

CHAÎNE DE PRODUCTION DE LA CARTE DE LA VÉGÉTATION ARBOREE FRANÇAISE

TOUZET T.(1), HALBECQ X.(2), LECORDIX F.(2)

(1) INVENTAIRE FORESTIER NATIONAL, NOGENT-SUR-VERNISSON, FRANCE ; (2) INSTITUT GEOGRAPHIQUE NATIONAL, SAINT-MANDE, FRANCE

RÉSUMÉ

L'inventaire forestier national (IFN) et l'Institut géographique national (IGN) ont chacun dans leur mission la responsabilité de décrire, d'un point de vue géométrique, physique et thématique, le territoire national français, indépendamment de toute question de propriété.

L'IFN se concentre sur les ressources forestières et décrit plus de 16 millions d'hectares de forêt et 2 millions d'hectares d'espaces naturels dont les caractéristiques sont le morcellement et la diversité des espèces. L'IGN réalise le Référentiel à Grande Echelle (RGE®), qui fournit, sous forme d'une base de données vecteur, un descriptif géométrique et sémantique du territoire : réseaux routiers, ferrés et hydrographiques, bâtiments, surfaces hydrographiques, occupation du sol... En 2006, ces deux institutions ont signé un partenariat afin de partager leurs connaissances et leurs expériences respectives. L'IFN et l'IGN collaborent pour mettre en place et exploiter en commun un nouveau processus entièrement numérique permettant, en s'appuyant sur le RGE®, de collecter l'ensemble des types de formation végétale avec une grande précision et une nomenclature nationale détaillée basée sur la définition internationale de la forêt. La présentation s'attachera à décrire le processus mis en œuvre pour produire cette information géographique forestière de référence et à faire un point d'avancement sur la production de cette information qui devrait s'achever d'ici 2014. L'exploitation de ces données au sein des deux instituts sera aussi présentée : réalisation de la nouvelle cartographie numérique topographique de référence au 1:25 000 sur tout le territoire national à l'IGN et utilisation pour les études environnementales et pour les gestionnaires des espaces forestiers à l'IFN.

1 PRÉSENTATION DES DEUX INSTITUTS

1.1 L'inventaire forestier national

L'Inventaire forestier national (IFN) a été créé en 1958 par ordonnance signée par le Général de Gaulle, pour mieux connaître les potentialités des forêts françaises. Il est chargé de l'inventaire permanent des ressources forestières nationales, indépendamment de toute question de propriété (article R 521-1 du code forestier). L'IFN est devenu établissement public à caractère administratif sous tutelle du ministre chargé des forêts en 1994. Sa direction est implantée à Nogent-sur-Vernisson (Loiret).

L'IFN réalise les opérations d'inventaire sur l'ensemble du territoire français au moyen d'un échantillon constitué pour une période de dix ans, dont un dixième est traité chaque année. Les données relatives aux points sont recueillies à partir de la BD Ortho® de l'IGN ainsi que d'observations et de mesures sur le terrain. Ces informations dendrométriques, écologiques et floristiques sont enregistrées dans des bases de données et mises à la disposition du public.

L'IFN utilise également les référentielles images de l'IGN pour réaliser une carte forestière par département administratif qui a pour objectif la représentation spatiale de la forêt française mais aussi des espaces naturels et semi-naturels que sont les landes, les garrigues et maquis, les pelouses alpines, quelle que soit la propriété, forêt publique et forêt privée et qui se veut exhaustive pour qualifier tous les ensembles boisés de plus de 0,5 ha. Cette couche d'information améliore de façon très sensible la précision des résultats statistiques et apporte une information indispensable à toutes les études spatiales.

1.2 L'Institut géographique national

Créée en juin 1940 en remplacement du Service géographique des armées, l'Institut géographique national est un établissement public à caractère administratif sous la tutelle du Ministère de l'Ecologie, de l'Energie, du Développement Durable et de la Mer. Par décret modifié du 22 novembre 2004, "l'Institut géographique national a pour vocation de décrire, d'un point de vue géométrique et physique, la surface du territoire national et l'occupation de son sol, d'en faire toutes les représentations appropriées et de diffuser les informations correspondantes.

Dans ce cadre, l'institut est chargé des missions d'intérêt général suivantes :

...

b) Réaliser, renouveler périodiquement et diffuser la couverture photographique aérienne de l'ensemble du territoire national ;

c) *Constituer et mettre à jour sur l'ensemble du territoire les bases de données géographiques et les fonds cartographiques*"

...

L'IGN constitue notamment, sur l'ensemble du territoire national, le Référentiel à Grande Echelle (RGE®), qui est composé de 4 bases de données : la BD Adresse®, la BD Parcellaire®, la BD Ortho® et la BD Topo®. Cette dernière base de données, fournissant la description topographique du territoire comporte une couche de végétation incluant la description de l'information forestière.

L'IGN réalise et diffuse aussi la carte topographique au 1:25 000, sous forme papier ou sous forme numérique avec des dalles de scan.

Dans le cadre de ses accords avec l'Etat, signés dans le Contrat d'objectifs et de Moyens pour la période 2003 - 2006 et renouvelés dans le Contrat d'objectif de performance pour la période 2010 – 2013, l'IGN vise à développer et renforcer les partenariats entre les services de l'Etat pour la collecte et l'échange des informations géographiques. Cet objectif a incité l'IGN à signer un partenariat avec l'IFN pour la réalisation d'une couche commune de végétation forestière.

2 DESCRIPTION DES COUCHES D'INFORMATIONS

L'IFN s'intéresse aux formations végétales forestières et naturelles (landes et formations herbacées) pour constituer la carte forestière version 2 (produit IFN).

La couche végétation de la BD TOPO®(produit IGN- « F- Occupation du sol : végétation » version 2) représente plus largement la thématique de l'arbre, elle inclut la forêt mais également les autres éléments arborés que sont les bosquets et certains arbres isolés (regroupés dans la même classe bois dans la nomenclature BD TOPO®), les haies, les alignements, les vergers. Elle représente également les landes et les formations herbacées.

La représentation de la forêt s'appuie sur la définition désormais le plus largement admise au niveau international, fournie par la FAO (Food and Agriculture Organization) en 2004 dans son rapport sur l'évolution des ressources forestières dans le monde (Global Forest resources Assessment 2005, FRA2005): "*La forêt est un territoire occupant une superficie d'au moins 50 ares avec des arbres pouvant atteindre une hauteur supérieure à 5 m à maturité in situ., un couvert arboré de plus de 10 % et une largeur moyenne d'au moins 20 mètres. Elle n'inclut pas les terrains dont l'utilisation du sol prédominante est agricole ou urbaine*".

Les définitions des autres terrains arborés sont en cohérence avec celle de la forêt. Une haie est un terrain arboré, d'une largeur inférieure à 20 m et d'une longueur supérieure à 25m. Un bosquet est un terrain arboré d'une largeur inférieure à 20 m et d'une surface comprise entre 5 et 50 ares, il est composé d'au moins 4 arbres. Un arbre isolé ou un bouquet d'arbres est composé de moins de 4 arbres. Pour la BD Topo®, la surface minimale de représentation (surface des houppiers) est arrêtée à 80 m². Un verger est un terrain composé d'arbres fruitiers d'une surface supérieure à 50 ares.

À l'intérieur des zones supérieures à 50 ares sont cartographiés des types de formation végétale. Un type de formation végétale est un ensemble forestier ou semi-naturel (landes ligneuses, formations herbacées) continu ou discontinu, qui présente une certaine unité pour la couverture et l'utilisation du sol, pour la densité du couvert et pour la composition en essences.

La couche forestière repose sur l'estimation par le photo-interprète des seuils de densité de couvert des arbres. Les plus utilisés sont :

- le seuil de 10 % de couvert absolu des arbres, qui permet de séparer la forêt des autres types de formations végétales, notamment les landes ;
- le seuil de 40 % de couvert absolu des arbres, qui sépare la forêt ouverte et la forêt fermée ;

le seuil de 75 % de couvert libre relatif des arbres, qui détermine la notion de pureté d'un peuplement selon sa composition ou son essence. Le couvert libre correspond en photo-interprétation avec la vision du dessus du peuplement, à la surface des houppiers ayant accès à la lumière. Il est qualifié de relatif quand on détermine la proportion d'un sous-peuplement par rapport à un autre.

3 DESCRIPTION SYNTHÉTIQUE DU PROCESSUS

Depuis 2006, l'IFN et l'IGN collaborent pour mettre en place un processus de saisie de l'information forestière qui permette d'exploiter aux mieux les nouvelles techniques de collecte et de traitement de l'information proposées notamment par les laboratoires du Service de la Recherche de l'IGN. Après une phase d'étude expérimentale réalisée par les laboratoires COGIT et MATIS, un processus complet a été progressivement développé pour être désormais exploité pleinement.

A partir des orthophotographies numériques départementales en infrarouge couleur et du RGE®, a été mis en place un processus de cartographie et une chaîne de production entièrement numérique avec le développement en interne de plusieurs applications dédiées faisant appel à plusieurs logiciels, Geoconcept, Clarity et ArcGis Engine pour la collecte et les traitements automatisés et le couple PostgreSQL-PostGis pour le stockage.

Le processus de saisie de l'information forestière se déroule par département qui constitue l'unité de production des orthophotos sur laquelle s'appuie l'essentiel du travail de collecte de l'information. Ce travail est réparti entre les deux instituts qui collaborent étroitement pour apporter chacun leur compétence respective pour dérouler le processus complet.

Le processus comporte les 6 phases suivantes, réparties dans les deux établissements: (Figure 1)

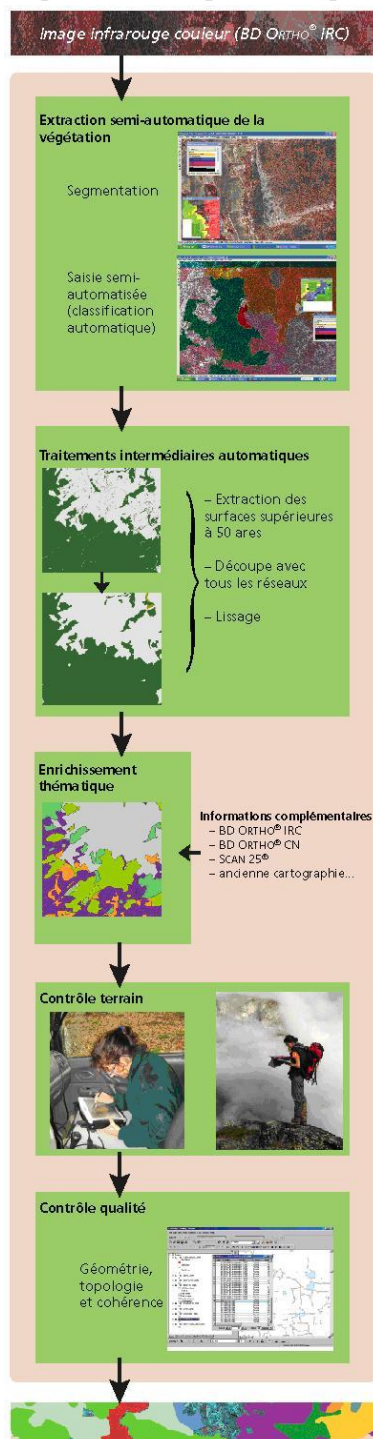


figure 1 – Processus de réalisation de la couche forestière

3.1 Préparation des données – extraction automatique des limites de végétation

A partir des orthophotos IRC, une segmentation automatique des images est réalisée avec trois niveaux de découpage, suivant une technique issue des recherches du laboratoire MATIS de l'IGN (Guigues, 2007). La méthode de segmentation multi-échelles d'images est utilisée pour individualiser les objets allant de l'arbre isolé aux massifs forestiers de grande surface. Ce processus permet de délimiter des zones relativement homogènes en intensité, en texture ou en couleur. Trois échelles significatives de description sont utilisées : une échelle à laquelle les arbres isolés sont individualisés, une échelle moyenne (haies, bosquets...) et une grande échelle à laquelle les massifs forestiers sont agglomérés. Cette grande échelle permet également des généralisations de zones de végétation clairsemée comme les forêts ouvertes ou les landes. À partir de l'image segmentée, il est alors possible de sélectionner les entités correspondant à la couche arborée sans avoir à redessiner les contours, la sélection s'appuyant sur les contours de la segmentation. Ce processus automatique pour chaque département, divisé en de nombreuses dalles de 1 km sur 1 km (en moyenne 6000 dalles), nécessite de nombreuses heures de calcul qui peuvent être distribuées sur plusieurs processeurs pour réduire les délais.

3.2 Saisie 1^{er} niveau

A partir des orthophotos et du découpage par segmentation automatique fournis par l'IGN, les photo-interprètes de l'IFN vont saisir une première classification thématique pour les zones de végétation hors zones urbaines. Les opérateurs de l'IGN compléteront le travail sur les zones urbaines. La saisie est réalisée avec le logiciel SEVE® (Système d'Extraction de la VEgétation) développé par l'IGN. Il permet l'accès à un grand nombre d'informations indispensables à la photo-interprétation par un affichage rapide de plusieurs couches images de référence sur l'ensemble d'un département : orthophotos en infrarouge couleurs et en couleurs naturelles, scan25® (carte topographique numérique au 1/25 000), couche forestière (version ancienne), orthophotos des missions antérieures.

La végétation arborée est classée selon plusieurs thèmes séparant la forêt, les landes, les formations herbacées et les vergers. La forêt est elle-même séparée en forêt fermée et forêt ouverte, selon la densité de couvert des arbres.

Le logiciel permet également une assistance à la classification selon deux méthodes issues de travaux du laboratoire MATIS de l'IGN (Trias-Sanz, 2006 ; Guigues et al, 2003). Le photo-interprète choisira la méthode la plus efficace selon le contexte paysager du territoire observé ou la qualité des orthophotos.

- La classification automatique : le processus de classification utilise la méthode d'apprentissage supervisé. Il est caractérisé par un apprentissage rapide (saisie manuelle de la végétation selon les thèmes définis sur quelques dalles kilométriques représentatives d'un ou plusieurs thèmes de végétation arborée) suivi de l'attribution automatique des classes par la méthode des kPPV (k Plus Proches Voisins, ou en anglais kNN k-nearest neighbors) pour les autres dalles kilométriques sur demande de l'opérateur. Les mêmes bases d'apprentissage peuvent être utilisées par plusieurs opérateurs et réalisent une pré-saisie qui est ensuite complétée manuellement.

- Pré-sélection à partir de couche pré-existante : ce processus fait intervenir les données de la couche végétation de la BDTOPO® déjà réalisée et pour laquelle une mise à jour est envisagée. Les données sont représentées par des masques de végétation en mode image (noir=bois, blanc=autre). SEVE® permet, si les données sont disponibles, de pré-saisir la végétation en faisant correspondre la couche végétation de la BDTOPO® avec les zones délimitées par la segmentation. Le but est d'importer un masque dans le secteur de travail ce qui aura pour effet de simuler une saisie. La saisie est ensuite complétée manuellement.

L'information à ce stade est dans la même géométrie que la source image (pixel de 0,5 m) stockée en format raster. À l'issue de ce travail, on dispose d'un masque de saisie sur l'ensemble du département, constitué de dalles de 1km sur 1 km. Figure 2

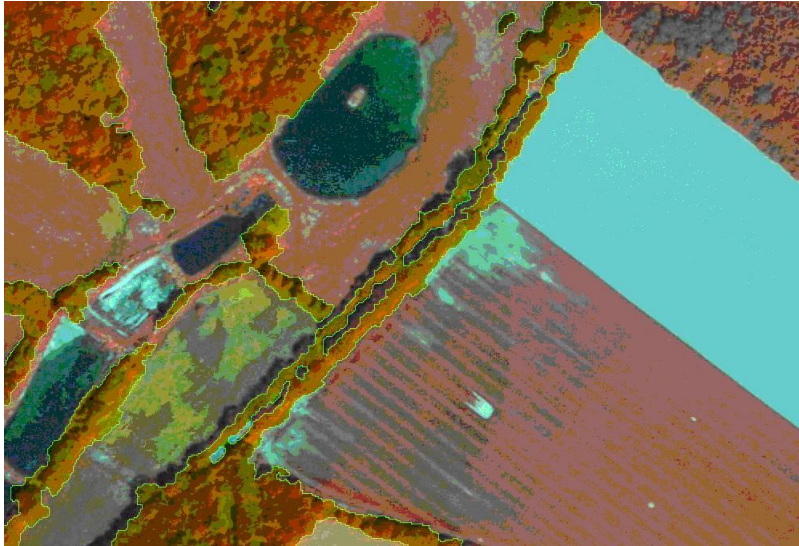


Figure 2 : Saisie de la végétation arborée sur une image IRC segmentée

3.3 Vectorisation des contours et extraction automatique des haies, bosquets et forêts

Les surfaces de végétation en mode vecteur sont traitées automatiquement à l'IGN afin, d'une part, de détecter en fonction de leurs géométries les zones de végétation qui correspondent à des haies ou des bosquets (Touya, 2007) et, d'autre part, les découper en fonction des réseaux présents dans le RGE.

Pour ce faire on considère une forêt, toute zone de végétation dont la plus petite largeur locale est supérieure à 20 m (cf. définition de la forêt, sous-section 2.2) en ayant exclu au préalable les haies attachées à la zone. Un bosquet est une zone de végétation petite et plutôt arrondie ou compacte, non attaché à une forêt. Une haie est une zone de végétation fine et allongée qui peut être attachée à une forêt. La finesse est fixée par le seuil de 20 m. Les houppiers des arbres fixent par segmentation automatique la limite de végétation ce qui a pour conséquence l'augmentation systématique des largeurs. Ces seuils sont augmentés pour les traitements automatiques.

La détection des éléments petits et fins se fait grâce à une transformation de morphologie mathématique, l'ouverture (Su *et al*, 1997). L'ouverture consiste en une érosion suivie d'une dilatation de même taille. Un seuil de 15 m est utilisé pour l'ouverture. Les éléments conservés par l'ouverture sont les véritables forêts. Il reste à classifier les éléments supprimés par l'ouverture en haies, bosquets ou autres tâches de végétation. Pour déterminer si des éléments sont fins et allongés, il faut combiner plusieurs mesures selon que les polygones sont convexes - on mesurera alors leur allongement -, ou concaves - et on mesurera alors la compacité.

Une surface est convexe [1] lorsque son taux de convexité est supérieur à 0,7, en dessous elle est concave.

Taux de convexité = Aire / AireEnv ou AireEnv = aire de l'enveloppe convexe) [1].

L'allongement [2] est utilisé pour les polygones convexes. Cette mesure utilise la notion de rectangle englobant d'aire minimale, appelée le MBR (acronyme anglais). Le MBR est le plus petit rectangle qui englobe complètement la surface. Quand la longueur du MBR est 2,5 fois plus grande que la largeur, on considère que la surface est allongée.

Elongation = L/l ou L = longueur du MBR et l = largeur du MBR [2]

La compacité [3] est utilisée pour les polygones concaves. Elle est calculée à partir de l'indice de Miller. Le seuil de compacité utilisé est 0,4 : en dessous les surfaces sont considérées comme non compactes et donc fines et allongées.

compacité = $4\pi \times \text{Aire} / \text{Périmètre}^2$ [3]

Afin de disposer par la suite de contours de surface en mode vecteur qui ne soient pas trop volumineux en nombre de points intermédiaires, les contours des surfaces sont lissés. La méthode retenue est un filtrage de Douglas & Peucker (Douglas *et al*. 1973)

3.4 Découpage et ajustement des contours aux réseaux routiers, ferrés et hydrographiques

L'ajustement aux différents réseaux est indispensable pour améliorer la précision de positionnement géographique de la lisière de la forêt. Il faut rappeler que la limite de la forêt est matérialisée par segmentation automatique au bord du houppier des arbres alors que la lisière sur le terrain est arrêtée par définition au tiers du houppier. La relation topologique « est au bord de » entre forêt et réseaux routiers, voies ferrées ou réseaux hydrographique est fréquente. Le seuil de proximité entre forêt et tronçon de

réseaux est fixé de façon empirique à 10 m. Lorsqu'un contour d'une forêt est à moins de 10 m d'un tronçon de réseau (route, chemin, voie ferrée ou cours d'eau), la forêt borde la portion de réseau. La géométrie de la forêt est alors modifiée pour forcer le partage de géométrie avec la portion de réseau. La transformation utilise les fonctionnalités topologiques du logiciel Radius ClarityTM.

La segmentation a comme conséquence l'augmentation systématique de la surface des forêts. Afin d'obtenir une plus grande précision de l'estimation de la surface, la forêt est découpée selon l'emprise des réseaux routiers principaux. Une sélection est faite sur les réseaux routiers et les voies ferrées selon leur importance. La largeur des réseaux est extraite de la base de données de la BD Topo® réseaux.



IGN / Direction Technique / Service de la recherche / Laboratoire COGIT

Figure 3 : Découpage et ajustement de la forêt aux réseaux

3.5 Saisie complémentaire des essences forestières

À l'issue de la première étape, l'IFN dispose d'un canevas thématique où les espaces forestiers et semi-naturels (landes et formations herbacées) de plus de 50 ares, relevant de la couche forestière, sont isolés des autres éléments arborés comme les bosquets ou les vergers.

Lors de cette seconde étape de saisie, en plus des référentiels images, des informations complémentaires peuvent fournir une aide précieuse à l'interprétation. Il s'agit des données statistiques IFN, de la cartographie forestière IFN version 1, d'autres types d'images comme la BD ORTHO® antérieure (de 5 ans) à la référence utilisée, des cartes géologiques, pédologiques, phytosociologiques, des guides de typologie des stations forestières ou encore du recueil d'informations sur l'écologie des espèces.

Enfin, des tournées de vérification sur le terrain restent nécessaires pour confirmer ou infirmer certains types de formations difficilement identifiables sur écran. Elles représentent environ un tiers du temps consacré à la réalisation de la couche forestière proprement dite.

L'enrichissement est réalisé avec une application métier développée par l'IFN sur la technologie ArcGis Engine d'Esri.

3.6 Contrôles

Les contrôles de cohérence entre types de formation végétale mais aussi entre la forêt et les autres objets de la couche végétation sont assurés pendant la saisie.

- concernant les contrôles topologiques : contrôle des auto-intersections, des superpositions, des vides ;
- concernant les contrôles sémantiques: contrôle des cohérences entre types, sélection d'échantillons tirés à l'intérieur d'une famille de types de formation végétale en vue d'une ré-interprétation...

Afin d'assurer la cohérence des données entre les deux partenaires, les contours forestiers avec le niveau de détail thématique contenu dans la BDTopo® sont livrés à l'IGN pour être contrôlés et stockés au niveau du Référentiel à Grande Echelle.

Le processus complet ci-dessus nécessite en moyenne un délai d'environ huit mois pour réaliser un département à partir de la date de disposition des orthophotos. Avec ce processus, il est prévu que l'ensemble de la France soit traité en 10 ans.

Cette nouvelle couche de végétation est mise progressivement à la disposition des utilisateurs au rythme de saisie des départements. Les deux instituts utilisent bien évidemment ces données pour leur propre besoin.

4 APPLICATIONS DES COUCHES D'INFORMATION VÉGÉTATION ET FORÊT

A l'IGN, le projet Nouvelle Carte de Base du Service des développements a mis en place une nouvelle chaîne de production des données cartographiques au 1:25 000, sous forme de carte papier ou de dalles numériques, à partir du RGE®. Les données de la couche végétation sont par conséquent exploitées directement : forêts, bosquets, landes et formations herbacées. Des traitements spécifiques sont nécessaires pour obtenir la représentation souhaitée, notamment pour les haies qui doivent être squelettisées pour être représentées sous forme de données linéaires plus régulières et donc plus lisibles. Les premiers prototypes de cette nouvelle carte ont été obtenus et permettront de disposer d'une cartographie totalement cohérente avec la végétation présente dans le RGE® (Maugeais, 2011)

L'IFN propose de réaliser différents traitements géomatiques basés sur cette carte, faisant intervenir ou non d'autres sources d'information : organisation spatiale du couvert forestier, quantification des lisières forêt/non forêt, part de la forêt protégée, surface de forêt par habitant ou sous influence urbaine, caractérisation des surfaces forestières incendiées, etc.

CONCLUSIONS ET PERSPECTIVES

Une autre perspective de recherche et d'évolution des spécifications concerne la végétation en zones urbaines. En effet, la description des espaces boisés urbains est très complexe et nécessite une grande précision du fait de l'imbrication avec les réseaux routiers et les bâtiments. La problématique de l'arbre en ville (alignements de bord de rue, espaces verts, jardins privés...) est très spécifique par rapport à l'objectif de réalisation de la couche végétation ou de la précision de la couche forestière. Des travaux de recherche sont menés pour isoler ces espaces boisés par croisement avec des zones d'extension urbaine, réalisées à partir de la couche des bâtiments. Ces zones ont été déjà calculées par l'IGN dans le cadre d'études sur le survol aérien des zones urbaines. Elles pourraient être réutilisées pour servir de référentiel.

La production de la carte forestière est désormais complètement opérationnelle, en numérique et donc sans papier. 30 départements ont été produits et archivés fin 2010. Le rythme nominal de production de 10 départements par an est désormais atteint. Les chaînes de production des deux établissements sont imbriquées et permettent des coûts optimisés après une collaboration exemplaire pour la mise en place du processus. La volonté des deux producteurs nationaux d'éviter la double saisie et la rationalisation des coûts de production sont désormais satisfaites. La production de données publiques est désormais plus cohérente et non redondante.

Parallèlement à cette production qui va se poursuivre, il est nécessaire désormais de réfléchir à l'objectif à la mise à jour de ces données forestières. La précision actuelle permet d'envisager une mise à jour par différence, mais nécessite à nouveau une collaboration poussée entre les deux instituts pour relever ce nouveau défi.

RÉFÉRENCES

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*, vol. 10, n°2, 1973, p. 112 – 122

Duprez M., Lucas L., *L'IF* n°20, Nouvelle cartographie forestière – De la production à l'utilisation, IFN, 8 pages, 3ème trimestre 2008

http://www.ifn.fr/spip/IMG/pdf/IF20_carto.pdf

Guigues L., 2007, Modèles Multi-Echelles pour la Segmentation d'Images. *Thèse de doctorat de l'Université de Cergy-Pontoise*, décembre 2003

E. Maugeais, F. Iecordix, X. Halbecq, A. Braun Dérivation cartographique multi échelles de la BDTOP de l'IGN France : Mise en oeuvre du processus de production de la nouvelle Carte de Base, 25th ICC , Paris, 2011

IFN, Guide technique, La cartographie forestière version 2, 52 pages

http://www.ifn.fr/spip/IMG/pdf/Guide_technique_cartographie_IFN1.11.pdf

Touya G., Duchêne C., Mustière S., 2010, Généralisation et intégration pour un fond vert commun entre l'IFN et l'IGN, *Revue Internationale de Géomatique* vol.20 n°1, mars 2010, pp 65-86

Trias-Sanz R., Mise à jour automatique de données d'occupation du sol à grande échelle par des méthodes d'analyse d'images, *Thèse de doctorat de l'université de Paris V René Descartes*, mars 2006