



HAL
open science

Généralisation et intégration pour un fond vert commun entre l'IFN et l'IGN

Guillaume Touya, Cécile Duchêne, Sébastien Mustière

► **To cite this version:**

Guillaume Touya, Cécile Duchêne, Sébastien Mustière. Généralisation et intégration pour un fond vert commun entre l'IFN et l'IGN. *Revue Internationale de Géomatique*, 2010, 20 (1), pp.65-86. 10.3166/rig.20.65-86 . hal-02303215

HAL Id: hal-02303215

<https://hal.science/hal-02303215v1>

Submitted on 2 Oct 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.

Généralisation et intégration pour un fond vert commun entre l'IFN et l'IGN

Guillaume Touya — Cécile Duchêne — Sébastien Mustière

Laboratoire COGIT, Institut Géographique National
2 Avenue Pasteur 94160 Saint-Mandé Cedex
{ prénom.nom}@ign.fr

RÉSUMÉ. Le travail présenté est une étude réalisée au laboratoire COGIT de l'IGN (Institut Géographique National) qui a pris place dans le contexte de la collaboration entre l'IGN et l'IFN (Inventaire Forestier National), organisme en charge de l'inventaire permanent de la ressource forestière nationale. L'IGN et l'IFN ont travaillé à mutualiser la production de leurs données vectorielles relatives aux zones de forêt pour constituer un "fond vert" commun à partir duquel chacun pourrait dériver les données répondant à ses propres besoins. Le laboratoire COGIT a étudié dans quelle mesure il était possible de passer automatiquement des données végétation produites à l'IGN, à ce fond vert commun, puis à des données répondant aux besoins propres de l'IFN et de l'IGN. Les techniques utilisées pour cela sont d'une part la généralisation pour simplifier les données, et d'autre part l'appariement pour mettre en correspondance des données saisies indépendamment. Cette étude testée sur trois zones représentatives a montré la faisabilité d'un processus de généralisation et d'intégration automatique répondant aux besoins des deux organismes. Elle a débouché sur une mise en production effective de certaines des propositions faites.

ABSTRACT. The presented paper deals with a study carried out at COGIT laboratory of IGN France in relation to a collaboration between IGN and IFN (Inventaire Forestier National), the French institute that manages national forests. For some time, IGN and IFN have tried to gather efforts of production of their own vector vegetation coverage through the creation of a "common green coverage". The common green coverage would be derived to meet IGN and IFN different specifications that are different in terms of level of detail and thematic content. COGIT laboratory studied if it was possible to automatically produce the common green coverage and then the derived products by generalisation and data matching. Tests were carried out on three different landscape areas and they allow to conclude that a partly automated process is possible to meet both institutes needs. It lead to effective production flow lines.

MOTS-CLÉS: généralisation, appariement, végétation.

KEYWORDS: generalisation, data matching, vegetation.

1. Introduction

Le travail présenté dans cet article est le résultat d'une étude réalisée par le laboratoire COGIT de l'Institut Géographique National (IGN) à la demande de son Service des Bases Vecteurs (SBV) dans le cadre de la collaboration entre l'IGN et l'Inventaire Forestier National (IFN) sur la saisie d'un "fond vert commun".

Jusqu'à récemment, l'IGN et l'IFN saisissaient et maintenaient tous deux des données vectorielles relatives aux zones de forêt. Les deux organismes ont étudié la possibilité de mutualiser la production de ces données pour constituer un "fond vert" commun, factorisant les besoins communs aux deux organismes et à partir duquel chacun d'eux pourrait dériver les données répondant à ses besoins propres : d'une part la couche végétation du Référentiel Grande Echelle (RGE) et les produits cartographiques dérivés à l'IGN et d'autre part la couche forestière de l'IFN. Les besoins des deux organismes, donc les spécifications des données forestières qu'ils saisissent et utilisent, sont très différents, notamment une résolution géométrique plus fine des données à l'IGN et un découpage thématique plus fin à l'IFN. La mutualisation des efforts a déjà commencé, puisque la saisie initiale de la végétation est faite à la fois à l'IGN et à l'IFN : les différentes tâches sont réparties entre les deux instituts. Dans ce scénario de mutualisation vers un fond vert commun, la création de ce fond vert et sa spécialisation pour l'IGN et l'IFN nécessitent des opérations et des transformations de généralisation et d'intégration qui seraient les plus automatiques possibles. L'objet de l'étude était d'étudier la faisabilité de ce scénario et d'identifier un processus optimal pour créer et adapter le fond vert commun. La faisabilité de l'automatisation de ces transformations a été également étudiée. Enfin, la généralisation de zones de forêts vecteurs est un sujet de recherche peu traité (Mackanness *et al.*, 2007) et un objectif de l'étude est aussi d'apporter une contribution sur ce sujet.

La deuxième partie de cet article traite de la problématique de la mutualisation et de la définition du fond vert commun. La troisième partie décrit le processus de généralisation automatique proposé pour obtenir le fond vert commun à partir des données IGN. Les quatrième et cinquième parties traitent des processus proposés pour les dérivations de produits spécifiques IFN puis IGN. Enfin, la sixième partie dresse le bilan de l'étude et expose quelques perspectives de développement et de recherche.

2. Définition du fond vert commun

2.1. Objectifs

L'objectif de la collaboration entre l'IFN et l'IGN est de mutualiser le plus possible les tâches de production tout en améliorant la cohérence de leurs données. Produites à partir de la même base, elles pourront être facilement manipulées et

visualisées ensemble par les utilisateurs. L'hypothèse de départ de l'étude technique est la saisie partagée entre l'IGN et l'IFN d'un même fond vert brut, grâce au logiciel SEVE développé à l'IGN à partir de travaux du laboratoire MATIS de l'IGN (Trias-Sanz, 2006, Guigues *et al* 2003). Ce logiciel permet la délimitation automatique de zones de végétation dans une ortho-image et leur classification semi-automatique en "forêt" ou "non forêt" (Figure 1). Une nouvelle version du logiciel SEVE permet une classification thématique plus poussée grâce au traitement d'une image infrarouge en distinguant par exemple feuillus, conifères et vignes. Cette dernière version du logiciel est actuellement utilisée par les deux instituts pour leur saisie du fond vert brut commun.

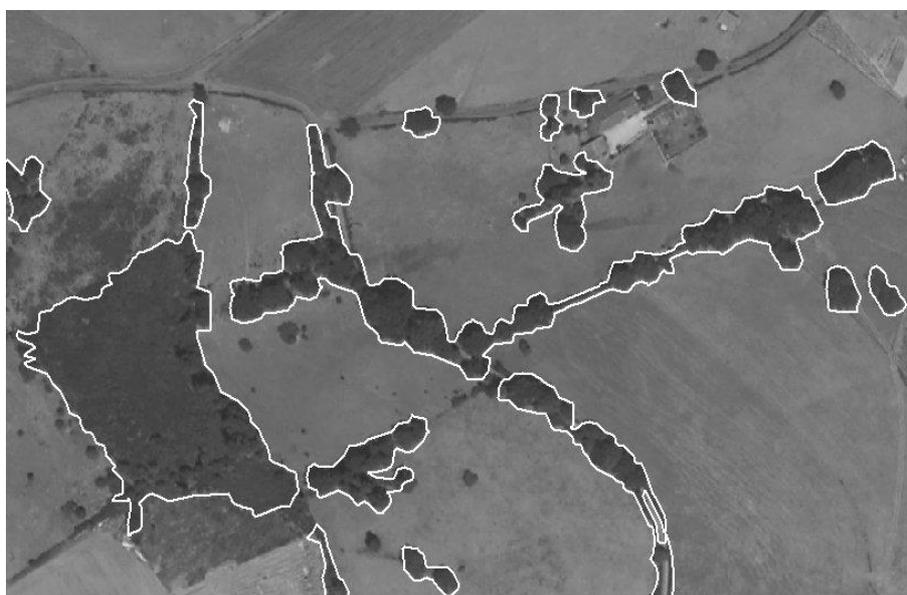


Figure 1. Exemple de zones de forêt saisies semi-automatiquement sous SEVE sur une ortho-photographie.

La couche ainsi obtenue correspond à l'actuelle couche végétation du RGE. Si les tests présentés dans les parties suivantes n'ont pas été réalisés sur le fond vert brut thématiquement obtenu par la nouvelle version de SEVE, les processus ont été conçus pour fonctionner indifféremment soit sur le fond vert ne distinguant que forêt et non forêt soit sur le fond vert thématiquement.

L'IGN souhaite utiliser le fond vert commun pour améliorer la qualité de la couche végétation du RGE et pour ses besoins de création de cartes topographiques. De son côté, l'IFN souhaite l'utiliser pour produire une couche de gestion thématique. Dans les deux cas, ils ne peuvent pas utiliser directement le fond vert brut qui est trop détaillé pour ce genre d'utilisations. L'idée est donc de traiter par

généralisation ce fond vert brut pour obtenir un fond vert commun utile aux deux instituts puis de réaliser des traitements spécifiques aux utilisations topographiques ou thématiques (voir Figure 2). Pour l'IFN, une saisie thématique complémentaire sera par exemple indispensable.

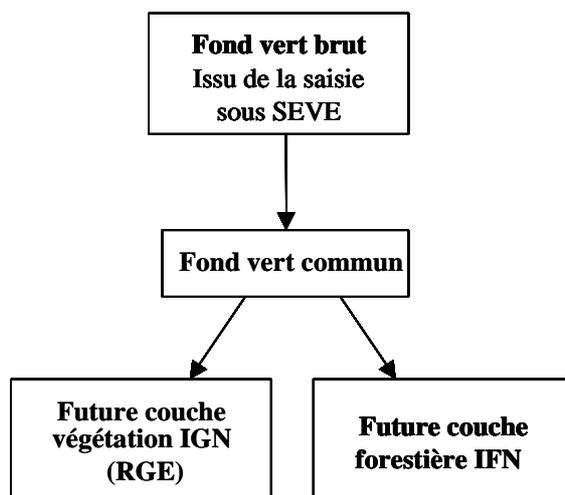


Figure 2. *objectifs de la chaîne de création du fond vert commun IGN-IFN.*

2.2. Spécifications du fond vert commun

Afin de définir les spécifications du fond vert commun, il faut identifier les besoins relatifs aux données de végétation pour les deux instituts. L'IGN, dans ses bases de données, a pour mission de recenser des ressources et de décrire leur position donc la conservation de la précision planimétrique ainsi que de l'exhaustivité du fond vert brut dans le fond vert commun serait intéressante pour la qualité de la couche. Néanmoins, le niveau de détail du contour des zones de végétations nécessite une simplification pour se rapprocher du niveau de détail du reste des données topographiques du RGE (Figure 3a). De plus, une différenciation des haies (Figure 3b), des bosquets (Figure 3c) et des forêts (Figure 3e) serait souhaitable pour des besoins de bases de données (les haies sont une information topographique intéressante) et pour des besoins cartographiques (les haies sont représentées sur une carte par un symbole linéaire). En outre, il semble important de restaurer la relation topologique entre les forêts ou les haies et les réseaux routiers, hydrographiques et ferrés (Figure 3d) : par exemple, le RGE devrait contenir l'information qu'une forêt borde une route sous la forme d'un partage de géométrie. Enfin, pour des besoins de gestion SIG des données, il apparaît important de réduire la taille des polygones représentant les forêts. Pour cela, nous proposons de les découper selon les réseaux routiers, hydrographiques et ferrés.

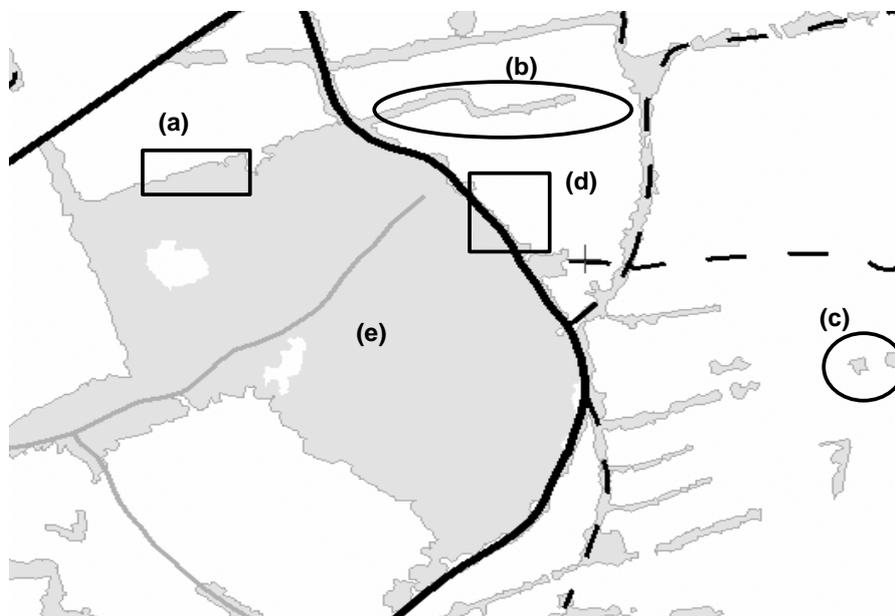


Figure 3. Identification des besoins pour le fond vert commun sur un extrait du fond vert brut : (a) niveau de détail trop important. (b) une haie à différencier de la forêt. (c) un bosquet à différencier des forêts. (d) une route le long d'une lisière de forêt sans partage topologique.

Les besoins de l'IFN sont un peu différents. Comme pour l'IGN, la précision (position des objets par rapport au monde réel) et l'exhaustivité (la quantité des objets du monde réel représentés) du fond vert brut conviennent alors qu'il faut simplifier le niveau de détail (« échelle » de la base de données). Le premier besoin est la possibilité d'enrichissement du fond pour conserver le découpage thématique des forêts qui fait la richesse de la couche forestière actuelle de l'IFN. L'autre besoin majeur de l'IFN est la simplification du niveau de détail de la végétation qui est bien trop important dans le fond vert brut. Enfin, comme pour l'IGN, il apparaît nécessaire de différencier haies et forêts et utile de rétablir la cohérence entre les forêts et le réseau routier.

L'expression de ces besoins nous ont permis de définir plus précisément et par convention certaines notions qui vont être utilisées dans la suite de cet article. Pour les deux instituts, une distinction forêt/haies/bosquets apparaît. Au vu des besoins exprimés et après examen des données, nous définissons ces notions de la manière suivante : tout d'abord, on appelle *haie* toute zone de végétation du fond vert brut fine et allongée. Nous utilisons un seuil empirique de 25 m pour déterminer la finesse d'une zone de végétation : il correspond à peu près à la largeur maximale d'une rangée d'arbre et est issu des spécifications de la BD Topo Standard® de

l'IGN. Cette définition regroupe les véritables haies (clôtures composées d'arbres, arbustes, branchages servant à délimiter un champ) mais également les alignement d'arbres tels ce que l'on trouve le long des routes ou des cours d'eau. On considère comme *forêt* toute zone de végétation dont la plus petite largeur locale est supérieure à 30 m en ayant exclu au préalable les éventuelles haies attachées à la zone. Un *bosquet* est une zone de végétation petite et plutôt arrondie ou compacte. Le seuil de taille utilisé, empirique lui aussi est le suivant : la plus petite dimension doit être inférieure à 30 mètres (par exemple un cercle de moins de 30 mètres de diamètre). Les bosquets correspondent à des petits groupes d'arbres ou d'arbustes isolés.

2.3. Modélisation multi-représentation de la végétation

Au vu des besoins exprimés au paragraphe précédent et des processus qui vont être mis en oeuvre lors de la dérivation du fond vert commun, représenter la végétation dans une seule classe d'objet n'est plus très pertinent : nous souhaitons distinguer les classes "forêt", "haie" ou "bosquet". De plus, pour pouvoir stocker les différentes représentations d'un même objet comme par exemple les haies surfaciques et les haies linéaires, nous avons décidé de représenter et stocker les données de végétation dans un schéma de données multi-représentation enrichi (Figure 4). Ce schéma est utile pour manipuler les données durant les différents processus de dérivation qui seront présentés dans les paragraphes suivants mais aussi peut-être plus tard pour des utilisations éventuelles du fond vert commun comme la généralisation de base de données qui consiste à dériver une base de données à partir d'une base de données plus détaillée (Weibel *et al.*, 1999).

Le formalisme utilisé dans la figure est celui de (Vangenot *et al.*, 2002), une extension à la représentation multiple du modèle MADS (Parent *et al.*, 1998). On reprend la notation de (Vangenot *et al.*, 2002) mais on l'applique au diagramme de classes d'UML et non au modèle "entité-relation". Le schéma prend en effet en compte plusieurs niveaux de représentation : le fond vert brut issu de la saisie initiale (classes étiquetées SI), le fond vert commun (FV) ou encore les applications IGN comme les haies linéarisées (IGN) et IFN (IFN). Cette multi-représentation est effective par la conservation des liens de généralisation comme le lien entre les classes "Tache végétation" et "Fond vert brut" qui est un lien de correspondance entre points de vue (notation (\Rightarrow) sur le schéma), ici SI et FV. De plus, ce schéma permet une explicitation des relations topologiques entre les différents éléments de la végétation. Ainsi, il existe un lien "est voisin par le réseau de" entre les objets de la classe "Forêt" une fois le découpage par le réseau effectué ou encore un lien "prolonge" entre les "Haies" et les "Forêts" auxquelles elles sont accolées.

Enfin, cette modélisation répond au besoin essentiel pour la généralisation de la reconnaissance de structures et patterns (Mackaness *et al.*, 2002). Des structures simples sont détectées comme les haies ou les bosquets par exemple qui permettent à la fois d'expliciter des informations implicites contenues dans les données géographiques et de faciliter le processus de généralisation. De plus, des structures

complexes issues d'agrégation peuvent également être reconnues comme les haies complexes qui sont des alignements de haies proches séparées par de petits espaces.

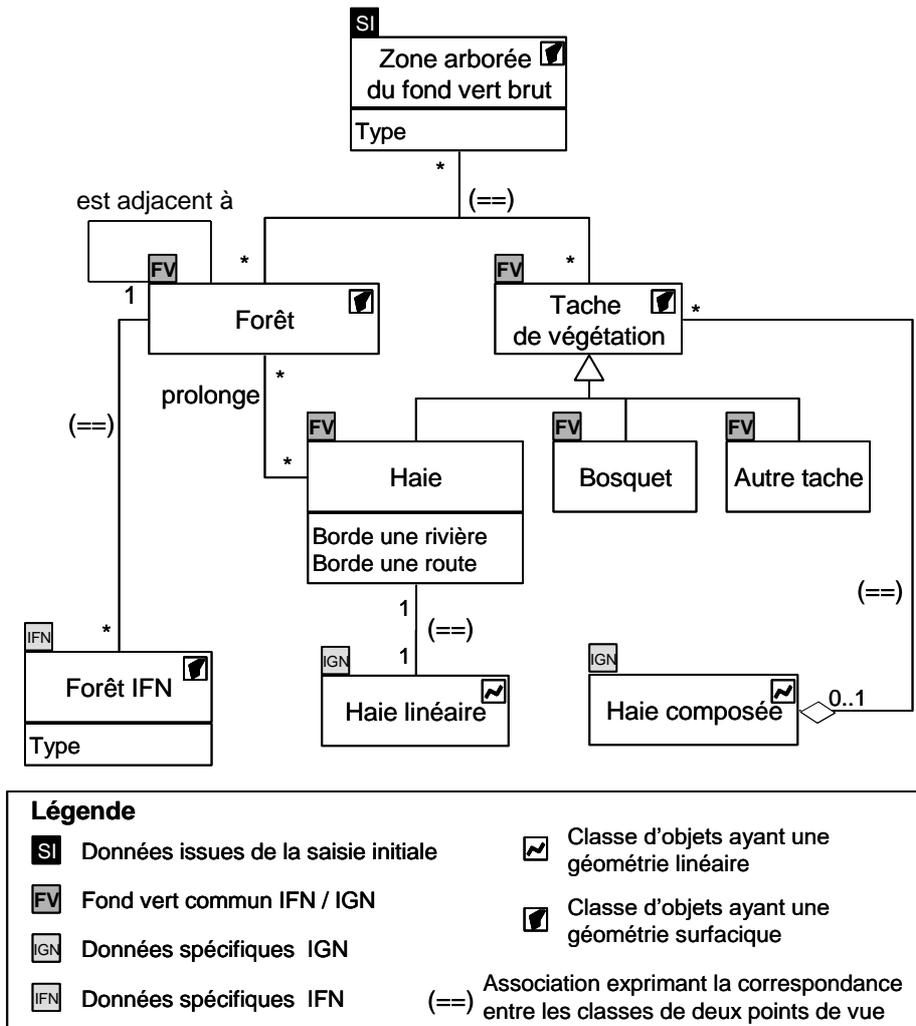


Figure 4. Schéma UML adapté du modèle enrichi et multi-représentation de la végétation pour la création et la gestion du fond vert commun.

Ce schéma répond donc aux exigences des traitements de généralisation et d'intégration de base de données géographiques (Mustière *et al.*, 2007) : modélisation orientée-objet contenant des classes explicites pour chaque niveau de représentation, stockage explicite des relations entre objets par des liens entre classes, explicitation

de la topologie et des structures composées comme les "haies composées" dans notre schéma.

3. Un processus de création du fond vert commun

Nous présentons dans cette partie un processus de création du fond vert commun en trois étapes pour répondre aux besoins identifiés dans la partie précédente : la première étape est le découpage et la classification des données brutes en forêts, haies et bosquets; la deuxième est la simplification du niveau de détail des forêts; et la troisième est le recalage et le découpage des forêts sur les réseaux. Ce processus a été développé sur un jeu de données 100 km² localisé dans le Calvados pour son paysage de bocage complexe à traiter. Le processus a ensuite été entièrement testé sur deux autres zones de la même taille mais aux paysages différents dans la Meuse et le Rhône. Nous commençons cette partie par un état de l'art des recherches en généralisation des zones de végétation.

3.1. Etat de l'art

La généralisation des zones de végétations est un sujet beaucoup moins traité que celle des routes ou des bâtiments qui constituent généralement les thèmes principaux d'une carte topographique. Il existe néanmoins quelques travaux sur le sujet. (Revell, 2005) est confronté à des zones de végétations membres d'une partition détaillée de l'espace en occupation du sol. Les différentes zones de forêt sont donc séparées par de fins polygones de routes ou de cours d'eau. Les zones proches sont regroupées par classification puis agrégées par des techniques de reposant sur la triangulation de Delaunay. (Revell, 2007) propose une méthode complémentaire pour généraliser la végétation en fonction de règles de proximité sémantique : on agrège les différents types de conifères quand ils sont adjacents.

(Mackaness *et al*, 2007) est confronté à un problème différent à savoir représenter des forêts sur des cartes au 1/250 000 à partir de données équivalentes à notre fond vert brut ce qui représente un très fort saut d'échelle. Il utilise ainsi des techniques morphologiques comme la dilatation et l'érosion pour simplifier fortement les zones de forêts principales et éliminer les plus petites.

Enfin, (Touya *et al*, 2006) présente un processus de généralisation de base de données pour passer automatiquement de la BD Topo® de l'IGN (donc du fond vert brut pour la végétation) à la BD Carto® de l'IGN, une base de données de résolution 10m. Le processus contient des phases de regroupement et de simplification importants.

Les travaux présentés ci-dessus sont assez spécifiques pour des problèmes données (type des données, résolution de départ, résolution d'arrivée) et aucun ne nous semble adapté à la problématique du fond vert commun présentée dans la partie 2. Nous avons donc décidé de développer un nouveau processus que nous détaillons

dans la suite de la partie 3 mais celui-ci reprend par moment des idées développées dans ces quatre travaux.

3.2. Classification du fond vert brut

Pour passer du fond vert brut au fond vert commun, la première étape est la classification du fond vert brut. Il s'agit de différencier les différents objets ou parties d'objets du fond vert brut en quatre nouvelles classes : forêts, haies, bosquets et autres. En se basant sur la définition des haies, des bosquets et des forêts de la partie 2.2, nous repérons les éléments petits ou fins du fond vert brut en segmentant les forêts. Une haie peut être isolée et constituer à elle seule un objet du fond vert brut ou être accolée à une forêt c'est-à-dire correspondre à une partie d'un objet du fond vert brut (Figure 3). Les bosquets sont toujours des objets à part entière du fond vert brut. Les grands éléments restants seront classifiés comme forêts. La classe autre est réservée aux éléments résidus du processus de classification. Elle est gardée car elle peut être utile pour la construction de haies complexes et de forêts ouvertes par exemple.

La détection des éléments petits ou fins se fait grâce à une transformation de morphologie mathématique, l'ouverture (Su *et al*, 1997). L'ouverture consiste en une érosion de l'objet de départ (un buffer "intérieur") puis une dilatation (buffer "extérieur") de même taille (Figure 5). L'ouverture a pour but de faire disparaître les petites structures et détails fins, enlevés par l'érosion et qui ne seront pas recréés par la dilatation. L'adaptation de ces opérateurs ensemblistes à des surfaces vecteurs revient à utiliser des buffers. L'érosion est la soustraction d'un buffer du contour à la surface initiale soit un "buffer intérieur". La dilatation est l'union de ce même buffer avec la surface initiale soit un "buffer extérieur". L'ouverture d'une surface est toujours incluse dans la surface initiale et elle élimine les protubérances petites ou fines. Un seuil de 15 m a été utilisé pour l'ouverture ce qui a pour effet d'éliminer de la surface des haies qui y étaient attachées car elle ont par définition moins de 25 m de largeur. On peut voir un exemple de cet effet sur les haies sur la Figure 5.

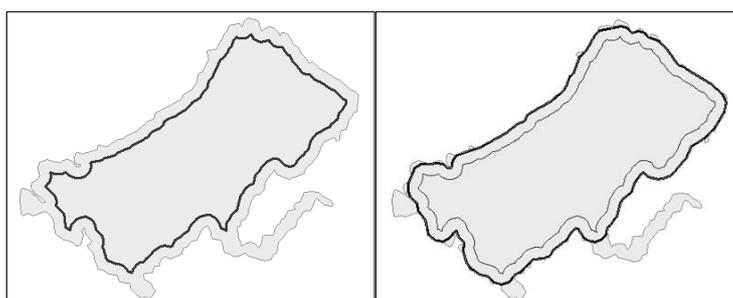


Figure 5. A gauche, étape 1 : érosion de la surface. A droite, étape 2 : dilatation de la surface érodée.

Les parties des forêts conservées par l'ouverture sont les véritables forêts du fond vert commun (elles appartiennent à la classe "forêt" de la Figure 4). Les parties supprimées par l'ouverture, que l'on obtient par soustraction géométrique du fond vert brut moins le résultat de l'ouverture, sont les haies, bosquets et autres taches de végétation qui doivent alors être classifiées. Un bosquet sera un élément arrondi ou compact, de petite taille et non attaché à une forêt. Une haie est un élément long et fin, attaché ou non à une forêt. Les éléments restants sont considérés comme les "autres taches". Nous cherchons donc à isoler les formes allongées et fines puis les formes compactes des formes restantes. Pour déterminer si les éléments sont fins et allongés, plusieurs mesures déjà utilisées pour la généralisation des bâtiments (AGENT, 1999) sont combinées. La Figure 6 présente la séquence de mesures appliquée avec les seuils choisis.

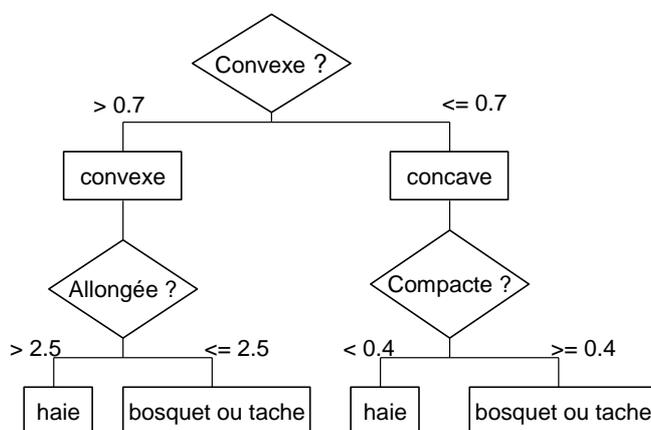


Figure 6. Schéma de séquence pour la classification des haies par les mesures de taux de convexité, d'allongement et de compacité.

La première mesure [1] est le taux de convexité du polygone. Elle est utile car nous ne mesurons pas l'allongement et la finesse d'un polygone convexe et d'un polygone concave de la même manière. Expérimentalement, nous considérons que la surface est convexe lorsque son taux de convexité est au-dessus de 0.7. En dessous, nous considérons qu'elle est significativement concave.

$$\text{taux convexité} = \frac{\text{Aire}}{\text{AireEnv}} \text{ où AireEnv} = \text{aire de l'enveloppe convexe} \quad [1]$$

La deuxième mesure concerne l'allongement du polygone [2]. Cette mesure utilise la notion de Rectangle Englobant d'Aire Minimale communément appelé par son acronyme anglais MBR. Le MBR est le plus petit rectangle qui englobe complètement la surface. Cette mesure permet de discriminer les haies parmi les surfaces éliminées après ouverture. Un seuil de 2.5 est utilisé : quand la longueur du

MBR est plus de 2,5 fois plus grande que la largeur, on considère que la surface est allongée (et fine car convexe et éliminée par ouverture).

$$\text{élongation} = \frac{L}{l} \text{ avec } L = \text{longueur du MBR et } l = \text{largeur du MBR} \quad [2]$$

Pour les taches de végétation concaves, l'élongation du MBR n'est pas une mesure efficace. Nous lui préférons dans ce cas l'utilisation de la compacité qui mesure à quel point une surface est compacte (un cercle est très compact). Nous utilisons comme formule de compacité l'indice de Miller [3] qui oscille entre 0 et 1. Il vaut 1 pour un cercle qui est la figure géométrique la plus compacte et tend vers 0 quand l'objet se rapproche d'une forme « linéaire ». Dans notre cas, les taches de végétation seront considérées comme haies si elles sont très peu compactes et sinon comme bosquets ou autres taches suivant que la tache est attachée à une forêt ou non. Le seuil de compacité utilisé est 0.4 : en dessous, les surfaces sont considérées comme non compacte donc fines et allongées.

$$\text{compacité} = 4\pi \times \frac{\text{Aire}}{\text{Périmètre}^2} \text{ où compacité} = 1 \text{ pour un cercle} \quad [3]$$

De plus, dans le cas où l'objet mesuré issu de la soustraction est accolé à une véritable forêt, un seuil de taille est utilisé (400 m² soit 20m sur 20m). En deçà de ce seuil, les objets ne peuvent être considérés comme des haies et quelques soient les résultats des trois mesures ci-dessus, ces objets sont considérés comme des petits artefacts de l'ouverture et sont placés dans la classe "autres taches". Enfin, en reprenant la séquence de la Figure 6, les bosquets sont distingués des autres tâches par la mesure de compacité avec un seuil de 0.7 (au-dessus, les surfaces sont des bosquets). Les seuils utilisés ont été déterminés de manière empirique après expérimentation et les résultats obtenus sont satisfaisants (Figure 7) sur l'ensemble de la zone test mais surtout sur les deux autres zones aux paysages différents.



Figure 7. Un résultat de classification dans le bocage normand : en gris, les forêts; en gris foncé, les haies et en noir les bosquets. Les autres taches sont ici éliminées.

3.3. Simplification du niveau de détail

Le niveau de détail du contour des forêts doit être légèrement simplifié pour répondre aux besoins IGN. Le fond vert brut est issu d'une segmentation d'image avec un pixel de deux mètres ce qui donne des contours à la fois trop détaillés et anguleux. Le processus de classification par ouverture a permis une première simplification du contour en gommant des petites irrégularités (Figure 5). La méthode retenue pour simplifier le niveau de détail des forêts est un filtrage de Douglas & Peucker (Douglas *et al*, 1973). Cet algorithme a l'avantage d'être rapide même pour de grands polygones comme peuvent l'être les forêts issues de la classification. La Figure 8 montre un résultat obtenu pour un pas de 5m.

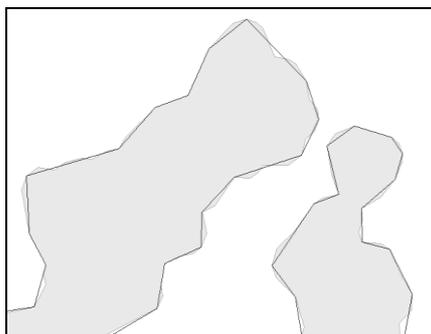


Figure 8. Des forêts avant et après filtrage de Douglas & Peucker à 5m.

3.4.Recalage sur les réseaux

La relation topologique "une forêt est au bord d'une route" est souvent présente sur le terrain car aujourd'hui les forêts sont principalement artificielles et délimitées par les réseaux de transport. Le fond vert brut étant issu d'une détection automatique sur des ortho-images indépendamment des réseaux RGE, cette relation n'a pu être conservée durant ce processus et il va falloir la recréer. Nous considérons que lorsque le contour d'une forêt est à moins de 10 mètres d'un tronçon de réseau (route, chemin, cours d'eau ou voie ferrée), la forêt borde la portion de réseau. La géométrie de la forêt est alors modifiée pour forcer un partage de géométrie avec la portion de réseau. Pour réaliser cette transformation (Figure 9), on utilise les fonctionnalités topologiques du logiciel Radius Clarity™, sur lequel est basée la plate-forme de développement en généralisation du COGIT. Les fonctions permettent d'assigner des seuils de partage topologique entre objets appartenant à des classes différentes.

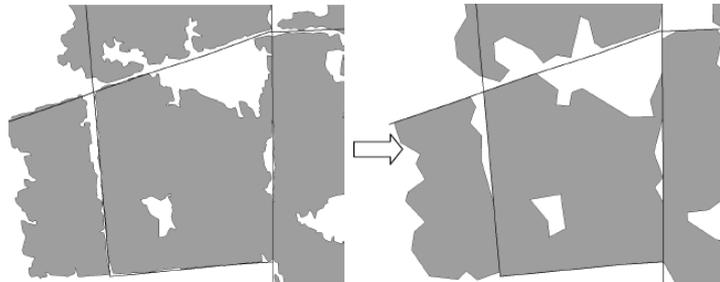


Figure 9. Exemple sur plusieurs forêts de l'application de la simplification et du recalage sur les réseaux.

Une fois ce recalage automatique effectué, une dernière transformation est effectuée pour obtenir le fond vert commun : le découpage des forêts selon les réseaux. Ce découpage a deux objectifs : tout d'abord il permet de réduire la taille des objets forêts dans la base de données ce qui permet de faciliter leur future manipulation; il permet ensuite de détecter des haies supplémentaires, en l'occurrence les rangées d'arbres le long d'une route longeant une forêt (voir Figure 10). Le découpage se fait grâce aux outils d'intersection géométrique présents dans tout SIG. La détection des rangées d'arbres se fait en utilisant le même processus que lors de la classification (partie 3.1), mais seules les parties découpées mesurées comme très allongées sont analysées et mesurées dans ce deuxième processus (mesure d'allongement supérieure à 4).

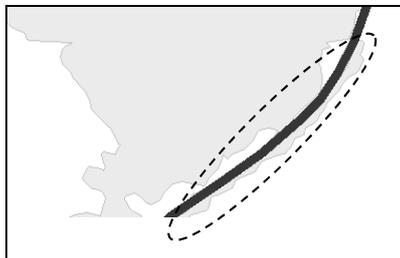


Figure 10. La partie entourée de la forêt de l'autre côté de la route est une rangée d'arbre que l'on doit séparer de la forêt par découpage et classer dans les haies.

Cette dernière étape permet d'obtenir les classes d'objets définitives contenues dans le fond vert commun à savoir les forêts (simplifiées, recalées et découpées), les haies et les bosquets. Les tâches de végétations "autres" peuvent être considérées comme une couche optionnelle du fond vert commun qui pourrait être utilisée au besoin dans certains traitements spécifiques. La Figure 11 est une illustration du fond vert commun obtenu automatiquement partir du fond vert brut.



Figure 11. A gauche le fond vert brut et à droite le fond vert commun avec les forêts en gris clair, les haies en gris foncé et les bosquets en noir.

3.5. Adaptation des traitements à un fond brut thématique

Pour faciliter la thématique du fond vert, il faut prévoir des algorithmes traitant correctement les données segmentées sans altérer leur segmentation. Ce que nous appelons ici données segmentées sont des surfaces partitionnées constituées de plusieurs surfaces adjacentes qui doivent le rester. Par exemple, une forêt peut être

segmentée en une surface de conifères et une surface de feuillus adjacentes. En particulier, deux algorithmes sont à modifier pour qu'ils ne génèrent pas d'erreurs : l'ouverture et le filtrage de Douglas & Peucker. En particulier l'ouverture génère des erreurs quand elle est appliquée séparément sur des zones adjacentes constituant un agrégat (Figure 12). Comme exposé dans la partie 3.1, l'ouverture a pour propriétés d'éliminer les petites protubérances d'une surface. Quand une protubérance est située au niveau d'une segmentation (situation 1 de la figure), un creux est au contraire créé par l'ouverture. Autre propriété de l'ouverture, les petites surfaces (celles dont la largeur maximum est inférieure à 2 fois le seuil d'ouverture) sont éliminées. Quand le découpage crée ce type de surfaces entourées de plus grandes (situation 2 de la figure), l'ouverture simple les fait disparaître. Enfin, si une segmentation génère des surfaces en bord de forêt avec des parties fines (situation 3 de la figure), l'ouverture simple va les éliminer abusivement. Il apparaît donc indispensable de ne pas considérer les contours partagés avec d'autres forêts dans le calcul de l'ouverture.

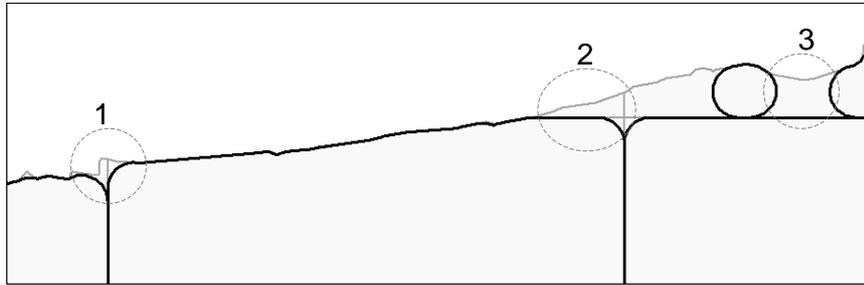


Figure 12. Une zone de forêt segmentée en plusieurs parties (limites grises). L'ouverture simple réalise l'ouverture sur chaque zone découpée indépendamment ce qui génère différents problèmes (cercles 1, 2 et 3).

Pour éviter ces problèmes, l'ouverture appliquée à une zone doit prendre en compte les zones voisines : on réalise donc l'ouverture sur l'union spatiale des zones. Puis, on retrouve ensuite les frontières initiales en réalisant l'intersection du résultat de l'ouverture avec les zones initiales. On utilise une des propriétés mathématiques de l'ouverture qui veut que l'ouverture d'un ensemble est forcément incluse dans l'ensemble de départ. La Figure 13 illustre le résultat obtenu avec cette méthode. Le même procédé est utilisé pour modifier le filtrage de Douglas et Peucker.

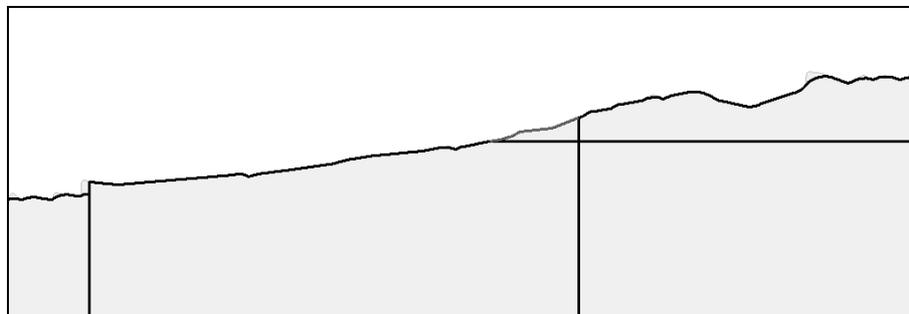


Figure 13. Résultat de la méthode d'ouverture (en gras) pour des données segmentées sur la même zone que la Figure 11 : les problèmes identifiés sont réglés.

4. Dérivation de produits IFN

La création de la nouvelle couche forestière de l'IFN nécessitera une saisie thématique complémentaire pour ajouter de nombreuses informations attributaires au fond vert commun. Cet aspect étant interactif (réalisé par des opérateurs sur un SIG), il n'a pas été pris en compte dans cette étude. Nous avons néanmoins envisagé quelques autres traitements du fond vert commun qui pourraient intéresser l'IFN. Nous avons proposé la possibilité d'un appariement automatique entre les actuelles données IFN et le fond vert commun pour faciliter la récupération des données actuelles IFN lors de la réfection prévue de ces données. L'appariement est un processus nécessaire à l'intégration de base de données hétérogènes (Sheeren *et al*, 2004) qui consiste à mettre en correspondance les éléments de différentes bases de données qui représentent la même entité du monde réel. Ici, ce sont les contours des zones de forêts qui vont être appariés (voir Figure 14). Un algorithme d'appariement d'objets linéaires est utilisé (Devogele 1997, Mustière 2002, Mustière 2006).

Si une forêt présente dans la couche IFN actuelle correspond à une forêt du fond vert commun nouveau, leurs contours linéaires sont appariés (un lien entre objets est instancié dans une base de données multi-représentation). Cela peut permettre de transférer automatiquement l'information attributaire thématique contenue dans la couche IFN actuelle par jointure dans le fond vert commun.

L'étape supplémentaire est de recalibrer les contours appariés de la couche IFN actuelle sur les contours du fond vert commun pour obtenir une cohérence géométrique entre les deux couches (voir Figure 14). Le recalage est fait dans ce sens car le fond vert commun est plus précis en terme de granularité et de positionnement planimétrique.

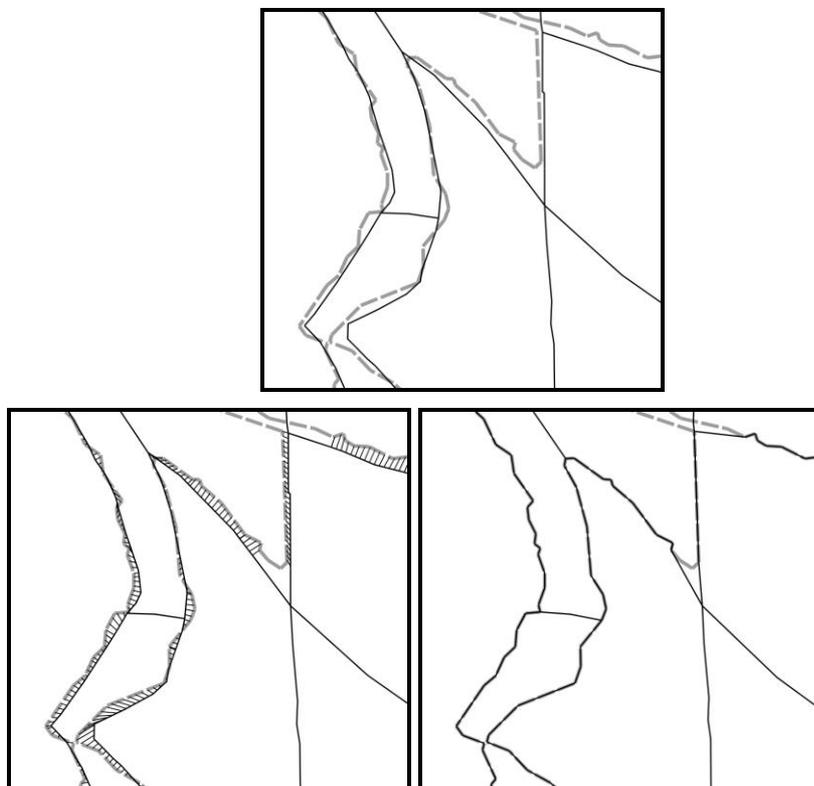


Figure 14. Données de l'IFN (trait continu noir) et du fond vert commun (pointillés gris), superposées (en haut), appariées (à gauche), puis recalées (à droite).

5. Dérivation de produits IGN

5.1. Linéarisation des haies

Comme cela a été expliqué dans la partie 2.2, un des besoins exprimés par l'IGN pour la production de la couche végétation de sa base de donnée topographique de référence, le RGE (Référentiel Grande Echelle), est la linéarisation ou squelettisation des haies.

La première étape de ce processus de linéarisation est de simplifier fortement la géométrie des haies pour accélérer le calcul de linéarisation et pour éviter les artefacts dans le squelette dus à un contour trop granuleux. Dans ce cas, l'algorithme le plus approprié semble être Whirlpool (Dougenik, 1980) car il est rapide et fournit des lignes plus lisses que Douglas & Peucker, utilisé précédemment, ce qui donne de

meilleurs résultats pour la squelettisation qui suit. Cet algorithme consiste à regrouper les points du contour d'un polygone qui sont proches les uns des autres et à remplacer chacun des clusters par le centroïde des points qui le constituent (voir Figure 15). Un test de contrôle de cohérence topologique est effectué en suivant (contour de la surface qui s'auto-intersecte). En cas d'échec à ce test, la surface est simplifiée avec d'autres paramètres jusqu'à obtenir une surface cohérente.

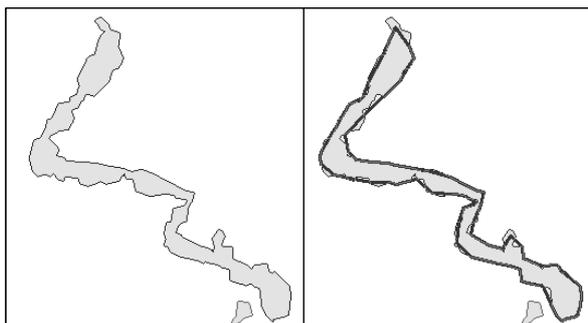


Figure 15. Une haie issue du fond vert commun à gauche avec son contour simplifié par whirlpool à droite.

Une fois la haie simplifiée, la linéarisation est effectuée grâce à un algorithme de squelettisation dont l'intérêt en généralisation a été montré par (Haunert *et al*, 2004). Le principe est d'effectuer des rétractions itératives du polygone jusqu'à ce que les bords se touchent et forment un squelette (Felkel *et al*, 1998). Cet algorithme a été implémenté au laboratoire dans le SIG Radius Clarity™ (Mangeat, 2005) et permet d'obtenir des squelettes plus réalistes pour des polygones complexes que les méthodes utilisant une triangulation de Delaunay. Le squelette ainsi obtenu est ensuite élagué pour ne garder que les branches significatives puis il est traité pour être transformé en graphe planaire et enfin il est simplifié par l'algorithme de Douglas & Peucker. La Figure 16 montre un exemple de résultat obtenu par cette méthode.

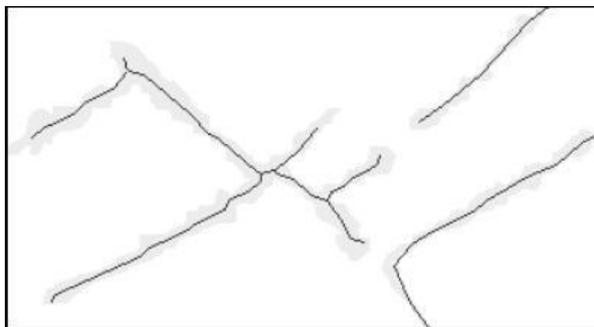


Figure 16. Exemples de haies squelettisées par l'algorithme du straight skeleton.

5.2. Applications cartographiques du fond vert commun

En plus des traitements nécessaires pour obtenir la couche végétation du RGE, des traitements automatiques supplémentaires sont nécessaires pour les applications cartographiques comme la production des cartes au 1:25000 et 1:50000 par généralisation automatique (Lecordix *et al*, 2006). Pour une meilleure représentation cartographique, la chaîne de production contient une phase de recalage plus forte que celle du fond vert commun : le partage de géométrie avec les réseaux est forcé pour des seuils plus importants. Pour réaliser cela, on utilise un processus d'appariement équivalent à celui décrit au paragraphe 4. On apparie le fond vert commun avec les réseaux (routes, cours d'eau, voies ferrées) à l'aide de l'algorithme d'appariement d'objets linéaires (Devogele 1997, Mustière 2002, Mustière 2006) et on recale la géométrie des forêts appariées sur la géométrie de leur homologue dans les réseaux. Les résultats sont très intéressants (Figure 17) et ce processus pourrait même être utilisé pour la création du fond vert commun en lieu et place du recalage topologique décrit au paragraphe 3.3.

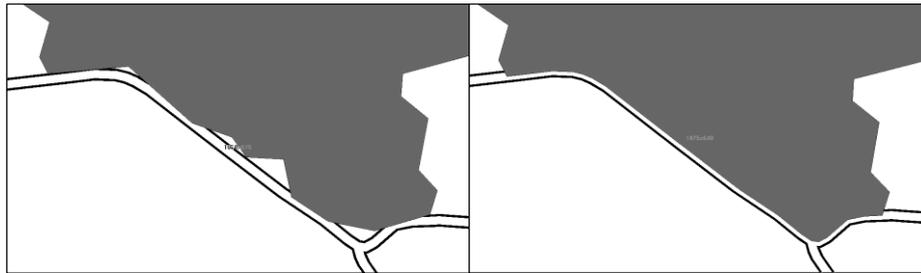


Figure 17. Recalage par appariement des forêts sur les réseaux pour la production de la carte au 1/25000 de l'IGN.

6. Conclusion et perspectives

Nous avons présenté une méthode pour intégrer les couches de végétation de l'IGN et l'IFN d'un point de vue technique dans le cadre de la mutualisation des efforts de production des deux instituts. Elle consiste en une modélisation et une création par généralisation à partir d'une couche très détaillée, le fond vert brut, d'un fond vert commun répondant aux besoins communs des deux instituts. Le processus proposé comprend également une phase de traitements (généralisation et appariement) pour répondre aux besoins de base de données et cartographiques spécifiques de chaque institut.

Des recherches supplémentaires pourraient être effectuées pour améliorer ou enrichir le processus. La Figure 4 prévoyait l'enrichissement du fond vert commun par des classes de "Haies complexes" et de "Forêts ouvertes". N'étant pas

prioritaires, ces enrichissements n'ont pas été testés. Par exemple, en ce qui concerne les haies complexes, des alignements de haies proches pourraient être repérés en utilisant un algorithme de calcul d'Arbre de Recouvrement Minimal (Regnauld *et al*, 2006). En outre, nous avons proposé une solution pour un besoin spécifique, notamment pour la généralisation de la végétation pour des échelles relativement grandes. Si on souhaitait étendre l'étude à la dérivation d'une couche forestière à petite échelle (1:250k par exemple), l'approche de (Mackaness *et al*, 2007) qui propose une méthode pour ce type de problème pourrait être utilisée ou intégrée à notre approche.

L'étude présentée dans cet article est une étude de faisabilité technique avec les limites que cela suppose. Les algorithmes développés ne sont pas tous complètement robustes et le processus n'a pas été conçu pour gérer efficacement les erreurs éventuelles. De plus, si les tests ont été effectués sur trois zones aux paysages différents, aucune mise à l'échelle n'a été évaluée pour pouvoir traiter des zones de travail plus importantes (de l'ordre d'un département français). Depuis, ces problèmes ont été traités dans une phase d'industrialisation du processus et celui-ci est passé en production.

Cette étude prouve qu'en utilisant des algorithmes d'appariement et de généralisation, il est dorénavant techniquement possible d'une part de croiser les données de producteurs différents afin de les enrichir mutuellement et d'autre part d'élargir la diversité des produits dérivés des bases de données par généralisation. De plus, ces processus permettent à des organismes ou services producteurs de données thématiques de transférer leurs informations thématiques de bases de données moyenne échelle (du type BD Carto® de l'IGN) à des bases de données plus précises (du type BD Topo® de l'IGN) sans perdre leur information et en minimisant le temps de transaction.

Bibliographie

- AGENT project, Selection of Basic Measures. Rapport technique du projet AGENT. Documentation en ligne disponible à l'URL : <http://agent.ign.fr/deliverable/DC1.pdf>. 1999.
- Devogele T., Processus d'intégration et d'appariement des bases de données géographiques. Application à une base de données routière multi-échelles. Thèse de l'Université de Versailles, IGN/Laboratoire COGIT. 1997.
- Dougenik J., « WHIRLPOOL: A geometric processor for polygon coverage data ». In *proceedings of Auto-Carto 4* (ACSM-ASPRS), 1980, p. 304-311
- Douglas D.H., Peucker T.K., « Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature », *The Canadian Cartographer journal*, 10(2), 1973 p.112-122.

- Felkel P., Obdrzalek S., « Straight Skeleton Implementation », *In proceedings of 14th Spring Conference on Computer Graphics*, 1998. p. 210-218.
- Guigues L., Le Men H., Cocquerez J.-P.. « Analyse et représentation ensembles-échelle d'une image ». *In Proc. 19th GRETSI Symposium on Signal and Image Processing*, Paris, France, September 2003.
- Hauert J.-H., Sester M., « Using the Straight Skeleton for Generalisation in a Multiple Representation Environment ». *In proceedings of the 6th CA Workshop on progress in automated map generalisation*, Leicester, 2004.
- Lecordix F. , Le Gallic J.-M., Gondol L, « Clarity™ experimentation for cartographic generalisation in production ». *In proceedings of ICA Workshop on Generalisation and Multiple Representation 2006*, 25 June 2006, Portland, USA.
- Mackaness W., Edwards, G., « The Importance of Modelling Pattern and Structure in Automated Map Generalisation ». *In Proceedings of Joint Workshop on Multi-scale Representations of Spatial Data*, Ottawa, Canada , 2002.
- Mackaness W., Perikleous S., Chaudhry O., « Representing Forested Regions at Small Scales : Automatic Derivation from the Very Large Scale». *In proceedings of the Geographical Information Science Research UK Conference (GISRUK'2007)*, Maynooth, Ireland, 2007.
- Mangeat C., Codage, adaptation et comparaison d'algorithmes de squelettisation d'une surface sur le SIG LAMPS2. Rapport de stage de DESS de Cartographie et SIG, Université Paris 1. Laboratoire COGIT, 2005.
- Mustière S., Description des processus d'appariement mis en oeuvre au COGIT. Rapport technique IGN, SR/2002.0072. 2002.
- Mustière S., « Results on experiments on automated matching of networks ». *In proceedings of the ISPRS Workshop on Multiple Representation and Interoperability of Spatial Data*, Hannover, 2006. p. 92-100.
- Mustière S., Van Smaalen J., « Databases Requirements for Generalisation and Multiple Representations ». Mackaness W., Ruas A., Sarjakoski T. (eds) : *The Generalisation of Geographic Information : Models and Applications*. Elsevier 2007.
- Parent C., Spaccapietra S., Zimanyi E., Donini P., Plazanet C., Vangenot C., Modelling Spatial Sata in the MADS Conceptual Model. *Proceedings of the 8th International Symposium on Spatial Data handling*, Vancouver, Canada, 1998, p. 138-150.
- Regnauld N., Mackaness W. « Creating a Hydrographic Network from its Cartographic Representation : A Case Study Using Ordnance Survey MasterMap Data ». *International Journal of Geographical Information Science (IJGIS)*. Vol. 20, No. 6, July 2006, p. 611-631.
- Revell, P., « Seeing the Wood from the Trees: Generalising OS MasterMap® Tree coverage Polygons to Woodland at 1:50 000 Scale ». *In proceedings of 8th ICA Workshop on Generalisation and Multiple Representation*, La Coruña (Spain), 2005.
- Revell, P., « Generic Tools For Generalising Ordnance Survey Base Scale Landcover Data ». *In proceedings of 10th ICA Workshop on Generalisation and Multiple Representation*. Moscow (Russia). 2007.

- Sheeren D., Mustiere S., Zucker J-D., « How to Integrate Heterogeneous Spatial Databases in a Consistent Way? » *Conference on Advanced Databases and Information Systems (ADBIS)*, Budapest, septembre 2004, paru en tant que LNCS 3255, Benezur A., Demetrovics J. et Gottlob G. (eds), p.364-378, Springer
- Su B., Li Z., Lodwick G., and Müller J.-C., « Algebraic models for the aggregation of area features based upon morphological operators ». *International Journal of Geographical Information Science (IJGIS)*, Vol. 11, 1997, p. 233-246.
- Touya G., Grosso E., Mustière S., Duchêne C., « Generalization or Data Matching for Integration? Insights from Experiments on Databases at Different Levels of Details ». *In Proceedings of the 4th International Conference on Geographic Information Science (GIScience'2006) – Extended Abstracts*. Raubal, Miller, Frank and Goodchild (eds), IFGI Prints, vol.28, pp.373-376.
- Trias-Sanz R., Mise à jour automatique de données d'occupation du sol à grande échelle par des méthodes d'analyse d'image. Thèse de doctorat de l'Université Paris V, René Descartes, mars 2006.
- Vangenot, C., Parent, C., Spaccapietra, S., « Modelling and manipulating multiple representations of spatial data ». *Proceedings of the 10th International Symposium on Spatial Data Handling*, 2002, p. 81-93.
- Weibel, R., Dutton, G., « Generalising Spatial Data and Dealing with Multiple Representations ». *Geographical Information Systems, Vol. 1, Principles and Technical Issues*. 1999.p. 125-150