

## **Modélisation des erreurs de position et d'attributs dans les bases de données géographiques.**

*Olivier BONIN*

Institut Géographique National, Laboratoire COGIT,

2-4 avenue Pasteur, F-94165 Saint-Mandé CEDEX, phone : 01.43.98.84.09, fax : 01.43.98.81.71, olivier.bonin@ign.fr

Les bases de données géographiques présentent des particularités qui les différencient nettement des bases de données classiques. Même si on peut considérer les bases de données géographiques comme des bases de données comportant un attribut géométrique pour le stockage de l'information spatiale proprement dite, elles contiennent une grande quantité d'information implicite qui doit être explicitée lors des analyses, comme les relations topologiques et les relations métriques.

L'attribut géométrique permet de stocker les trois primitives principalement utilisées : le point, la ligne polygonale et le polygone. Ces primitives sont définies par deux ou trois coordonnées pour un point, et une liste de coordonnées pour une ligne polygonale ou pour un polygone. L'information implicite la plus naturelle à dériver de la représentation géométrique des objets géographiques est la structure de graphe, dans laquelle les points deviennent des nœuds, les lignes polygonales des arêtes, et les cycles formés par les arêtes des faces. On parle alors de topologie de graphe. Cette topologie de graphe est complétée par des relations topologiques entre les objets (« touche », « intersecte », etc.), et par des relations métriques (« est proche de », « longe », etc.).

Comme n'importe quelle source de données, une base de données géographique peut être entachée d'erreurs. L'évaluation de ces erreurs est nécessaire pour de nombreuses applications. Nous montrons que lors de la conception de modèles d'erreurs adaptés à l'information contenue dans les bases de données géographiques il est souhaitable de prendre en compte une partie de l'information implicite. Ces modèles d'erreurs portent sur des objets différents de ceux couramment manipulés en statistique spatiale (champ, ou processus ponctuel par exemple), et nécessitent d'utiliser d'autres techniques.

Nous proposons des modèles d'erreurs adaptés aux erreurs de position des objets géographiques, ainsi qu'aux erreurs d'attributs décrivant les caractéristiques de ces objets. Ces modèles font intervenir des mélanges, et des processus. Nous discutons l'opportunité d'utiliser des modèles à variables cachées, et des techniques non paramétriques.