

ETUDE DE LA RESOLUTION DANS UNE BASE DE DONNEES GEOGRAPHIQUE

Par Oumrane. Nasser

SGT/ Service géographique et de télédétection de l'ANP

ملخص: تعتبر قواعد المعطيات الجغرافية اليوم من الأدوات الأساسية و اللازمة لمعالجة المعطيات المتعلقة بالوضعية. التقدم الذي يشهده العالم في ميدان الإعلام الآلي سهل بصورة واضحة إتقاط المعطيات، حفظها و معالجتها.

لكن من جهة أخرى، هذا التغيير أحدث مشاكل أخرى لا سيما تلك المتعلقة بنوعية المعطيات المستعملة. هذه الدراسة، المنجزة في إطار تحضير دبلوم الدراسات المعمقة في العلوم الجغرافية بالمدرسة الوطنية للعلوم الجغرافية - فرنسا، موجهة أساسا لدراسة التمييز الهندسي في قاعدة للمعطيات الجغرافية و كيفية تحسينه. التمييز الرقمي، و الذي له علاقة بدقة القياسات هو في غالب الأحيان كافي. الهدف من هذا العمل هو ضمان توازن بين هذين النوعين من التمييز، الهندسي و الرقمي، الشيء الذي يتضمن تكامل في القاعدة.

Résumé:

Les bases de données géographiques sont devenues de nos jours des outils indispensables pour la gestion des données localisées. La mise en forme numérique des données géographiques a facilité pleinement l'acquisition des données, leur stockage et leur analyse. Mais d'autre part, cet aspect numérique a soulevé d'autres problèmes liés notamment à la qualité des données que manipulent ces bases. Cette étude, réalisée dans le cadre de préparation d'un DEA en sciences de l'information géographique à l'ENSG-France, est orientée dans un but de maîtriser et améliorer la résolution géométrique dans une base de données géographique. La résolution numérique dans une base, liée essentiellement à la précision des mesures, est jugée généralement suffisante. L'objectif de ce travail est d'assurer la correspondance entre ces deux types de résolution, à savoir, la résolution géographique et la résolution numérique. Ainsi, ceci permet d'assurer une cohérence dans la base et évite la manipulation de données lourdes.

Mots clés: Qualité des données, cohérence logique, qualité géométrique, arrondi, résolution numérique, résolution géographique.

1- Introduction:

Les systèmes d'information géographique connaissent un développement considérable du fait de leur importance comme outil d'aide à la décision. Afin de répondre à ce besoin, il est important de maîtriser les données que manipulent ces systèmes. Ces données utilisées dans le processus d'analyse spatiale doivent être correctes. Cette nécessité exige du producteur des données une parfaite maîtrise de ses produits afin qu'ils puissent répondre au mieux aux besoins. Les modèles de données utilisés dans les systèmes d'information géographique existants représentent souvent de manière grossière les propriétés géométriques et topologiques des objets géographiques (ex. erreurs dues à la digitalisation). Ce constat a pour conséquence de rendre plus difficiles et moins fiables les raisonnements spatiaux (erreurs dues aux changements de coordonnées, généralisation, arrondis informatiques).

La saisie de l'information géographique se fait à une échelle donnée et la géométrie saisie a une résolution non nulle. On entend par résolution la plus petite valeur chargée de sens dans l'évaluation d'une variable.

Dans une base de données géographique, la notion de résolution est très importante. En effet, sans maîtrise de la résolution, on ne peut pas gérer la topologie par la géométrie, ni faire de l'analyse spatiale. De plus, pour passer les données en mode maillé, la résolution est indispensable pour éviter d'obtenir une topologie faussée par un excès de confiance en la résolution des données.

C'est dans cette optique que s'inscrit cette étude, pour maîtriser la notion de résolution dans une base de données géographique, et assurer la mise en accord entre la résolution géographique et la précision géométrique. On pourra avoir une gestion optimale de la résolution si on la définit comme l'unité entière dans laquelle on exprime les coordonnées. D'autre part, cette définition nous permet d'éviter de manipuler des données volumineuses du point de vue traitement informatique. La résolution doit donc avoir le même ordre de grandeur que la précision.

2- Définition de la qualité:

Les bases de données géographiques sont devenues aujourd'hui un outil indissociable des systèmes d'information qui les englobent et exploitent les données qu'elles contiennent. Ces systèmes d'information sont au cœur d'applications variées tant au niveau de la finalité (systèmes d'aide à la décision) que des organismes consommateurs de données (collectivités locales). Du fait de l'importance des données géographiques, leur qualité est essentielle. Cette qualité doit assurer la fiabilité des processus s'appuyant sur elles (par la connaissance de la qualité traduite dans les données), ainsi que la capacité du système à remplir les fonctions attendues exprimées par les spécifications. Ces deux notions complémentaires se trouvent dans la définition de la qualité donnée par l'Organisation Internationale de Standardisation [ISO 8402-94]: "Ensemble des propriétés et caractéristiques d'un produit ou d'un service qui lui confère l'aptitude à satisfaire des besoins exprimés ou implicites".

2.1- Composantes de qualité:

L'actualité: représente le décalage entre le jeu de données à une date T1 et le terrain nominal à une date de référence T2. Pour le producteur, l'actualité des données implique une politique de mise à jour décrivant la validité du jeu de données avant la prochaine modification. Pour l'utilisateur, l'actualité des données doit lui permettre de vérifier la validité de ses données.

La généalogie (lineage): cette composante doit contenir une description des procédés d'acquisition ainsi que toutes les informations ayant conduit au résultat final. La généalogie assure que les standards de production sont maintenus pour le producteur.

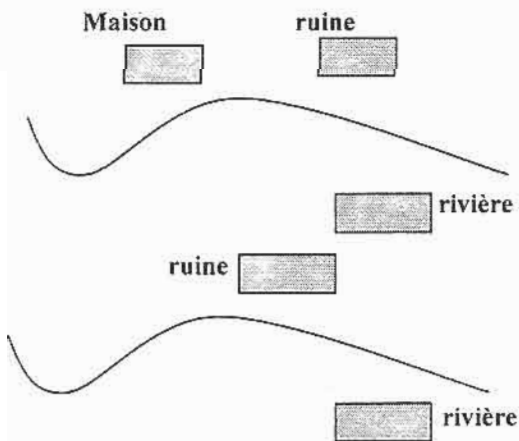
La cohérence logique (logical consistency): rend compte du nombre de caractéristiques, relations ou attributs qui ont été correctement encodés, en accord avec les contraintes d'intégrité données dans les spécifications.

La précision géométrique (positional accuracy): Elle décrit l'écart de géométrie entre l'objet dans le terrain nominal et son homologue dans le jeu de données. On parle de précision de position ponctuelle, linéaire et surfacique, qui nous renseigne sur l'écart entre la position planimétrique ou altimétrique des objets dans le terrain nominal et ceux dans le jeu de données. Quant à la précision de forme, elle donne l'écart de forme géométrique.

La précision sémantique et exhaustivité: La précision sémantique décrit la différence entre les valeurs descriptives des éléments du jeu de données et les valeurs de leur homologues dans le terrain nominal. Elle porte essentiellement sur la classification des objets, la codification des attributs et les relations entre les objets.

L'exhaustivité quant à elle, décrit la conformité de la présence ou l'absence des éléments du jeu de données par rapport au terrain nominal.

Ces composantes décrites ci-dessus se recouvrent partiellement rendant quelque fois difficile le classement d'une erreur. La figure illustre un cas:



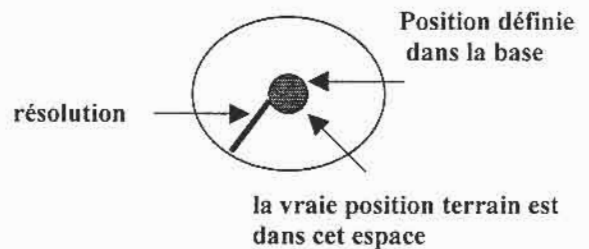
En comparant ces deux figures, on constate:

- une erreur de précision géométrique (la ruine se trouve à l'ouest) plus une erreur d'exhaustivité (la maison manque),
 - une erreur de classification, donc de précision sémantique (la maison a été classée en ruine) plus erreur d'exhaustivité (la ruine manque) plus erreur de précision géométrique (maison décalée à l'est).
- Mais la définition de la qualité reste même avec cette décomposition complexe, ce qui a poussé à subdiviser ces composantes en indicateurs permettant de les estimer. Nous citerons pour la précision géométrique par exemple: EMQ (erreur moyenne quadratique).

3- La résolution dans une base de données géographique:

La résolution est la plus petite différence distinguée entre deux valeurs mesurées. La résolution numérique est considérée comme la plus petite distance qu'on peut mesurer dans une base de données géographique. Dans une carte, on peut définir cette notion par hypothèse, par la largeur minimum d'un trait. Pour les données vecteur, la résolution est la taille de la plus petite entité géographique qu'on peut distinguer à une échelle donnée. Pour les données raster, la résolution réfère à la taille d'un pixel de la grille.

Une résolution d'un mètre par exemple, signifie que la vraie position terrain d'un point se trouve à un mètre de part et d'autre de la position sauvegardée.



Dans la littérature, deux types de résolution interviennent:

- numérique: qui caractérise la précision avec laquelle les mesures ont été réalisées
- spatiale: qui caractérise la plus petite grandeur qui a un sens géographique.

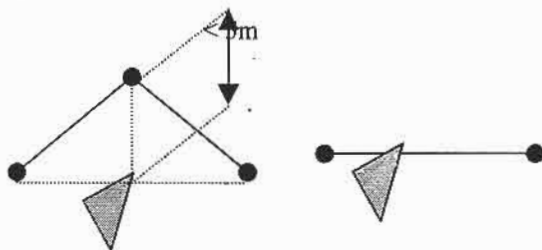
En général, la résolution numérique est plus importante que la résolution spatiale. La tendance actuelle est à l'amélioration de cette dernière.

En plus du nombre de chiffres utilisés pour enregistrer des données, la résolution dépend d'autres facteurs tels que: le précision des équipements de saisie et l'échelle des documents utilisés pour la saisie.

La résolution dans une base de données géographique peut être définie par le respect des spécifications. Ex pour les primitives linéaires:

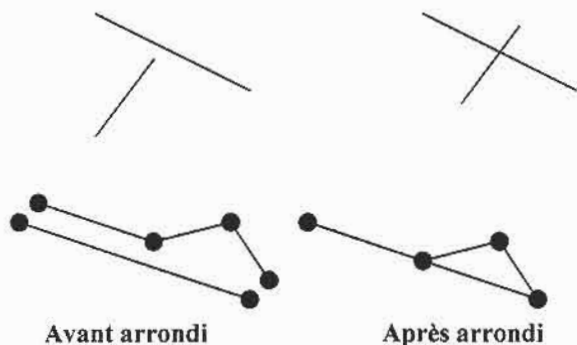


Dans certains cas, le respect des spécifications (de la résolution donc) peut engendrer des conflits et interactions entre les objets géographiques. C'est le cas de la figure suivante:



Dans cet exemple, le respect d'un critère a engendré la violation d'une autre contrainte, interdisant le passage d'une route dans une construction par exemple.

Selon les types d'objets manipulés et les méthodes d'acquisition des données, la résolution diffère d'une couche à l'autre. Cette marge de différence en résolution entre les différentes couches empêche de garder une cohérence entre les couches. Ce qui a poussé à penser à définir un même seuil de résolution pour toutes les couches. Une des solutions à ce cas de figure, consiste à arrondir les coordonnées en tenant compte d'un certain nombre de problèmes d'ordre topologique qui peuvent se poser.



Une série de tests a été effectuée sur un jeu de données en arrondissant les coordonnées et modéliser les types de problèmes qui peuvent surgir après l'arrondi. Ces tests ont montré qu'il y a des intersections parasites qui sont générées. D'autres objets peuvent disparaître. Néanmoins, la liste de ces problèmes n'est pas exhaustive.

Ainsi, à l'heure actuelle, les différents résultats présentés sont une contribution à la recherche en la matière qui peuvent être complétés dans l'avenir.

Conclusion:

L'avènement des Bases de Données Géographiques et des SIG a mis en évidence l'importance de la qualité des données que manipulent ces systèmes, et l'influence de la qualité sur les différents traitements faisant intervenir ces données géographiques. Chose qui nécessite du producteur des données une parfaite maîtrise de ses produits.

L'objectif de ce travail est l'étude de la résolution de la BD-Carto®, et la façon de l'améliorer en assurant une correspondance entre sa résolution géographique et sa précision de mesure.

Cette étude nous a permis d'étudier tous les problèmes liés au respect des spécifications de la base relatives à la notion de résolution, et mettre en évidence les problèmes qui peuvent se poser après un changement de la résolution actuelle en arrondissant les coordonnées. Les différents tests réalisés, nous ont permis d'étudier concrètement cette solution, mais aussi, la capacité de définir les différents problèmes, après un arrondi des coordonnées et la manière de les détecter.

Ainsi, à l'heure actuelle, les différents résultats présentés sont une contribution à la recherche en la matière qui peuvent être complétés dans l'avenir.

Bibliographie:

[Vauglin-97]: "Modèles statistiques des imprécisions géométriques des objets géographiques linéaires" Thèse de doctorat de l'université de Marne-La-Vallée. 286p.

[Omrane-98]: "Etude de la résolution de la BD-CARTO®; DEA-Sciences de l'Information Géographique, ENSG-France, 52p.