



HAL
open science

Détection d'incohérences de niveau de détail dans des données collaboratives

Guillaume Touya, Carmen Brando

► **To cite this version:**

Guillaume Touya, Carmen Brando. Détection d'incohérences de niveau de détail dans des données collaboratives. *Cartes & géomatique*, 2013, 217, pp.59-72. hal-02288752

HAL Id: hal-02288752

<https://hal.science/hal-02288752v1>

Submitted on 16 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Détection d'incohérences de niveau de détail dans des données collaboratives

par Guillaume Touya¹ et Carmen Brando^{1,2}

¹ Université Paris-Est, IGN, COGIT

73 avenue de Paris

94160 Saint-Mandé

² CRHM - Université Paris 1 Panthéon-Sorbonne

17 rue de la Sorbonne

75005 Paris

guillaume.touya@ign.fr

carmen.brand@gmail.com

Alors qu'il est possible de définir le niveau de détail de bases de données géographiques classiques, cela n'est pas le cas pour les données collaboratives, qui contiennent fréquemment des niveaux de détail hétérogènes. Cette hétérogénéité est un véritable frein pour la réalisation de cartes, notamment quand on utilise les techniques traditionnelles, comme la généralisation, pour produire une carte à partir d'une base de données. Cet article propose une méthode pour estimer le niveau de détail de données issues de sources collaboratives. Ensuite, les incohérences entre objets de la carte à différents niveaux de détail, qui empêchent la réalisation de cartes lisibles, sont détectées automatiquement. Des propositions sont faites pour, ensuite, harmoniser le niveau de détail des objets de la carte. La méthode a été implémentée et des résultats sont présentés sur des données OpenStreetMap.

1 Introduction

Les données géographiques collaboratives sont de plus en plus utilisées dans la communauté de recherche en géomatique mais aussi par l'ensemble des utilisateurs de données géographiques. En effet, à l'image de la communauté OpenStreetMap, la qualité de ces données ne cesse de s'améliorer (Girres and Touya 2010, Haklay 2010), encourageant leur utilisation dans les entreprises et administrations. Même les agences de cartographie nationales qui produisent des données faisant autorité et de grande qualité, commencent à intégrer des informations collaboratives dans leurs données (par exemple, les Etats-Unis et le Canada). Au-delà de la simple qualité des données, cette intégration pose des questions de validation pour une réelle intégration dans des données faisant autorité (Du et al, 2012), ainsi que des questions d'évaluation de la fiabilité des contributeurs (Skarlatidou et al, 2011). De plus, il paraît nécessaire de vérifier que ces données collaboratives peuvent être utilisées dans les applications classiques des données géographiques. L'application principale des données géographiques est la cartographie, et nous allons donc vérifier que des cartes lisibles peuvent bien être produites à partir de ces données.

Les données collaboratives et particulièrement OpenStreetMap ont du succès car ils sont simples d'utilisation (Brando, 2012) : n'importe quelle personne capable d'utiliser un géonavigateur peut contribuer, car les spécifications ne sont pas complexes à gérer. Le défaut de cette simplicité est l'hétérogénéité des données produites (Girres et Touya, 2010). Le niveau de détail de ces données est particulièrement hétérogène. Par exemple, des bâtiments faits pour une visualisation de type 'plan de ville' cohabitent avec des parcelles d'occupation du sol extraites à partir d'images satellitaires. Les cartes produites à partir de données collaboratives montrent que des incohérences de niveau de

détail entre objets de la carte reliés spatialement (par exemple un bâtiment et la zone bâtie qui le contient) provoquent des conflits cartographiques (Das et al, 2012). Ces conflits gênent fortement la réalisation de cartes de bonne qualité. Le but du travail présenté dans cet article est de proposer une méthode pour identifier les incohérences de niveau de détail dans les données collaboratives. En effet, si on arrive à identifier des incohérences de niveau de détail dans la carte, il sera possible plus tard de corriger ces défauts et de rendre ces cartes plus lisibles et plus compréhensibles. Le point clé de notre proposition pour l'identification des incohérences est une méthode en deux étapes : tout d'abord, on estime le niveau de détail des objets de la carte pris individuellement, puis on identifie des relations spatiales entre objets de la carte présentant des niveaux de détail très différents.

La partie 2 définit précisément la notion de niveau de détail et son importance en cartographie. La partie 3 présente notre proposition pour identifier le niveau de détail des objets d'une carte et détecter les incohérences. La partie 4 montre des résultats obtenus sur des données OpenStreetMap. La partie 5 explore brièvement des stratégies pour corriger les incohérences et la partie 6 conclut et décrit des pistes de travail futures.

2 L'Importance du niveau de détail

2.1 Echelle et niveau de détail

L'échelle d'une carte a une définition simple : c'est le facteur d'homothétie la taille terrain des objets et leur taille sur la carte. En cartographie, le mot d'échelle peut aussi avoir un aspect représentation : tous les éléments du monde réel ne peuvent pas être cartographiés à la même échelle mathématique, par exemple, les bâtiments peuvent être cartographiés individuellement à grande échelle, par exemple au 1 : 25 000, mais sont représentés par une zone bâtie aux petites échelles, par exemple au 1 : 200 000.

Cependant, l'échelle ne peut pas être utilisée pour caractériser une base de données géographique, car les données ne sont pas seulement dédiées à la représentation cartographique, et le terme niveau de détail est préféré. Ruas et Bianchin (2002) définissent le niveau de détail dans une base de données géographique comme la conjugaison de plusieurs facteurs : le **schéma conceptuel** des données, la **résolution sémantique**, la **résolution géométrique**, la **précision géométrique** et la **granularité**.

Le **schéma conceptuel** est la manière de représenter le terrain nominal (Ruas et Bianchin, 2002) dans une base de données géographique : représenter les forêts comme des objets surfaciques ou avec des objets ponctuels représentant chaque arbre sont des schémas conceptuels qui sous-tendent différents niveaux de détail.

La **résolution sémantique** est la quantité de détails dans les données attributaires liées à chaque objet géographique. Cette composante n'étant pas directement liée à la cartographie, nous n'y faisons pas référence dans le reste de l'article.

Par analogie avec la résolution des images, la **résolution géométrique** des objets vectoriels est approximativement la distance minimum entre deux points de la géométrie de l'objet. La fig. 1a montre qu'une telle définition doit être manipulée avec précaution car la distance entre deux points est toujours plus courte pour des géométries sinueuses. Comme Ruas et Bianchin (2002), nous considérons que la valeur de résolution d'une base de données géographique devrait être utilisée comme une estimation grossière de la résolution de ses objets.

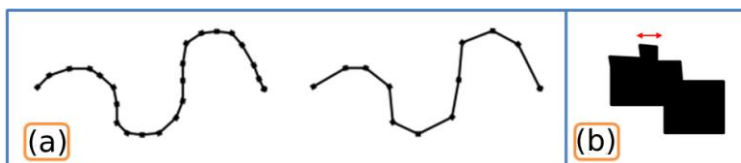


Figure 1 : (a) deux lignes à différentes résolutions. (b) la granularité d'un bâtiment (Touya et Brando, 2013).

La **précision géométrique** est simplement l'écart de positionnement entre la vérité terrain et la représentation d'un objet dans la base de données. La **granularité** décrit la taille des plus petites formes d'un objet, comme par exemple la taille de la plus petite protubérance d'un bâtiment (fig. 1b).

Dans le reste de l'article, nous considérons le niveau de détail comme l'agrégation des quatre composantes décrites ci-dessus. Nous appelons un haut niveau de détail, un niveau de détail avec une grande précision et une grande résolution (c'est-à-dire une petite distance entre deux points), etc., et nous appelons un faible niveau de détail un niveau de détail avec une faible précision, etc. Malgré cette définition, le niveau de détail est une notion relative qui reste difficile à mesurer quantitativement ou qualitativement. La partie 3.1 de l'article décrit notre proposition pour mesurer un niveau de détail haut ou faible.

2.2 Cas d'étude issu d'OpenStreetMap

Afin d'illustrer le problème posé par l'incohérence de niveau de détail et afin de tester nos propositions, trois zones d'études ont été extraites des données OpenStreetMap (OSM) françaises, car OSM est une source de données collaboratives dont un des objectifs est la réalisation de cartes. Les zones d'études couvrent différents types de paysages : urbain, rural, périurbain and montagneux (fig. 2).

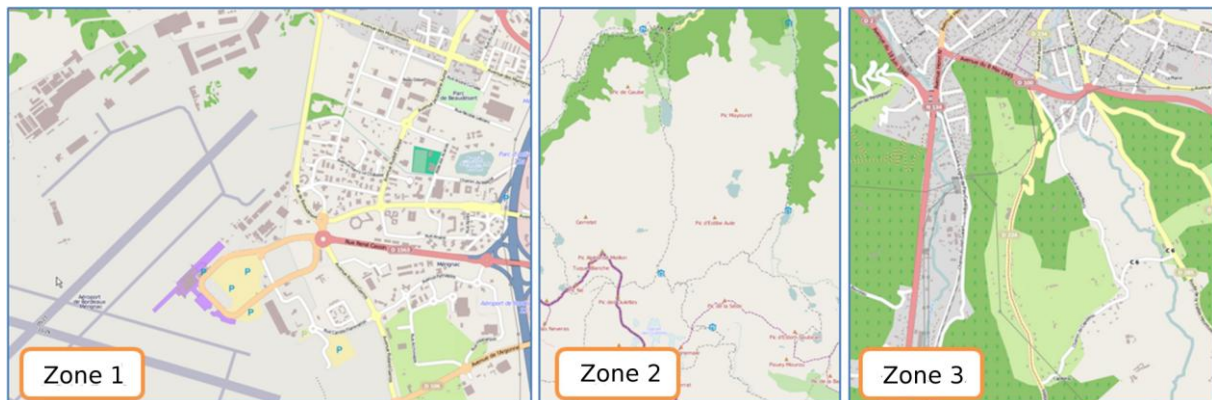


Figure 2 : Extraits des trois zones d'étude: une zone urbaine (Mérignac), une zone de montagne (Pyrénées), et une zone périurbaine (Jurançon) ©OpenStreetMap (Touya et Brando, 2013).

Les trois zones d'études contiennent une grande quantité d'incohérences de niveau de détail, car OSM cherche à rassembler indifféremment des objets très détaillés comme les pistes cyclables et des objets moins détaillés comme les routes maritimes ou les zones bâties. La quantité d'incohérence augmente naturellement à cause de l'hétérogénéité des méthodes et des sources de saisie utilisées par les différents contributeurs. Par exemple, on peut saisir un bâtiment à partir d'un outil automatique traitant des feuilles cadastrales françaises mises à disposition, ou à partir de traces issues d'un géonavigateur standard, ce qui ne permet clairement pas d'obtenir le même niveau de détail.

Le type d'incohérence le plus courant et le plus évident dans les zones d'étude est la cohabitation de bâtiments et de parcelles d'occupation du sol. Dans la fig. 3, des bâtiments intersectent une limite de forêt, ce qui est très improbable, ou se retrouvent juste à l'extérieur d'une zone bâtie dont ils devraient faire partie.



Figure 3 : Exemples d'incohérences de niveau de détail : les bâtiments en haut à gauche intersectent la limite de la forêt et les bâtiments sur la droite sont juste à l'extérieur de la zone bâtie ©OpenStreetMap (Touya et Brando, 2013).

Même dans les zones rurales où OSM est moins complet (Girres et Touya, 2010), les objets existants montrent des incohérences de niveau de détail. Dans la fig. 4, on peut voir des toilettes publiques et leur chemin d'accès connecté à des chemins moins détaillés, tout cela à côté d'objets beaucoup moins détaillés représentant un lac et des cours d'eau. Un chemin intersecte même le lac car celui-ci est très peu résolu (il n'y a pas de pont sur le terrain).

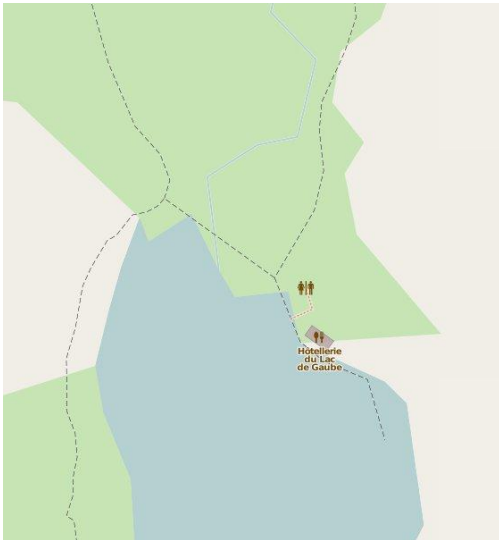


Figure 4 : Exemples d'incohérences de niveau de détail : un symbole de toilettes publiques et son chemin d'accès cohabitent avec des chemins moins détaillés (tirets) et un lac encore moins détaillé ©OpenStreetMap (Touya et Brando, 2013).

2.3 Cartes avec des incohérences de niveau de détail

Des incohérences de niveau de détail dans les données sources d'une carte peuvent entraîner la création de cartes de mauvaise qualité pour deux raisons principales :

- cela cause des problèmes de lisibilité à de nombreuses échelles,
- cela peut conduire à de fausses interprétations de la carte.

Tout d'abord, des incohérences de niveau de détail peuvent rendre des cartes à petite échelle illisibles, car des objets à haut niveau de détail ne devraient pas y figurer. En effet, la connaissance des limites de la perception visuelle nous permet de définir des seuils de lisibilité (par exemple seuil de perception, de séparation, de différenciation, etc.) : en dessous de ces seuils, l'œil ne perçoit que du bruit qui gêne la lecture de la carte, quelque soit le choix de légende (Duchêne et al, 2011). La fig. 5a montre un extrait des cartes standard OSM où la plupart des bâtiments sont trop petits pour être lisibles à cette échelle et cela gêne la lecture de la carte. Dans ce cas, il y a simplement besoin de généralisation, par exemple en augmentant la taille des bâtiments. Par ailleurs, des couches de données trop denses peuvent générer des problèmes de superposition de symbole, comme des

points d'intérêt sur un fond de carte lisible (fig. 5b). Bien qu'OSM ne soit pas conçu pour faire des cartes avec l'ensemble de ses données à chaque échelle, des problèmes de lisibilité apparaissent à la plupart des échelles, même les plus grandes.

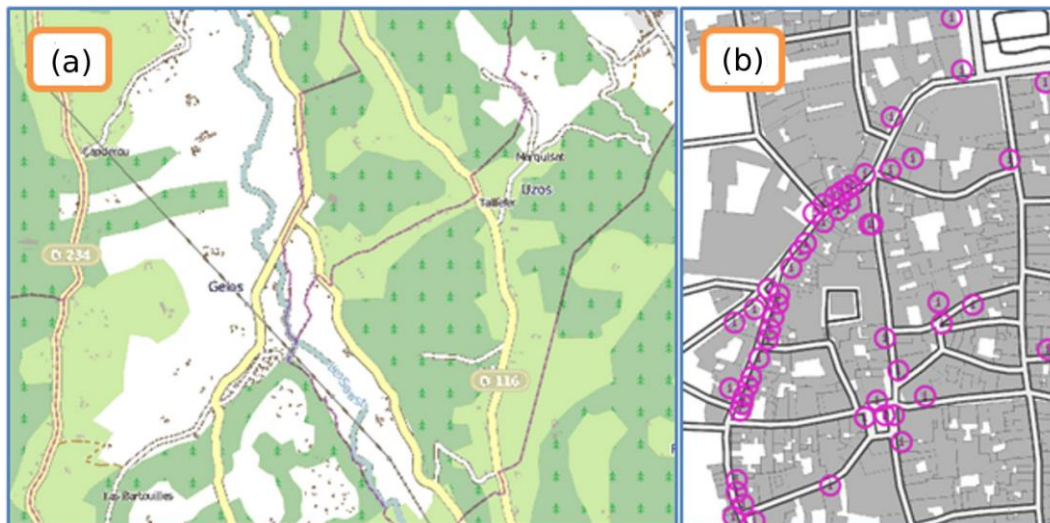


Figure 5 : (a) le niveau de détail du bâtiment est incohérent avec celui des routes, rivières et de l'occupation du sol, ce qui génère un bruit pour la lecture de la carte à cette échelle (©OpenStreetMap) (b) les symboles de points d'intérêt se superposent à cette échelle car trop nombreux (Touya et Brando, 2013).

Des incohérences de niveau de détail peuvent aussi transmettre des informations trompeuses au lecteur de la carte qui n'est pas conscient du niveau de détail particulier de chaque objet, et qui interprète la carte comme un tout cohérent. La fig. 6 montre un exemple où une incohérence entre le niveau de détail de bâtiments et d'une forêt dans OpenStreetMap trompe le lecteur sur la présence des bâtiments dans la forêt, la carte IGN montrant la présence d'une large clairière au niveau des bâtiments. Le niveau de détail de la forêt devrait limiter son utilisation aux cartes à petite échelle où les bâtiments ne sont pas représentés. De plus, le lecteur d'une carte s'attend par habitude à rencontrer un niveau de détail cohérent et son interprétation cognitive de la carte peut être biaisée par ces incohérences.



Figure 6 : le niveau de détail de la forêt est incohérent avec les bâtiments dans OSM entraînant une fausse interprétation du lecteur de la carte car les bâtiments ne sont pas dans la forêt mais bien dans une clairière (Touya et Brando, 2013).

Enfin, la cartographie automatique nécessite l'utilisation de processus complexes comme la généralisation pour rendre les données géographiques lisibles à une échelle donnée. Les processus de généralisation sont majoritairement conçus pour traiter des données issues d'agences de cartographie nationales, données homogènes en niveau de détail (Stanislowski et al, 2012). La généralisation cartographique de données collaboratives n'a pas encore été abordée en recherche : des projets de recherche débutent mais ne se concentrent pas sur l'hétérogénéité de niveau de détail. Par exemple, Klammer et Burghardt (2012) s'intéressent à la généralisation de cartes « tuilées » et

des projets basés sur OSM comme Mapnik (mapnik.org) ou CloudMade (cloudmade.com) ne fournissent que des opérations très simples de généralisation, comme le filtrage de ligne et la sélection d'objets, ce qui est insuffisant pour généraliser une carte. Ainsi, la détection des incohérences de niveau de détail permettrait de grandement améliorer les méthodes de création de cartes à partir de données collaboratives et la manière dont ces cartes sont perçues par un lecteur.

3 Détection des incohérences de niveau de détail

Cette partie décrit notre proposition pour détecter les incohérences de niveau de détail : on estime d'abord le niveau de détail des objets géographiques, puis on identifie des relations spatiales entre objets de niveaux de détail très différents. La première sous-partie présente brièvement une classification des niveaux de détail utilisée pour mesurer le niveau de détail. La deuxième sous-partie traite notre méthode d'estimation du niveau de détail d'un objet géographique, fondé sur une analyse multicritères. La dernière sous-partie décrit la détection des incohérences à partir des relations spatiales.

3.1 Classification des niveaux de détail

Le niveau de détail est une notion assez relative et floue car ses cinq composantes ne peuvent pas être définies clairement et quantitativement. Nous faisons donc le choix de mesurer le niveau de détail qualitativement grâce à une classification prenant la forme d'une échelle de Likert (Likert, 1935). Nous définissons cinq catégories de niveau de détail dérivées des niveaux de zooms du géoportail français (geoportail.gouv.fr), de la plus détaillée à la moins détaillée : rue, ville, département, région et pays. Le niveau de détail « rue » comprend les objets représentés pour la gestion cadastrale ou l'orientation dans des rues. La base de données anglaise OS MasterMap® est un exemple de ce niveau de détail. Le niveau de détail « ville » comprend les objets représentés pour décrire ce qui est visible sur le terrain (bâtiments, routes, rivières, forêts, etc.). Le niveau de détail « département » comprend les objets qui permettent les déplacements dans une petite zone, comme ceux d'un touriste (randonnée, visites, cyclotourisme par exemple). Le niveau de détail « région » est lié à la représentation de grandes régions et le niveau de détail « pays » est encore moins détaillé, pour la représentation de pays ou de très grandes régions.

La méthode d'estimation du niveau de détail présentée dans la partie suivante classe les objets géographiques dans une des cinq catégories ci-dessus, grâce à des critères quantitatifs et qualitatifs.

3.2 Estimation du niveau de détail

3.2.1 Critères pour déterminer le niveau de détail des objets géographiques

Les composantes du niveau de détail présentées en 2.1 permettent la définition de plusieurs critères pouvant être calculés pour les objets géographiques d'un jeu de données collaboratif :

- critère de **nature d'objet** (composante de schéma conceptuel)
- critère de **densité de points** (composante de résolution)
- critère de **longueur médiane d'une arête** (composante de résolution)
- critère de **source de saisie** (composante de précision géométrique)
- critère de **plus petite arête** (composante de granularité)
- critère de **taille** (composante de granularité)
- critère de **coalescence** (composante de granularité)

Chaque critère est conçu pour donner une valeur normalisée entre 0 et 1, 0 représentant le niveau de détail le plus faible, le niveau « pays ». La normalisation empirique de chaque critère implique des imprécisions et du flou, qui devra être pris en compte par la méthode de décision multicritère choisie.

Le critère de **nature d'objet** analyse la nature des objets géographiques en relation avec les spécifications et une ontologie des concepts géographiques. La présence d'objets de certaines natures dans la carte révèle la composante « schéma conceptuel » du niveau de détail. Par exemple, les bâtiments et les points d'intérêt ont un plus haut niveau de détail que les parcelles de zones bâties.

Le critère de **densité de points** analyse le nombre de points dans la géométrie d'un objet, par rapport à sa longueur, pour donner un indice de la résolution géométrique de l'objet, défini de manière floue. Afin de normaliser empiriquement cette densité, des valeurs de densité pour différents types d'objets, dans différents jeux de données venant de l'IGN, ont été étudiés. Le critère est calculé à l'aide de la formule suivante (Eq. 1) :

$$\text{valeur_de_densité} = (1 - \text{densité_de_points})^8 \text{ (Eq. 1)}$$

Le critère de **longueur médiane d'une arête** analyse la médiane des longueurs d'arêtes entre deux points des géométries pour compléter l'estimation de la résolution. Girres (2011) affirme qu'il s'agit de la meilleure mesure pour évaluer la résolution géométrique des routes et des rivières. Une fonction affine par morceaux donnant pour chaque longueur médiane une valeur entre 0 et 1 a été déterminée empiriquement après analyse des valeurs dans des données IGN et d'après les conclusions de Girres (2011).

Le critère de **source de saisie** analyse les métadonnées sur la source de saisie des objets (par exemple « traces GPS » ou « images satellitaires ») pour estimer la précision géométrique, à partir de connaissances sur ces sources. Les valeurs prises par le tag « source » dans OpenStreetMap ont été analysées pour les données françaises, et une catégorie de niveau de détail a été assignée pour chaque valeur possible du tag (en fait, une valeur entre 0 et 1 correspondant aux catégories). Par exemple, des objets saisis à partir du cadastre français ont une valeur de 0,1 pour ce critère alors que des objets importés dans OSM depuis CORINE Land Cover ont une valeur de 0,7.

Le critère de **plus petite arête** analyse de la plus petite arête entre deux points de la géométrie pour estimer la granularité. C'est une mesure classique pour mesurer la granularité d'un bâtiment en généralisation cartographique (Stöter et al, 2010). Le critère a été calibré en utilisant des données OSM et IGN pour obtenir une fonction affine par morceaux donnant pour chaque longueur une valeur dans [0,1].

Le critère de **taille** analyse la taille des objets pour compléter l'estimation de la granularité. En effet, la présence de très petits objets traduit une forte granularité. Nous avons utilisé les seuils de perceptibilité des objets pour relier les valeurs de taille aux catégories de niveau de détail.

Enfin, le critère de **coalescence** est dédié à l'estimation de la granularité des objets linéaires. Le critère se base sur le principe que si le symbole d'un objet linéaire se chevauche à une certaine largeur (quand la ligne est sinueuse fig. 7), cela signifie que l'objet ne peut pas être affiché à cette échelle (Girres 2011). Pour relier la largeur du symbole à l'échelle et aux catégories de niveau de détail, nous avons étudié les largeurs de symboles utilisées dans les cartes IGN aux différentes échelles.

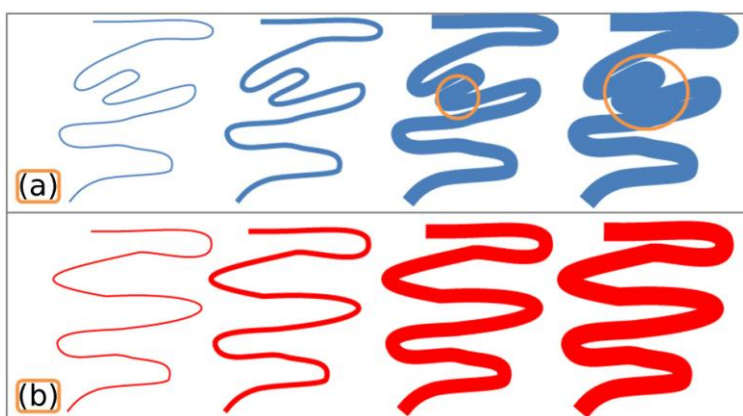


Figure 7 : granularité linéaire estimée par des tests de coalescence: la ligne (a) a une granularité plus grande que la ligne (b) car la coalescence arrive plus vite quand la largeur du symbole grandit (Touya et Brando, 2013).

Afin d'estimer le niveau de détail d'un objet à partir des ces critères, des techniques de décision multicritères sont nécessaires. La partie suivante présente la méthode que nous avons choisie, ELECTRE TRI (Figueira et al, 2005).

3.2.2 Estimation du niveau de détail avec ELECTRE TRI

L'analyse multicritères est un domaine de l'informatique qui développe des méthodes pour prendre des décisions éclairées par de multiples critères. La décision en elle-même peut être de plusieurs ordres : déterminer le meilleur choix, classer les choix ou bien les classer. Dans notre cas, la décision que nous voulons prendre est de classer un objet géographique dans la bonne catégorie de niveau de détail. ELECTRE TRI est une méthode de décision multicritères qui classe des choix, reprenant les principes généraux des méthodes de décision multicritères ELECTRE (Figueira et al, 2005). Les méthodes ELECTRE comparent seulement les choix critère par critère et ne comparent pas des valeurs issues de différents critères (on ne compare pas des choux et des carottes pour parler trivialement). Cette méthode a déjà été utilisée dans le cadre de la recherche en généralisation cartographique par Taillandier et Taillandier (2012).

La méthode ELECTRE TRI devrait être utilisée dans les conditions suivantes (Figueira et al, 2005):

- Classification en catégories ordonnées (comme de 'rue' à 'pays' dans notre cas).
- Il y a plus de trois critères (sept pour nous).
- Des critères hétérogènes (par exemple, il est difficile de comparer les critères de granularité et de source de saisie).
- Pour certains critères, une petite différence peut ne pas être significative, alors que l'addition de petites différences pour chaque critère peut devenir significative. Cela permet de compenser l'imprécision ou le flou sur les critères.

ELECTRE TRI classe des vecteurs de valeurs pour chaque critère (la dimension du vecteur est le nombre de critères), valeurs qui correspondent aux mesures de chaque critère pour un objet donné. Pour chaque catégorie, on définit deux vecteurs représentant la borne inférieure et la borne supérieure de la catégorie, et on compare notre objet aux bornes des catégories pour trouver dans laquelle il se trouve. Par exemple, pour la catégorie 'rue', le critère de densité de points (Eq. 1) a une valeur de 0 dans le vecteur de borne inférieure et de 0.2 pour le vecteur de la borne supérieure.

Le principe d'ELECTRE TRI est de comparer le vecteur pour un objet donné au vecteur d'une borne de catégorie avec une relation d'ordre : un objet est classé dans une catégorie si son vecteur est « meilleur » que la borne inférieure et si la borne supérieure est « meilleure » que le vecteur. La comparaison de deux vecteurs, soit le calcul de la relation d'ordre, est déduite de la comparaison des valeurs des vecteurs pour chaque critère pris indépendamment, deux valeurs de critères n'étant jamais directement comparées (fig. 8). Pour un critère donné, une valeur peut être « meilleure » qu'une autre ou les deux peuvent être considérées équivalentes pour ce critère (la décision globale étant alors faite sur les autres critères). En utilisant cette relation de préférence, chaque critère vote alors pour ou contre l'affirmation "le vecteur u est meilleur que le vecteur v" (Figueira et al, 2005).

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \end{pmatrix} \quad \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \end{pmatrix} \quad \left\{ \begin{array}{l} u_1 \succ v_1 \\ u_2 \succ v_2 \\ u_3 \approx v_3 \\ u_4 \succ v_4 \Rightarrow u \succ v \\ u_5 \approx v_5 \\ u_6 \succ v_6 \\ u_7 \approx v_7 \end{array} \right.$$

Figure 8 : dans ELECTRE TRI, la comparaison entre deux vecteurs (i.e. la relation d'ordre) est déduite des comparaisons pour chaque critère (Touya et Brando, 2013).

Afin de permettre cette comparaison des valeurs pour un critère donné malgré leur imprécision et leur flou, quatre propriétés doivent être définies pour chaque critère :

- Le poids : il traduit l'importance d'un critère par rapport aux autres.
- Le seuil de préférence : si la différence entre des valeurs pour ce critère est plus grande que ce seuil, le vecteur contenant cette valeur est préféré pour ce critère. Le seuil doit s'ajuster à la normalisation utilisée sur le critère (valeur entre 0 et 1 pour nous). Par exemple, le critère

de densité de point a un seuil de préférence de 0.2 alors que le critère de taille a un seuil de 0.5.

- Le seuil d'indifférence : si la différence entre des valeurs pour ce critère est inférieure à ce seuil, les deux vecteurs sont considérés comme équivalents pour ce critère. Par exemple, pour la plus petite arête, une différence inférieure à 0.25 n'est pas significative.
- Le seuil de veto : si la différence entre des valeurs pour ce critère est plus grande que ce seuil, alors le vecteur contenant la plus grande valeur sera toujours préféré globalement, quelque soient les valeurs pour les autres critères. Par exemple, le critère de taille a un grand seuil de veto car il n'est pas possible d'estimer le niveau de détail uniquement avec ce critère.

Le résultat de la méthode ELECTRE TRI pour un objet géographique donné est une catégorie de niveau de détail avec un indice de confiance sur la classification, par exemple, "l'objet route n°64 a un niveau de détail ville avec 80% de confiance".

3.3 Détection des incohérences à partir de relations spatiales

Tenir compte des relations spatiales est fondamental pour la dérivation automatique de cartes et notamment la généralisation (Duchêne et al, 2012). (Brando et al, 2011) affirment que l'identification de relations spatiales implicites aide aussi à la définition de bonnes spécifications dans le cadre de la saisie collaborative. (Brando 2012) suit cette idée en proposant d'ajouter une vérification de contraintes d'intégrité portant sur des relations spatiales dans les outils de saisie collaborative comme ceux utilisés par OSM. Elle propose par exemple de vérifier que les routes existantes ne croisent pas un nouveau bâtiment, ou que le nouvel arrêt de bus est bien le long d'une route existante. Nous pensons qu'une même approche est applicable au problème de l'incohérence de niveaux de détail : les relations spatiales peuvent nous aider à identifier les endroits où des incohérences existent et peuvent gêner la cartographie. Si deux objets qui devraient respecter une relation spatiale d'intégrité ne la respectent pas, et qu'en plus, ces deux objets ont des niveaux de détails très différents, il est fort probable que la non-intégrité soit due à l'hétérogénéité du niveau de détail (fig. 9a). Inversement, si une relation spatiale interdite existe dans les données entre deux objets dont le niveau de détail est très différent, il y a probablement aussi une incohérence de niveau de détail entre ces deux objets (fig. 9d).



Figure 9 : (a) le bâtiment devrait être dans la zone bâtie. (b) pas d'intersection entre lacs et chemins. (c) les arrêts de bus doivent être le long de la route. (d) un groupe de maison devrait être dans une clairière ©OpenStreetMap (Touya et Brando, 2013).

Donc, la méthode proposée utilise comme paramètres un ensemble de relations spatiales permettant de vérifier l'intégrité des données, et traite automatiquement les données pour identifier les instances de ces relations spatiales entre objets dont le niveau de détail est différent. Par exemple, si l'ensemble de relations d'intégrité contient les quatre relations représentées dans la fig. 9, la méthode recherche les instances de bâtiments justes à l'extérieur d'une zone bâtie, etc. La réification des relations suit l'idée de définir des instances de relations spatiales en généralisation de (Duchêne et al, 2012).

4 Application aux données OpenStreetMap

Les critères pour l'estimation du niveau de détail individuel ont été calibrés avec les données de la zone d'étude n°2 (voir 2.2) et testées sur les deux zones d'études restantes. La méthode a été implémentée sur la plateforme de généralisation open source CartAGen (Renard et al, 2010). La fig. 10 montre des résultats d'estimations automatiques du niveau de détail sur des routes, des bâtiments, des chemins ou des parcelles d'occupation du sol. Comme cela était attendu, les parcelles d'occupation du sol sont classées en « région » car elles sont directement importées depuis la base de données européenne CORINE Land Cover. Les bâtiments, importés depuis des feuilles cadastrales, sont classés en niveau « rue », sauf certains grands bâtiments, dont les détails sont moins marqués. Les routes sont classées de manière hétérogène ce qui s'explique par des modes de saisies eux aussi hétérogènes : des objets très détaillés comme les pistes cyclables sont classés en niveau « rue » alors que sont classées en niveau département (fig. 10). Après évaluation de la zone, les résultats correspondent bien à une évaluation visuelle du niveau de détail, à quelques très rares exceptions près.

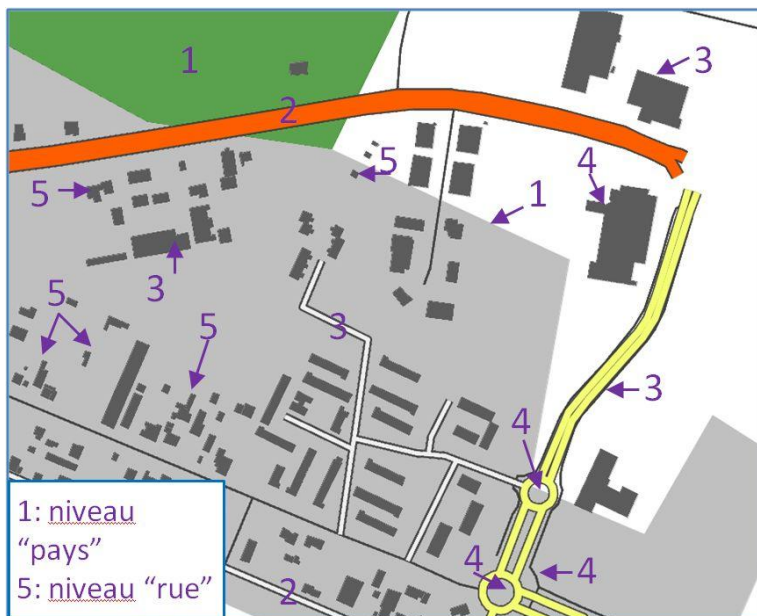


Figure 10 : des résultats de détection individuelle dans la deuxième zone d'étude (Touya et Brando, 2013).

La fig. 11 montre comment les bâtiments de la zone d'étude n°2 sont classés différemment suivant leurs caractéristiques (seulement la granularité, la résolution et la précision varient ici suivant les bâtiments).

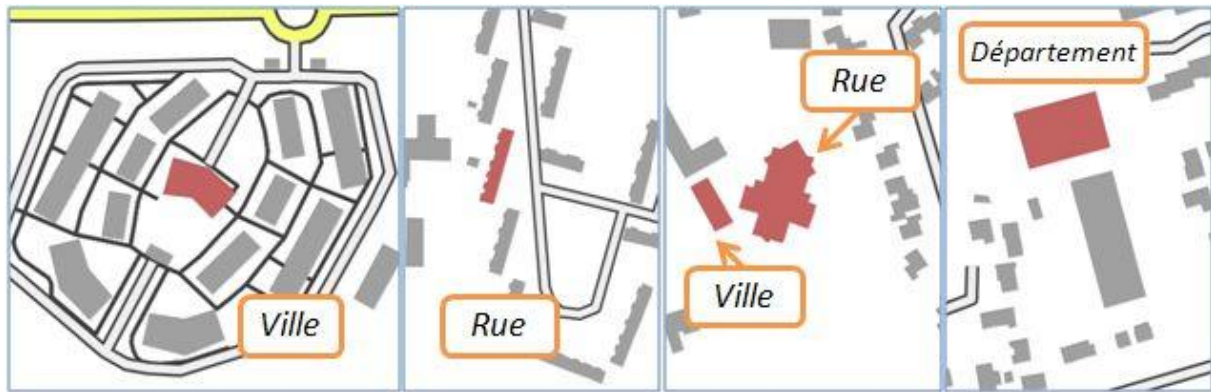


Figure 11 : Illustration de la détection du niveau de détail avec des bâtiments choisis dans la 2^{ème} zone d'étude (Touya et Brando, 2013).

Ensuite, les quatre relations spatiales de la fig. 9 sont choisies comme l'ensemble des relations d'intégrité à tester pour identifier les incohérences : (i) un bâtiment ne peut pas être juste à l'extérieur d'une zone bâtie; (ii) les lacs et les chemins ne peuvent s'intersecter; (iii) les arrêts de bus doivent être le long d'une route; (iv) un groupe de maisons proches ne peut pas être dans une forêt. Des instances de relations entre objets de différents niveaux de détail ont été automatiquement détectées dans les zones d'étude. La fig. 12a montre un groupe de maisons dans une forêt dont le faible niveau de détail explique l'absence de clairière autour du groupe de maisons. La fig. 12b montre des bâtiments identifiés comme justes à l'extérieur d'une zone bâtie peu détaillée. La fig. 12c montre avec la croix rouge trop loin pour être considéré comme le long d'une route au contraire des deux autres en bas de la figure. La fig. 12d montre un chemin détaillé qui intersecte un lac peu détaillé.



Figure 12 : détection automatique des incohérences de niveau de détail pour les quatre relations spatiales sélectionnées (en rouge) (Touya et Brando, 2013).

5 Harmonisation du niveau de détail

Puisque les incohérences de niveau de détail empêchent la production de cartes de qualité à partir de données collaboratives, des stratégies doivent être trouvées pour résoudre ces incohérences. Si l'objectif est de construire une carte à une échelle donnée à partir des données collaboratives, trois stratégies peuvent être adoptées. Cette partie discute seulement ces stratégies qui n'ont pas encore été testées.

La première stratégie serait de construire une base de données multi-représentations à partir des données collaboratives comme proposé dans (Müller et Wiemann, 2012). Cela permettrait l'adaptation du contenu de la carte à partir des représentations attachées à un niveau de détail donné. Construire une base de données multi-représentations nécessite des outils d'appariement de données, qui pourraient bénéficier du fait que les données collaboratives sont souvent aussi des données liées (Hahmann et Burghardt, 2010).

La deuxième stratégie serait une harmonisation des niveaux de détails vers celui qui est le plus proche de l'échelle de la carte. Lorsque l'on souhaite réaliser une carte à petite échelle, l'harmonisation est en fait une généralisation des objets dont le niveau de détail est plus élevé que celui de la carte. Cette généralisation devrait prendre en compte, en plus de ses contraintes habituelles, la résolution des incohérences de niveau de détail. Lorsque l'on souhaite réaliser une carte à grande échelle, l'harmonisation correspond cette fois à une amélioration des objets dont le niveau de détail est plus faible que celui de la carte et qui sont en incohérence avec des objets plus détaillés. La fig. 13 montre un outil automatique simple, que nous avons développé sur la plateforme CartAGen (Renard et al, 2010), qui améliore la géométrie des zones bâties pour inclure les bâtiments situés juste à l'extérieur : des zones tampons sont calculées autour des bâtiments situés juste à l'extérieur et une union de ces zones tampons avec la géométrie de la zone bâtie est réalisée. Cela montre qu'il est envisageable de développer d'autres outils similaires pour harmoniser d'autres incohérences de niveau de détail pour des cartes à grande échelle.

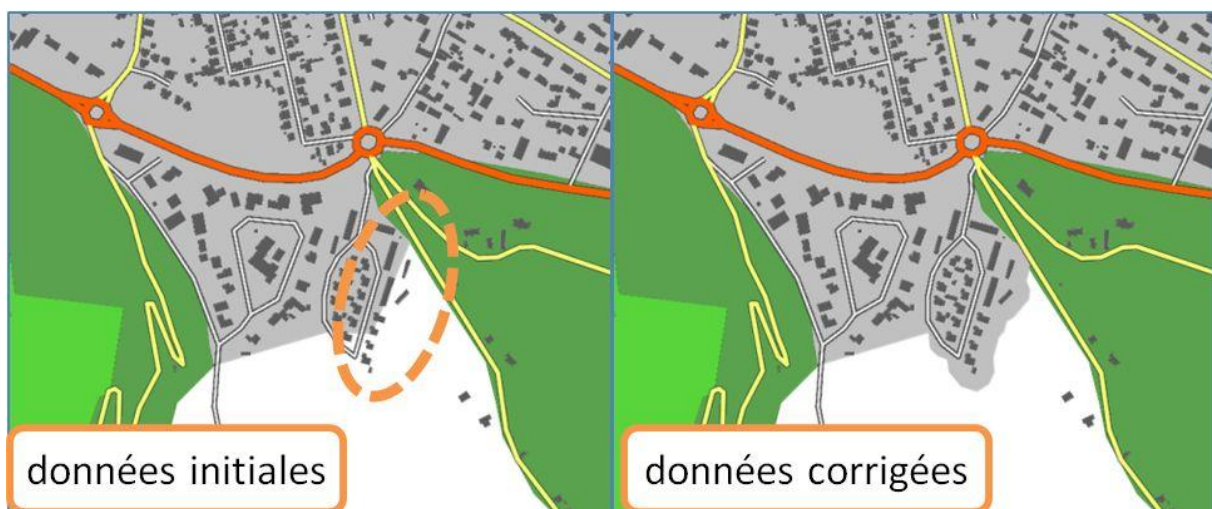


Figure 13 : données initiales (©OpenStreetMap) avec des bâtiments en dehors de la zone bâtie. La correction automatique étend la zone bâtie pour inclure les bâtiments à l'origine de l'incohérence (Touya et Brando, 2013).

La dernière stratégie d'harmonisation consiste à agir au moment de la contribution collaborative, en proposant des outils automatiques de réconciliation avec les objets déjà présents dans le jeu de données collaboratif. Brando (2012) propose un outil de réconciliation pour éviter la violation de contraintes d'intégrité pendant la contribution collaborative : par exemple, l'outil propose de déplacer un bâtiment hors d'une route si la contribution introduit ce bâtiment sur la route déjà présente dans les données. Des outils d'harmonisation pourraient réconcilier des contributions de la même manière pour des incohérences de niveau de détail, par exemple, étendre la zone bâtie quand un bâtiment est ajouté juste à l'extérieur de sa limite.

6 Conclusion et perspectives

Pour conclure, cet article présente les problèmes posés par l'hétérogénéité de niveau de détail dans les données collaboratives, si l'on souhaite réaliser des cartes. Il propose une méthode pour détecter automatiquement le niveau de détail des objets dans un jeu de données collaboratif ce qui est un point de départ pour la résolution du problème pour la cartographie à partir de données collaboratives. La méthode a été testée avec succès sur des données françaises OpenStreetMap.

Comme cela a été mentionné à plusieurs reprises dans l'article, il reste beaucoup de problèmes si l'on souhaite faire des cartes lisibles à partir de données hétérogènes en niveau de détail. Tout d'abord, il pourrait être intéressant d'intégrer la résolution sémantique dans les critères de détection du niveau de détail en reprenant les idées de Mooney et Corcoran (2012) sur les objets OSM lourdement modifiés. Il est ensuite indispensable de s'intéresser à une modulation du niveau de détail individuel en fonction de la population de l'objet ou de sa famille (l'ensemble des bâtiments par exemple) : l'effet de 'clutter' (c'est-à-dire la surcharge visuelle, Rosenholtz et al, 2007, Stigmar et Harrie, 2011) de fortes densités locales en objets doivent pouvoir augmenter le niveau de détail individuel des objets concernés. Par ailleurs, les stratégies présentées pour l'harmonisation du niveau de détail ne sont que des pistes de recherche qui méritent d'être approfondies. Enfin, il serait intéressant d'étudier la sensibilité des techniques actuelles de généralisation automatique aux données comportant des hétérogénéités de niveau de détail. Cela permettrait d'identifier les faiblesses de ces processus par rapport à l'homogénéité de niveau de détail, et de les adapter, comme l'ont fait Stanislawski et al (2012) pour le processus automatique de généralisation des cours d'eau.

Bibliographie

Brando C. « Edition collaborative d'un contenu géographique: une approche fondée sur des spécifications formelles et des relations spatiales explicites », dans *Actes des Journées de la Recherche IGN*, Saint-Mandé, France, 2012.

Brando C., Bucher B., Abadie N. « Specifications for user generated spatial content », dans *Advancing Geoinformation Science for a Changing World*, Berlin, Springer, 2011, p. 479-495.

Das T., van Elzakker C.P.J.M., Kraak M.J. « Conflicts in neogeography maps », dans *Actes d'AutoCarto 2012*, Columbus, OH, USA, 2012.

Du H., Anand S., Alechina N., Morley J., Hart G., Leibovici D., Jackson M., Ware M., « Geospatial information integration for authoritative and crowd sourced road vector data », *Transactions in GIS*, n°16-4, 2012, p. 455-476.

Duchêne C., Christophe S., Ruas A. « Generalisation, symbol specification and map evaluation: feedback from research done at COGIT laboratory, IGN France », *International Journal of Digital Earth*, n°4-1, 2011, p. 25-41.

Duchêne C., Ruas A., Cambier C. « The CartACom model: transforming cartographic features into communicating agents for cartographic generalisation », *International Journal of Geographical Information Science*, n°26-9, 2012, p. 1533-1562.

Figueira J., Mousseau V., Roy B. « ELECTRE methods », dans *Multiple Criteria Decision Analysis: State of the Art Surveys*, Berlin, Springer, 2005, p. 133-162.

Girres J.F. « An evaluation of the impact of cartographic generalisation on length measurement computed from linear vector databases », dans *Actes de la 25^{ème} Conférence de Cartographie Internationale (ICC'11)*, Paris, France, 2011.

Girres J.F., Touya G. « Quality assessment of the french OpenStreetMap dataset », *Transactions in GIS*, n°14-4, 2010, p. 435-459.

Haklay M. « How good is volunteered geographical information? a comparative study of OpenStreetMap and ordnance survey datasets », *Environment and Planning B: Planning and Design*, n°37-4, 2010, p. 682-703.

- Hahmann S., Burghardt D. « Connecting LinkedGeoData and geonames in the spatial semantic web », dans *Proceedings of extended abstracts, 6th International GIScience Conference*, Zurich, Switzerland, 2010.
- Klammer R., Burghardt D. « Approaches for enhancing tile-based mapping with cartographic generalisation », dans *Proceedings of 15th ICA Workshop on Generalisation and Multiple Representation*, Istanbul, Turkey, 2012.
- Likert R. « A technique for the measurement of attitudes », *Archives of Psychology*, n°22-140, 1932, p. 1-55.
- Mooney P., Corcoran P. « The annotation process in OpenStreetMap », *Transactions in GIS*, n°16-4, 2012, p. 561-579.
- Müller M., Wiemann S. « A framework for building multi-representation layers from OpenStreetMap data », dans *Proceedings of 15th ICA Workshop on Generalisation and Multiple Representation*, Istanbul, Turkey, 2012.
- Renard J., Gaffuri J., Duchêne C. « Capitalisation problem in research - example of a new platform for generalisation: CartAGen », dans *Proceedings of 11th ICA Workshop on Generalisation and Multiple Representation*, Zurich, Switzerland, 2010.
- Rosenholtz R., Li Y., Nakano L. « Measuring visual clutter », *Journal of Vision*, n°7-2, 2007.
- Ruas A., Bianchin A. « Echelle et niveau de détail », dans *Généralisation et représentation multiple*, Paris, Hermès Lavoisier, 2002, p. 25-44.
- Skarlatidou A., Haklay M., Cheng T. « Trust in web GIS: the role of the trustee attributes in the design of trustworthy web GIS applications », *International Journal of Geographical Information Science*, n°25-12, 2011, p. 1913-1930.
- Stanislowski L.V., Doumbouya A., Miller-Corbett C., Battenfield B., Arundel-Murin S. « Scaling stream densities for hydrologic generalization », dans *Proceedings of GIScience 2012*, Columbus, OH, USA, 2012.
- Stigmar H., Harrie L.E. « Evaluation of analytical measures of map legibility », *The Cartographic Journal*, n°48-1, 2011, p. 41-53.
- Stoter J.E., Baella B., Blok C., Burghardt D., Duchêne C., Pla M., Regnauld N., Touya G., 2010, *State-of-the-art of automated generalisation in commercial software*, Amsterdam, EuroSDR series on research projects.
- Taillandier P., Taillandier F. « Multi-criteria diagnosis of control knowledge for cartographic generalisation », *European Journal of Operational Research*, n°217-3, 2012, p. 633-642.
- Touya G., Brando C., « Detecting Level-of-Detail Inconsistencies in Volunteered Geographic Information Data Sets », *Cartographica*, n°48-2, 2013, p. 134-143.