

# L'EXPLICATION DES COMPORTEMENTS DE VENTE DES ARTICLES TEXTILES PAR UNE METHODE DE DATA-MINING

François BOUSSU\* et Jean Jacques DENIMAL†

---

Résumé. - L'idée principale développée dans cet article peut se résumer par cette interrogation : comment exploiter les données de ventes des saisons précédentes pour en comprendre les comportements en vue d'en extrapoler les futurs profils ? Pour y répondre, nous proposons une méthode d'analyse statistique des données en plusieurs étapes fondée sur la régression PLS nous permettant de dégager les classes d'articles dont les ventes s'expliquent par un regroupement des caractéristiques similaires. Une application réalisée sur des données réelles nous a permis de constater, en moyenne, que 80% des ventes des articles textiles peuvent être reconstitués à partir de la seule connaissance de leurs caractéristiques. Cette base de connaissance des ventes textiles peut être utilisée pour estimer les profils de vente des articles de la nouvelle collection uniquement à partir de leurs caractéristiques.

Mots-clés : Data-mining, Régression PLS, Comportement de vente, Prévision des ventes, Articles textiles.

## 1. Introduction

Nos précédentes analyses (Denimal *et al.*, 1999) effectuées sur des données de vente d'articles textiles nous ont permis de constater les points suivants :

---

\* Maître de Conférences au laboratoire GEMTEX (GENie et Matériaux TEXtiles) de PENSAT (Ecole Nationale des Arts et Industries Textiles), Ingénieur Textile, francois.boussu@ensait.fr

† Maître de Conférences au laboratoire de Statistiques et Probabilités de l'Université des Sciences et Technologie de Lille, jean-jacques.denimal@univ-lille.fr

- D'une saison à l'autre le renouvellement de la collection des articles textiles est partiel ou total. Selon le degré de renouvellement, l'influence sur les ventes est plus ou moins forte.
- D'une saison de vente à l'autre, en respectant les saisonnalités équivalentes (saison été année n-1 et n), la définition précise des articles évolue fortement en raison des matières textiles utilisées, des coloris et des tissus nouveaux. Cette évolution reflète en grande partie le phénomène de mode lié à la définition des articles textiles. Cependant, nos précédentes études ont montré que les ventes d'articles similaires peuvent être regroupées dans une catégorie ou classe représentée par son centre de classe. Le fait le plus marquant provient du fait que les profils de ces centres de classe évoluent peu d'une saison de vente à l'autre en dépit du fort renouvellement des articles les caractérisant.

Face à toutes ces remarques, Time des clés du "succès commercial" d'une collection d'articles textiles repose sur l'élaboration de prévisions de ventes aussi fiable que possible. Celles-ci permettent d'anticiper les achats en matières premières et la planification des charges de production. Deux solutions se présentent aux responsables de l'entreprise textile pour établir des prévisions : soit ils s'appuient sur leur intuition et leur connaissance du marché, soit ils ont recours à des modèles mathématiques simples ou compliqués. La première solution accentue la responsabilité du décideur et peut être à l'origine de bons ou de mauvais choix. La deuxième solution permet de conforter les intuitions premières et peut fournir de bonnes prévisions, à condition que le modèle utilisé soit suffisamment fiable en milieu fortement perturbé et incertain. Une étude bibliographique (Boussu, 1998) sur les modèles de prévisions montrent le peu de fiabilité des outils habituellement utilisés dans l'environnement de vente des articles textiles.

Les auteurs de cet article propose une méthode d'analyse des données de ventes — plus précisément de data-mining — qui permet, dans un premier temps d'expliquer les regroupements des ventes par les caractéristiques descriptives des articles et, dans un second temps, d'estimer les profils des ventes des articles de la nouvelle saison. Cette méthode permet d'affiner les intuitions du décideur de l'entreprise textile et fournit une aide à la planification des ressources financières et des charges de production à engager bien avant le début de la saison des ventes.

La construction de la méthode s'est déroulée en plusieurs phases successives optimisant la précision des allures prévisionnelles des articles.

Une premier traitement des données de vente a été proposé en premier lieu (Boussu *et al.*, 1996), permettant d'identifier le nombre optimal d'allures prévisionnelles représentatives de l'ensemble des ventes de la base. Cet identification repose sur l'utilisation d'un critère de classification spécifique permettant de rapprocher les allures quasi-identiques et d'éloigner les allures différentes (Xie and Beni, 1991). La méthode d'identification des relations entre deux ensembles de données corrélées ou non a été rendue possible par l'utilisation première d'une méthode d'analyse multivariée fondée sur des variables qualitatives (Denimal *et al.*, 1999). Puis, face à la contrainte d'utilisation de données de vente avec des données manquantes ou fortement perturbées, le recours à la régression PLS nous a permis d'identifier plus précisément les relations entre les variables qualitatives et les profils de vente (Boussu and Denimal, 2000). La connaissance de ces relations nous a permis de fournir une estimation des profils de vente

des articles pour les comparer aux profils réels, et de valider l'utilisation de notre méthode par les données fournies par l'industriel textile (Denimal and Boussu, 2000).

Au delà de la description de la construction de la méthode, celle-ci peut être englobée dans les méthodologies de data-mining permettant d'apporter une plus-value informationnelle d'une donnée brute jusqu'à la prise de décision.

De façon plus générale, l'objectif d'une méthode de data-mining peut se décliner en trois points. Dans un premier temps, l'identification des relations complexes existant entre les données sans effectuer de recherche approfondie d'un modèle de relation. Dans un second temps, l'établissement des règles de comportement permettant d'en comprendre les fondements. Puis, dans un troisième temps, la mise à jour des modèles fonctionnels permettant de prévoir le comportement futur (Lefébure and Venturi, 1999 ; Jambu, 1999).

## 2. Description de la méthode

Après avoir partagé l'ensemble des articles en deux échantillons base et test, une analyse des correspondances (Benzecri, 1976) appliquée à l'ensemble des caractéristiques qualitatives mesurées sur l'échantillon de base permet de résumer ces dernières en un ensemble réduit de composantes principales associées aux axes factoriels les plus significatifs. Les relations entre les profils de vente et les caractéristiques descriptives sont ensuite mises à jour par l'utilisation de la régression PLS (Partial Least Squares) (Wold, 1985 ; Tenenhaus, 1998) expliquant les profils de vente à partir des composantes principales obtenues précédemment. Cette méthode d'analyse des données a été initialement conçue pour comprendre et décrire les relations souvent très complexes entre les variables de sortie Y et les variables d'entrée X d'un système dont le modèle théorique est inconnu.

Celle-ci permet de construire un modèle explicatif des allures de ventes des articles de l'échantillon de base à partir de ces axes factoriels. Ce modèle, construit à partir de l'échantillon de base, est utilisé pour estimer les allures de vente des individus de l'échantillon test. Une comparaison de ces estimations avec les allures réelles nous permet de valider le modèle obtenu.

Quatre étapes majeures nous permettent d'illustrer notre démarche méthodologique.

La première étape permet de constituer des classes d'articles fondées sur l'optimisation d'un critère permettant à la fois d'agréger les profils de vente similaires et d'éloigner les profils de vente différents. Un tirage aléatoire est alors effectué au sein de chaque classe d'article pour élaborer un échantillon de base et un échantillon test des profils de vente.

La deuxième étape consiste à extraire les variables les plus explicatives des ventes par l'application de la régression PLS aux articles de l'échantillon de base.

La troisième étape aboutit à l'estimation des profils de vente des articles par l'application d'une méthode de régression à partir de la base extraite des profils de vente et des caractéristiques qualitatives de l'échantillon test.

La quatrième et dernière étape autorise la comparaison, pour chaque article de l'échantillon test, des profils estimés de vente par la méthode de régression et des profils réels de vente.

La représentation suivante reprend les quatre étapes principales de la méthode en respectant l'ordre chronologique des imbrications nécessaires pour l'extraction des connaissances à partir des données brutes.

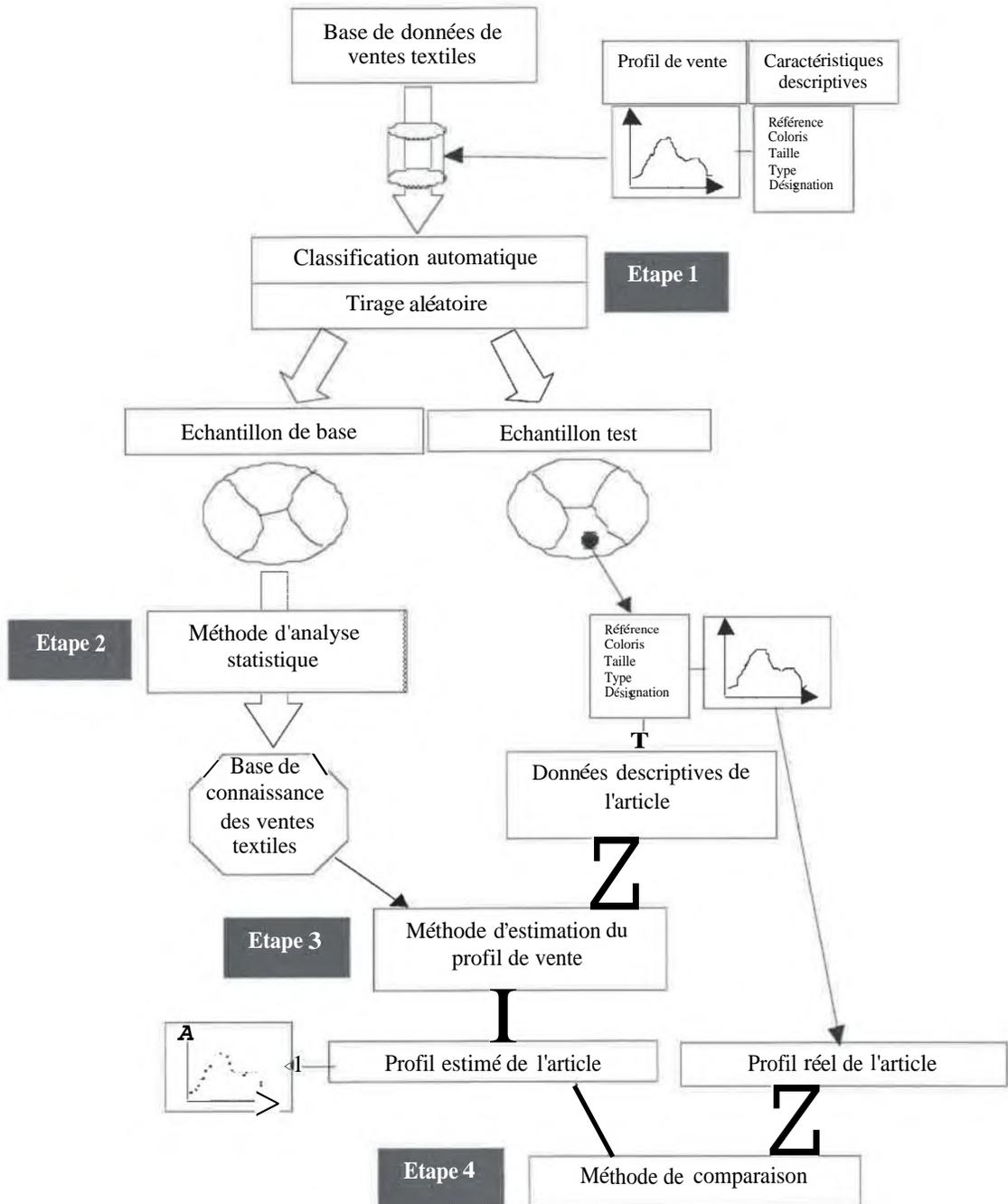


Figure 1. Représentation simplifiée de la méthode d'analyse statistique.

Afin d'illustrer la méthodologie proposée, nous utiliserons dans l'application des données de vente issues d'une entreprise de confection spécialisée dans la fabrication de vêtement de sport. Les données sont ici extraites d'une base de données spécifiques à une classe d'article unique représentant des maillots de bain, des shorts, des collants, des justaucorps et des body. Nous avons volontairement restreint l'échantillon des données à considérer afin de ne pas alourdir les représentations graphiques des résultats obtenus.

### 3. Applications

Une application de la méthode précédente est réalisée sur des données de vente d'articles textiles et permet d'illustrer son intérêt au regard des résultats obtenus.

#### 3.1 Description des données

On trouvera dans le Tableau 1 un descriptif des articles par leurs caractéristiques qualitatives et le nombre de modalités correspondantes.

Caractéristiques qualitatives	Exemples	Nombre de modalités
Genre	Piscine ou Gym.	2
Type	Enfant, femme ou homme	3
Désignation	Slip de bain, justaucorps, etc.	11
Numéro tissu	1345, 1226, etc.	16
Numéro modèle	2014, 2045, etc.	23
Couleur	Blanc, Noir, Jaune, Rose, etc.	28
	Total	83

Tableau 1. Description des caractéristiques qualitatives des articles textiles.

A chaque article textile est associé un ensemble de 6 variables qualitatives ou descriptives (respectivement le genre, le type, la désignation, le numéro de tissu, le numéro de modèle et la couleur) avec un ensemble de valeur de vente sur 52 semaines.

La première étape de la méthode nous permet de distinguer, parmi les 59 articles traités dans la base de données, 4 allures de vente majoritaires représentatives de l'ensemble des ventes représentées sur 52 semaines de vente.

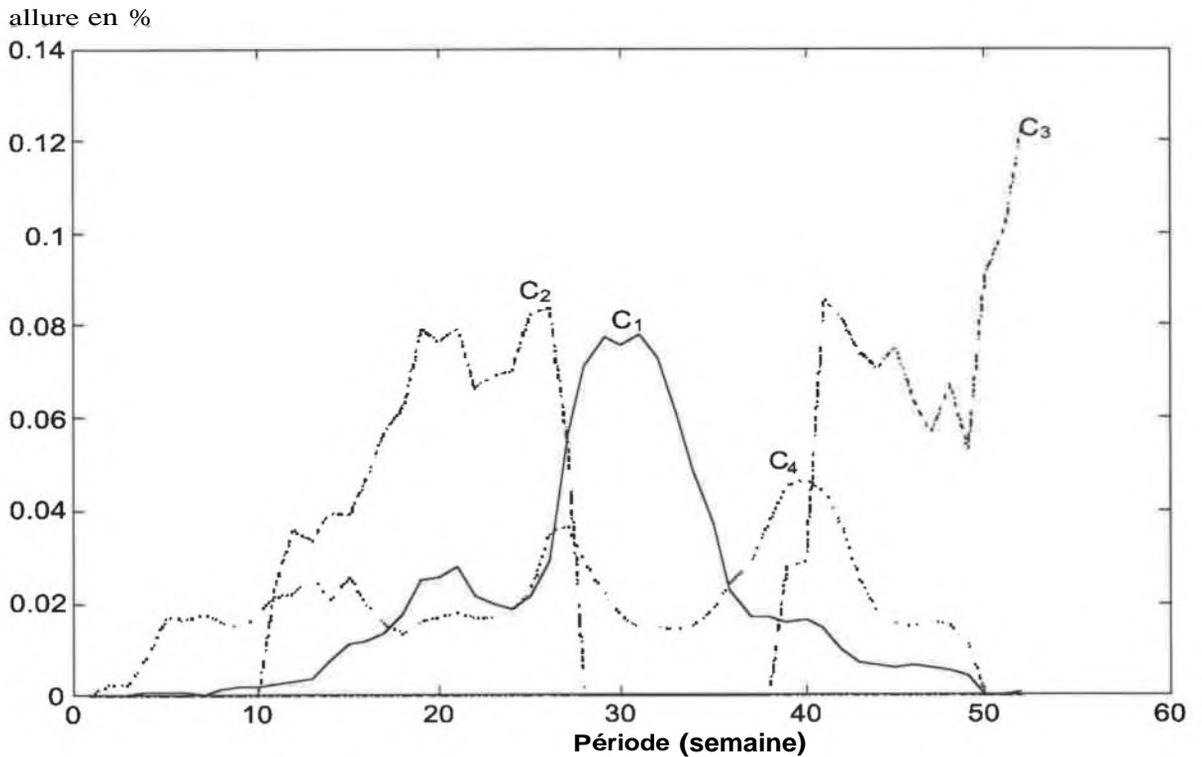


Figure 2 : Représentation des allures des 4 principaux regroupements des ventes d'articles textiles.

Nous pouvons constater les différences de durée de vie des classes d'articles. En effet, les articles des classes C2 ou C3 possèdent des durées de vie moyennes de 15 à 18 semaines et les articles des classes C1 ou C4 englobent des articles de durée de vie moyenne plus longue, de l'ordre de 40 à 52 semaines.

### 3.2 Résultats obtenus

La représentation de la répartition des semaines de vente de l'ensemble des articles de l'échantillon de base dans le plan PLS principal permet de distinguer 5 périodes sur la saison de vente.

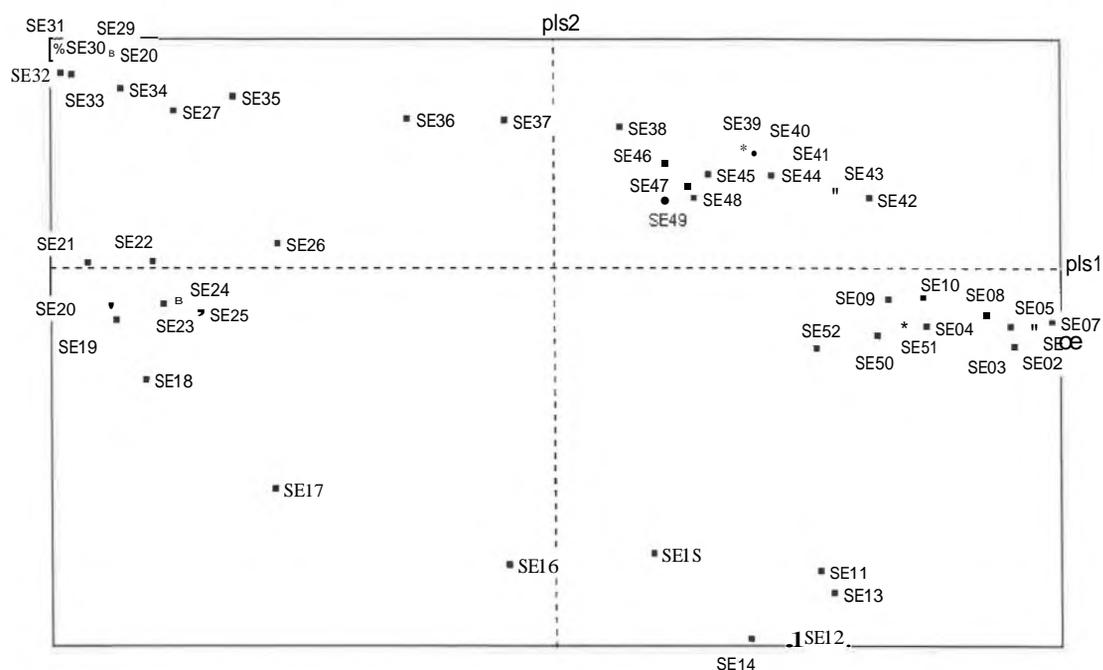


Figure 3 : Répartition des périodes de vente au cours de la saison extraite de la régression PLS.

5 périodes de vente des articles textiles au cours des 52 semaines de vente considérées sont alors identifiées.

La période 1 (de SE 50 à SE 10) représente le début du mois de décembre et se prolonge jusqu'à la moitié du mois de mars, englobant ici les fêtes de fin d'année ainsi qu'une majeure partie de l'hiver.

La période 2 (de SE 11 à SE 17) débute de la moitié du mois de mars pour finir au mois d'avril, englobant principalement la période des vacances de Pâques et le début du printemps.

La période 3 (de SE 18 à SE 26) s'étale du début du mois de mai jusqu'à la fin du mois de juin, correspondant à la deuxième partie du printemps et juste avant les grandes vacances d'été.

La période 4 (de SE 27 à SE 35) comprend essentiellement les mois de juillet et d'août.

La période 5 (de SE 36 à SE 49) commence au début du mois de septembre et s'achève au début du mois de décembre, incluant à la fois le phénomène calendaire de la rentrée des classes et la totalité de la période automnale.

Les deux premières variables explicatives extraites par la régression PLS mettent en valeur majoritairement l'effet saisonnier qui est donc pris en compte par nos variables qualitatives. Les trois variables explicatives suivantes permettent d'expliquer les allures de vente en fonction de phénomènes autres que saisonniers.

3.3 Validation des profils estimés

La comparaison du profil estimé par la méthode d'analyse (étape 3) et du profil réel de chaque article de l'échantillon test est effectuée au moyen de la valeur du Chi2. Ainsi, tous les numéros d'articles dont la valeur moyenne se situe en-dessous de la valeur théorique du Chi2 au seuil de 5% à 50 degrés de liberté possèdent un profil estimé statistiquement acceptable.

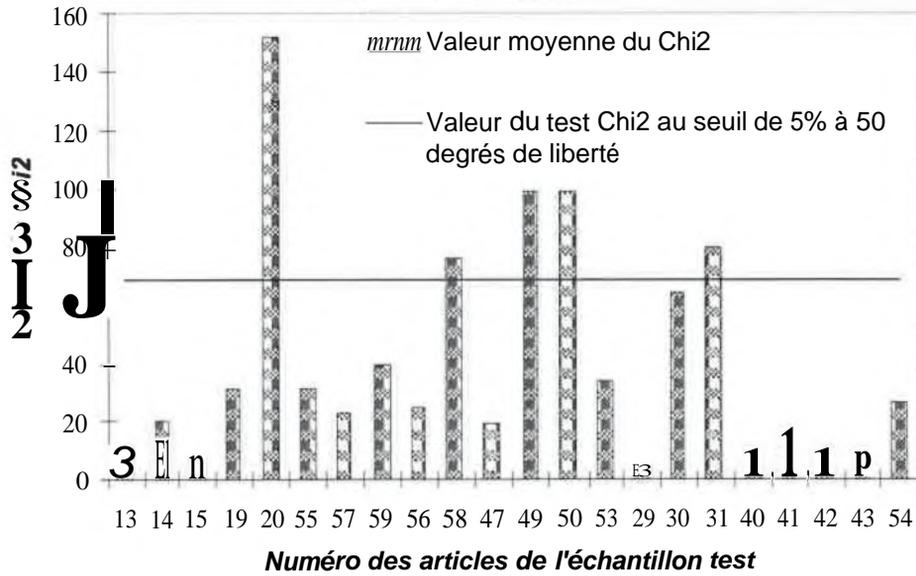


Figure 4 : Evaluation des profils estimés de vente par la valeur moyenne du Chi2.

Par exemple, les représentations des profils estimés et réels de vente des articles n°47 et n°40 permettent d'illustrer la précision obtenue des estimations par notre méthode d'analyse.

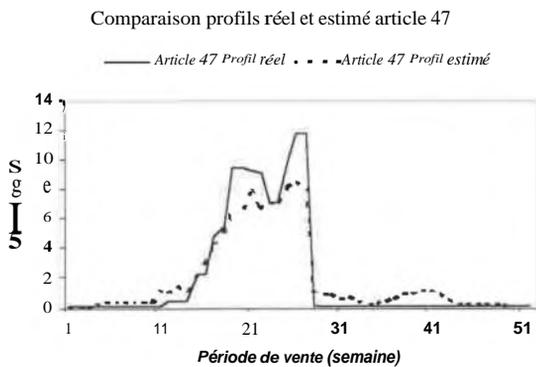


Figure 5 : Représentation des profils réel et estimé de l'article n°47 de l'échantillon test.

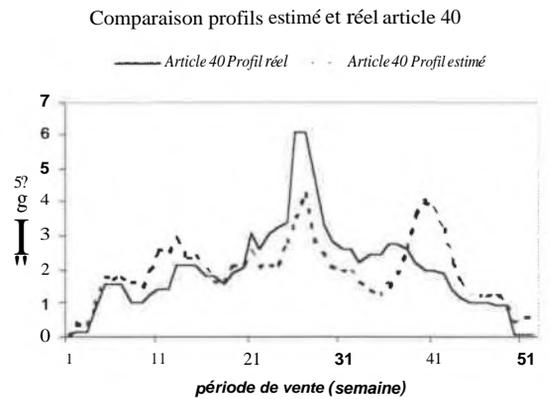


Figure 6 : Représentation des profils réel et estimé de l'article n°40 de l'échantillon test.

Dans notre cas de figure, seuls 5 articles sur les 22 de l'échantillon test (23%) ne possèdent pas une estimation de leur profil de vente par rapport à leur profil réel statistiquement acceptable. Par exemple, l'hypothèse selon laquelle le profil de vente de l'article 20 peut être expliqué par les variables qualitatives semble insuffisante et conduit à une estimation peu précise de son allure prévisionnelle. L'apport d'autres facteurs exogènes (météorologie, environnement économique, indice calendaire...), complétant les méthodes d'estimation, permettrait d'expliquer l'allure de vente de l'article n°20. Une meilleure connaissance du profil de vente d'un article textile conduit à une estimation plus précise de son allure prévisionnelle.

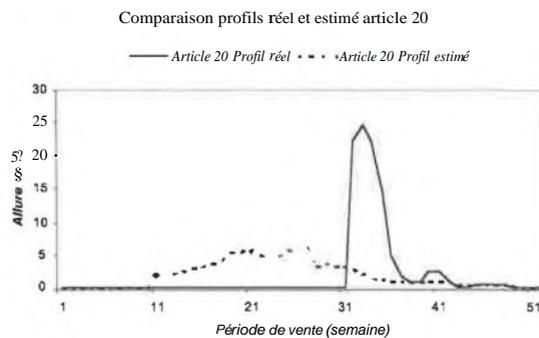


Figure 7 : Représentation des profils réel et estimé de l'article n°20 de l'échantillon test.

Cependant, les résultats obtenus dépendent étroitement du tirage aléatoire initial permettant la répartition des articles au sein des échantillons de base et de test. Pour conforter nos résultats, nous avons effectué 1000 tirages aléatoires de répartition pour comptabiliser le nombre d'articles "mal classés", c'est-à-dire les articles dont les estimations de profils sont statistiquement inacceptables par rapport à leurs profils réels. Le graphique suivant illustre ce résultat et en moyenne 4,51 articles de l'échantillon test sont "mal classés".

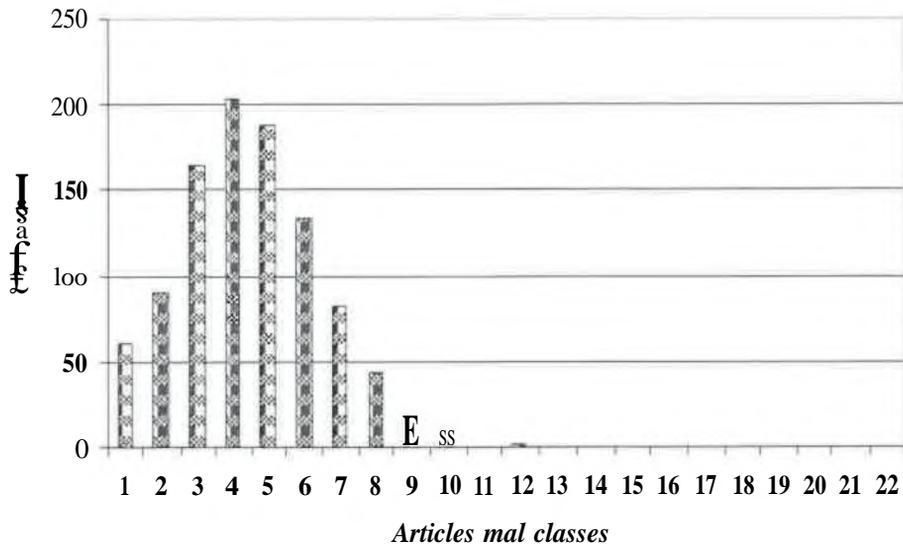


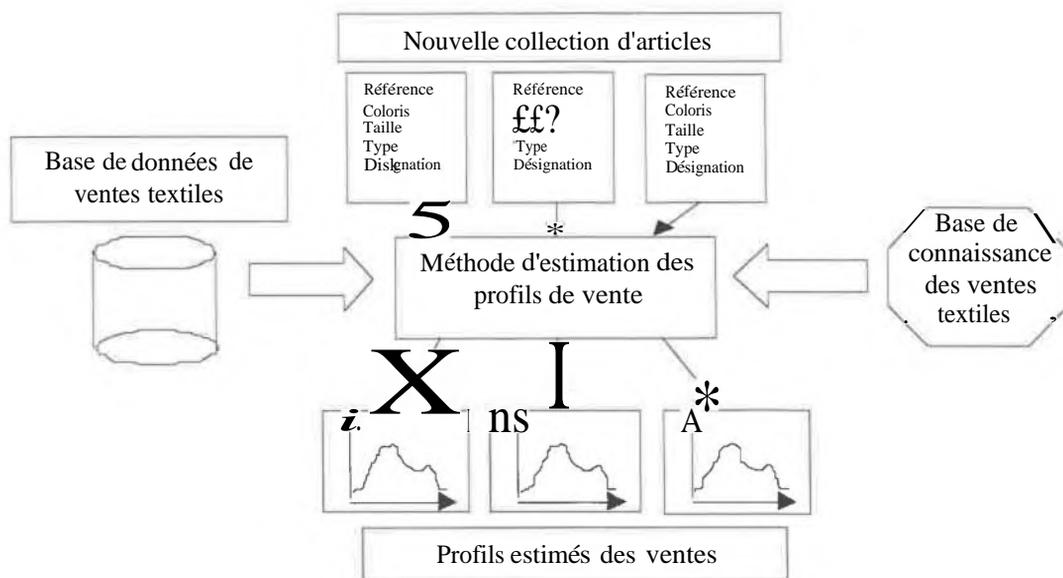
Figure 8 : Représentation de la distribution de la loi des erreurs.

De façon plus générale, notre méthode nous permet d'expliquer 80% en moyenne (79,5% exactement) des ventes des articles textiles par un regroupement de variables qualitatives similaires.

#### 4. Perspectives

Au delà de la compréhension d'une majorité des comportements de vente fournie par notre méthode d'analyse des données, nous pouvons également utiliser ces résultats à des fins prévisionnelles pour la mise en place de nouvelles collections.

En effet, à partir de la base de connaissances précédemment acquise et de la base de données de vente des articles textiles des saisons précédentes, un profil de vente peut être estimé pour chaque article de la nouvelle collection en fonction de ses caractéristiques qualitatives. Cependant, la principale difficulté réside dans le rapprochement ou l'appréciation de la similitude des variables qualitatives d'une saison de vente à une autre.



## 5. Conclusion

Notre étude a permis de confirmer l'idée selon laquelle les profils de vente peuvent être expliqués et estimés à partir des caractéristiques qualitatives connues *a priori* des articles. La méthode ainsi développée permettra à l'industriel d'estimer ces allures de vente de nouveaux articles à partir de leurs caractéristiques qualitatives.

L'hypothèse d'une possible construction des allures prévisionnelles de vente d'une majorité d'articles textiles uniquement à partir des variables descriptives est ici vérifiée et justifie que l'on conserve les données de vente des saisons précédentes pour en retirer non plus une simple information mais plutôt des règles de connaissance précises.

Cependant, pour certains articles (en nombre limité), l'utilisation seule de ces variables s'avère insuffisante. Des améliorations sont actuellement envisagées : incluant l'introduction d'autres paramètres exogènes (tels que : les indices météo ou calendaires...) ou la prise en compte de l'auto-corrélation éventuelle entre deux semaines consécutives.

## 6. Bibliographie

- Benzecri J.P., (1976), *L'analyse des correspondances*, Dunod, 2ème édition.
- Boussu F., Flappiette M. and Rabenasolo B., (1996), « Sales partition for forecasting into textile distribution network », *IEEE/SMC International Conference on Systems, Man and Cybernetics*, vol. 4, pp 2868-2873, Beijing, China.
- Boussu F., (1998), « Simulation de la filière Textile/Habillement/Distribution : Réduction de la complexité en vue d'une meilleure prévision de ventes », Thèse de doctorat, Université des Sciences et Techniques de Lille.

- Boussu F. and Denimal JJ. (2000), « The PLS regression : a logistic tool to the textile sales estimation », *Workshop RIRL 2000*, Acts on CD-ROM, Trois-Rivières, Canada.
- Denimal J.J., Diallo O.W., Boussu F., (1999), « Estimation of textile sales behaviour », *IMACS/IEEE CSCC'99 International Multiconference*, Acts on CD-ROM, Athens, Greece.
- Denimal JJ. and Boussu F. (2000), « Contribution to the logistic of a textile firm : Validation of a sale profiles estimation method », *IFAC/IFIP/IEEE Second Conference on Management and Control of Production and Logistics*, Grenoble, France.
- Jambu M., (1999), *Introduction au Datamining - Analyse intelligente des données*, Collection technique et scientifique des télécommunications, Eyrolles.
- Lefébure R. and Venturi G., (1999), *Le data Mining*, Eds Eyrolles-Informatiques.
- Tenenhaus M., (1998), *La régression PLS*, Editions Technip.
- Wold H., (1985), Partial Least Squares, *Encyclopedia of Statistical Sciences*, vol 6, Kotz S. & Jhonson N.L. (Eds), John Wiley & Sons, New York, pp 587-599.
- Xie X.L. and Beni G. (1991), « A validity measure for fuzzy clustering », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol 13, n°8, pp 841-847.