

# ESTIMATION INDIRECTE EN SCIENCES HUMAINES UNE MÉTHODE BAYÉSIENNE

**Henri Caussinus** (Institut de Mathématiques de Toulouse, henri.caussinus@wanadoo.fr)

**Résumé** - *Nous partons d'un problème de paléodémographie : l'estimation de la structure par âge d'une population ancienne à partir de caractéristiques biologiques relevées sur des restes osseux. Le problème statistique semble simple mais, d'une part, si les caractéristiques relevées sont corrélées à l'âge au décès, cette corrélation n'est pas très élevée, d'autre part les échantillons dont on dispose sont le plus souvent de taille réduite. Les méthodes statistiques les plus usuelles sont alors très instables et fournissent des résultats peu fiables. Pour les améliorer nous avons proposé une méthode bayésienne qui tient compte au maximum de la nature du problème concret et des informations peu quantifiables que possède le paléodémographe [1]. Nous décrivons ce problème et les grandes lignes de la solution proposée en essayant de mettre en relief en quoi des questions analogues sont susceptibles de se poser dans diverses disciplines (anthropologie, histoire,...) et en quoi une étroite collaboration entre le chercheur spécialisé et le statisticien est tout à fait nécessaire.*

**Mots clés** -

## Références

- [1] Caussinus, H., Courgeau, D., 2010, « Estimating Age without Measuring it : A New Method in Paleodemography », *Population-E*, 65 (1), p. 117-144.

# LOIS D'ÉCHELLE ET DIFFUSION DES INNOVATIONS DANS LES SYSTÈMES URBAINS

**Denise Pumain**

Géographie-Cités, pumain@parisgeo.cnrs.fr

**Résumé** - *Les lois d'échelle révèlent l'existence de contraintes pesant sur l'évolution des systèmes complexes. Appliquées à des ensembles de villes très interconnectées par de multiples réseaux, elles mesurent des relations non linéaires mais systématiques entre la population de ces villes et divers autres attributs urbains quantitatifs. D'après des ajustements entre population et nombre d'emplois dans des activités économiques ou des catégories socio-professionnelles, nous mettons en évidence que les exposants de ces fonctions puissance varient selon le stade atteint par le développement des activités dans les vagues d'innovation économiques et traduisent la diffusion hiérarchique de ces innovations dans le système des villes. Nous expliquons ainsi par une théorie urbaine évolutive les inégales concentrations de certaines activités et des niveaux de qualification relevés entre les métropoles et les plus petites villes.*

**Mots clés** -

# UNE MÉTHODE D'ESTIMATION ADAPTATIVE DES DIMENSIONS FRACTALES DE RÉSEAUX DE TALWEGS DE RAVINES

**Jean-Marc Bardet**

SAMM, Université Paris 1

90 rue de Tolbiac, 75013 Paris, France

**bardet@univ-paris1.fr**

Travail joint avec J.S. Bailly, B. Kaiser, C. Puech et N. Thommeret.

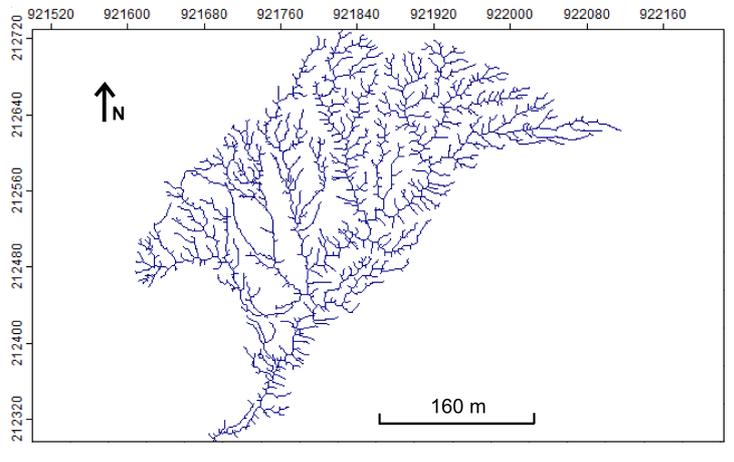
**Résumé** - *Dans cette contribution, nous considérons des réseaux de talwegs de ravines situés dans les Alpes de Hautes-Provence, et cherchons à calculer deux dimensions fractales “empiriques”, la dimension de boîte et la dimension de corrélation. Plusieurs difficultés apparaissent alors. En premier lieu le calcul de ces deux dimensions demandent de déterminer un intervalle contenant les différents rayons utilisés. Cela nous a conduit à proposer deux méthodes permettant d’estimer de manière adaptative la plus “grande” zone de linéarité dans un nuage de point, appelée intervalle de fractalité. Il est à noter que le coefficient  $R^2$  de la régression linéaire simple sur cet intervalle est compris entre 0.997 et 1 pour les réseaux de talwegs considérés. En second lieu, nous proposons une technique de bootstrap pour obtenir des intervalles de confiances sur l’estimation des intervalles de fractalité et sur le calcul des dimensions.*

**Mots clés** - Réseaux hydrographiques ; Réseaux de talwegs de ravines ; Dimension de boîte ; Dimension de corrélation ; Estimation adaptative

## 1 Réseaux de talwegs de ravines

Le réseau exploré ici (voir Figure 1) est un réseau de talwegs de ravines dans les secteurs très disséqués des PréAlpes de Digne (badlands de Draix). Nous nous intéressons au réseau dun bassin (le Moulin) extrait à partir de MNT, acquis dans le cadre de l’Observatoire de recherche en environnement ORE Draix, à deux résolutions différentes : 1 m et 2 m (Puech et al., 2009). A la résolution 1 m, le bassin du Moulin possède une densité de drainage de  $120 \text{ km}^{-1}$  (soit  $120 \text{ km.km}^{-2}$ ), un rapport moyen des longueurs selon les ordres de Strahler (c’est-à-dire la moyenne des rapports de longueur entre les segments d’ordre  $i$  et  $i + 1$ ) de 2.05. A la résolution 2 m, il possède une densité de drainage de  $85 \text{ km}^{-1}$ , un rapport moyen de longueur de 2.11. De tels réseaux existent également en plusieurs autres lieux sur la planète (notamment aux USA, en Chine,...). Mme si ce ne sont pas des réseaux hydrographiques mais des réseaux liés à l’érosion, leur étude peut se faire suivant les outils usuels appliqués aux réseaux hydrographiques. Parmi ces outils on s’intéressera ici au calcul de dimensions dites fractales.

FIGURE 1 – Réseau du Moulin extrait à partir dun MNT à 1 m (bassins expérimentaux de Draix, système de coordonnées Lambert 3)



## 2 Dimension de comptage de boîtes et dimension de corrélation

Pour un réseau ou un ensemble de points mathématiques (par exemple, un ensemble de Cantor, un flocon de Koch, un attracteur de Hénon,...) il est parfois possible de déterminer théoriquement une dimension fractale. Il existe en fait plusieurs dimensions fractales “théoriques”, dont notamment la dimension de Hausdorff, la dimension de Kolmogorov,... (voir Falconner, 1997).

Ces dimensions portent sur le comportement asymptotique (quand on zoom à l’infini) du réseau et ne peuvent tre utilisées telles quelles sur des réseaux ou ensembles de points issus de la nature. Nous présentons ici deux dimensions que l’on peut calculer à partir d’un ensemble de points  $A$  dans  $R^p$ .

### Dimension de comptage de boîtes

La dimension de comptage de boîtes se calcule de la manière suivante : pour  $r > 0$ , on considère  $N(r)$  le nombre minimum de boîtes de diamètres inférieurs à  $r$  permettant de recouvrir  $A$  (le diamètre étant calculé à partir une certaine distance). En pratique on choisit la forme des “boîtes”. Une manière simple de le faire est de recouvrir la partie de  $R^p$  contenant  $A$  d’une grille composée de pavés réguliers de longueur  $r$  (attention le choix de la forme de la boîte n’est pas anodin et influe sur le résultat final).

La dimension de Kolmogorov est obtenue comme  $D_K := \lim_{r \rightarrow 0} -\frac{\log(N(r))}{\log r}$  si cette limite existe, ce qui signifie qu’asymptotiquement  $N(r) \sim C r^{-D_K}$ . En pratique on choisira  $r_1, \dots, r_p$ , des diamètres “suffisamment” petits et à partir du nuage de points  $(\log(r_i), \log(N(r_i)))_i$ , la dimension de comptage de boîtes sera l’opposée de la pente de la droite de régression linéaire simple par moindres carrés. On pourra trouver quelques éléments asymptotiques sur le comportement de la dimension de boîtes dans Hall et Wood (1993).

### Dimension de corrélation

La dimension de corrélation avait déjà été introduite sous un autre nom par Alfred Rényi (voir Falconner, 1997), mais a été fortement popularisé par l'article de Grassberger et Proccacia (1983). Si on suppose que l'ensemble  $A$  est composé des points  $(X_1, \dots, X_n)$ , et si on choisit une norme  $\| \cdot \|$  (typiquement la norme euclidienne sur  $R^p$ , mais cela n'a pas d'importance), pour  $r > 0$  on calcule :

$$C(r) = \frac{2}{N(N-1)} \sum_{1 \leq i < j \leq N} \mathbf{1}_{\|X_i - X_j\| \leq r}$$

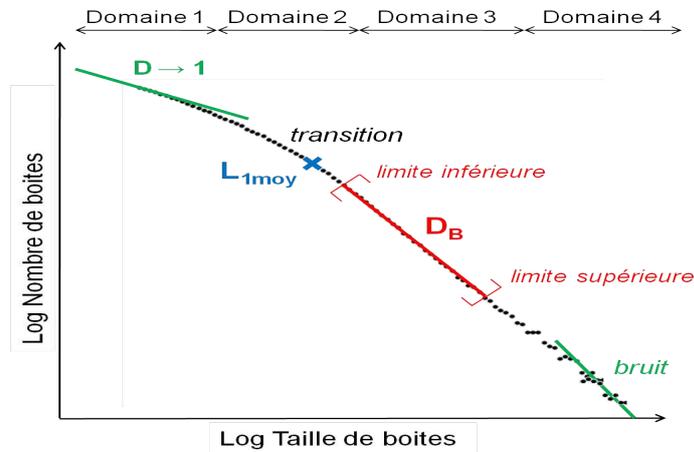
Ainsi  $C(r)$  est la proportion de couples  $(X_i, X_j)$  dont l'inter-distance est inférieure à  $r$ . Une nouvelle fois, pour une fractale théorique, asymptotiquement on peut escompter que  $C(r) \sim Cr^{-a}$ , o  $C$  est une constante et  $a$  un nombre réel positif (appelé dimension de corrélation théorique). Mais pour une fractale de la nature, on procédera comme précédemment : on choisit  $r_1, \dots, r_p$ , des rayons "suffisamment" petits et à partir du nuage de points  $(\log(r_i), \log(C(r_i)))_i$ , la dimension de corrélation  $D_C$  sera l'opposée de la pente de la droite de régression linéaire simple par moindres carrés.

### 3 Intervalle de fractalité

Nous nous plaons ici dans le cadre d'un réseau de Talwegs, donc  $p = 2$  et les dimensions  $D_B$  et  $D_C$  appartiennent à l'intervalle  $[1, 2]$ .

Les deux dimensions présentées demandent le choix de  $(r_1, \dots, r_p)$ . Or ce choix n'est en général pas simple. Si on choisit  $r$  trop petit, on risque d'obtenir  $C(r) = 0$  et  $N(r)$  qui varie comme  $1/r$ . En revanche, si  $r$  est trop grand  $C(r) = 1$  et  $N(r) = 1$ . Voici par exemple sur Figure 2 un exemple de ce que l'on obtient pour le calcul de  $D_B$  en ayant choisi une très grande plage d'échelles (entre 0.1 m et 200 m, sachant que la résolution est de 1 m et que la taille totale du bassin est d'environ 500 m). On voit ainsi apparaître plusieurs zones

FIGURE 2 – Nuage de points  $(\log(r), \log(N(r)))$  pour  $r$  variant de 0.1 m à 200 m



sur ce graphique. La zone qui nous intéresse est celle de linéarité notée Domaine 3, que nous appellerons également intervalle de fractalité. C'est pour des rayons choisis dans cet intervalle que nous calculerons  $D_B$ . La question que l'on se pose est la suivante : comment estimer cet intervalle de fractalité, qui d'une certaine manière correspond à la plus grande zone de linéarité intermédiaire du nuage de points ?

Pour ce faire nous allons utiliser une technique de détection de ruptures dans un modèle composé de 4 régimes différents. Ainsi estimera-t-on les instants de ruptures  $t_1, t_2$  et  $t_3$  comme suit :

$$(\hat{t}_1, \hat{t}_2, \hat{t}_3) = \text{Argmin}_{t_1 < t_2 < t_3} \left( \sum_{t=1}^{t_1} (\log(N(r_t)) - \hat{a}_0 + \log(r_t))^2 + \sum_{t=t_1+1}^{t_2} (\log(N(r_t)) - \hat{b}_0 - \hat{b}_1 \log(r_t) - \hat{b}_2 \log^2(r_t))^2 \right. \\ \left. + \sum_{t=t_2+1}^{t_3} (\log(N(r_t)) - \hat{c}_0 - \hat{c}_1 \log(r_t))^2 + \sum_{t=t_3+1}^p (\log(N(r_t)) - \hat{d}_0 - \hat{d}_1 \log(r_t))^2 \right).$$

Les différents coefficients  $\hat{a}_0, \dots, \hat{d}_1$  sont obtenus par moindres carrés dans la zone  $[t_i + 1, t_{i+1}]$  correspondante. L'intervalle de fractalité sera ainsi  $[r_{t_2+1}, r_{t_3}]$ .

Le mme type de graphe apparaît si on considère  $\log(C(r))$  au lieu de  $\log(N(r))$ , à ceci près que le Domaine 1 est remplacé par une fonction croissante en escalier. Nous proposons donc une autre technique pour estimer l'intervalle de fractalité.

Les résultats obtenus sont les suivants :

	$D_B$	Intervalle de fractalité	$R^2$	$D_C$	Intervalle de fractalité	$R^2$
Moulin 1 m	1.70	[5.3 m, 44.3 m]	0.9998	1.81	[15.6 m, 42.4 m]	0.9999
Moulin 2 m	1.57	[8.7 m, 40.5 m]	0.9997	1.78	[16.7 m, 46.8 m]	0.9999

On obtient donc des résultats assez proches pour les 2 résolutions, les intervalles de fractalité étant également assez proches pour les 2 dimensions calculées. On remarque également l'incroyable qualité du modèle linéaire simple, avec des coefficients  $R^2$  quasiment égaux à 1 : la fractalité de ces réseaux est édifiante...

### 3.1 Robustesse des résultats

Le calcul des dimensions demande l'estimation de l'intervalle de fractalité. Ainsi la robustesse du calcul de ces dimensions est essentiellement liée à celle de l'estimation de ces intervalles. Il n'est pas possible d'obtenir un intervalle de confiance théorique sur ces intervalles de fractalité. La technique que nous avons choisie est celle du bootstrap : on sélectionne plusieurs fois aléatoirement une partie du réseau sur laquelle on effectue les estimations et calculs de dimension.

Les résultats obtenus confirme la robustesse de la technique proposée : les intervalles de fractalité sont à  $\pm 2$  m de ceux proposés dans le tableau et on aboutit ainsi à une fourchette  $\pm 0.02$  pour le calcul des deux dimensions.

## Références

- [1] Dodds P.S., Rothmann D.H. (1999), Unified view of scaling laws for river networks, *Physical Review E.*, **59**, p. 4865-4877.
- [2] Falconer K. (1997), *Techniques in fractal Geometry*, John Wiley & Sons, Chichester, 256 p.
- [3] Forriez M., Martin P., Nottale L. (2010), Lois d'échelle et transitions fractal-non fractal en géographie, *L'Espace géographique*, **39**, p. 97-112.
- [4] Frankhauser P. (1994), *La fractalité des structures urbaines*, Paris, Anthropos, 291 p.
- [5] Grassberger P., Procaccia I. (1983), Measuring the strangeness of strange attractors, *Physica D*, p. 189-208.
- [6] Hall P., Wood A. (1993), On the performance of box-counting estimators of fractal dimension, *Biometrika*, **80**, p. 246-251.
- [7] Hauchard E. Delahaye D. Frankhauser P. (1999), Analyse morphologique de talwegs et comportement scalant. Application des dimensions non-entières à la géographie physique, *L'Espace géographique*, **3**, p. 215-224.
- [8] Horton R.E. (1945), Erosional development of streams and their drainage basins : hydro-physical approach to quantitative morphology, *Geological Society of America, Bulletin*, **56**, p. 275-370.
- [9] Moussa R., Bocquillon C. (1993), Morphologie fractale du réseau hydrographique, *Hydrological Sciences Journal*, **38**, p. 187-201.
- [10] Puech C., Thommeret N., Kaiser B., Bailly J.S., Jacome A., Rey F., Mathys N. (2009), MNT à très haute résolution dans les modelés fortement disséqués : des données aux tests d'application, *Géomorphologie : relief, processus, environnement*, **2**, p. 141-152.
- [11] Rodriguez Iturbe I., Rinaldo A. (1997), *River basins ; chance and self-organisation*, New-York, Cambridge University Press, 547 p.
- [12] Schumm S.A. (1956), Evolution of Drainage Systems and Slopes in Badlands at Perth Amboy, *Geological Society of America. Bulletin*, **67**, p. 597-646.
- [13] Tarboton D.G. (1996), Fractal river networks, Horton's laws and Tokunaga cyclicity, *Journal of Hydrology*, **187**, p. 105-117.
- [14] Thommeret N. et Puech C. (2010), Réseaux de talwegs en secteurs érodés ; de la détection spatiale à l'analyse fractale, *Revue Internationale de Géomatique*, **20**, p. 201-219.
- [15] Turcotte D.L. (2007), Self-organized complexity in geomorphology : Observations and models, *Geomorphology* **91**, p. 302-310.

SPATIAL CORRELATION IN BIPARTITE NETWORKS : THE IMPACT OF  
THE GEOGRAPHICAL DISTANCES ON THE RELATIONS IN A CORPUS OF  
MEDIEVAL TRANSACTIONS

Nathalie Villa-Vialaneix (SAMM, Université Paris 1, nathalie.villa@univ-paris1.fr)

Bertrand Jouve (ERIC, Université Lyon 2, bertrand.jouve@univ-lyon2.fr)

Fabrice Rossi (SAMM, Université Paris 1, fabrice.rossi@univ-paris1.fr)

Florent Hautefeuille (Traces, Université toulouse 2, florent.hautefeuille@univ-tlse2.fr)

**Résumé** - *Nous nous intéressons à l'influence de la localisation spatiale des individus sur la manière dont ils interagissent : le cas d'étude auquel nous nous intéressons est celui d'un corpus d'actes notariés du Moyen-Âge. Dans ce corpus, les individus interagissent par le biais de transactions communes qui sont géo-localisées. Ce travail cherche à quantifier l'impact de la localisation géographique sur le réseau de relations des individus. Comme positionnement géographique et relations des individus sont construits à partir de la même source (la liste des transactions auxquelles chaque individu a participé), un test de Mantel classique est inadéquat : nous proposons une méthodologie basée sur une adaptation du test de comparaison de distances que nous illustrons dans ce contexte.*

**Mots clés** -

# CARTOGRAPHIE DE LA CHRONIQUE D'HENRI DE LIVONIE

**Nicolas Bourgeois**

(ESSEC, nbourgeo@phare.normalesup.org)

**Résumé** - *La chronique d'Henri de Livonie est la principale et la plus ancienne source écrite à notre disposition en ce qui concerne l'évangélisation des actuels Pays Baltes. Nous nous proposons d'étudier par des méthodes simples issues des statistiques ou de la théorie des graphes la cartographie de la Livonie que cette chronique nous dessine, dans l'espoir d'en tirer quelques intuitions tant sur la rédaction du texte que sur les événements qu'il relate.*

**Mots clés** -

# HDclassif: An R Package for Model-Based Clustering and Discriminant Analysis of High-Dimensional Data

L. Bergé<sup>a</sup> and C. Bouveyron<sup>b</sup> and S. Girard<sup>c</sup>

<sup>a</sup>Laboratoire GREthA  
Université Bordeaux IV  
laurent.berge@u-bordeaux4.fr

<sup>b</sup>Laboratoire SAMM, EA 4543  
Université Paris 1 Panthéon-Sorbonne  
charles.bouveyron@univ-paris1.fr

<sup>b</sup>Equipe Mistis  
INRIA Rhône-Alpes & LJK  
stephane.girard@inrialpes.fr

**Mots clefs** : Model-based classification and clustering, high-dimensional data, subspaces.

This paper presents the **R** package *HDclassif* which is devoted to the clustering and the discriminant analysis of high-dimensional data. The classification methods proposed in the package result from a new parametrization of the Gaussian mixture model which combines the idea of dimension reduction and model constraints on the covariance matrices. The supervised classification method using this parametrization is called high dimensional discriminant analysis (HDDA). In a similar manner, the associated clustering method is called high dimensional data clustering (HDDC) and uses the expectation-maximization algorithm for inference. In order to correctly fit the data, both methods estimate the specific subspace and the intrinsic dimension of the groups. Due to the constraints on the covariance matrices, the number of parameters to estimate is significantly lower than other model-based methods and this allows the methods to be stable and efficient in high dimensions. Two introductory examples illustrated with **R** codes allow the user to discover the *hdda* and *hddc* functions. Experiments on simulated and real datasets also compare HDDC and HDDA with existing classification methods on high-dimensional datasets. *HDclassif* is a free software and distributed under the general public license, as part of the **R** software project.

The **R** package *HDclassif* (currently in version 1.2) implements these two classification methods for the clustering and the discriminant analysis of high-dimensional data. The package is available from the CRAN at <http://CRAN.R-project.org/package=HDclassif>.

## Références

- [1] L. Bergé, C. Bouveyron and S. Girard, *HDclassif : an R Package for Model-Based Clustering and Discriminant Analysis of High-Dimensional Data*, Journal of Statistical Software, vol. 42 (6), pp. 1-29, 2012.

## URBAN-SEGREGATION MODELS

**Aurélien Hazan**(Lissi, Université Paris-Est Créteil aurelien.hazan@u-pec.fr)  
& **Julien Random-Furling** (SAMM, Université Paris 1,  
Julien.Randon-Furling@univ-paris1.fr)

**Résumé** - *Combinations of multi-agent stochastic simulations and analytical solutions from statistical physics spin models have recently provided new insights into segregation phenomena described via Schelling's classic model. We explore various types of segregation dynamics on the square lattice, with a view to identifying courses of action leading to less-segregated environments which are stable in the long-run.*

**Mots clés** - **Schelling; Segregation; Multi-Agent Models; Statistical Physics; Complex Systems.**

## SOCIO-SEMANTIC FRAMEWORKS FOR TECHNO-SOCIAL SYSTEMS

**Camille Roth** (EHESS, roth@ehess.fr)

**Résumé** - *Socio-semantic systems involve agents who create and process knowledge, exchange information and create ties between ideas in a distributed manner : webloggers, communities of scientists, software developers and wiki contributors are, among others, examples of such networks. The state-of-the-art in this regard focuses on two main issues which are generally addressed in an independent manner : the description of content dynamics and the study of social network characteristics and evolution. This talk will review recent endeavors to merge both types of dynamics into co-evolutionary frameworks, where social and semantic aspects are being jointly appraised. Case studies featuring socio-semantic networks, socio-semantic hypergraphs and automatic socio-semantic history reconstruction will be notably discussed.*

**Mots clés** -

# MODÉLISATION DU TEMPS DE VENTE DES VÉHICULES D'OCCASION

**Solohaja-Faniaha DIMBY**

Université Paris 1

90 rue de Tolbiac, Paris, France

**Solohaja-Faniaha.Dimby@malix.univ-paris1.fr**

**Résumé** - *En France, le secteur automobile fournit au total environ 2,5 millions d'emplois. En 2007, la production de véhicules a atteint près de 6,2 millions, soit 8,8% de la production mondiale. Plusieurs politiques ont été mises en place par les constructeurs afin d'inciter les consommateurs pour l'achat de véhicules neufs tout en réduisant la durée de détention des véhicules achetés. Cela fait que la fluidité du marché automobile est basée fortement sur le marché de l'occasion. Nous présentons ici la manière dont nous avons modélisé, le temps qu'un véhicule d'occasion se met à vendre. Cela consiste à créer à partir des durées de vente, une variable catégorielle indiquant le segment de rotation et appliquer par la suite une régression logistique avec les différents niveaux de cette variable. Les variables explicatives ont été choisies à partir d'un modèle qui permettait de prédire le prix des véhicules. Les performances du modèle de prédiction ont été estimées au moyen de simulations. Les résultats ont confirmé les informations données par les experts du marché, entre autre, la durée de vente dépend du modèle, de la carrosserie et de la motorisation du véhicule.*

**Mots clés** - *régression logistique, rotation de marché, véhicule d'occasion.*

# VISUALISATION DE BANQUES D'IMAGES PAR DES TECHNIQUES DE RÉDUCTION DE DIMENSION

**John A. Lee**

(Machine Learning Group, UCL, john.lee@uclouvain.be)

**Résumé** - *Comment visualiser un ensemble d'images sur une feuille de papier ou un écran d'ordinateur? Comment afficher les images qui se ressemblent le plus de manière à ce qu'elles soient proches les unes des autres? Des techniques de réduction de dimension peuvent résoudre ce problème. En effet, les images sont des objets qui évoluent dans des espaces de très grandes dimensions, de l'ordre du nombre de pixels dans les images. Une représentation compacte en deux dimensions passe donc obligatoirement par une réduction drastique du nombre de variables considérées.*

*Les techniques décrites dans cet exposé opèrent par force brute en agissant directement sur les images vectorisées, mais elles restent applicables à d'autres formes de données multidimensionnelles et donc aussi à des caractéristiques ou propriétés extraites des images. L'exposé partira de techniques bien connues, telles que l'analyse en composantes principales, plusieurs variantes de l'échelonnage multidimensionnel, et les cartes auto-organisées. Viendra ensuite une analyse plus détaillée des techniques récentes basées sur la préservation de similarités. Nous montrerons en particulier comment ces méthodes parviennent à déjouer élégamment le phénomène de concentration des normes. Quelques exemples clôtureront l'exposé.*

**Mots clés** -

# CURVE ALIGNMENT : THEORY AND APPLICATIONS

**Benoît Frenay**

(Machine Learning Group, UCL, Benoit.Frenay@uclouvain.be)

**Résumé** - *Curves are ubiquitous in data analysis, where many quantities of interest are represented as a time series. Hence, it is often necessary to compare curves, e.g. in order to analyse the curves or to cluster them. However, the number of points in each curve to be compared is not necessarily identical and measures are not necessarily taken at the same time. Moreover, slight time shifts may appear in the curves. These phenomena make the comparison of curves difficult. In such cases, it is necessary to preprocess the curves in order to align them. This presentation shows how to perform curve alignment using dynamic time warping. Several examples of successful applications of curve alignment are given.*

**Mots clés** -

# CLUSTERING PATTERNS OF URBAN BUILT-UP AREAS WITH CURVES OF FRACTAL SCALING BEHAVIOUR

**Benoît Frenay**

(Machine Learning Group, UCL, Benoit.Frenay@uclouvain.be)

**Résumé** - *Fractal dimension is an index which can be used to characterize urban areas. The use of the curve of scaling behaviour is less common. However, its shape gives local information about the morphology of the built-up area. This work suggests a method based on a  $k$ -medoid for clustering these curves. It is applied to forty-nine wards of European cities, and shows that the curves add interesting intraward information to our knowledge of the spatial variation of the urban texture. Moreover, morphological similarities are observed between cities : living, architectural, and planning trends are not specific to individual cities.*

**Mots clés** -