

UN PROCESSUS DE SÉLECTION DU RÉSEAU HYDROGRAPHIQUE, BASÉ SUR LA DÉTECTION DE STRUCTURES

par Guillaume Touya

Laboratoire COGIT
Institut géographique national,
2-4 avenue Pasteur 94165 Saint-Mandé Cedex
guillaume.touya@ign.fr

Introduction

La généralisation est un processus qui vise à réduire le niveau de détail d'une base de données géographique dans le but de satisfaire de nouvelles spécifications. Depuis des années, beaucoup de travaux ont traité de son automatisation. La plupart de ces travaux concernent la généralisation cartographique dont l'objectif est la production de cartes, mais cet article se concentre uniquement sur la généralisation de base de données ou de modèle. L'article traite un point spécifique et important de la généralisation de base de données, à savoir la généralisation des réseaux hydrographiques et plus particulièrement le problème de la sélection de ces réseaux : quels cours d'eau ou tronçons de cours d'eau sont assez importants ou significatifs pour faire partie de la base de données généralisées ? Ce sujet a déjà été traité plusieurs fois dans le passé. La particularité de la solution proposée est qu'elle est basée sur la détection de structures à l'intérieur du réseau hydrographique.

La deuxième partie de cet article présente un état de l'art de la généralisation de réseaux hydrographiques. La troisième partie expose la problématique, les hypothèses et les pré-traitements nécessaires. La quatrième partie décrit précisément les différentes étapes du processus de sélection proposé. La cinquième partie présente des résultats obtenus avec ce processus lors de tests sur des données réelles de l'IGN. Enfin, la dernière partie est une conclusion avec la présentation de quelques perspectives de recherche.

1 La généralisation

Puisque la généralisation est un problème clé pour la représentation et l'intégration des données

géographiques, c'est un sujet qui a été particulièrement abordé. On parle de généralisation cartographique lorsque l'objectif est la production de cartes et, dans ce cas, les principaux problèmes à traiter sont liés à la lisibilité et aux symboles cartographiques. Ce papier traite de généralisation de bases de données dont les objets ne sont pas symbolisés [Weibel & Dutton 1999], et donc de problèmes différents. L'objectif de la généralisation est alors de réduire le niveau de détail d'une base de données géographique dans son ensemble, en modifiant ainsi à la fois les données et le schéma de données, sans prendre en considération une possible symbolisation de ces données.

Le processus de généralisation de base de données peut être décomposé en plusieurs étapes [McMaster & Shea 1992] : la sélection, la généralisation attributaire et la généralisation géométrique. Cet article ne traite que de la sélection qui est en général l'étape la plus ardue du processus. [Mackaness & Edwards 2002] insiste sur l'importance de l'analyse spatiale pour détecter les structures implicites d'une base de données géographique et enrichir celle-ci en les explicitant. Cela permet de guider et d'améliorer le processus de généralisation et c'est particulièrement le cas dans le cadre de la sélection. [Heinzle et al. 2006] et [Touya 2007] montrent bien l'importance de cette reconnaissance des structures dans la généralisation des réseaux routiers et ce postulat peut être étendu aux réseaux hydrographiques, bien que les structures dans des réseaux naturels, tels que les réseaux hydrographiques, soient différentes de celles présentes dans des réseaux créés par l'homme comme les réseaux routiers.

2 Généralisation des réseaux hydrographiques

Les principes de l'organisation perceptive ont été énoncés par les psychologues du Gestalt [Wertheimer 1938]. L'organisation perceptive décrit le phénomène par lequel le cerveau humain organise tous les éléments de son champ visuel ou de son environnement. Les lois de l'organisation perceptive jouent un rôle clé dans la compréhension des images en deux dimensions de scènes ou paysages en trois dimensions, et donc dans la compréhension des cartes et de la sémiologie graphique [McEachren 1995]. Ces lois jouent aussi un rôle clé en généralisation cartographique et sont ainsi utilisées depuis de nombreuses années [McMaster & Shea 1992]. Dans le cas de réseaux linéaires, la loi d'organisation perceptive dite de "bonne continuation" est particulièrement intéressante. Elle permet en effet de grouper des tronçons d'un réseau en "strokes", ensembles de tronçons qui semblent se suivre de manière continue comme si on les avait dessinés d'un seul trait de crayon [Thomson & Richardson 1999]. Dans les cartes ou les bases de données géographiques, les "strokes" de routes ou de cours d'eau ont une vraie signification géographique car ce sont des phénomènes assez continus. Par exemple, dans un réseau routier, un boulevard périphérique est un stroke. [Thomson & Brooks 2000] a appliqué ce principe de bonne continuation pour construire automatiquement les strokes dans des réseaux routiers et hydrographiques dans le but de réaliser leur sélection dans un processus de généralisation.

2.1 Sélection de réseaux hydrographiques

Parmi les premiers travaux sur la généralisation de réseaux hydrographiques, [Richardson 1994] affirme que le problème est principalement une question de sélection qui peut être résolue par une hiérarchisation des cours d'eau, couplée à des seuils de longueur des cours d'eau. Les principales méthodes de classification des cours d'eau sont celles de Strahler [Strahler 1957], Horton [Horton 1945] (fig.1) et Shreve [Shreve 1966].

[Mauger 1997] compare les différentes méthodes de hiérarchisation et les méthodes pour déterminer le cours d'eau principal à un point de confluence. Cet article éclaire également sur les problèmes posés par ces méthodes. Par exemple, déterminer le cours d'eau principal en utilisant la longueur du cours d'eau

depuis la source [Horton 1945] sur des réseaux qui ne sont que des extraits d'un bassin versant (la source réelle de certains cours d'eau n'apparaît pas dans le réseau) peut causer de grosses erreurs.

Pour ce qui est des autres travaux sur le sujet, [Richardson 1994] et [Martinez Casanovas & Molenaar 1995] proposent un modèle de généralisation basé sur l'agrégation et l'abstraction hiérarchique. [Thomson & Brooks 2000] utilise le principe de continuation des lois d'organisation perceptive pour construire des "strokes" de cours d'eau ou "cours d'eau continus". La sélection consiste en une hiérarchisation de ces cours d'eau continus par la classification de Horton. Sur la figure 1, les cours d'eau en gras dans la classification de Horton sont en fait des cours d'eau continus hiérarchisés.

Cet article présente une méthode de sélection de réseaux hydrographiques, déjà introduite dans [Touya 2006], basée sur la hiérarchisation de cours d'eau continus. La valeur ajoutée de cette méthode tient dans sa prise en compte des structures implicites du réseau comme les îles ou les zones d'irrigation pour une meilleure sélection, mais aussi dans le développement de pré-traitements ou la gestion de réseaux ne représentant pas un bassin versant complet.

2.2 Contexte et données

Le problème traité dans cet article est la sélection de réseaux hydrographiques linéaires dans le contexte de la généralisation de base de données. L'application de la méthode proposée est liée à deux bases de données géographiques produites par l'Institut géographique national (IGN) : la BD TOPO®, qui a pour résolution 1 m et pour échelle de référence le 1:15 000 environ, et la BD CARTO®, qui a pour résolution 10 m et pour échelle de référence le 1:100 000 environ (fig.2). Mais l'objectif est de concevoir un processus de sélection générique pour les réseaux hydrographiques linéaires permettant de répondre à différents types de spécifications de base de données généralisée en changeant simplement les paramètres. En général et dans le cas de la BD CARTO®, les spécifications d'une base de données sont assez floues car elles ne détaillent pas parfaitement quels objets géographiques doivent être présents dans la base de données. Par exemple, des spécifications peuvent mentionner : "les impasses ne devraient pas être dans la base sauf si elles sont très longues". Déduire des paramètres de généralisation à partir de telles spécifications est assez difficile et

devient par là même une tâche primordiale de toute généralisation de base de données.

Les hypothèses faites sur les données dans ce travail correspondent aux spécifications de la BD TOPO® pour permettre les tests. Dans la BD TOPO®, le réseau hydrographique est modélisé comme un réseau linéaire planaire complet. Donc le processus ne traitera que des données linéaires. De plus, dans la BD TOPO®, l'information sur le sens d'écoulement des cours d'eau est contenue dans la géométrie des tronçons de cours d'eau. L'eau s'écoule dans le sens de vectorisation de la polyligne. Pour utiliser ce processus avec des données différentes, des pré-traitements sont nécessaires pour obtenir un réseau complètement linéaire et planaire avec une information sur le sens d'écoulement de l'eau de chaque tronçon.

Même si un algorithme générique ne devrait pas prendre en compte des paramètres aussi variables d'un jeu de données à un autre, tels que les informations attributaires, la sélection de réseaux hydrographiques est bien meilleure quand des attributs, même incomplets comme le nom des cours d'eau, sont utilisés. On peut donc considérer que les attributs présents dans les données initiales sont des paramètres du processus.

3 Pré-traitements

L'information sur le sens d'écoulement de l'eau sur les tronçons de cours d'eau n'est pas toujours complètement fiable. Donc, il est indispensable de nettoyer les données initiales pour qu'elles contiennent le moins d'erreurs possible sur le sens d'écoulement de l'eau. Pour cela, on peut utiliser la topologie et la géométrie comme dans [Paiva & Egenhofer 2000] pour recalculer le sens d'écoulement de chaque tronçon du réseau. Mais cet algorithme ne génère pas de résultats assez fiables pour une bonne sélection. Ainsi, deux processus ont été développés pour corriger les quelques erreurs dans le réseau hydrographique. Le premier processus n'est pas générique et dépend ici de la BD TOPO®. Il consiste en une utilisation des informations d'altitude de la base de données : si l'altitude du point final du tronçon est supérieure à celle du point initial, c'est que le tronçon a été vectorisé dans le mauvais sens. Le second processus est lui générique car il consiste en une analyse du sens d'écoulement des tronçons du voisinage. Analyser localement les tronçons entrants et sortants peut permettre de détecter des incohérences comme dans les cas simples de la figure 3a.

Aux sources ou pertes du réseau, l'analyse de voisinage peut aussi détecter les anomalies (fig. 3b).

Pour les données initiales de ce travail, ces deux processus sont suffisants. Le premier corrige la plupart des erreurs à condition que la cohérence des données d'altitude soit assurée. Le deuxième processus corrige presque toutes les erreurs restantes : il reste moins de 0,1% de tronçons contenant une erreur sur le sens d'écoulement. De plus, les processus de généralisation nécessitent souvent des données topologiquement correctes. La plupart des données produites contiennent des erreurs et l'utilisation d'un processus de vérification et de correction de la topologie permet de s'affranchir de ces erreurs et de travailler sur un graphe réellement planaire. Une fois ces pré-traitements effectués, le processus de sélection peut être appliqué.

4 Processus de sélection proposé

Le § 4.1 traite de la première étape du processus, l'enrichissement de la base de données par détection de structures ; le 4.2 décrit le processus de hiérarchisation utilisé ainsi que les critères de sélection appliqués en utilisant les enrichissements apportés à la base de données.

4.1 Phase d'enrichissement de la base de données

4.1.1 Schéma de données enrichi

La généralisation automatique et donc la sélection de réseaux hydrographiques sont des processus très complexes qui nécessitent l'utilisation d'information contextuelle. Par conséquent, il est souvent nécessaire d'enrichir les données en explicitant dans la base de données des structures implicites qu'un œil humain pourrait aussi détecter [Mackaness & Edwards 2002]. Le schéma de données de la figure 4 résume tous les enrichissements à effectuer pour rendre possible le processus de sélection. Le schéma de données initial ne contenait que la classe "tronçon de cours d'eau". Comme dans [Thomson & Brooks 2000], une classe "cours d'eau continu" est ajoutée pour stocker les "strokes" qui seront la base de la sélection. Les classes "source" et "perte" sont ajoutées pour stocker les nœuds de début et de fin de chaque cours d'eau continu afin qu'il soient utilisés facilement comme des objets de la base de données dans le processus de sélection. Deux classes sont aussi ajoutées dans le but de gérer les îles sur les cours d'eau. Enfin, les classes "zone d'irrigation"

et “stroke d’irrigation” sont ajoutées pour permettre la gestion de ces structures particulières du réseau. Les sous-parties 4.1.2 à 4.1.5 décrivent comment ces classes sont automatiquement instanciées.

4.1.2 Sources et pertes

Pour construire correctement les zones d’irrigation et les cours d’eau continus, il est indispensable d’identifier d’abord les sources et pertes du réseau. Une source est détectée comme un nœud du réseau qui a un tronçon sortant mais aucun tronçon entrant, et inversement pour les pertes. [Thomson & Brooks 2000] utilisait déjà les sources et pertes comme des objets de la base de données, mais travaillait sur un bassin versant entier, ce qui est rarement le cas avec la plupart des données. Ce processus permet de travailler sur seulement un extrait d’un ou plusieurs bassins versants sans engendrer d’erreurs dans la sélection grâce à l’attribut “type” des sources et des pertes. Cet attribut détermine si une source (ou une perte) du réseau est naturelle (elle représente une source dans la réalité géographique) ou si elle est issue du découpage de la zone de travail (fig.5).

Puisque la longueur des cours d’eau d’un point de confluence à la source peut être utilisée comme critère de détermination du cours d’eau principal, il faut s’assurer que les sources en question sont bien naturelles et non issues du découpage de la zone. La distinction entre ces sources et pertes est donc primordiale et permet une sélection de bonne qualité sur n’importe quel type de réseau hydrographique initial.

4.1.3 Construction des “strokes” ou “cours d’eau continus”

La construction des strokes de cours d’eau ou cours d’eau continus est ensuite l’enrichissement essentiel à la sélection. Il s’agit d’un algorithme descendant : on commence au niveau des sources et l’algorithme se termine lorsque l’ensemble des nœuds du réseau (sources, pertes, points de confluences ou nœuds simples) ont été traités. L’algorithme parcourt le réseau en utilisant le sens d’écoulement de l’eau, d’où l’importance de sa justesse et du pré-traitement présenté en 3. Comme dans [Thomson & Brooks 2000], les cours d’eau continus sont créés pour correspondre à la représentation et à la classification réelle des cours d’eau : un cours d’eau continu commence à une source et se termine à une perte ou à un point de confluence avec un cours d’eau continu plus important. “Plus impor-

tant” signifie que, dans ce cas, l’autre cours d’eau continu est “mieux continué” par le tronçon en aval du point de confluence.

La partie périlleuse de la création des cours d’eau continus est donc la détermination du chemin de meilleure continuation à chaque point de confluence. Si on suit géométriquement le principe de bonne continuation du Gestalt, c’est le chemin sans rupture de courbure ou avec la meilleure courbure qui est le plus continu. Mais pour avoir des résultats meilleurs et plus réalistes, ce principe de bonne continuation est contraint par plusieurs règles, les cours d’eau continus différant de strokes purement géométriques. De nombreuses règles peuvent être utilisées pour déterminer le chemin principal. [Horton 1945] utilise le chemin le plus long et droit, [Thomson & Brooks 2000] utilise prioritairement le nom du cours d’eau puis la continuité de courbure, et d’autres peuvent utiliser la plus grande surface de drainage.

Voici les règles utilisées dans ce travail:

- La continuité de nom de cours d’eau est prioritaire.
- Toute chose étant égale par ailleurs, un cours d’eau de régime “permanent” est prioritaire sur un cours d’eau de régime “intermittent”.
- Toute chose étant égale par ailleurs et les sources des cours d’eau continus en amont étant toutes “naturelles”, le cours d’eau continu amont le plus long a la priorité. La longueur à la source est le meilleur critère dans ce cas selon [Mauger 1997] par rapport à la continuité de courbure. La courbure est plutôt utilisée si un des angles de confluence est supérieur à 60° et que la différence de longueur est peu significative (500 m).
- Toute chose étant égale par ailleurs et une des sources des cours d’eau continus amont étant de type “ limite de zone”, le chemin avec la meilleure continuité de courbure a la priorité.

Par ailleurs, les réseaux hydrographiques sont souvent composés de deltas ou de rivières à plusieurs bras. Tous ces bras annexes doivent être traités de façon appropriée par l’algorithme de création des cours d’eau continus et le processus de sélection. Donc en premier lieu, deux types de cours d’eau continus sont distingués : les cours d’eau continus “principaux” et les cours d’eau continus “annexes” comme dans [Thomson & Brooks 2000]. Quand un cours d’eau se sépare en plusieurs bras, un bras est considéré comme le principal et continue le cours d’eau continu principal ; les autres bras sont utilisés

comme premiers tronçons de nouveaux cours d'eau continus annexes. Le bras aval principal est celui avec la meilleure continuité de courbure (fig.6). Le plus court chemin vers la perte peut être utilisé comme critère mais il n'a pas été choisi dans ce travail pour des questions de rapidité de calcul, sachant que l'apport en qualité de ce critère ne compensait pas la perte d'efficacité. Ce critère serait très approprié pour de petits jeux de données.

4.1.4 Les îles et les îles complexes

Dans les réseaux hydrographiques, les bras annexes correspondent souvent à la présence d'îles sur les rivières. De plus, on trouve souvent plusieurs îles adjacentes sur une même rivière (fig.7). Ces îles sont des structures significatives du réseau et il peut être intéressant de sélectionner leur contour dans la base généralisée pour maintenir l'information sur la présence d'une île ou d'îles, mais de manière généralisée. Durant la phase d'enrichissement, l'objectif est donc de détecter automatiquement ces îles simples ou complexes et de les caractériser afin de permettre leur sélection. On définit une île complexe comme un agglomérat d'îles simples. Cette idée d'agréger des éléments pour augmenter le niveau d'abstraction correspond à la philosophie de généralisation de [Richardson 1994]. Dans la base de données enrichie, chaque île simple ou complexe est associée au tronçon formant son contour (fig. 7). Si les spécifications de sélection requièrent un niveau d'abstraction différent pour les îles, les îles simples formant une île complexe peuvent être agrégées en plusieurs îles complexes adjacentes afin de maintenir l'information d'îles adjacentes après la sélection par exemple. Cela explique la présence du lien de la classe "île complexe" sur elle-même dans le schéma de la figure 4.

Afin de construire automatiquement de telles îles, les faces topologiques du réseau sont utilisées. Seules les petites faces, celles qui représentent vraiment des îles sont considérées en utilisant un seuil déterminé empiriquement. Les îles complexes sont construites par clustering sur les îles simples : on agrège petit à petit les îles simples adjacentes. Pour permettre la création d'îles complexes représentant un plus petit niveau d'abstraction, un clustering par classification ascendante hiérarchique est préférable.

4.1.5 Les zones d'irrigation

Les réseaux hydrographiques contiennent un

autre type de structure particulière, les zones d'irrigation agricole. Il s'agit d'ensembles de petits canaux artificiels utilisés dans des zones plates pour faciliter l'irrigation des champs (fig. 8). La détection de telles zones est importante pour plusieurs raisons : d'abord ces zones sont sources d'erreurs pour la création des cours d'eau continus car les tronçons n'ont pas de réel sens d'écoulement, étant artificiels et plats. Dans la base de données, un sens d'écoulement arbitraire et hétérogène est attribué à ces tronçons. Ajouté à cela, il serait intéressant de pouvoir maintenir l'information de la présence de ces zones après sélection en choisissant, par exemple, une version simplifiée de ces zones.

Les zones d'irrigation sont donc caractérisées par un sol à peu près plat, une forte densité de tronçons petits et droits, et une forte densité de sources et pertes. Ces caractéristiques peuvent être utilisées pour la détection automatique de ces zones et leur construction dans la base de données enrichie. Ensuite, une mesure de compacité est utilisée pour écarter les zones sur-détectées. La figure 8 montre un exemple de détection automatique réussie de zone d'irrigation.

L'étape suivante est la reconnaissance automatique des tronçons artificiels et naturels à l'intérieur de chaque zone d'irrigation afin de les enlever du processus de création des cours d'eau continus. On remarque que les cours d'eau artificiels sont plutôt courts et droits alors que les naturels sont plutôt longs, sinueux, et intersectent beaucoup d'autres cours d'eau. Pour traduire cette différence en mesures d'analyse spatiale, on utilise encore la notion de strokes mais cette fois de manière purement géométrique car seule la différence de courbure est utilisée. Ainsi, des strokes sont créés dans chaque zone d'irrigation et caractérisés par des mesures de longueur, sinuosité et nombre d'intersections avec d'autres strokes. La figure 9 montre les résultats de cette différenciation automatique.

La détection des cours d'eau naturels et artificiels dans les zones d'irrigation termine la phase d'enrichissement de données du processus de sélection de réseaux hydrographiques proposé. L'étape suivante concerne la hiérarchisation des cours d'eau continus en vue de leur sélection.

4.2 La phase de sélection

Une fois la base de données initiale enrichie, la

sélection des cours d'eau peut enfin être effectuée. Cette sélection est complètement basée sur ces enrichissements. La majeure partie de cette sélection concerne le choix des cours d'eau continus à garder. Comme ces cours d'eau continus sont construits pour représenter les rivières entières, la sélection concerne plutôt ces cours d'eau continus que les tronçons de cours d'eau initialement contenus dans la base de données. Tous les tronçons d'un cours d'eau continu sont sélectionnés ou aucun ne l'est. Afin de déterminer quels seront les cours d'eau continus à sélectionner, on utilise une classification hiérarchique de ces cours d'eau. Comme dans [Thomson & Brooks 2000], on utilise la classification de Horton (fig. 1) sur les cours d'eau continus. Pour réaliser cette classification automatiquement, une première étape est de réaliser la classification de Strahler sur tous les tronçons.

Afin de gagner du temps de calcul, l'ordre de Strahler des tronçons est assigné durant l'algorithme de création des cours d'eau continus. Quand un tronçon étend un cours d'eau continu après un point de confluence, l'ordre de Strahler de ce tronçon est calculé comme dans la figure 1 : l'ordre est le même que le maximum des tronçons précédents sauf si deux tronçons amonts ont le même ordre, auquel cas celui du tronçon aval est incrémenté de 1. Les tronçons associés à un stroke de type "bras annexe" ne se voient pas assigner d'ordre car leur importance dépend surtout de l'ordre du cours d'eau continu principal auquel ils sont associés, et de leur propre longueur.

Ensuite, l'ordre de Horton de chaque cours d'eau continu est finalement calculé comme étant le maximum des ordres de Strahler des tronçons qui le composent. Ainsi, on peut déterminer quels sont les cours d'eau continus les plus importants et quels sont les moins importants. Cette hiérarchisation est le critère de base de la sélection.

Donc, on sélectionne d'abord les cours d'eau continus par un seuil sur l'ordre de Horton et sur la longueur totale du cours d'eau continu. Par exemple, si on se place dans notre cas d'étude qui est la généralisation de la BD CARTO® à partir de la BD TOPO®, on dérive ces seuils des spécifications de la BD CARTO® : tous les cours d'eau continus d'ordre supérieur à 3 ou d'une longueur supérieure à 1 km sont sélectionnés. Dans le cas de cours d'eau continus de type "bras annexe", la sélection est un peu

différente. Ils ne sont sélectionnés que si le cours d'eau continu principal associé est lui-même sélectionné et ensuite un seuil de longueur est imposé. Par exemple, dans l'application BD TOPO®/BD CARTO®, le seuil est de 1 km. Ces paramètres sont, à l'évidence, facilement modifiables pour être appliqués à d'autres cas de bases de données.

Les premières expérimentations ont montré que généralement le critère portant sur les bras annexes ne pouvait généraliser correctement le phénomène des îles sur les cours d'eau. Pour cette raison, la base de données a été enrichie avec les îles simples et complexes et leurs contours. Ainsi, en plus de la première phase de sélection des cours d'eau continus, un seuil de taille est appliqué sur les îles et les tronçons de cours d'eau appartenant au contour d'une île assez grande sont ajoutés à la sélection. Cette sélection ne dépend pas cette fois des cours d'eau continus. Un deuxième seuil peut être ajouté sur la taille que peuvent avoir deux îles adjacentes dans la base de données généralisée. Ainsi, certaines îles complexes peuvent être agrégées partiellement en respectant ce seuil pour maintenir dans la base de données généralisée, l'information de la présence d'une île complexe.

Finalement, dans les zones d'irrigation, les tronçons naturels sont sélectionnés grâce aux processus précédents et les tronçons artificiels sont éliminés. Mais il est possible d'envisager aussi une typification de ces zones. La typification signifie que l'on réduit arbitrairement la quantité de tronçons artificiels pour correspondre à la résolution de la nouvelle base de données mais que les tronçons restants permettent toujours d'identifier la zone d'irrigation en gardant ses propriétés.

5 Résultats

Le processus complet d'enrichissement et de sélection décrit dans cet article a été testé sur le réseau hydrographique d'un département français entier, les Pyrénées-Atlantiques (7600 km²) qui possède une diversité de paysages (montagnes, plaines, côte maritime et zones d'irrigation) en plus d'être le découpage d'un plus grand bassin versant. Le processus a été implémenté sur le SIG Radius Clarity™ de la société 1Spatial qui est utilisé à l'IGN pour les processus de généralisation. Comme expliqué dans la partie 3 de cet article, le processus a été appliqué au cas de la généralisation de la BD

CARTO® (10 m de résolution) à partir de la BD TOPO® (1 m de résolution). Plusieurs résultats obtenus complètement automatiquement sont présentés en figure 10.

La figure 10 montre que le jeu de données sélectionné est très proche de la base de données cible (BD CARTO) et même légèrement de meilleure qualité avec des problèmes de connectivité du réseau corrigés (fig.10(a)). Les données généralisées et les données cibles ont la même densité et quantité de cours d'eau et les cours d'eau principaux sont les mêmes. Les résultats présentés en figure 10(b) permettent également d'expliquer la notion de "sélection équivalente". Comme les données initiales et cibles sont issues de bases de données complètement indépendantes dans leur gestion et leur création, des différences de mise à jour et des biais dans les spécifications de saisie empêchent de pouvoir sélectionner automatiquement et exactement les données cibles. L'idéal est alors d'obtenir une sélection équivalente aux données cibles. Une sélection équivalente consiste en une sélection exacte des éléments (ici cours d'eau) les plus importants, une sélection donnant à peu près la même quantité et la même densité d'objets, et enfin une sélection acceptant quelques différences dans le choix des éléments secondaires sélectionnés (fig.10(b)). Dans cet exemple, la sélection est même de meilleure qualité que les données cibles car l'actualité des données est meilleure tout comme la cohérence par rapport aux spécifications. Enfin, la figure.10(c) montre des résultats intéressants sur la sélection des îles. Sur un réseau parfait en terme de sens d'écoulement, les résultats sont très bons mais dans le cas contraire, les pré-traitements de correction présentés dans la partie 3 sont indispensables.

6 Conclusion et perspectives

Cet article propose un processus complet de sélection des réseaux hydrographiques linéaires. Ce processus est basé sur des travaux précédents en généralisation de réseaux hydrographiques comme [Thomson & Brooks 2000] qui introduit la notion de "strokes" ou cours d'eau continus, mais aussi [Mauger 1997] pour la hiérarchisation des cours d'eau du réseau. Les apports de ce travail sont mul-

tiples : proposition d'une méthode pour gérer les extraits de bassins versants, mais surtout introduction d'une dynamique prenant en compte les structures caractéristiques d'un réseau hydrographiques (îles, zones d'irrigation) ainsi que des méthodes pour les détecter ; enfin, proposition d'une méthode de correction contextuelle du sens d'écoulement de l'eau. Le processus se déroule en plusieurs étapes : correction des données initiales, enrichissement des données, hiérarchisation des cours d'eau et enfin sélection. L'implémentation du processus sur le SIG Radius Clarity™ donne des résultats encourageants sur des données réelles et importantes en taille.

Plusieurs perspectives existent pour améliorer ce processus. Tout d'abord, les structures géomorphologiques des bassins versants sont un sujet important (fig.11). Savoir que les cours d'eau appartiennent à une structure spécifique de bassin versant pourrait permettre de contraindre la sélection pour préserver les caractéristiques de ces structures. Le laboratoire COGIT essaie actuellement de développer des mesures pour détecter automatiquement de telles structures. Il serait donc intéressant de pouvoir les appliquer au cas de la sélection de réseaux hydrographiques.

Ensuite, davantage de travail pourrait être effectué sur la détection de structures dans les réseaux hydrographiques. Par exemple, les méandres pourraient être détectés pour être sûr que le cours d'eau continu principal suit bien le tracé du méandre. Les méandres pourraient aussi être caractérisés automatiquement (taille, homogénéité, etc..) pour enrichir la base de données généralisée et aussi faciliter une future généralisation géométrique de ces méandres. De plus, il serait intéressant d'améliorer encore la correction contextuelle du sens d'écoulement de l'eau pour obtenir de très bonnes corrections sans utiliser de données d'altitude. Des règles plus complexes pourraient également être introduites dans le choix du chemin de meilleure continuité. Enfin, dans le cas d'un processus complet de généralisation du réseau hydrographique, il serait intéressant de considérer des algorithmes de simplification tels que [Christensen 2003].

Bibliographie

- Christensen A.H.J., 2003**, "Two Experiments on Stream Network Generalisation", dans *Proceedings of the 21th ICC*. ICA, Durban, South Africa.
- Deffontaines B., Chorowicz J., 1991**, "Principles of Drainage Basin Analysis from Multisource Data: Application to the Structural Analysis of the Zaire Basin", *Tectonophysics*, 194, p. 237-263.
- Heinzle F., Anders K.-H., Sester M., 2006**, "Pattern Recognition in Road Networks on the Example of Circular Road Detection", dans *Proceedings of the 4th International Conference GIScience 2006*, Münster, Germany, p.153-167.
- Horton R.A., 1945**, "Erosional development of Streams and their Drainage Basins: Hydrophysical approach to Quantitative Morphology", *Geo. Soc. America Bull.*, vol. 56, p. 275-370.
- Howard A.D., 1967**, "Drainage Analysis in Geologic Interpretation: a summation", *Bull. Am. Assoc. Pet. Geol.*, 51(11), p. 2246-2259.
- MacEachren A.M., 1995**, *How Maps Work*, The Guilford Press, p. 51-150.
- Mackness W., Edwards G., 2002**, "The Importance of Modelling Pattern and Structure in Automated Map Generalisation", dans *Proceedings of Joint Workshop on Multi-scale Representations of Spatial Data*, Ottawa, Canada.
- MacMaster R.B., Shea K.S., 1992**, *Generalisation in Digital Cartography*, Washington D.C., Association of American Cartographers.
- Martinez Casasnovas J.A., Molenaar M., 1995**, "An Aggregation Hierarchy for Multiple Representation of Hydrographic Data Within the Context of Conceptual Generalisation", dans *Proceedings of the 17th ICC*. ICA, Barcelona, vol. 1.
- Mauger F., 1997**, *Hiérarchisation d'un réseau de talwegs en vue de sa généralisation*, mémoire de master, Laboratoire COGIT Laboratory.
- Paiva J., Egenhofer M., 2000**, "Robust Inference of the Flow Direction in River Networks", *Algorithmica*, 26 (2).
- Richardson D., 1994**, "Generalisation of Spatial and Thematic Data Using Inheritance and Classification and Aggregation Hierarchies", dans *Advances in GIS Research 2*, London, Taylor and Francis, p. 957-972.
- Shreve R.L., 1966**, "Statistical law of stream numbers", *Journal of Geology*, vol.74, p. 17-37.
- Strahler A.N., 1957**, "Quantitative Analysis of Watershed Geomorphology", *American Geophys. Union Trans.*, vol. 38 (6), p. 913-920.
- Thomson R., Brooks R., 2000**, "Efficient generalisation and abstraction of network data using perceptual grouping", dans *Proceedings of the 5th GeoComputation*, University of Greenwich, Kent U.K.
- Thomson R., Richardson D., 1999**, "The "Good Continuation" Principle of Perceptual Organization Applied to the Generalisation of Road Networks", dans *Proceedings of the 19th ICC*. ICA, Ottawa, Canada.
- Touya G., 2006**, " ", Münster, Germany, p.

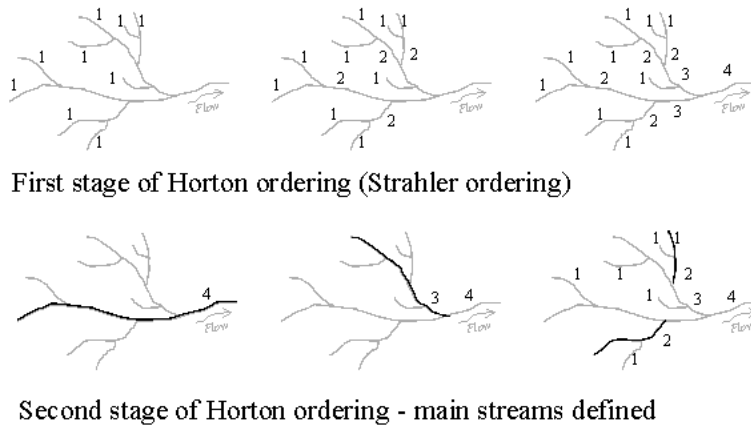


Figure 1 : Description de la hiérarchisation de Strahler et de Horton utilisée pour la sélection. La première étape est la classification de Strahler de chaque tronçon. Ensuite, la classification de Horton d'un cours d'eau est le maximum des ordres de Strahler des tronçons qui le composent.. Extrait de [Thomson & Brooks 2000]

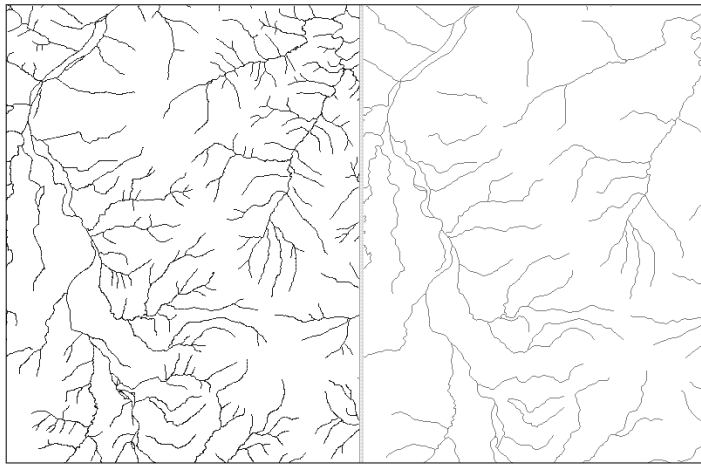


Figure 2 : Deux extraits des réseaux hydrographiques utilisés : à gauche, un extrait de la BD TOPO® et à droite, un extrait de la même zone dans la BD CARTO®



Figure 3 : (a) le tronçon en gras a un mauvais sens d'écoulement car il y a incohérence avec ses voisins ; (b) le tronçon en gras a un mauvais sens d'écoulement car il y a incohérence avec ses voisins et il est le premier tronçon d'une rivière, issu d'une source

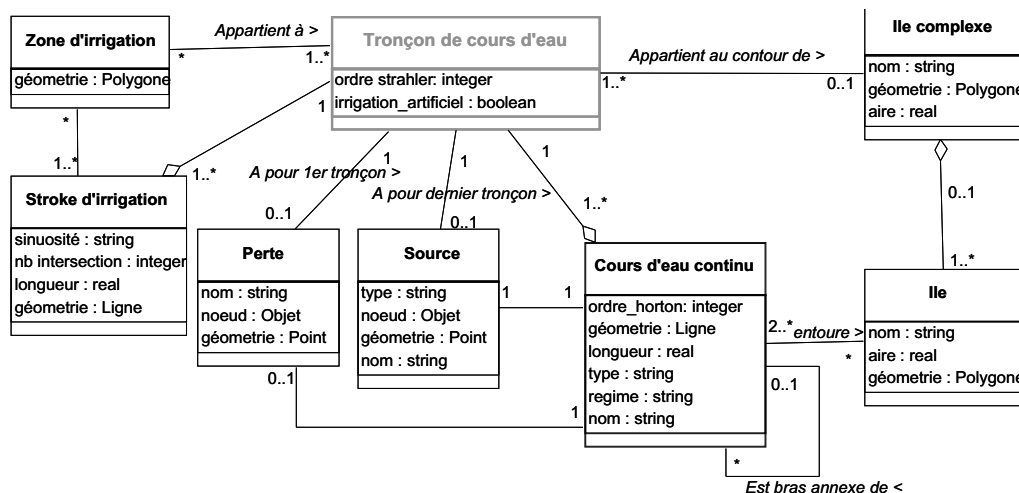


Figure 4 : Schéma de données enrichi pour la sélection de réseaux hydrographiques

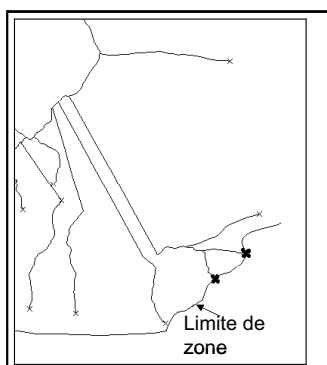


Figure 5 : Les 2 types de sources : les croix simples sont des sources naturelles alors que celles en gras sont des sources de type "limite de zone"

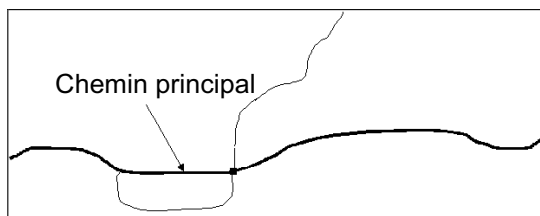


Figure 6 : Choix du cours d'eau principal à un point de confluence, point avec 2 tronçons amont et 2 tronçons aval

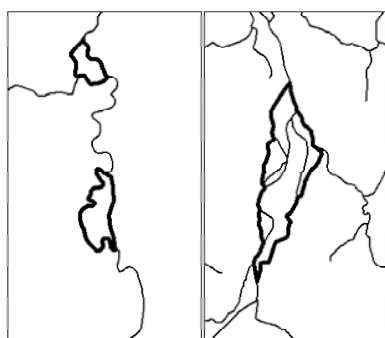


Figure 7 : Des îles simples à gauche et une île complexe composée d'îles simples adjacentes à droite. Les contours sont automatiquement créés par le processus



Figure 8 : Résultats de la création automatique de zones d'irrigation : le polygone en gras marque le contour de la zone d'irrigation

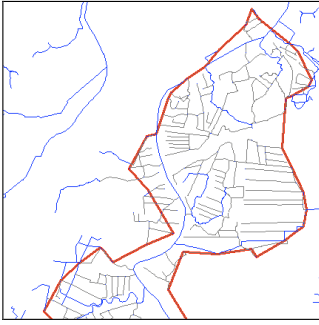


Figure 9 : Résultats de la détection automatique des tronçons artificiels dans une zone d'irrigation. En bleu, les tronçons naturels et en gris, les tronçons artificiels

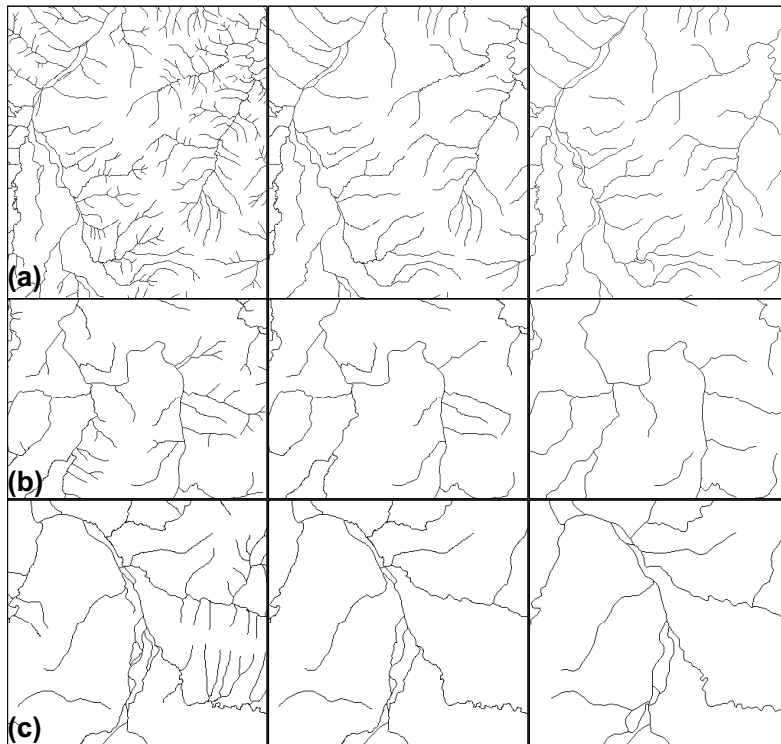


Figure 10 : Résultats du processus de sélection dans 3 zones différentes : à gauche, les données initiales de la BD TOPO®, au milieu, les données généralisées et à droite, les données de comparaison de la BD CARTO®

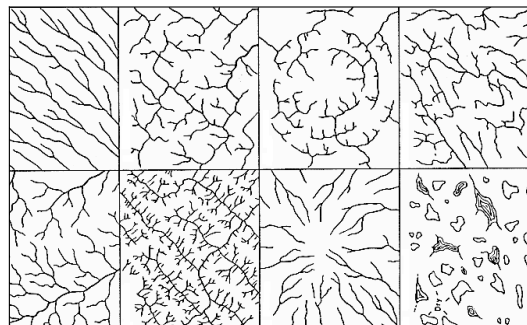


Figure 11 : Classification des structures géomorphologiques des bassins versants [Howard 1967]