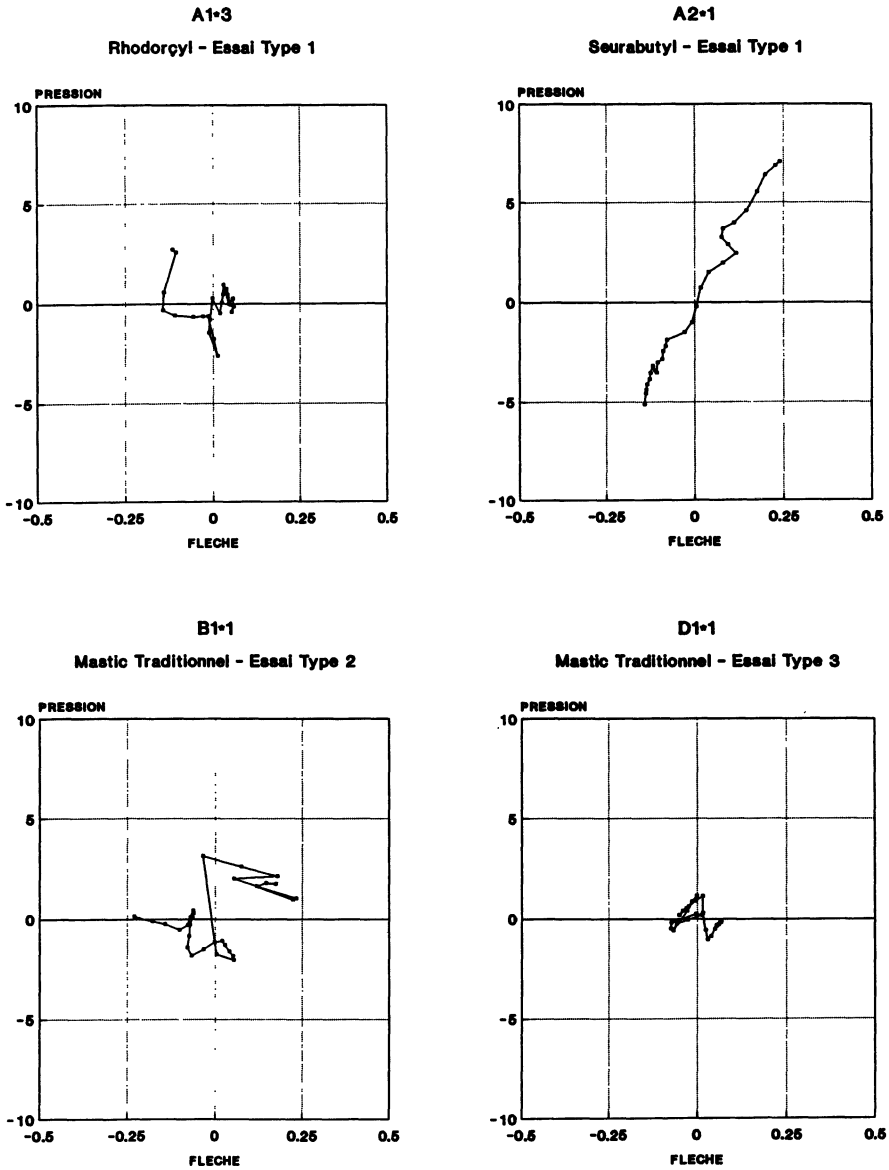

FIGURE 1

Dans le tableau 1 on donne les caractéristiques statistiques de ces 7 variables, ainsi que la matrice des corrélations.

Les 7 paramètres de forme	Moyenne	Ecat-type	Minimum	Maximum
F*S	.09	.05	.01	.22
P*S	1.92	1.26	.12	5.61
CORR	.43	.62	−.95	.99
FD5C	−.09	.11	−.28	.20
FF5C	.09	.11	−.15	.36
PD5C	−1.90	2.56	−8.09	3.11
PF5C	1.91	2.40	−1.74	8.36

	Matrices des corrélations						
F*S	1.00						
P*S	.49	1.00					
CORR	.35	.41	1.00				
FD5C	−.61	−.34	−.40	1.00			
FF5C	.77	.50	.53	−.89	1.00		
PD5C	−.43	−.79	−.64	.37	−.56	1.00	
PF5C	.49	.88	.61	−.36	.61	−.93	1.00
	F*S	P*S	CORR	FD5C	FF5C	PD5C	PF5C

TABLEAU 1

Pour étudier les différentes formes de trajectoires on effectue l'analyse en composantes principales des 7 paramètres choisis pour caractériser les courbes. Le tableau 2 donne l'histogramme des valeurs-propres de cette analyse ainsi que les coordonnées des variables sur les 5 premiers axes.

Les figures 2 et 3 montrent la localisation des 7 variables actives de l'analyse en composantes principales dans les plans principaux (1,2) et (1,3) de l'analyse. On y a placé les variables illustratives suivantes, qui sont davantage des indicateurs de localisation des courbes :

F*M la moyenne des flèches,
FD5 la moyenne des 5 premières valeurs des flèches,
FF5 la moyenne des 5 dernières valeurs des flèches,
P*M la moyenne des pressions,
PD5 la moyenne des 5 premières valeurs des pressions,
PF5 la moyenne des 5 dernières valeurs des pressions.

Num.	Valeur propre	Pourc.	Pourc. cumulé	Analyse des 7 paramètres de forme
1	4.4597	63.71	63.71	**********
2	1.2698	18.14	81.85	**********
3	.6509	9.30	91.15	******
4	.3658	5.23	96.37	***
5	.1582	2.26	98.63	**
6	.0773	1.10	99.74	*
7	.0184	.26	100.00	*

Variables	Coordonnées				
	1	2	3	4	5
F*S	.73	–.42	.29	–.45	.03
P*S	.81	.38	.35	.07	–.27
CORR	.70	.14	–.66	–.20	–.13
FD5C	–.70	.62	.08	–.32	.07
FF5C	.87	–.45	–.02	.09	.08
PD5C	–.86	–.42	.03	–.09	–.23
PF5C	.89	.41	.08	.04	.08

TABLEAU 2

Pour compléter l'analyse, on effectue une classification des 56 points en groupes de trajectoires semblables, en travaillant dans l'espace des premiers axes factoriels. On construit un arbre hiérarchique en utilisant les 6 coordonnées factorielles de l'analyse précédente pour calculer les distances entre points. L'arbre est coupé pour donner une partition en 10 classes représentées dans le premier plan factoriel de l'analyse. Cette partition est représentée sur la figure 4.

En se reportant au catalogue des 56 courbes (P. Alevizos, 1990), on a pu constater que le premier axe oppose les trajectoires très courtes à gauche (courbes en «pelotes d'épingles») aux trajectoires longues à droite. Le deuxième axe oppose en bas les trajectoires correspondant à de grandes flèches à celles correspondant à de grandes pressions en haut.

On notera sur la figure 2 la position du point «corrélation» à droite du premier axe, rappelant que la corrélation augmente en général avec la longueur des trajectoires. Cependant la figure 3 montre que c'est sur l'axe 3 que cette variable a la position la mieux marquée.

Le catalogue des 56 courbes (non publié ici pour des raisons de volume) permet aussi de vérifier que les classes obtenues par la classification automatique contiennent effectivement des courbes aux formes très semblables. Ceci nous assure empiriquement que le choix des 7 paramètres utilisés pour décrire les formes était convenable.

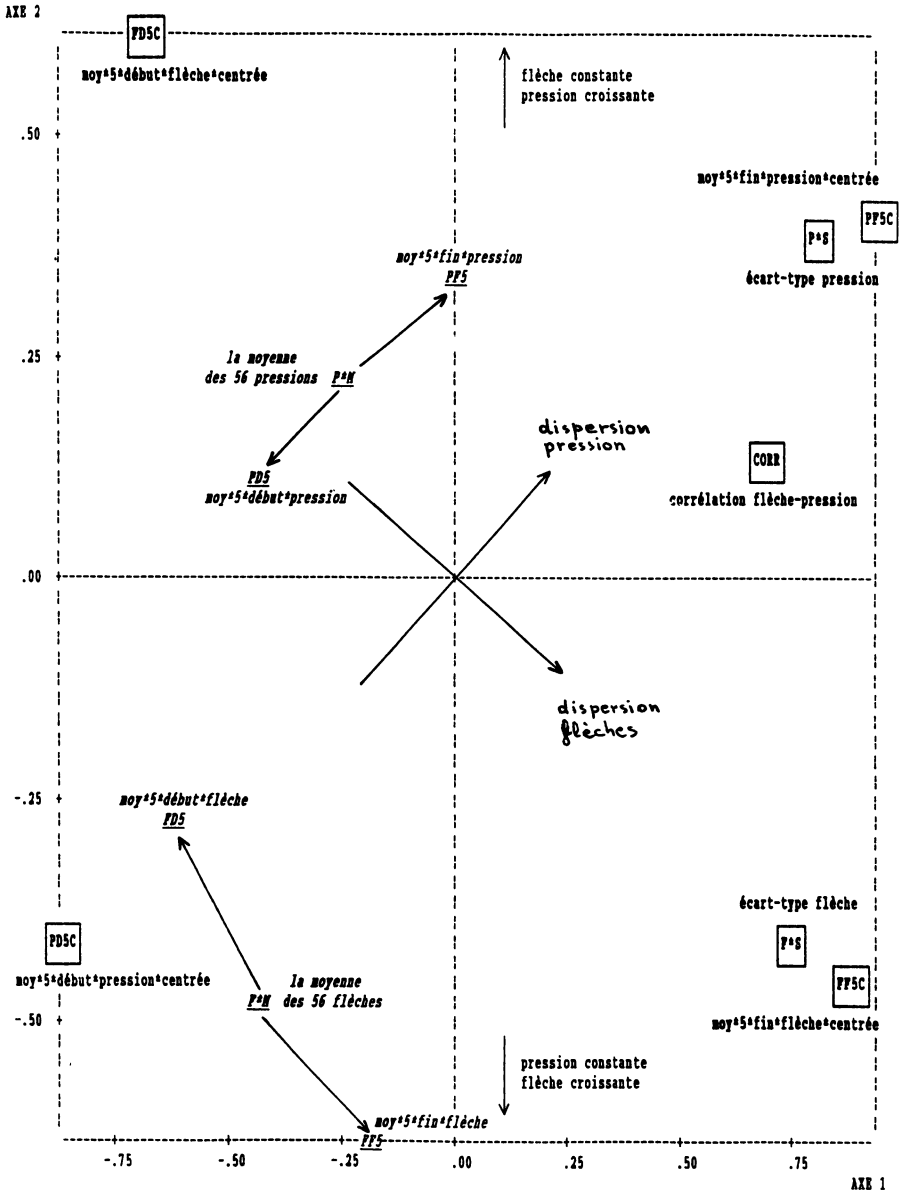

FIGURE 2
Analyse des 7 paramètres de forme : plan (1,2)
les variables «encadrées» sont les 7 actives

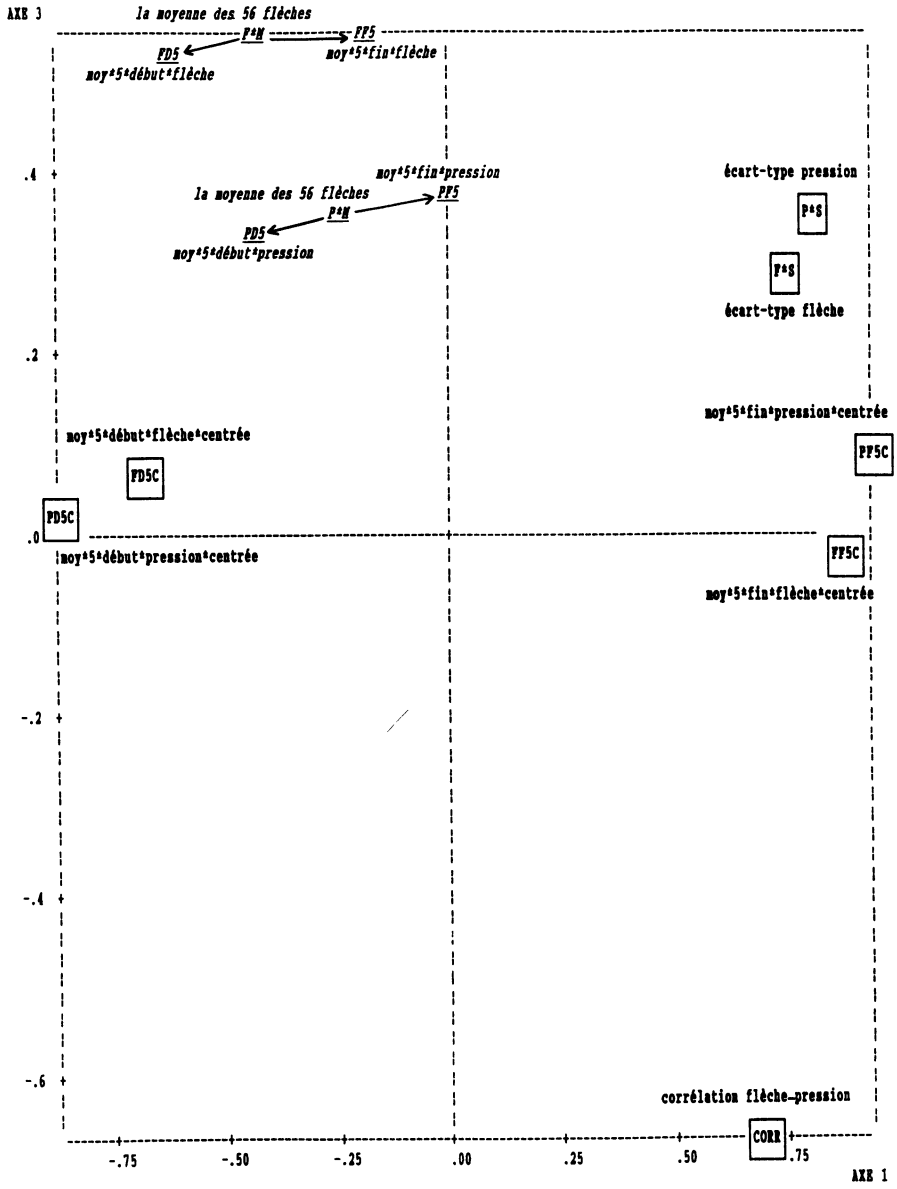

FIGURE 3
Analyse des 7 paramètres de forme : plan (1,3)
les variables «encadrées» sont les 7 actives

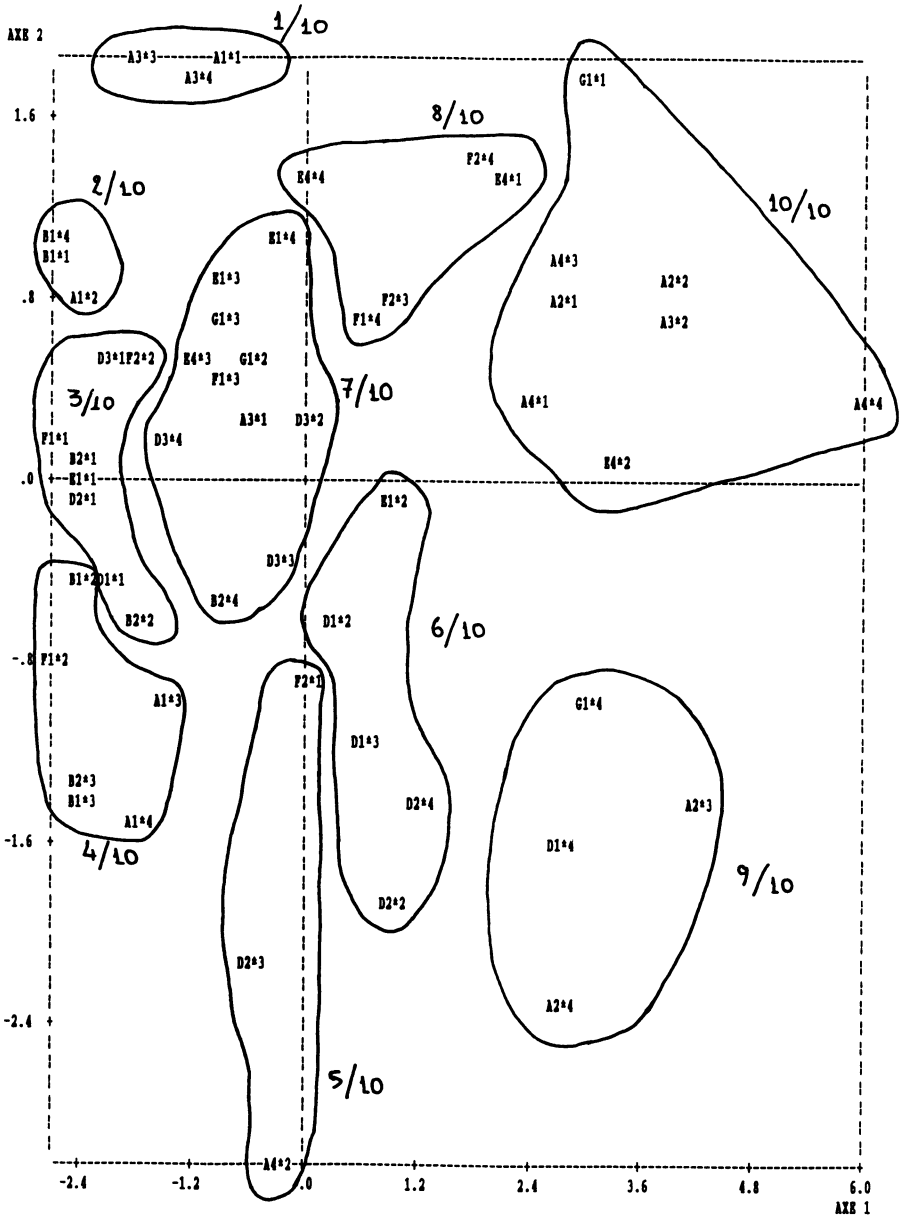

FIGURE 4
Position des courbes dans le plan (1,2)
On a identifié les 10 classes avec leurs numéros

2.1.1 Simulations et tests

Grâce aux doublons, répliques et reprises, on a pu s'assurer que la localisation des courbes dans le plan «flèche-pression» n'était pas le fait du hasard (P. Alevizos, A. Morineau, 1992). Rappelons les définitions utilisées ici. Il y a *doublon* si 2 panneaux identiques sont soumis à la même épreuve. Il y a *réplique* si l'épreuve est identique mais pas réalisée en même temps. Il y a *reprise* quand c'est le même panneau qui subit deux épreuves successives. Ainsi les doublons devraient donner des formes de trajectoires très semblables; les trajectoires des répliques pourraient différer davantage, ainsi que celles des reprises.

Peut-on dire maintenant que les formes des trajectoires sont, comme les localisations, liées globalement aux facteurs de l'expérience ?

On va faire une étude analogue à celle qui a été réalisée sur la localisation dans le plan «flèche-pression». Pour effectuer cette étude, on utilise les résultats de l'analyse en composantes principales réalisée sur les 7 paramètres de forme.

La proximité entre les points (dans l'espace factoriel) est fonction de la similitude de forme des trajectoires. On calcule les $(56 \times 55)/2 = 1540$ distances entre couples de points dans l'espace à 7 dimensions où se trouve le nuage et on effectue des simulations analogues à celles faites pour l'étude des localisations : 1000 puis 4000 tirages dans chacun des cas suivants :

8 couples pour l'étude des doublons (il y a 8 doublons),
12 couples pour l'étude des répliques (il y a 12 répliques),
4 couples pour l'étude des reprises (il y a 4 reprises).

Les résultats sont consignés dans le tableau 3.

Distance	Moyenne	Ecart-type	Comptage
1540 couples	14.255	(12.460)	–
DOUBLONS : 8 couples	8.953	–	–
1000 tirages de 8 couples	14.053	4.320	98/1000
4000 tirages de 8 couples	14.319	4.271	373/4000
REPLIQUES : 12 couples	5.558	–	–
1000 tirages de 12 couples	14.117	3.549	1/1000
4000 tirages de 12 couples	14.208	3.501	3/4000
REPRISES : 4 couples	4.390	–	–
1000 tirages de 4 couples	14.241	6.200	23/1000
4000 tirages de 4 couples	14.346	6.153	77/4000

TABLEAU 3
Tests sur les doublons, répliques et reprises

Il apparaît ici que près de 10 % des moyennes de 8 couples pris au hasard sont aussi extrêmes que la moyenne observée sur les 8 doublons. Ainsi les formes

des trajectoires de doublons ne sont pas «réellement» ressemblantes : elles se ressemblent à peine plus que des trajectoires tirées au hasard.

Par contre les trajectoires des répliques, assez curieusement, sont très semblables puisque seulement dans 3 cas sur 4000 on obtient une moyenne des 12 distances plus faible que la moyenne des distances entre les 12 répliques observées. Les formes des trajectoires de reprises paraissent également «assez similaires» (2 % seulement des simulations donnent des résultats plus extrêmes que l'observation).

On arrive donc à un résultat assez surprenant. L'analyse factorielle et la classification, donnant des résultats vérifiables sur les courbes elles-mêmes, nous assurent que la distance entre points dans l'analyse en composantes principales est une bonne mesure de la similitude des trajectoires. Les simulations entreprises à partir de ces distances montrent cependant que, contrairement à ce qu'on pouvait attendre, les trajectoires associées aux doublons ne sont pas très semblables et (curieusement) celles associées aux répliques sont assez analogues. Le résultat est moins tranché dans le cas des reprises. Aucune explication satisfaisante n'a été trouvée.

2.2 Deuxième étude des formes de trajectoires

L'objet de cette analyse est d'effectuer une étude des formes des trajectoires à partir des trajectoires proprement dites (et non à partir de paramètres de forme), en s'arrangeant pour que les comparaisons ne tiennent pas compte des localisations des trajectoires dans le plan «flèche-pression». Les résultats seront ensuite confrontés aux résultats de l'analyse précédente.

Chaque trajectoire est définie par 30 couples de valeurs flèche-pression, numérotés de 1 à 30. Le tableau des données est constitué de 56 lignes représentant les 56 trajectoires et de 2 fois 30 colonnes définissant la forme de la trajectoire. Afin d'éliminer l'effet de localisation, on amène chaque trajectoire à avoir pour origine son propre centre de gravité. On remplace donc les 30 valeurs de flèche d'une trajectoire par les valeurs centrées sur leur propre moyenne. On fait de même pour les 30 valeurs de pression.

Le tableau ainsi obtenu est soumis à une analyse en composantes principales effectuée sur la matrice des corrélations. Les variables actives de l'analyse sont les 60 paramètres quantitatifs. On utilisera aussi comme variables illustratives les paramètres de l'analyse précédente.

Le tableau 4 donne l'histogramme des valeurs propres et montre l'importance du premier plan factoriel. Les 10 premiers axes factoriels contiennent plus de 96 % de l'inertie totale des nuages.

Les figures 5 et 6 concernent les variables actives de l'analyse. Elles situent d'une part les 30 mesures de pression, d'autre part les 30 mesures de flèche. Les points sont numérotés et reliés dans l'ordre chronologique des mesures.

On constate la forte corrélation entre les premières et entre les dernières mesures, aussi bien pour la pression que pour la flèche. Les deux groupes extrêmes sont reliés par un chemin «chronologique» assez régulier (bien que le temps, naturellement, ne soit pas pris en compte explicitement dans l'analyse). L'observation de deux figures nous montre que, dans les deux cas, le temps s'écoule

N°	Valeur propre	Pourc.	Pourc. cumulé	
1	29.4286	49.05	49.05	**********
2	10.4862	17.48	66.52	**********
3	5.0559	8.43	74.95	********
4	3.6753	6.13	81.08	******
5	3.1986	5.33	86.41	****
6	2.0858	3.48	89.88	***
7	1.2673	2.11	92.00	**
8	1.1110	1.85	93.85	*
9	.8251	1.38	95.22	*
10	.5204	.87	96.09	*
11	.3922	.65	96.74	*
12	.2944	.49	97.23	*
13	.2728	.45	97.69	*
14	.2024	.34	98.03	*
15	.1786	.30	98.32	*

TABLEAU 4

dans la direction du premier axe. Le second axe sépare la pression (première bissectrice) de la flèche (deuxième bissectrice).

Dans le premier plan factoriel de cette analyse, on représente aussi (figure 7) la projection en variables supplémentaires des paramètres de forme utilisés dans l'analyse précédente. Cette figure est très semblable à la configuration des mêmes variables dans le premier plan factoriel de la première analyse (voir la figure 2).

Dans ces conditions, tous les graphiques dans le premier plan factoriel réalisés dans l'une ou l'autre analyse seront très semblables. On en conclut que les 7 paramètres choisis pour caractériser les formes fournissent essentiellement la même information que les 30 couples flèche-pression.

On notera également que le choix des paramètres «moyennes des 5 valeurs extrêmes» comme caractéristiques de forme est justifié ici par la forte corrélation entre ces valeurs extrêmes. En prenant ces moyennes, on a stabilisé (ou «robustifié») l'estimation des extrémités des trajectoires.

La classification qui suit cette analyse, effectuée sur les 10 premiers facteurs, fournit une partition en 9 classes qui ressemble beaucoup au classement des courbes obtenu après l'analyse des 7 paramètres de forme. Le tableau 5 montre la correspondance entre les classes des deux classifications. On remarquera que la plupart des classes se retrouvent intégralement dans les deux partitions. La similitude des résultats ne concerne donc pas seulement les deux premiers axes factoriels mais l'ensemble de la configuration.

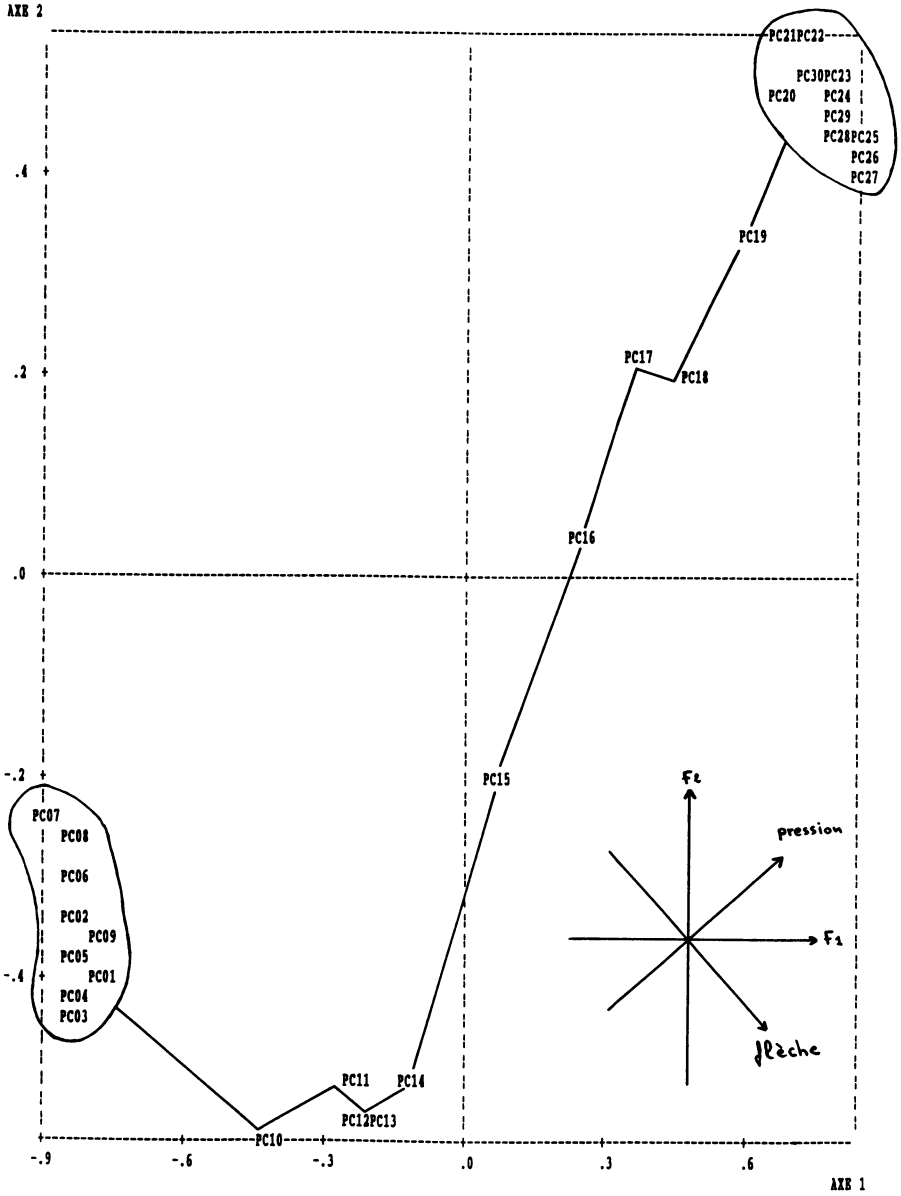

FIGURE 5
Deuxième analyse des formes :
les 30 pressions actives

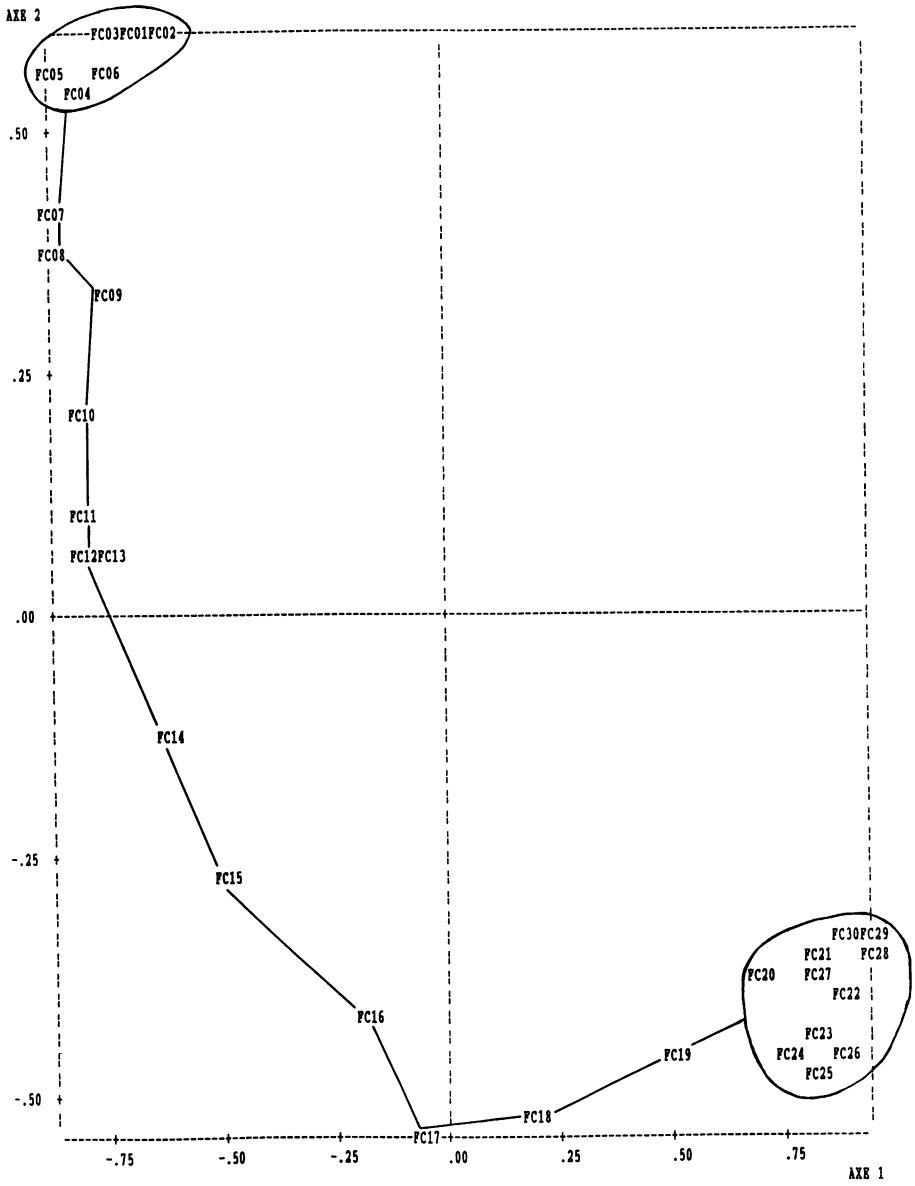

FIGURE 6
Deuxième analyse des formes :
les 30 flèches actives

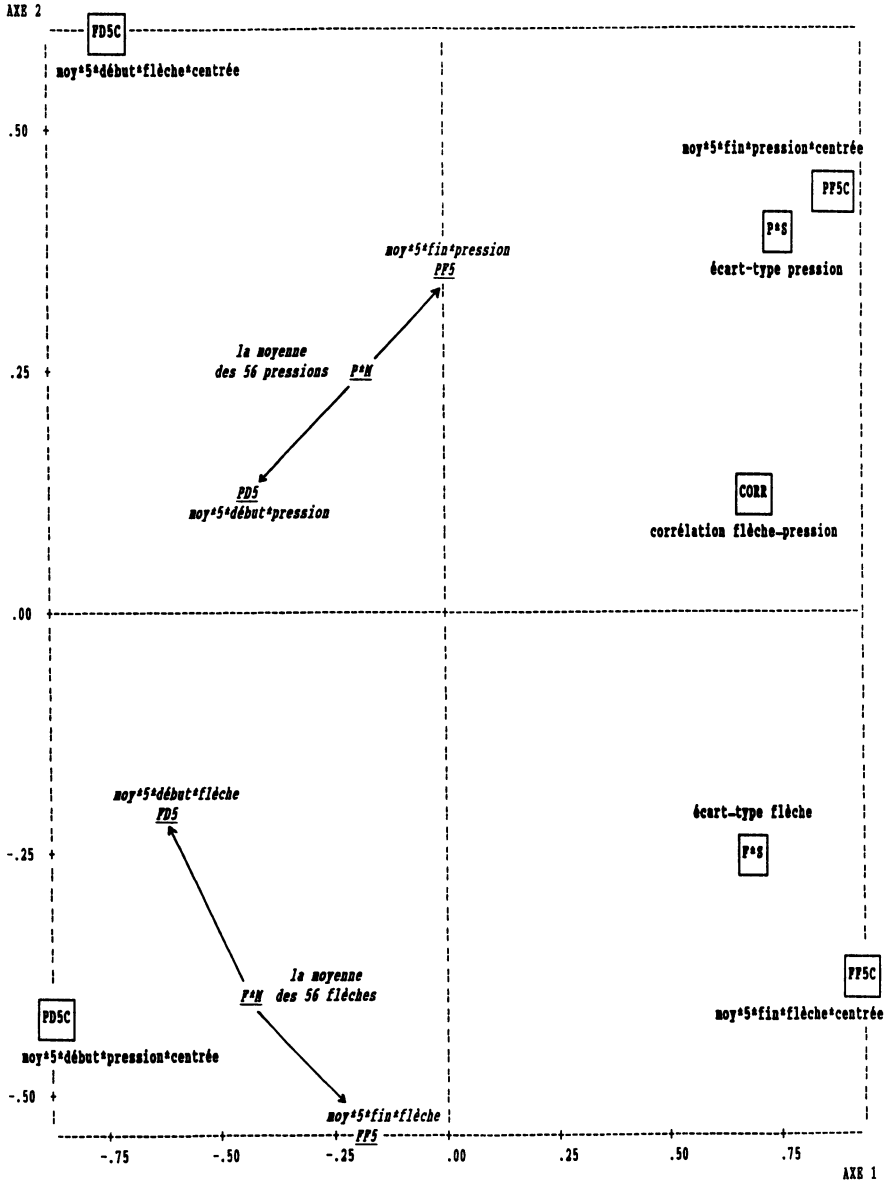

FIGURE 7

Deuxième analyse des formes : les 13 variables illustratives
les 7 variables «encadrées» sont les actives de la première analyse

	C1/9	C2/9	C3/9	C4/9	C5/9	C6/9	C7/9	C8/9	C9/9	Total
C01F	**3**									3
C02F	1	**2**								3
C03F			1	**5**		2				8
C04F						**6**				6
C05F			1		1	1				3
C06F					**5**					5
C07F				**11**						11
C08F			1	1					**3**	5
C09F							**4**			4
C10F								**3**	**5**	8
Total	4	2	3	17	6	9	4	3	8	56

TABLEAU 5

Croisement des partitions issues des 2 analyses sur les formes
en ligne : les10 classes de la première analyse
en colonne : les 9 classes de la seconde analyse

2.3 Troisième étude des formes des trajectoires

On peut simplifier encore la description des trajectoires en utilisant deux paramètres seulement : la longueur de la trajectoire et sa direction dans le plan «flèche-pression».

- la longueur de la trajectoire est mesurée par la distance entre les «points moyens extrêmes». Ces points ont pour coordonnées dans le plan «flèche-pression» les moyennes centrées des 5 valeurs extrêmes,
- la direction de la trajectoire est mesurée par l'angle de la droite joignant les «points moyens extrêmes» avec l'axe horizontal (qui est ici la flèche).

On jugera ces paramètres suffisants et satisfaisants s'ils permettent de retrouver la typologie des courbes établie par les analyses précédentes.

Ces deux paramètres peuvent êtres considérés comme les *coordonnées polaires* d'un point dans un plan, la position de ce point représentant alors la forme de la trajectoire. Dans ce plan, les points proches représentent des courbes qui se ressemblent (comme dans un plan factoriel d'analyse).

Ramené à des coordonnées cartésiennes, ce plan est donné par la figure 8. Le résultat est remarquablement proche des plans factoriels issus des deux analyses précédentes. On s'en aperçoit aisément en figurant par exemple les 10 classes de la typologie réalisée avec les 7 paramètres de forme. Le plan obtenue ici apparaît

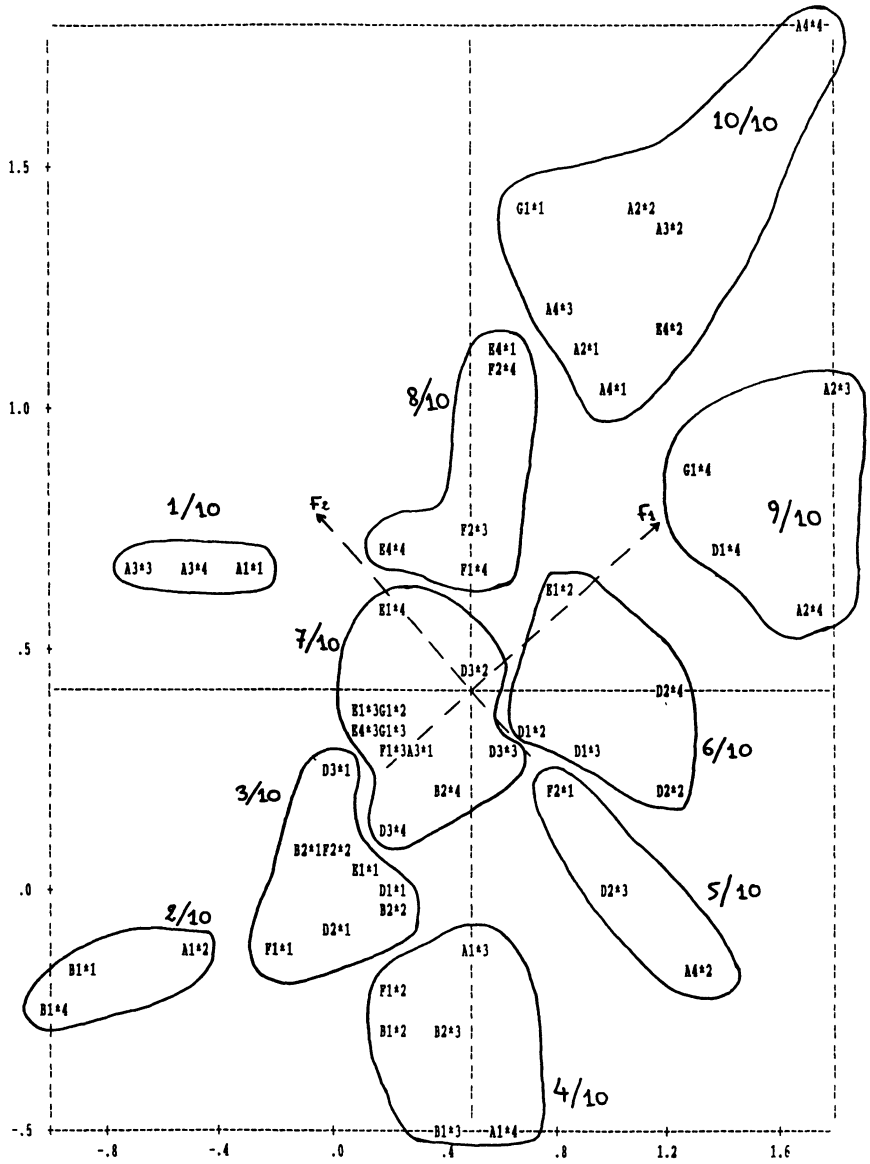

FIGURE 8

Plan «longueur-angle» : on représente les 10 classes obtenues après la première étude sur les 7 paramètres des formes

analogue (à une rotation près de 45 degrés) au premier plan factoriel de la première analyse (voir figure 4).

On en concluera que les deux paramètres choisis pour décrire la forme des courbes fournissent essentiellement la même information que les 7 paramètres de formes (semblable donc aux 60 variables initiales après centrage).

A ce stade, on peut résumer les analyses de la façon suivante. Pour chaque vitrail, le résultat de l'épreuve est synthétisable par 4 valeurs :

- 2 valeurs caractérisent la «localisation» : moyenne sur les 30 mesures de la pression et de la flèche,
- 2 valeurs caractérisent la «forme» : la longueur et la direction de la trajectoire mesurées comme on l'a indiqué.

Remarque

On a décrit deux analyses : avec 7 paramètres pour la première et 60 paramètres pour la deuxième. On a réalisé également l'analyse globale forme et localisation sur les 4 paramètres ci-dessus (qui n'est pas décrite en détail ici). On a constaté que ces trois analyses donnent des résultats très proches. Pour confirmer les résultats obtenus ci-dessus, on a récupéré les coordonnées des individus (les 56 courbes) sur les axes principaux issus des trois analyses. Le tableau 6 donne la matrice des corrélations entre les trois premiers facteurs des analyses (désignés par «iF1», «iF2» et «iF3» avec $i = 1, 2, 3$).

	1F1	1F2	1F3	2F1	2F2	2F3	3F1	3F2	3F3
1F1	1.00								
1F2	.00	1.00							
1F3	.00	.00	1.00						
2F1	−.96	.05	.01	1.00					
2F2	−.04	−.91	−.09	.00	1.00				
2F3	−.01	−.14	.15	.00	.00	1.00			
3F1	.71	.18	−.43	−.68	−.20	−.06	1.00		
3F2	.61	−.11	.54	−.56	−.01	.10	.00	1.00	
3F3	.06	−.54	.25	−.01	.54	.10	.00	.00	1.00
	1F1	1F2	1F3	2F1	2F2	2F3	3F1	3F2	3F3

TABLEAU 6

On observe que le premier plan factoriel de la première analyse est très semblable à celui de la deuxième, et correspond à une rotation près de 45 degrés au premier plan de la troisième analyse (la troisième analyse n'est pas conditionnelle par rapport à la localisation).

3. Conclusion

La motivation principale de cette étude était la description des formes d'objets assez complexes : des trajectoires planes quelconques décrites par 30 points. On a mis en concurrence une analyse réalisée sur 7 paramètres choisis pour caractériser *a priori* les formes et une seconde analyse sur les données brutes. Pour que l'analyse sur les données brutes ne décrive que les formes, indépendamment de la localisation de la trajectoire dans le plan, on a réalisé une analyse «conditionnelle à la localisation» en centrant chaque trajectoire sur son point moyen.

La comparaison des résultats est réalisée par l'intermédiaire des coordonnées des points sur les axes factoriels des deux analyses. Les analyses sont «identiques» dans le sous-espace des premiers facteurs si les coefficients de corrélation sur les axes de même rang sont proches de 1 (en valeur absolue) et si les coefficients sont proches de 0 sur les axes de rangs différents.

On a pu contrôler de cette façon que les 7 paramètres caractéristiques de la forme contenaient bien toute l'information utile pour décrire les trajectoires. Dans le souci d'une économie encore plus grande de paramètres, on a pu déterminer que 2 variables étaient suffisantes pour synthétiser globalement la forme des trajectoires (longueur et direction).

Cette étude empirique peut être considérée comme un exemple simple d'utilisation des techniques exploratoires multidimentionnelles pour aborder le problème du choix «optimal» d'un nombre minimum de paramètres décrivant correctement un phénomène assez complexe.

Bibliographie

ALEVIZOS P. (1990). *Analyse Factorielle Conditionnelle et Analyse de la Structure*, Thèse de Doctorat de l'Université Paris VI, Juin 1990, Paris.

ALEVIZOS P., MORINEAU A., (1992). *Tests et Valeurs-Tests. Application à l'Etude des Mastics Utilisés dans la Fabrication des Vitraux*, Revue de Statistique Appliquée, Vol. 40, n° 4, pp. 27-43.

BETTEMBOURG J.M., BURCK J.J., DRIVIERE J., HENRY J.P. (1984). *Influence des Mastics sur le Comportement des Panneaux de Vitraux*, 10ème congrès international de l'IIC, Paris, Septembre 1984.

LEBART L., MORINEAU A., TABARD N. (1977). *Techniques de la Description Statistique, Méthodes et Logiciels pour l'Analyse des Grands Tableaux*, Dunod.

LEBART L., MORINEAU A., FENELON J.P. (1982). *Traitement des Données Statistiques, Méthodes et Programmes*, Dunod, Paris (1ème édition 1979).