
Exploitation de données brutes de trafic routier urbain issues d'un réseau de capteurs géoréférencés

M. Joliveau* — F. De Vuyst* — G. Jomier** — C. Medeiros***

* MAS, École Centrale Paris, France
{marc.joliveau,florian.de-vuyst}@ecp.fr

** LAMSADE, Université Paris Dauphine, France
genevieve.jomier@dauphine.fr

*** Institute of Computing, UNICAMP, Brazil
cmbm@ic.unicamp.br

RÉSUMÉ. Les données de trafic provenant de réseaux de capteurs ont été la source de nombreuses recherches reliées aux SIT (Systèmes d'Information en Transport). Ces données se représentent généralement par de grands ensembles de séries spatio-temporelles corrélées. Ce papier présente une nouvelle approche pour manipuler des données brutes issues de capteurs fixes géoréférencés. Notre travail est fondé sur la combinaison de méthodes analytiques pour préparer les données des capteurs et sur la proposition d'une architecture pour un système d'information dédié au trafic routier. Il a été conduit dans le cadre d'un projet utilisant des données réelles générées par 1000 capteurs pendant 3 ans dans une grande agglomération française.(ville et couronne)

ABSTRACT. Traffic data coming from sensor networks have prompted a wide range of research issues related with Transportation Information Systems. These data are usually represented by large and complex spatio-temporal series. This paper presents a new approach to manage rough data coming from static georeferenced sensors. Our work is based on combining analytic methods to process sensor data and proposing an architecture for an information system dedicated to road traffic. It is being conducted within a project which uses real data, generated by 1000 sensors, during 3 years, in a french big city.

MOTS-CLÉS : Trafic routier, capteurs géoréférencés, algorithme STPCA, motifs spatio-temporels, système d'information de trafic routier, intelligibilité.

KEYWORDS: Road traffic, georeferenced sensors, STPCA algorithm, spatio-temporal pattern, road traffic information system, intelligibility.

1. Introduction

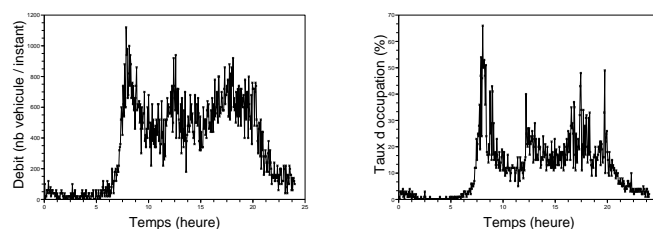
De par la place prédominante occupée par le transport dans notre environnement socio-économique et les enjeux d'environnement et de santé publique, de nombreuses recherches actuelles visent à fournir des outils de précision d'aide à la gestion et d'aide à la décision. Les travaux peuvent par exemple concerner l'étude des comportements des usagers, la prévision et le traitement des congestions ou l'influence de la météo sur le trafic. Ce papier se rapporte au traitement, à l'extraction de connaissances et au stockage de données de trafic routier urbain issues d'un réseau de capteurs géoréférencés.

2. Contexte du problème

Nos recherches s'inscrivent dans le cadre du projet ACI Masse de données CADDY (Contrôle de l'Acquisition de Données, stockage et modèles DYnamiques) [CAD 07]. Ce projet réunit une équipe scientifique pluridisciplinaire de différents instituts. Le but de CADDY est de développer un outil d'aide à la décision pour la gestion du trafic routier.

Les capteurs utilisés sont géoréférencés, fixés le long des voies de circulation. Ils collectent différentes informations sur le trafic durant la journée à fréquence régulière. Deux variables principales sont mesurées, produisant deux séries spatio-temporelles distinctes (mais interdépendantes) :

- le débit de véhicules (q), i.e., le nombre de véhicules passé devant le capteur durant une période donnée (3 minutes dans notre cas) ;
- le taux d'occupation (τ), i.e., l'espace moyen entre les véhicules durant un intervalle de temps donné. Un taux d'occupation de 100% signifie qu'il n'y a aucun espace entre les véhicules, alors que 0% indique qu'il n'y a pas de véhicules.



(a) Débit de véhicules

(b) Taux d'occupation

Figure 1. Débit et taux d'occupation journalier pour un capteur donné.

La figure 1 illustre les séries temporelles journalières de débit et de taux d'occupation mesurées à un capteur donné, un jour de semaine. Nos données sont collectées depuis 1000 capteurs sur 3 ans, les prises de mesure s'effectuant toutes les 3 minutes. Ceci représente un total de $420 \cdot 10^6$ valeurs. Nos données proviennent du système de supervision du trafic CLAIRE [SCE 04] développé par l'INRETS (Institut National de REcherche sur les Transports et leur Sécurité) au laboratoire GRECIA (G. Scemama). CLAIRE modélise le réseau routier urbain par l'intermédiaire d'un graphe orienté, où chaque arc correspond à une portion de route. Les localisations des capteurs mesurant le trafic sont associées à ces arcs.

Les données doivent être préparées en suivant plusieurs étapes cruciales afin que q et τ soient combinés pour offrir aux experts une vue globale du problème. L'interdépendance entre nos variables complexifie l'analyse mathématique. Les données sont mesurées et stockées séparément pour chaque capteur, chaque jour. Cela signifie qu'il existe au moins deux moyens de les représenter sous la forme de longues séries temporelles. Une première représentation consiste à représenter les données par capteur au cours du temps, et une deuxième par instant de mesure tous capteurs confondus. Chacune de ces représentations admet une analyse distincte. La première permet d'étudier le comportement, pour un capteur au cours du temps (point dans l'espace fixe, temps variant) alors que la seconde offre une vue de l'intégralité du réseau à un moment donné (temps fixe, espace variant). Dans la section 3 nous proposons de tenir compte simultanément de ces deux analyses pour calculer des résumés spatio-temporels des données.

La qualité des données représente une autre source de problèmes. Les données transmises par les capteurs sont très bruitées, parfois pas cohérentes, et contiennent beaucoup de valeurs manquantes. Celles-ci sont principalement la conséquence de pannes des capteurs ou de coupure totale du réseau. Les données nécessitent donc d'être nettoyées si on cherche à effectuer des calculs dessus. La section suivante présente les solutions proposées afin de réduire la dimension des données tout en réglant le problème des données manquantes. Elle explique comment les séries peuvent être nettoyées et représentées par une faible quantité de descripteurs avec une erreur d'estimation très faible.

3. Résumés spatio-temporels de données issues de capteurs

Cette section présente nos solutions pour traiter les données spatio-temporelles issues de capteurs, afin de les résumer. On commence en présentant une méthode qui diminue très fortement la dimension des séries temporelles (de débit et de taux d'occupation) pour l'ensemble des capteurs tout en conservant au mieux l'information au sens de l'énergie. Ensuite, une adaptation de cette méthode à un ensemble de données contenant des valeurs manquantes est proposée. Finalement, on combine sans perte les informations contenues dans les deux séries (débit et taux d'occupation) à l'aide d'une variable à valeurs continues dans l'intervalle $[0, 1]$ pour décrire le comportement du trafic à un capteur. Cet intervalle peut être divisé en plusieurs classes représentant des états du trafic définis par des experts (par exemple "saturé" ou "calme").

3.1. Réduction de la dimension - l'analyse STPCA

Dans [DEV 07], nous avons introduit l'algorithme de réduction appelé STPCA (*Space-Time Principal Component Analysis*) pour réduire la dimension d'un ensemble de données spatio-temporelles. Cette méthode consiste à appliquer simultanément une Analyse en Composantes Principales (ACP) [JOL 86] sur les dimensions spatiale et temporelle.

Supposons que nous possédons les données de trafic mesurées sur N jours. Les valeurs du débit ou du taux d'occupation peuvent être stockées dans une matrice \mathbf{X}^n , où n est un label de jour. De cette manière, la série temporelle discrète correspondant aux mesures faites par le capteur i le jour n est donnée par x_i^n . Les étapes décrites dans la suite sont appliquées séparément au débit et au taux d'occupation, calculant deux ensembles de séries temporelles résumées par l'analyse STPCA.

1) Dans le cadre d'une analyse spatiale, les matrices journalières sont assemblées horizontalement dans une unique matrice \mathbf{Y} et verticalement dans une matrice \mathbf{Z} . Dans la matrice \mathbf{Y} , chaque colonne contient toutes les valeurs mesurées à un moment et chaque ligne correspond à un capteur du réseau, les valeurs variant dans le temps (localisation fixe). Dans la matrice \mathbf{Z} , les colonnes représentent les capteurs et chaque ligne correspond aux mesures de tous les capteurs à un instant donné (temps fixe).

2) Une décomposition en valeur singulière est ensuite appliquée aux matrices \mathbf{Y} et \mathbf{Z} . A partir de la matrice de corrélation spatiale $\mathbf{M}^s = \mathbf{Y}\mathbf{Y}^T$, les K premiers vecteurs propres spatiaux $(\Psi^k)_{k=1\dots K_M}$ sont déterminés, avec $K \ll K_M$ et stockés dans une matrice \mathbf{P} :

$$\mathbf{P} = \text{col} \left(\Psi^1, \Psi^2, \dots, \Psi^K \right).$$

Quant à la matrice de corrélation temporelle $\mathbf{M}^t = \mathbf{Z}^T\mathbf{Z}$, elle permet de calculer les L premiers vecteurs temporels $(\Phi^l)_{l=1\dots L_M}$, avec $L \ll L_M$, qui sont ensuite stockés dans la matrice \mathbf{Q} :

$$\mathbf{Q} = \text{col} \left(\Phi^1, \Phi^2, \dots, \Phi^L \right).$$

3) Finalement, l'estimé par réduction STPCA $\hat{\mathbf{X}}^n$ de la matrice \mathbf{X}^n est défini par :

$$\hat{\mathbf{X}}^n = \mathbf{P}\mathbf{P}^T \mathbf{X}^n \mathbf{Q}\mathbf{Q}^T.$$

Remarquons que la matrice d'ordre réduit est donnée par :

$$\mathbf{X}_n^r = \mathbf{P}^T \mathbf{X}^n \mathbf{Q},$$

de taille $K \times L$, où les valeurs de K et L sont choisies pour être petite. Les valeurs de ces paramètres peuvent être déterminés à partir de la proportion d'énergie saisie par les vecteurs propres. L'énergie est obtenue par la trace de la matrice de corrélation \mathbf{M} définie par :

$$\text{tr}(\mathbf{M}) = \sum_{j=1}^J \lambda_j(\mathbf{M}),$$

où $\lambda_j(\mathbf{M})$ représente la j -ième valeur propre de \mathbf{M} et J le nombre de vecteurs propres. Les expériences numériques réalisées avec de très petites valeurs de ces paramètres, précisément $K = L = 3$ correspondant à un facteur de réduction de l'ordre de 10^4 , ont démontrées les aptitudes de la STPCA à proposer une bonne estimation des données -le lecteur se rapportera à [DEV 07] pour plus de détails.

Adaptation de la STPCA aux valeurs manquantes. La méthode STPCA ne peut pas être appliquée directement à un ensemble de données contenant des valeurs manquantes. Notre solution à ce problème a été définie dans [JOL 07] où on propose une adaptation de la STPCA aux valeurs manquantes. On utilise l'algorithme Expectation Maximization (EM) [DEM 77] afin d'estimer séparément les matrices de corrélation spatiale M^s et temporelle M^t . D'un autre côté, on calcule une estimation complète des données en utilisant les k plus proches séries de chaque série temporelle (au sens de la distance euclidienne). Cette estimation est finalement projetée sur les principales composantes des estimations de M^s et M^t .

Les expériences menées dans [JOL 07] ont montrées que les résultats obtenus par la STPCA sur des données contenant plus de 20% de valeurs manquantes demeuraient extrêmement proches de ceux obtenus par la STPCA sur l'ensemble de données complet correspondant.

3.2. Une sémantique pour les variables de trafic

La théorie du transport décrit la relation entre le débit de véhicules q et le taux d'occupation τ comme la loi fondamentale du trafic routier. Cette relation se représente graphiquement par le diagramme fondamental issu de la Théorie du Transport. La courbe en gras sur la figure 2 illustre l'allure globale du diagramme fondamental dans un repère où q et τ sont respectivement représentés par les axes des abscisses et des ordonnées.

Nous proposons une nouvelle variable, l'état de circulation E , dérivée du diagramme

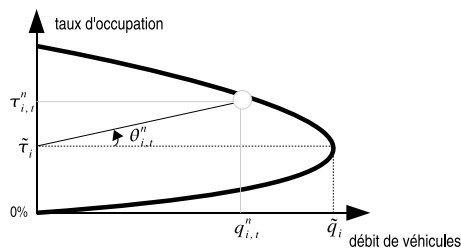


Figure 2. Diagramme fondamentale et variable d'état de circulation E .

fondamental et combinant sans perte d'information les données de débit et de taux d'occupation saisies par un capteur à un moment donné. Cette variable se base sur le calcul de deux informations retournées par chaque capteur i :

– le taux d’occupation maximum moyen (\tilde{q}_i) est donné par la moyenne du taux d’occupation maximal journalier mesuré au capteur i . Il correspond à un débit quasi-optimal respectivement au trafic devant le capteur i ;

– la variable $\tilde{\tau}_i$ mesure la valeur moyenne du taux d’occupation lorsque la circulation devant le capteur i atteint son débit journalier maximal.

A un instant t fixé, le jour n , le débit et le taux d’occupation au capteur i sont respectivement donnés par $q_{i,t}^n$ et $\tau_{i,t}^n$. Pour chacune de ces mesures, on calcule l’angle $\theta_{i,t}^n$ formé par la droite $y = \tilde{\tau}_i$ et la droite joignant les points $(0, \tilde{\tau}_i)$ et $(q_{i,t}^n, \tau_{i,t}^n)$. A partir de la valeur de $\theta_{i,t}^n$, nous définissons ensuite $e_i^n(t)$ la valeur de l’état de circulation E au capteur i , le jour n à l’instant t :

$$e_i^n(t) = \frac{1}{2} + \frac{1}{\pi} \theta_{i,t}^n = \frac{1}{2} + \frac{1}{\pi} \arctg\left(\frac{(\tau_{i,t}^n - \tilde{\tau}_i) \tilde{q}_i}{100 q_{i,t}^n}\right).$$

Précisons que les valeurs de E sont normalisées entre 0 et 1. Si la valeur de E est proche de zéro, le trafic est calme et fluide (débit et densité faibles). Un état de circulation dont la valeur est 1 représente une circulation congestionnée immobile (*jam traffic*) alors qu’une valeur proche de 0.5 symbolise une situation quasi-optimale du trafic avec un débit fort. Ainsi, en plus de combiner sans perte d’information le débit et le taux d’occupation mesurés par les capteurs, l’état de circulation offre également une représentation normalisée et intelligible du trafic.

Une fois les valeurs normalisées obtenues (par capteur, par jour), on considère l’information de manière experte pour définir sept états quantifiés à partir des valeurs de $e_i^n(t)$ - "saturé", "calme", "grande circulation", etc. . . Par cette procédure, on passe d’une description numérique des données à une description symbolique [HUG 03]. La mise en symbole de nos données de trafic est abordée de manière plus détaillée dans [JOL 08].

4. Extraction de connaissances sur les données de trafic

La méthode STPCA et l’introduction de la variable d’état de circulation E présentés dans la section précédente ont permis de préparer les données et notamment de passer de séries temporelles incomplètes à des résumés symboliques. Ces éléments combinés à d’autres outils permettent aussi de fouiller les données et d’en extraire des connaissances.

4.1. Fouilles sur les données de trafic par STPCA

L’analyse STPCA a démontré ses facultés à réduire fortement la dimension d’un ensemble de données spatio-temporelles de trafic urbain tout en conservant au mieux l’information au sens de l’énergie. Dans [DEV 07], on propose également différentes utilisations des résumés calculés par la méthode afin de dégager des informations sur les données. Par exemple, on peut mesurer pour chaque journée la distance moyenne

entre les séries temporelles de débit ou de taux d'occupation issues des capteurs et leur estimé par STPCA. De cette manière, on obtient une valeur indiquant l'erreur d'estimation de chaque jour à partir de laquelle on peut détecter les journées pour lesquelles le trafic est globalement atypique et en conséquence procéder à une classification sur les jours. En suivant cette idée, les validations réalisées dans [DEV 07] nous ont notamment permis d'identifier automatiquement et sans connaissances a priori les jours "spéciaux", tels que Noël ou le 14 Juillet (avec une erreur d'estimation élevée), ainsi que les jours correspondant aux congés scolaires d'été (avec une erreur d'estimation moyenne).

4.2. Motifs spatio-temporels de situations atypiques

L'analyse STPCA appliquée aux données d'état de circulation E offre un moyen de résumer très fidèlement l'information au sens de l'énergie dans un espace de dimension très faible (plus de 98% de l'énergie capturée par le premier mode spatial et le premier mode temporel). A partir de cette propriété, dans [JOL 08] nous avons proposé une méthode pour identifier des motifs spatio-temporels de propagation de situations atypiques au sens de l'occurrence sur le réseau. La propagation d'une situation atypique est définie comme une notion de voisinage locale en espace et en temps indiquant sur quels points du réseau on observe un comportement atypique suite à la réalisation d'un événement inhabituel à l'instant précédent. La réalisation d'une situation atypique au sens de l'occurrence à un capteur i , le jour n , à l'instant t est détectée en mesurant l'écart entre la valeur réelle d'état de circulation correspondante $e_i^n(t)$ et son estimé par réduction STPCA $\hat{e}_i^n(t)$. Les motifs de propagation sont ensuite utilisés pour faire des prédictions sur le trafic. Ces motifs sont calculés en combinant l'information mutuelle [SHA 63] calculée sur les données symboliques avec l'algorithme de réduction de dimension Isomap [TEN 00].

Les motifs spatio-temporels identifiés permettent de prédire plus de 80% des situations atypiques à court terme (3 minutes) et environ 70% à moyen terme (15 minutes) avec une proportion de faux positifs très acceptable. La distinction entre deux types de situations atypiques (sous-estimation et surestimation du trafic) a augmenté la précision des motifs tout en diminuant la proportion de faux positifs lors des prévisions. Les validations ont également démontrées les capacités d'apprentissage de la méthode calculant les motifs. D'autres expériences ont également prouvé qu'en connaissance du type de situation atypique observée à un capteur, en plus de savoir vers quels endroits du réseau elle se propagerait, il est également possible de prédire, dans une grande partie des cas, le type de chaque situation atypique consécutive à la propagation.

5. Un système d'information pour le trafic routier

Dans [MED 06] nous présentons l'architecture du système d'information pour le trafic routier que nous proposons. Celle-ci est construite autour d'un entrepôt de données dont le schéma en étoile est illustré sur la figure 3(a). Ce modèle ajoute notam-

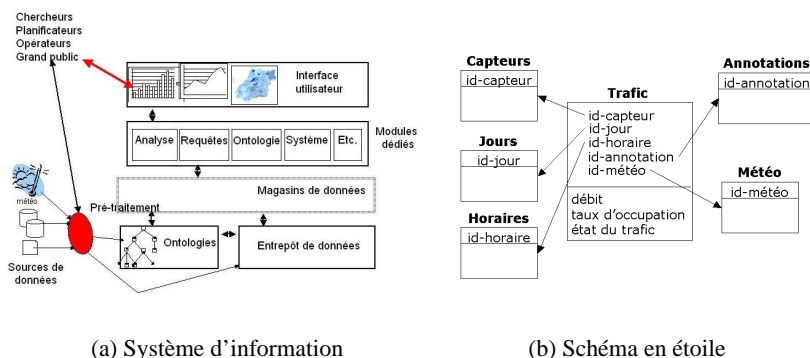


Figure 3. Architecture du système d'information pour le trafic routier et schéma en étoile de l'entrepôt de données proposés.

ment au schéma standard des ontologies au niveau du stockage des données, comme on peut le voir dans la partie inférieure de la figure 3(b). L'ensemble des ontologies permet d'organiser les définitions et la terminologie utilisées par le domaine de l'application, aide à la construction et à la maintenance de l'entrepôt. Il est également utilisé pour la création et la maintenance des magasins de données.

6. Comparaison aux travaux existants

Nos recherches combinent des travaux sur la préparation et l'extraction de motifs et des travaux sur l'entreposage de données de trafic issues d'un réseau de capteurs géoréférencés. Plusieurs techniques de réduction de dimension des séries temporelles ont été proposées pour aider à la recherche de similitudes. La majorité de ces méthodes, comme les *Discrete Fourier Transformations* [AGR 93] ou les *Discrete Wavelet Transformations* [CHU 99] calculent une estimation des séries temporelles en combinant différentes fonctions de même type. La représentation symbolique permet également de représenter les séries par une séquence de segments dont la longueur est prédéfinie (*Piecewise Aggregate Approximation* [KEO 01a]) ou adaptable aux fluctuations de la série (*Adaptive Piecewise Constant Approximation* [KEO 01b]), voire de segments mesurés en fonction des angles formés par les lignes composant la série temporelle (TIDES [MAR 07]). De son côté, l'algorithme STPCA propose d'estimer les séries en combinant les vecteurs propres spatiaux et temporels calculés sur l'intégralité des données.

En ce qui concerne l'entreposage des données spatio-temporelles, le couplage entre Système d'Information Géographique (SIG) et les entrepôts de données a conduit à la définition du SOLAP (*Spatial On-Line Analytic Processing*) ou OLAP spatial [BED 05]. Les outils SOLAP ont ensuite été étendus pour inclure des opérateurs spatio-temporels

topologiques qui permettent de poser des requêtes incluant des opérateurs tels que "en même temps", "au même endroit", "en même temps, au même endroit" [MAR 04]. En outre, l'utilisation d'entrepôts de données spatiales pour le trafic routier a été notamment faite dans [BER 05, LU 05] où un entrepôt est établi à des fins de stockage et de visualisation.

7. Conclusion et perspectives

Ce papier présente les recherches effectuées dans un projet pluridisciplinaire relié au traitement de données de trafic routier urbain issues d'un réseau de capteurs. Il combine la description de méthodes analytiques pour réduire la dimension des données, traiter les valeurs manquantes, proposer des résumés symboliques des différentes variables mesurées et les recherches sur une structure complexe permettant de stocker ces données et de les manipuler de manière efficace. Une partie du travail élaboré au sein du projet CADDY a conduit à la construction d'un prototype logiciel, illustré sur la figure 4 permettant d'explorer visuellement les données des capteurs dans les deux dimensions spatiale et temporelle.

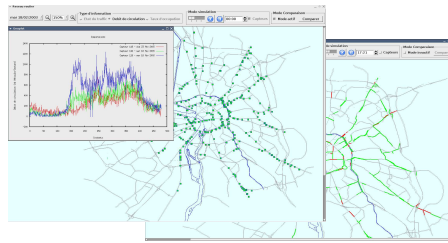


Figure 4. Copie d'écran du prototype.

Les travaux futurs concernent des problèmes théoriques et d'implantation, notamment par le développement d'autres outils tels que le prototype de visualisation. On pourrait également étendre l'algorithme STPCA à des agrégations spatiales afin de pouvoir effectuer des requêtes spatio-temporelles complexes sur les données telles que la recherche de motifs de congestion sur des zones géographiques.

8. Bibliographie

- [AGR 93] AGRAWAL R., FALOUSTOS C., SWAMI A., « Efficient similarity search in sequence databases », *Proc. of the 4th International Conference on Foundations of Data Organization and Algorithms*, 1993, p. 69 - 84.
- [BED 05] BEDARD Y., PROULX M. J., RIVEST S., « Enrichissement du OLAP pour l'analyse géographique : exemples de réalisation et différentes possibilités technologiques », *Entrepôts de Données et l'Analyse en ligne (EDA 2005)*, 2005, p. 1-20.

- [BER 05] BERTINI R., MATTHEWS S., HANSEN S., DELCAMBRE A., RODRIGUEZ A., « ITS Archived Data User Service in Portland, Oregon : Now and Into the Future », *8th Int. IEEE Conf. on Intel. Transport. Sys.*, Vienna (Austria), Sept. 2005, p. 13-16.
- [CAD 07] CADDY, *the CADDY website* - <http://norma.mas.ecp.fr/wikimas/Caddy>, 2007.
- [CHU 99] CHU K., WONG M., « Fast time serie searching with scalling and shifting », *Proc. 18th ACM Symposium on Principles of Databases Systems*, 1999, p. 237 - 248.
- [DEM 77] DEMPSTER A., LAIRD N., RUBIN D., « Maximum likelihood for incomplete data via the EM algorithm », *Journal of the Royal Stat. Society series B*, vol. 39, 1977, p. 1-38.
- [DEV 07] DE VUYST F., JOLIVEAU M., « Space-Time Principal Component Analysis for Multivariate Time Series », *Submitted to Computational Statistics & Data Analysis*, , 2007.
- [HUG 03] HUGUENEY B., « Représentations symboliques de longues séries temporelles », PhD thesis, Université Paris 6, 2003.
- [JOL 86] JOLLIFFE I., *Principal Component Analysis*, Springer-Verlag, New York, 1986.
- [JOL 07] JOLIVEAU M., DE VUYST F., « Space-Time summarization of multisensor time series. Case of missing data. », *Proc. of Int. Workshop on Spatial and Spatio-Temporal Data Mining SSTDM - IEEE ICDM*, 2007, p. 631-636.
- [JOL 08] JOLIVEAU M., DE VUYST F., « Recherche de motifs spatio-temporels de cas atypiques pour le trafic routier urbain », *Extraction et Gestion de Connaissances EGC 08, Revue des Nouvelles Technologies de l'Information - RNTI - EII*, 2008, p. 523-534.
- [KEO 01a] KEOGH E., CHAKRABARTI K., PAZZANI M., MEHROTRA S., « Dimensionality reduction for fast similarity search in large time series databases », *Knowledge and Information Systems*, vol. 3, n° 3, 2001, p. 263-286.
- [KEO 01b] KEOGH E., CHAKRABARTI K., PAZZANI M., S M., « Locally adaptive dimensionality reduction for indexing large time serie databases », *Proc. of the international IEEE Conference on Data Mining (ICDM)*, 2001, p. 151 - 162.
- [LU 05] LU C., SRIPADA L., SHEKHAR S., LIU R., « Transportation Data Visualization and Mining for Emergency Management », *Int. Journal of Critical Infrastructures (Inderscience)*, vol. 1, n° 2/3, 2005, p. 170-194.
- [MAR 04] MARCHAND P., BRISEBOIS A., BÉDARD Y., EDWARDS G., « Implementation and evaluation of a hypercube-based method for spatio-temporal exploration and analysis », *Int. Society for Photogrammetry and Remote Sensing Journal*, vol. 59, n° 1-2, 2004, p. 6-20.
- [MAR 07] MARIOTTE L., MEDEIROS C., TORRES R., « Diagnosing similarity of oscillation trends in time series », *Proc. of Int. Workshop on Spatial and Spatio-Temporal Data Mining SSTDM - IEEE ICDM*, 2007, p. 643 - 648.
- [MED 06] MEDEIROS C., CARLES O., DE VUYST F., HUGUENEY B., JOLIVEAU M., JOMIER G., MANOUVRIER M., NAIJA Y., SCEMAMA G., STEFFAN L., « Vers un entrepôt de données pour le trafic routier urbain », *Proc. of EDA' 06*, 2006, p. 119-137.
- [SCE 04] SCEMAMA G., CARLES O., « CLAIRE-SITI, public and road traffic network management control : a unified approach », *Proc. of the 12th IEEE Road Transport Information and Control Conference (RTIC 04)*, 2004, p. 11-18.
- [SHA 63] SHANNON C., WEAVER W., *The Mathematical Theory of Communication*, University of Illinois Press, 1963.
- [TEN 00] TENENBAUM J., LANGFORD J., « A Global Geometric Framework for Nonlinear Dimensionality Reduction », *Science*, vol. 290, 2000, p. 2319-2323.