

Méthodes robustes pour le croisement de données portant sur des partitions spatiales faiblement compatibles

Thèse en Sciences de l'Information Géographique proposée par l'Université Paris Est - IFSTTAR (Institut français des sciences et technologies des transports, de l'aménagement et des réseaux).

Contexte de la thèse

Le projet de réseau GEBD (Grand Équipement de Bases de Données) porté par l'IFSTTAR pour la DRI (MEEDTL) propose de mettre en réseau des laboratoires travaillant sur la ville, la mobilité et l'environnement, pour échanger et disséminer des informations pratiques sur les bases de données utilisées, ainsi que des développements méthodologiques associés à l'utilisation de ces données. Cette création d'un réseau d'échanges s'accompagne d'un travail méthodologique de socle sur l'exploitation de référentiels spatiaux.

Problématique

Dans le contexte du développement durable, il est donc de plus en plus fréquent de chercher à regrouper et croiser des données portant sur un territoire et relatives aux dimensions sociales, économiques et environnementales de ce territoire. La clé spatiale est naturellement une clé privilégiée de croisement de données : des données sont en relation si elles décrivent des phénomènes qui ont lieu au même endroit, éventuellement à des moments différents.

Cependant, cette vision purement spatiale se heurte au fait que de nombreuses données ne sont disponibles qu'à des niveaux agrégés, en faisant référence à des partitions irrégulières, parfois évolutives dans le temps (IRIS, communes, départements, bassin versants, etc.) ou régulières (carroyage) du territoire. La problématique du croisement spatial de données soulève le problème du croisement de partitions variées du territoire, avec des niveaux de granularité (des échelles géographiques) très différentes, et des imprécisions sur les tracés géométriques de ces frontières. Il apparaît donc utile de développer des méthodes robustes de croisement de données portant sur des partitions spatiales plus ou moins compatibles entre elles, en contribuant ainsi aux domaines de la géomatique et de l'économétrie spatiale.

Objectifs

Cette thèse propose d'effectuer un travail de ressourcement scientifique sur l'utilisation des référentiels spatiaux (partitions spatiales et réseaux), en étudiant deux problèmes connexes :

- la description formelle et la manipulation de référentiels spatiaux faiblement compatibles entre eux (carroyage régulier, zonage communal et zonage cadastral par exemple), et de référentiels qui évoluent, plus ou moins modérément, dans le temps (tel que le découpage administratif de la France) ;
- la mise au point de méthodes robustes de changement de référentiel spatial, et de croisement de données rattachées à des référentiels différents, c'est-à-dire des méthodes les moins sensibles possibles aux effets de précision des frontières des partitions, d'écart de granularité entre les partitions, et d'éventuelles erreurs dans les données.

Pour atteindre ces objectifs, les méthodes proposées pourront s'appuyer entre autres sur les approches suivantes :

- proposition de modèles mathématiques et informatiques adaptés à la description des évolutions de partitions spatiales et de réseaux (tels que ceux qui existent en géologie numérique ou encore en traitement d'images) ;
- définition de mailles « synthétiques » définies comme les mailles les plus proches possibles (au sens d'une métrique à définir) d'un ensemble de maillages différents ou des différentes versions temporelles d'un même maillage ;
- caractérisation des partitions spatiales (distribution des tailles des cellules, propriétés géométriques des tracés des frontières des cellules, arrangement spatial des cellules) ;
- sur-échantillonnage ou sous-échantillonnage des informations portées par les partitions pour régulariser l'information à croiser
- méthodes de statistique robuste étendues au cas des données spatiales

Les méthodes formelles développées seront confrontées à des applications réelles sur des territoires, comme la mesure de la dépendance à l'automobile ou la diffusion spatiale de comportements sociaux par croisement de données d'enquêtes.

Cadre de travail

La thèse sera dirigée par Jean-Paul Hubert (IFSTTAR-DEST) et encadrée par Olivier Bonin (IFSTTAR-LVMT). Son financement est assuré par la DRI du MEETL et par le Carnot Vitres. Elle se déroulera dans le cadre de l'école doctorale VTT du PRES Université Paris Est, dans le pôle « Ville, Ingénierie, Environnement ».

Contact : olivier.bonin@enpc.fr