

Apport des SIG dans l'analyse de données Climato Urbaine

M. Bendenia¹; A. Saidi²; A. Mebrek³; Z.F.
Makranfar⁴; Y.Z. Oulhaci⁵

^{1,2,3,4,5} Agence Spatiale Algérienne, Centre des Techniques Spatiales.
Arzew, Oran, Algérie

Résumé :

Un système d'information (SI) est un mécanisme qui permet d'acquérir, stocker, traiter et communiquer de l'information. On le retrouve dans différents domaines d'application dont le domaine géographique. Les Systèmes d'Informations Géographiques (SIG) en plus des fonctionnalités des SI, permettent également de manipuler tous les types de données spatiales et géographiques. On trouve les SIG dans de nombreux domaines tels que la santé, l'agriculture, le transport, et l'urbanisme. Dans le cadre de cet article, nous allons mettre en relief l'apport des nouvelles technologies d'analyse et de fouille de données géographiques dans la recherche et l'extraction d'informations implicites, de corrélations spatiales ou d'autres propriétés cachées sous une masse de données importante. L'objectif est d'appliquer l'opérateur "Pivot Spatial", sur un ensemble de données Climato Urbaine [1], afin de déceler des corrélations susceptibles d'être statistiquement et objectivement significatives. Cette solution vient en appui à des solutions déjà établies, en complémentarité avec les résultats du projet "Mise en œuvre d'outils d'analyse spatiale multidimensionnelle issus des outils de l'informatique géo-décisionnelle: Application aux problématiques environnementales (ASM)" finalisé au sein du service SPD.

Mots clés : SIG; Système d'information; Fouille de données; SOLAP; Climatologie; Urbanisme.

I. Introduction

Dans une époque où le volume de données est en croissance exponentielle, l'exploration et la manipulation de ces derniers sont des tâches problématiques et fastidieuses. Afin de pallier ces problèmes, les SI proposent des fonctionnalités et solutions telles que l'entreposage de données (ED), la fouille de données (DM), et les outils de traitement analytique en ligne (OLAP). Sachant que la présence de l'information géographique accroît les difficultés de l'exploitation des données, les solutions et fonctionnalités SI citées précédemment s'avèrent caduc et obsolète. Par conséquent, nous devons utiliser les fonctionnalités SIG qui permettent de gérer et manipuler la donnée géographique. De cela, des solutions variées des ED, DM et

OLAP ont émergé, nommé, Entrepôt de données spatial (EDS), Fouille de données spatial (SDM), et les outils de traitement analytique spatial en ligne (SOLAP). De l'ensemble des opérateurs SOLAP, on a choisi au cours de cet article, de nous concentrer sur le "Pivot Spatial", afin de montrer que la puissance de cette solution nous permet de distinguer des corrélations susceptibles de démontrer des relations causales. Le présent document comprend les sections suivantes : la section 1 est bien l'introduction ; la section 2 met en évidence la problématique ; la section 3 décrit la méthodologie ; la section 4 présente les résultats ; la section 5 discute les résultats ; et pour conclure, la section 6.

II. Problématique

À l'heure actuelle, il est presque impossible de fournir un environnement décisionnel parfait, sans y intégrer des données géographiques. Ces derniers sont essentiels et représentant un volume important. SOLAP permet de manipuler les données stockées dans le SDW via des opérations classiques (Slice, Dice, Drill-Down, Roll-up, etc.), agissant sur la granularité (Roll-Up et Drill-Down), et des opérations sur la structure (Pivot) [2]. L'outil SOLAP fait partie de l'approche SDM. Cette technique permet aux utilisateurs d'extraire des connaissances intéressantes et non évidentes à partir des données stockées. L'outil « Spatial Pivot » nous permet de visualiser les six (6) faces d'un Géocube. Pour mettre en place ce dernier, la difficulté de la cartographie, qui change toujours selon le sujet, doit être surmontée. Le problème ne se limite pas à cela, il est primordial que toutes les cartes aient un sens et une information à mettre en évidence. Alors, que se passe-t-il lorsque l'utilisateur fait une rotation sur un axe ? Quelles informations doivent être mises en évidence ? Quelle symbologie doit être utilisée ? Quelle nuance de couleur est la plus adéquate ? Surtout, comment gérer différentes thématiques, alors que chacune nécessite une cartographie spéciale ? En outre, pour avoir la possibilité de réaliser une telle opération, une structure adéquate est nécessaire. Les données contenues dans la base de données spatiale sont organisées et manipulées en tant que dimensions d'un SDW, afin de permettre aux utilisateurs de construire facilement un Géocube tridimensionnel à partir de ces dimensions, et de permettre à d'autres processus SDM d'extraire également des connaissances [3]. Il s'agit d'une solution « SIG-Dominant », où les fonctionnalités OLAP sont intégrées dans une solution SIG. En combinant la puissance des deux, une puissante structure décisionnelle émerge. L'outil « Spatial Pivot » en fait partie. Les processus SDM agissent avant la visualisation, où à partir d'une masse de données brutes et floues, différentes informations sont extraites. Il s'agit des classes associées à leurs chronologies et géométries que l'outil affichera. Tout au long de cet article, vous découvrirez comment traiter ces problèmes afin d'intégrer l'outil dans QGIS.

III. Méthodologie

Le SDW est la structure sur laquelle les outils développés sont greffés. Principalement, c'est une GDB avec des entités interprétées comme des dimensions. Pour assurer le bon fonctionnement des outils, chaque dimension ne doit contenir que des données stables, pertinentes et cohérentes « Fig. 1 » [4].

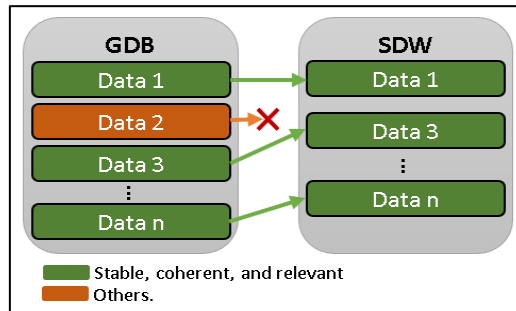


Figure 1 : Construire un SDW à partir d'une GDB.

Le Géocube est construit à partir de trois dimensions choisies à partir du SDW « Fig. 2 ». Une seule d'entre elles devrait être géographique pour garantir le bon fonctionnement de l'outil. Afin d'impliquer d'autres dimensions après le chargement des trois premières, l'utilisateur peut remplacer autant de dimensions que nécessaire, pourvu que nous obtenions trois dimensions à la fin.

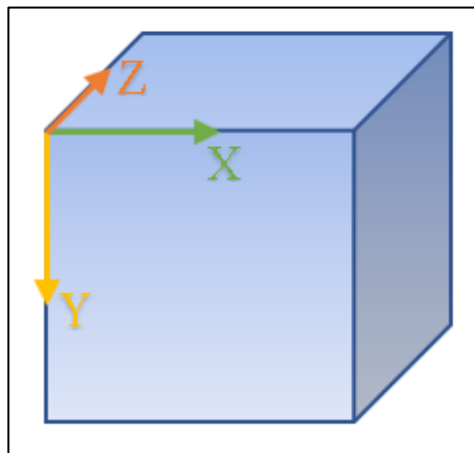


Figure 2 : Un Géocube à 3 dimensions.

Pour implémenter l'outil « Spatial Pivot » avec ce Géocube, nous avons utilisé « QGIS 3.16.11- Hannover », avec un environnement « Python 3.9 ». L'exécution dépend de plusieurs paramètres qui doivent être initialisés par l'utilisateur. En particulier, les paramètres d'environnement tels que la géodatabase et la fenêtre où afficher la carte. Lorsque la vue principale est affichée, l'utilisateur doit attribuer une dimension à chaque axe (x, y et z) via des listes déroulantes. Notez que la dimension spatiale doit être présente, de sorte que le Géocube prenne une position initiale comme le montre l'exemple suivant « Fig. 3 ».

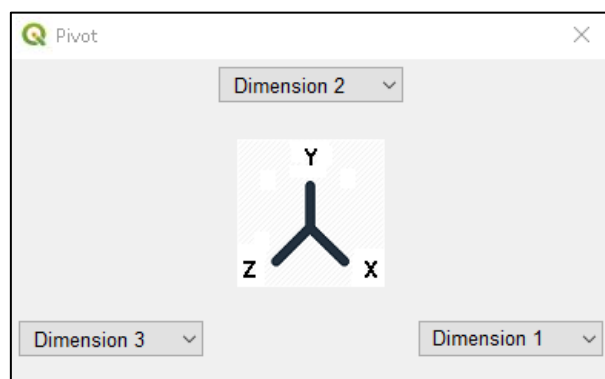


Figure 3 : Initialisation du Géocube.

L'outil doit satisfaire l'exigence fonctionnelle suivante qui sera exécutée par le système :
Pivoter sur un axe : Pour visualiser le Géocube dans les différentes positions possibles, l'utilisateur doit cliquer sur l'un des axes pour le faire pivoter.

Algorithme. Pivoter sur un
axe

Entrée: Nouvelle position du Géocube

Sortie: Rendu visuel de la nouvelle position

- 1: **Si** (les trois dimensions sont supportées) **Alors**
 - 2: Effacer la symbologie précédente
 - 3: Appliquer une nouvelle symbologie
 - 4: **Si** (la position supporte l'affichage d'un graph) **Alors**
 - 5: Afficher un graph
 - 6: **Fin Si**
 - 7: **Si** (la position peut supporter un curseur temporel) **Alors**
 - 8: Configurer un curseur temporel
 - 9: **Fin Si**
 - 10: **Sinon**
 - 11: Afficher un message d'erreur
 - 12: **Fin Si**
-

Cet algorithme est compatible avec n'importe quelle thématique. Cependant, l'application des symbologies est spécifique à la thématique « Climato Urbaine ». Pour appliquer de telles opérations avec divers thèmes, l'intervention de thématiciens est nécessaire pour gérer les paramètres de symbologie.

IV. Résultat

Dans cette section, nous exposerons les résultats du « Pivot Spatial » sur la cartographie et l'affichage graphique.

Appliqué sur un Géocube tridimensionnel construit à partir des données de Londres (2006-2012). Tout d'abord, nous initialisons les trois premières dimensions, pour construire le Géocube « Fig. 4 ». Dans cette position, nous avons le temps, les données météorologiques et la géométrie.

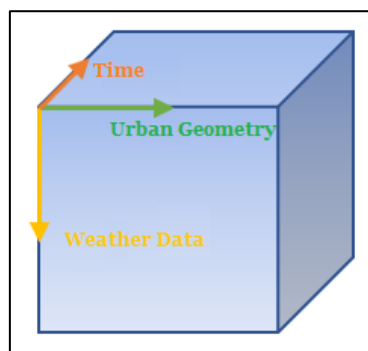


Figure 4 : La position initiale.

Nous avons choisi pour cette position de départ de mettre la géométrie (urbain) dans l'axe des X, les données météorologiques (Température, Point de rosée et Visibilité) dans l'axe des Y et le temps (2006/2012) dans l'axe des Z. La figure 5 ci-dessous illustre le résultat.

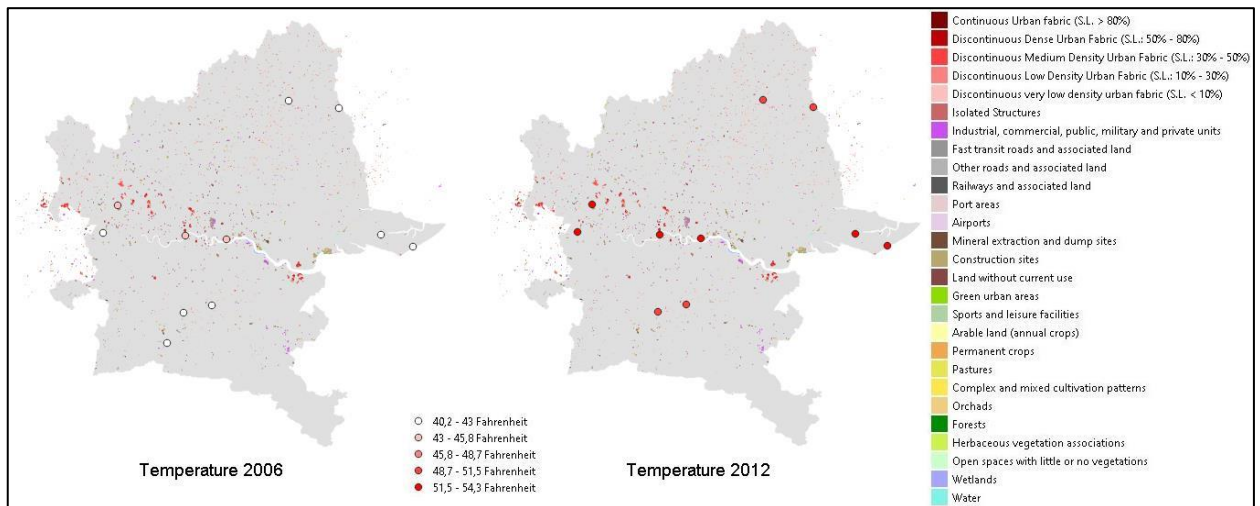


Figure 5 : Variation de la moyenne des températures de 2006 à 2012.

On constate à partir de cette cartographie que la moyenne des températures est fortement corrélée avec la densité urbaine. Chaque résultat est propre à un couple thématique/position. Dans les figures suivantes, nous allons voir des cartographies qui diffèrent au niveau de l'axe contenant les données météorologiques.

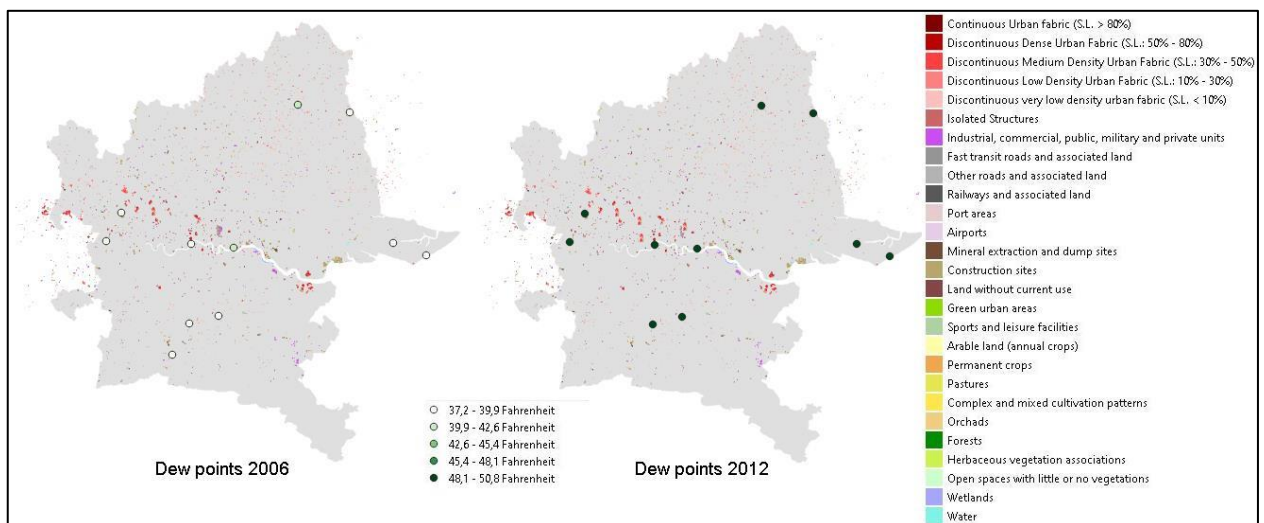


Figure 6 : Variation de la moyenne des points de rosée de 2006 à 2012.

Comme dans tous les processus de fouille de données, certains résultats peuvent ne mener à aucune conclusion pertinente. Comme dans la figure précédente « Fig. 6 », on ne peut établir un lien entre les différentes données.

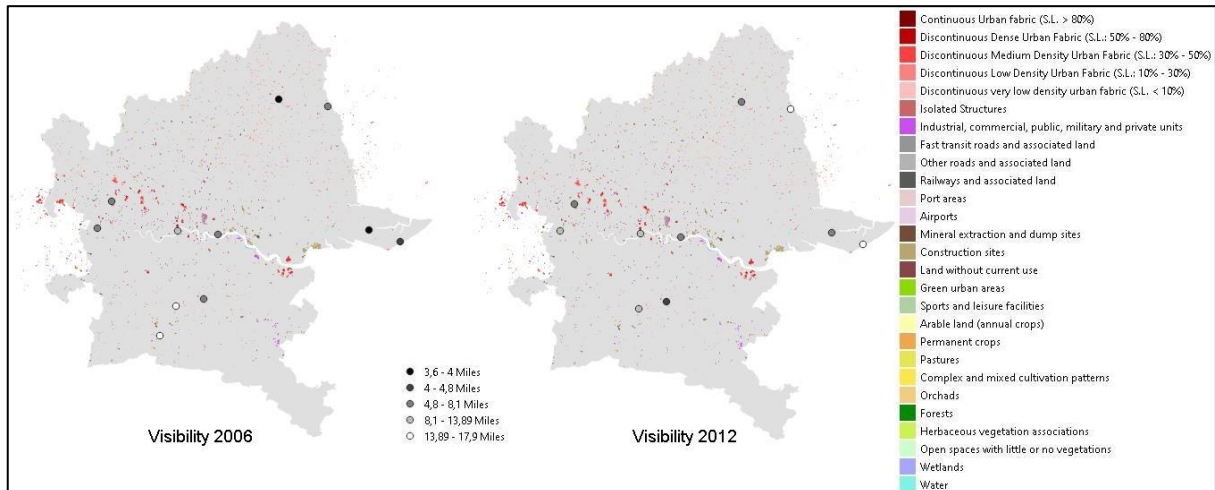


Figure 7 :Variation de la moyenne de la visibilité de 2006 à 2012.

Contrairement à la température, la visibilité est la densité urbaine sont inversement corrélés. Comme l'explique l'algorithme précédemment vu (page 3), pivoter sur un axe peut générer un graphe si la position le permet. Ci-dessous « Fig. 8 » un exemple ou les catégories urbaines sont dans l'axe des X et le temps dans l'axe des Y.

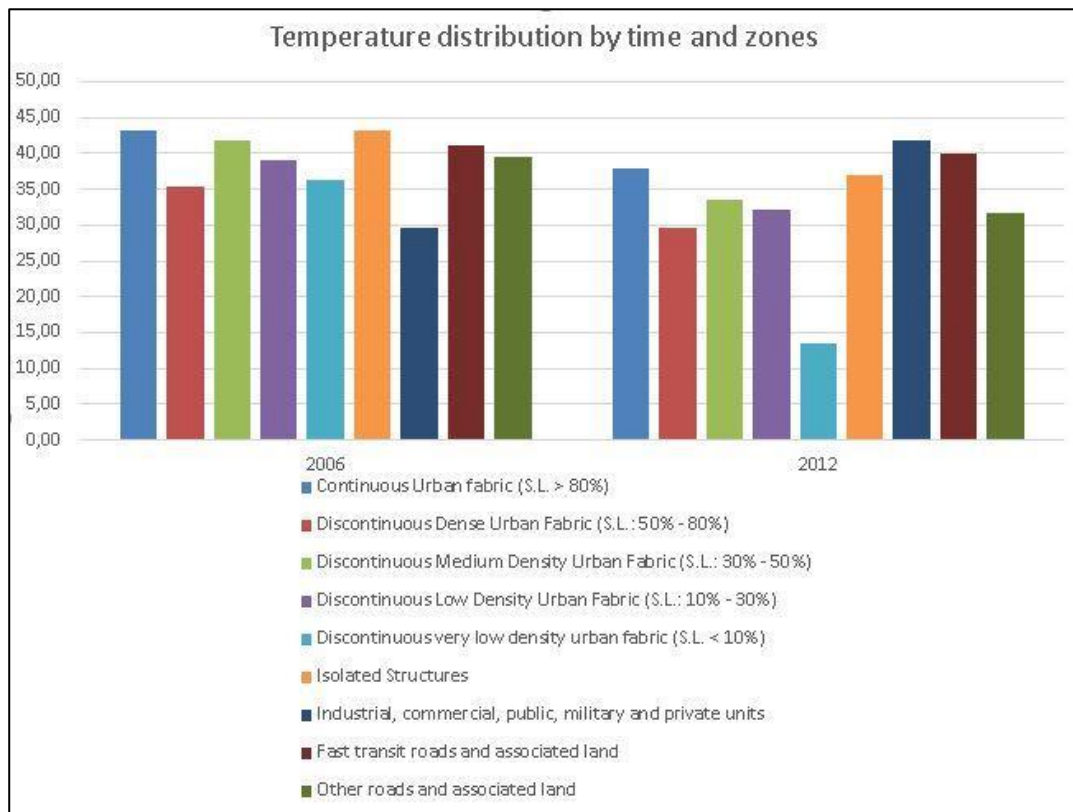


Figure 8 : Graph montrant la distribution de la température par temps et zones

V. Discussion

L'outil PIVOT assimile les concepts SOLAP et SDM. De tous les résultats précédents, on peut déduire que l'utilisation d'un Géocube tridimensionnel est plus susceptible d'obtenir des informations décentes à partir d'un volume important de données. L'objectif de creuser dans

une grande masse de données est souvent incertain, il est nécessaire de commencer le processus d'extraction avant de voir quelles données sont susceptibles d'apporter des connaissances. Après de nombreuses tentatives, on a choisi les positions précédentes pour mettre en évidence l'impact urbain sur le climat. Certains chiffres fournissent plus d'informations que d'autres, présentent des informations plus importantes que d'autres et ont un rendu plus significatif que d'autres. Une telle solution est spécifique à une thématique unique, pour accroître son efficacité, l'intervention d'un thématicien est indispensable. Ce dernier peut gérer quelles données sont appropriées pour quel affichage, et comment l'opération de cartographie devrait procéder. Actuellement, l'outil développé ne prend en charge que ces trois dimensions, pour parvenir à une solution universelle, les fonctionnalités de l'outil doivent être améliorées. Selon une thématique aléatoire, l'outil devrait pouvoir détecter l'information intrinsèque de la dimension spatiale, pour afficher à chaque fois une carte significative. Notez que l'outil ne prend en charge que 3 dimensions à la fois. D'autres peuvent être intégrées en remplaçant celles déjà existantes. Notre objectif de recherche est de mettre en place une extension QGIS qui inclut le Spatial Pivot et toutes les autres fonctionnalités développées par notre équipe. En particulier, la création, l'alimentation et la mise à jour du SDW, ainsi que le Drill-Down et le Roll-up. La boîte à outils devrait pouvoir supporter certaines thématiques environnementales.

VI. Conclusion

La force de cette solution est sa capacité à mettre en évidence des corrélations non évidentes entre des objets. De tels résultats exigent la mise en œuvre d'un système de SDM crédible. Ce dernier implique des mathématiques, des statistiques, de la classification, de l'interpolation et de la segmentation. En raison de l'indisponibilité d'un moteur OLAP à notre niveau, les fonctionnalités sont assimilées aux actions SIG. Cela induit certaines limites et contraintes. Malgré cela, les résultats que nous fournit cette solution demeurent très satisfaisants. Cela ouvre une nouvelle perspective de recherche pour notre groupe, vers la mise en œuvre d'une solution SOLAP SIG dominant universelle. Pour les travaux ultérieurs, nous cherchons à trouver un moyen de dériver un modèle géométrique abstrait d'objets spatiaux, pour atteindre un outil qui ne dépend pas de la thématique. Nous restons convaincus de la puissance des SDW, SOLAP et SDM en tant qu'approches innovantes pour l'analyse spatiale multidimensionnelle.

VII. Reference

- [1] The NCDC Climate Services Branch (CSB) « ncdc.orders@noaa.gov »
- [2] B. Yvan, R. Sonia, P. Marie-Josée, N. Martin, H. Frederic et P. Julien, «SOLAP technology: Merging business intelligence with geospatial technology for interactive spatio-temporal exploration and analysis of data» ISPRS, 2005.
- [3] H. Jaiwei, K. Krzysztof et S. Nebojsa, «GeoMiner: A System Prototype for Spatial Data Mining,» SIGMOD Rec., 1997.
- [4] Inmon, W. H. «Building the Data Warehouse, 4rd Edition». Wiley Publishing, 2005.