

SIPV

Système d'Information du Patrimoine Vert

Apports de la télédétection et des modèles statistiques pour l'automatisation de la cartographie des milieux naturels

Introduction

Cette méthode de cartographie s'inscrit dans le cadre du Système d'Information du Patrimoine Vert (SIPV) (voir *Géomatique Expert* n°58), qui se veut le référentiel régional concernant la biodiversité végétale.

Le projet SIPV a été initié par la ville de Genève. Il est logiquement intégré au Système d'Information du Territoire Genevois (SITG, <http://www.sitg.ch>). Il regroupe au niveau métier tous les partenaires des administrations concernées, tant au niveau municipal que cantonal, ainsi que des acteurs externes issus notamment du Centre du Réseau Suisse de Floristique (CRSF, <http://www.crsf.ch>), de l'Université de Genève ou du Réseau Mondial des Centres d'Information sur l'Environnement (GRID-PNUE, <http://www.grid.unep.ch>). Au niveau technique, ce projet reçoit le soutien de la Direction des Systèmes d'Information et de Communication (Ville de Genève), du Service de l'Organisation et des Systèmes

d'Information (État de Genève) et du Service Informatique de l'Association de Communes Genevoises. Ses principaux objectifs sont :

- La récolte, l'intégration et la synthèse de l'information portant sur les milieux, les espèces et les individus constituant la biodiversité végétale ;
- La mise à disposition de l'information pour la recherche, la gestion et l'aide à la décision ;
- La création de méthodes communes.

Quatre modules métiers sont mis en œuvre, encadré par le module technique assurant la cohérence informatique au sens large :

- Module **Flore sauvage** : actuellement plus de 120 000 données ponctuelles géoréférencées disponibles, tant sur la flore commune, que sur des groupes plus spécifiques (espèces des prairies sèches, espèces menacées...), complétées par des référentiels biologiques, écologiques et législatifs;
- Module **Arbre isolés** : un référentiel taxonomique a été développé. L'inventaire historique

(1976) a été numérisé et les relevés ultérieurs ont été harmonisés et intégrés. Enfin des outils de saisie nomade ont été mis à disposition des partenaires afin d'assurer la compatibilité des inventaires actuels. Ainsi, nous avons plus des 235 000 individus saisis dans le système d'information, dont environ 18 % ont été observé récemment (après 2000) ;

- Module **Espace vert** : pour le moment, seuls les espaces entretenus par la ville de Genève sont intégrés au système. L'objectif visé est un inventaire surfacique afin d'optimiser la gestion différenciée de ces espaces ;

- Module **Milieux naturels** : nous disposons de cartes réalisées dans les années 1980-1990 couvrant chaque commune du canton. Par la suite, certaines communes ont fait réaliser, dans le cadre de leurs Agenda 21, des cartes d'occupation du sol à partir d'orthophotographies et de visites sur le terrain. Finalement, le Service de la Mensuration Officielle (SEMO, cadastre) a réalisé une

Nicolas WYLER, Pascal MARTIN ■ Conservatoire et Jardin botaniques de la Ville de Genève
Karin ALLENBACH ■ GRID-PNUE
Anthony LEHMANN ■ GRID-PNUE - Université de Genève

carte d'occupation du sol (haute et basse) à l'échelle du canton, avec une trentaine de catégories, dont environ la moitié concerne de la végétation.

Dans le cadre de ce dernier module, une méthode originale de cartographie a été développée qui couple la segmentation d'orthophotographies à des modèles statistiques de prédiction dans un système expert. L'objectif est d'automatiser la production et d'assurer les mises à jour grâce à la reproductibilité de la méthode. C'est cette méthode qui fait l'objet de cet article.

Les partenaires

Les Conservatoire et Jardin botaniques de la Ville de Genève (CJBG) remplissent aujourd'hui cinq missions dont l'objectif principal est une meilleure connaissance du monde végétal :

- Explorer ;
- Conserver ;
- Rechercher ;
- Transmettre ;
- Protéger.

En outre, les CJBG sont les coordinateurs du programme « Système d'Information Patrimoine Vert » (SIPV) : un système d'informations de la biodiversité végétale.

Le GRID-Europe du Programme des Nations-Unies pour l'Environnement (PNUE) a pour mission de fournir une expertise géomatique aux ONG et collectivités publiques. Il met particulièrement l'accent sur des services dans le domaine de l'analyse d'images satellitaires et aériennes, d'analyses spatiales en vue de la cartographie environnementale.

L'Université de Genève à travers son nouveau laboratoire d'analyses spatiales de l'environnement (*enviroSPACE*) bénéficie de plusieurs expériences dans le domaine de la cartographie des milieux naturels.

Méthodologie

Différents travaux de cartographie déjà réalisés ont servi de modèles pour ce travail, à savoir le projet Landspot, Landscape Potential (http://www.unine.ch/cscf/PROJETS/landspot/tab_content.htm) et le projet Habitalp, Alpine Habitat Diversity (<http://www.habitalp.de>). Ces deux projets visaient à optimiser la cartographie automatisée des milieux naturels à partir de différentes sources de données (orthophotographies, occupation du sol, gradients environnementaux...). Nous allons développer et adapter ces méthodes pour les appliquer à la cartographie des milieux naturels existants dans le canton de Genève.

4. La caractérisation spectrale des milieux ;
5. La modélisation-prédiction des habitats.

La figure 1 représente le modèle conceptuel du processus mis en œuvre.

Segmentation

Pour le canton de Genève, nous disposons de la Carte de Couverture du sol (CCSol) haute (CCSH) et sol, pour basse (CCSS) réalisée à partir des orthophotographies à 25 cm (figure 2). Cette cartographie permet de distinguer clairement six grands types d'affectation (bâtiments, eau, surfaces à revêtement dur, surfaces vertes, surfaces boisées et surfaces sans

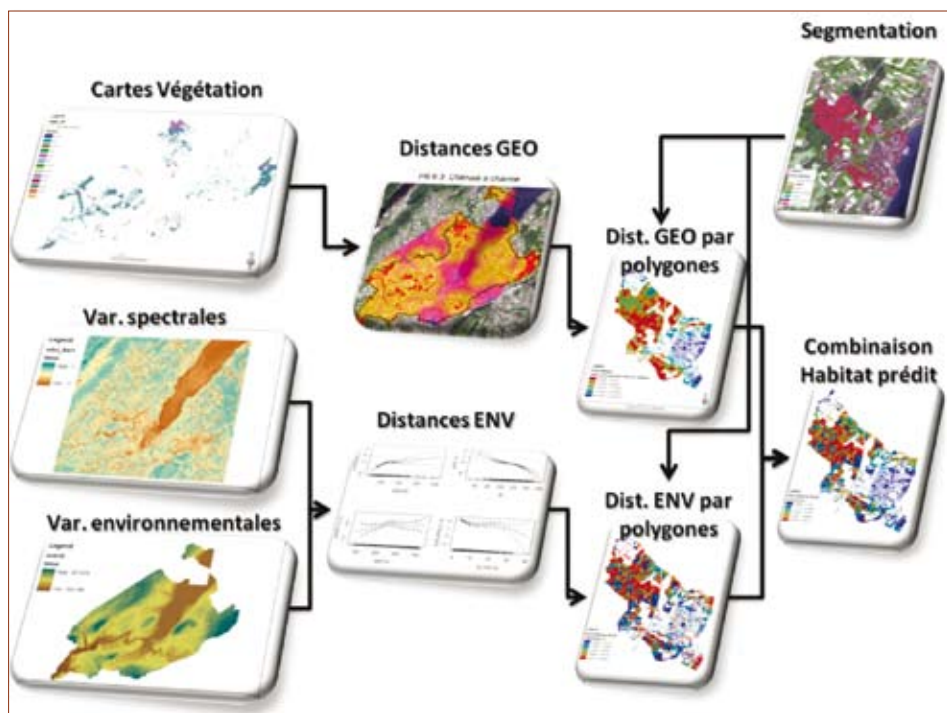


Figure 1 : Modèle de cartographie automatisée.

Cette méthodologie repose sur cinq points importants :

1. La segmentation d'orthophotographies aériennes en polygones du couvert ;
2. Le calcul des distances aux habitats connus ;
3. Le calcul de l'enveloppe environnementale des milieux naturels ;

végétation) qui forment 35 classes d'occupation du sol.

Notre objectif est d'affiner les classes qui concernent les surfaces vertes et boisées et d'attribuer l'un des 53 types de milieux naturels présents potentiellement dans le canton de Genève. Les données sources sont les orthophoto-

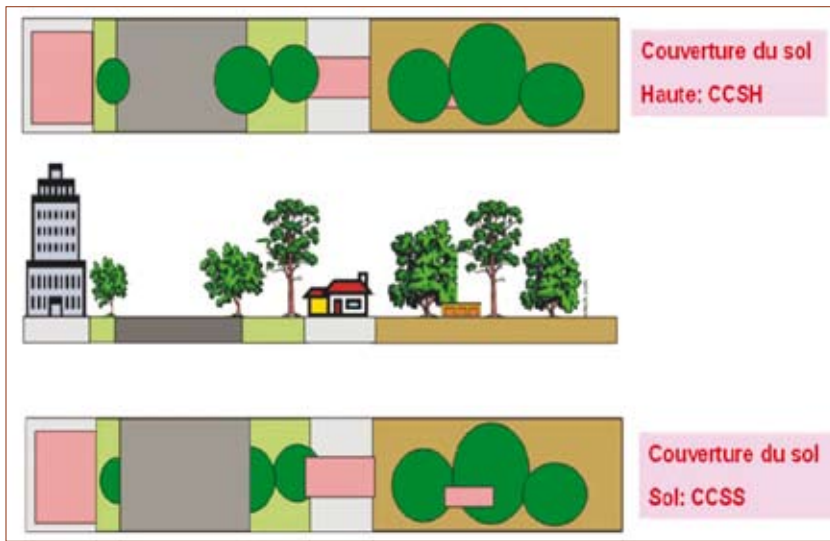


Figure 2 : Différence entre la Couverture du sol haute (CCSH) et la couverture du sol basse (CCSS).

graphies mosaïquées du canton (bandes Rouge, Verte, Bleue (RVB) et Proche InfraRouge PIR) et les contours de la CCSol basse pour les masques.

Contrairement à une approche d'analyse d'image classique (par pixel), nous procédons à une segmentation de l'image. Les signatures spectrales, permettant de classifier différents types de milieux, seront donc calculées sur des objets (regroupement de pixels homogènes) définis selon certains critères.

Le logiciel utilisé, *Definiens* (anciennement *eCognition*, proposé par *Definiens AG*) propose différents



Figure 3 : Exemple de segmentation.

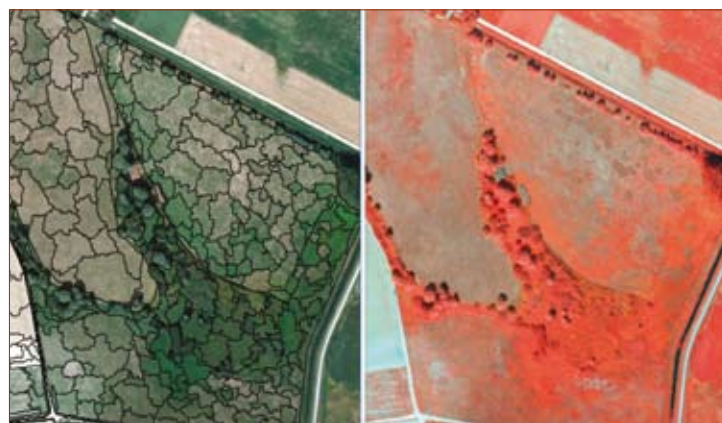


Figure 4 : À gauche résultat de la segmentation dans les marais de Sionnet avec fusion des polygones à une échelle de 15, à droite l'orthophotographie (combinaison de bandes (PIR, vert, bleu)).

algorithmes de segmentation, utilisant une approche ascendante, fusionnant de petits objets (objet original : pixel) pour en obtenir de plus grands. L'algorithme retenu est le *Multiresolution segmentation*, qui s'applique au niveau des pixels ou sur/sous d'autres niveaux de segmentation permettant ainsi une analyse à différentes échelles. Cet algorithme offre la possibilité de combiner des couches images (*rasters*) et des couches thématiques (*vecteurs*). Un poids peut également être attribué aux différentes couches images utilisées. Ainsi par exemple, le PIR, qui permet une meilleure distinction des formations végétales, peut se voir attribuer un poids plus important dans l'analyse.

Nous définissons la taille des objets avec la fonction *scale parameter*. C'est un indice de dissemblance maximum entre les pixels ou les objets. Il permet notamment de définir l'échelle de travail (longueur maximum de frontière entre polygones). Plus la valeur est grande plus les objets seront grossiers.

Les différents critères d'homogénéité à définir procèdent par paires :

- *Couleur/Forme* permet de favoriser la couleur (spectrale) au détriment de la forme (ou inver-

sement). Cependant, pour une segmentation de qualité, il est fortement recommandé de favoriser l'information spectrale (source d'information principale) aux dépends des critères de forme ;

- *Flou/Compacité* est utilisé pour favoriser des objets aux contours flous dans une image très hétérogène ou au contraire des objets compacts séparés par une patine à faible contraste spectral. Ces deux critères de forme ne sont pas antagonistes. Par exemple, un objet compact peut très bien avoir des frontières floues. Le choix des critères dépendra fortement des objets à cartographier.

Pour garder les frontières des objets de la carte de couverture du sol basse (CCSS), nous avons intersecté les nouveaux polygones créés par une sélection de classes de la CCSol (surfaces vertes et surfaces boisées). Cela a engendré de nombreux petits objets aux frontières de thèmes, qu'il a fallu nettoyer. Les segments obtenus représentent bien les variations observées sur le terrain. Voir les figures 3 et 4 qui représentent des exemples de segmentation.

Calcul des distances aux habitats connus

Afin d'alimenter les modèles statistiques, dans un premier temps, il est nécessaire de regrouper toutes les informations existantes sur les milieux naturels à l'échelle du canton de Genève. Différents services publics (Direction Générale de la Nature et du Paysage (DGNP), Conservatoire et Jardin botaniques (CJB), Service de l'organisation et des systèmes d'information (SOSI), Département de l'Agriculture, ainsi que des entreprises privées mandatées par les collectivités) disposent de références spatialisées sur les formations végétales. En outre, certains milieux naturels font l'objet de protection à l'échelle fédérale et sont donc bien connus. D'autre part, des sites d'intérêt particulier pour les



Figure 5 : Carte de végétation de Werdenberg et Hainard (2000).

espèces qu'ils abritent font l'objet de plan de gestion par la DGNP et ont été cartographiés précisément. Ce sont par exemple les surfaces de compensation écologiques, les réserves forestières, les zones alluviales, les prairies sèches, les bas-marais, les zones de roselières. Une étape importante a été de regrouper l'ensemble de ces données et d'uniformiser la typologie utilisée pour décrire les milieux naturels.

Nous utilisons notamment la carte de végétation de Werdenberg et Hainard (2000) pour calculer la distance aux milieux déjà cartographiés dans le passé (figure 5).

Cette démarche nous permet ensuite d'intégrer une probabilité d'existence d'un milieu naturel dans telle ou telle zone. Il y a en effet plus de chances d'avoir également de la chênaie à proximité de chênaies connues historiquement que d'autres milieux. Dans ArcGIS (ESRI), chaque type d'habitat a été transformé en une couche polygone, puis un simple calcul de distance en mode *raster* (figure 6) a été effectué à une résolution très faible (2 m), afin de permettre l'interrogation des couches rasters obtenues par la suite par des statistiques zonales à partir de polygones assez petits issus de la segmentation.

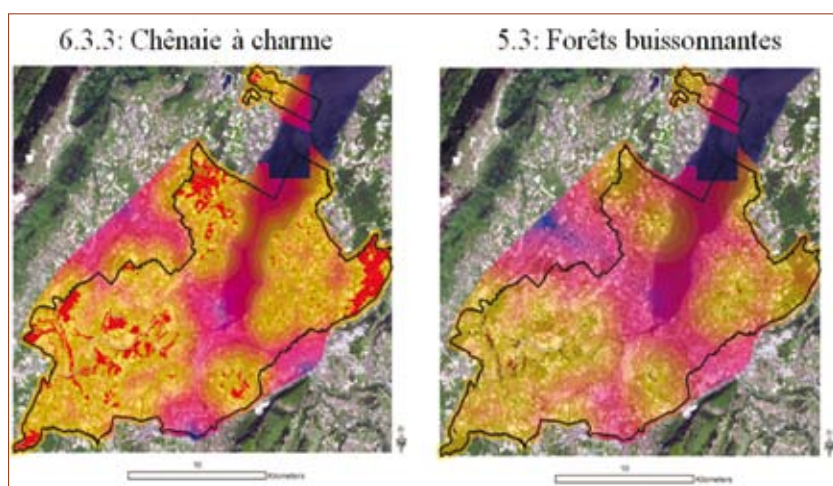


Figure 6 : Calcul des distances à deux milieux déjà cartographiés.

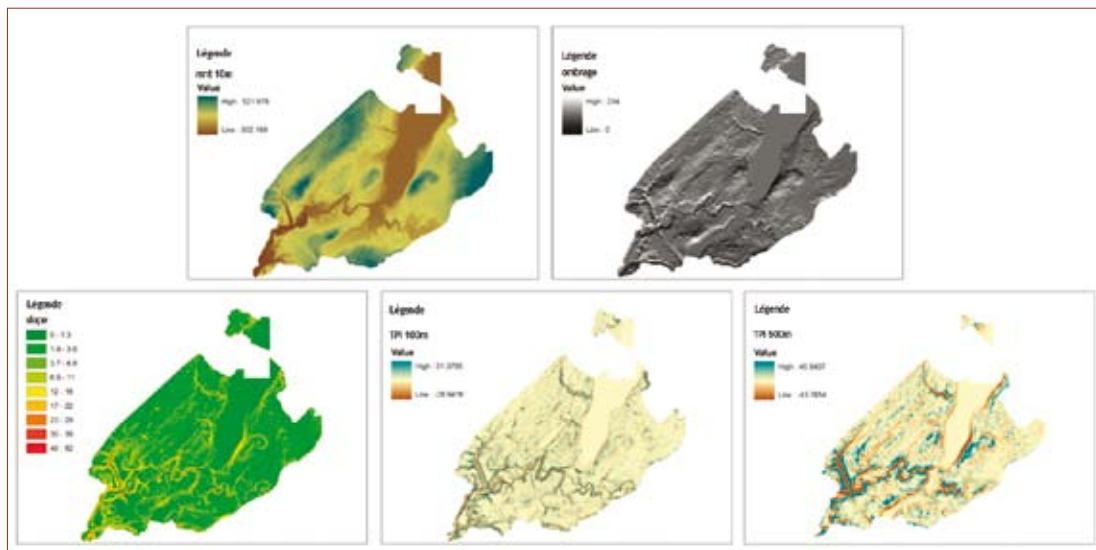


Figure 7 : Variables environnementales utilisées dans le modèle statistique GRASP : altitude, ombrage, pente, TP 100, TP 500.

Calcul des distances environnementales

Les variables environnementales choisies dans le cadre de nos travaux sont pour l'instant toutes dérivées des modèles numériques d'altitude (MNA) et de hauteur (MNH), agrégés à une résolution de cinq mètres. Il s'agit de l'altitude, de l'ombrage, de la pente, de l'indice topographique à 100 m et 500 m (figure 7). Ce dernier permet de qualifier la géomorphologie du terrain en combinant la pente et les indices topographiques. Ces variables naturelles permettent de définir

une enveloppe environnementale qui caractérise les zones de végétation. Par exemple, à l'échelle du canton de Genève, les prairies sèches se rencontrent sur les versants exposés sud, alors que les prairies humides à populage sont limitées aux bas de pente plutôt exposés nord. Il est évidemment possible d'intégrer une plus large palette de variables écologiques, la présence de nappes phréatiques, la distance aux cours d'eau... Toutes les données supplémentaires peuvent permettre de caractériser au mieux les conditions environnementales dans lesquelles

les différents habitats naturels existent.

Calcul des caractéristiques spectrales

Plusieurs variables spectrales ont été préparées afin de tester leur relation avec la présence des habitats naturels cartographiés. Il s'agit de la bande proche infra rouge (PIR), rouge, verte et bleue, ainsi que l'indice classique de *Normalized Difference Vegetation Index (NDVI)* qui est le rapport $(PIR - rouge) \div (PIR + rouge)$ ou indice de végétation normalisé (figure 8). Ces valeurs de réflectance, le PIR

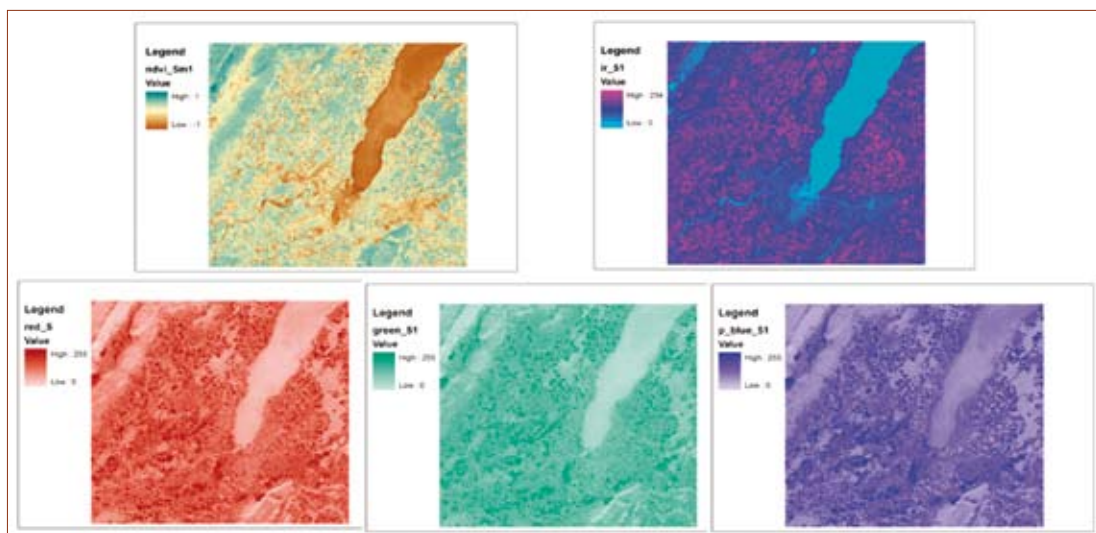


Figure 8 : Variables spectrales utilisées dans le modèle statistique GRASP : NDVI, IR, rouge, vert, bleu.

et le NDVI en particulier, permettent de bien caractériser l'état de la végétation via la quantité de chlorophylle active.

Modèles statistiques

L'idée principale de l'approche proposée est de combiner la proximité géographique et spectro-environnementale (figure 9) pour évaluer la probabilité d'occurrence de chaque habitat au sein de chaque polygone issu de la segmentation. Ainsi, plus un polygone est proche d'un type d'habitat cartographié dans le passé ou existant actuellement, et plus il devrait ressembler spectralement et environnementalement à ce milieu. Le score final (la probabilité renvoyée par le modèle) de cet habitat dans ce polygone sera alors élevé. La comparaison des scores pour chaque habitat dans chaque polygone permettra ensuite de proposer d'attribuer l'habitat le plus probable au polygone tout en maintenant l'information sur les autres habitats prédits. Il sera possible aussi de ne pas attribuer d'habitat si le polygone est trop éloigné géographiquement et/ou spectralement (figure 10).

Pour la modélisation et les prédictions nous utilisons l'outil *Generalized Regression Analysis and Spatial Prediction (GRASP)* dans *Splus* pour sélectionner le modèle statistique décrivant au mieux la relation entre chaque habitat et son environnement. Ces modèles additifs permettent d'intégrer un grand nombre de variables explicatives, une procédure itérative de soustraction-addition des variables abouti au meilleur modèle. C'est-à-dire celui qui explique le plus la variable à expliquer. Par exemple, pour la modélisation de la chênaie-charmaie (la variable dépendante à expliquer), toutes les variables spectrales et environnementales (les dix variables explicatives) sont dans un premier temps testées ; après plusieurs boucles de combinaison de ces variables, le meilleur modèle

n'a conservé que sept de ces variables (vert, IR, NDVI, MNT à 10 m, pente à 10 m, TPI à 100 et 500 m). De plus les variables explicatives sont hiérarchisées en fonction de leur contribution à la modélisation de la variable dépendante. La validité (la confiance que l'on peut attribuer aux modèles sélectionnés) est testée par validation simple et croisée avec les statistiques de corrélation (r) ainsi que l'aire sous la

courbe ROC. Une fois le modèle établi, pour un milieu naturel donné, le résultat final est la production d'une probabilité d'appartenance à ce milieu pour chacun des segments du territoire.

le terrain permet de renseigner les polygones présentant des réponses confuses et surtout de valider les prédictions. Notre étude pilote a traité successivement deux aspects : d'une part, la modélisation-prédiction des zones forestières pour lesquelles les résultats sont très prometteurs, étant donné que la méthode permet de distinguer les principales variations existantes dans le canton de Genève ; d'autre part, la modélisation des zones ouvertes (prairies, pelouses, pâtures...) qui s'est avérée plus difficile en raison de la date de prise de vue des orthophotographies.

Actuellement cette méthode est employée pour cartographier un corridor biologique (Vuache-Sion-Laire) du plan d'aménagement de l'agglomération franco-valdo-genevoise (<http://www.projet-agglo.org>), voir ci-après.

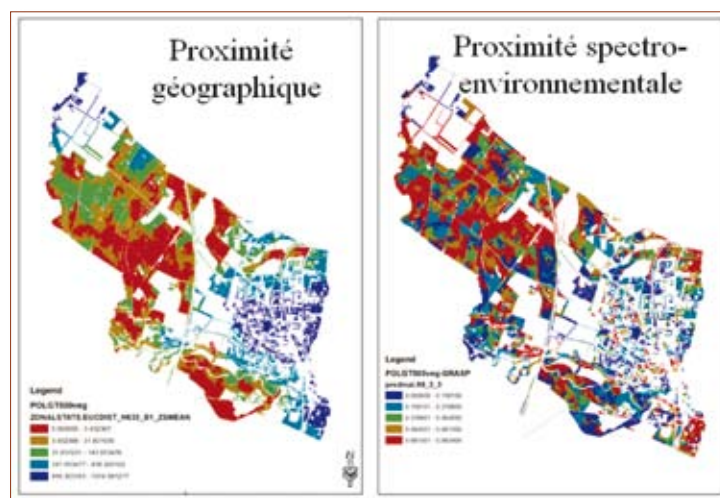


Figure 9 : Distances GÉographiques et ENVironnementales calculées pour chaque polygone identifié par la segmentation du départ pour le milieu 6.3.3 Chênaie à Charmes dans la commune de Versoix.

Résultats

L'étude s'est, dans un premier temps, concentrée sur le bassin versant de la Versoix pour les milieux forestiers, puis sur les communes de Vandoeuvres-Choulex-Puplinge pour les milieux ouverts. Actuellement l'étude est terminée, elle s'est principalement concentrée sur le développement de la méthodologie, seuls quelques résultats concrets en terme de cartographie ont été fournis aux institutions mandantes.

Cette méthode automatisée de cartographie haute résolution

Développements et perspectives

Sur l'ensemble de l'étude, de nombreux tests ont permis de faire évoluer conceptuellement la méthode. C'est surtout la procé-

ture générale qui a été présentée ici. Néanmoins certains détails méthodologiques additionnels méritent d'être mentionnés.

Les différentes tentatives d'utilisation de la CCSS ont dévoilé de nombreux problèmes qui la rendent difficilement utilisable à l'échelle du canton. Au sein même d'une petite région comme la commune de Versoix ou le long de l'Allondon, de nombreux problèmes ont fortement compliqué l'élaboration de la procédure. Cependant, dans la mesure où la carte de couverture du sol pourrait être améliorée, en corrigeant les problèmes topologiques (superpositions d'objets, etc.) et en diminuant son poids par un lissage des polygones, la mise en œuvre de notre démarche gagnerait en rapidité.

Il est ressorti de l'étude que la date de prise de vue des orthophotographies est un facteur primordial pour la reconnaissance des milieux ouverts. En effet, les zones de prairie de fauche présentent une forte confusion avec les zones de pâturages, car les orthophotographies ont été acquises en août 2005, date à laquelle les prairies ont déjà toutes été fauchées.

Les variables environnementales utilisées pour la modélisation peuvent être enrichies par diverses informations (géologie, pédologie, distance aux cours d'eaux, présence de nappe phréatiques superficielles...), ce qui donnera encore plus de validité aux prédictions des modèles. En effet, comme les modèles additifs généralisés sélectionnent les meilleures variables explicatives, il est certain que des variables secondaires puissent être très pertinentes pour quelques milieux particuliers (par exemple la proximité aux cours d'eau avec les chênaies-charmaies à herbe-aux-goutteux : *Aegopodium podagraria*).

Cette méthode sera appliquée à l'échelle du périmètre de l'agglomération franco-valdo-genevoise (environ 2 300 km²) avec une cartographie finale au 1:25 000 prévue pour 2012. La diversité des milieux naturels, les gradients environnementaux beaucoup plus importants qu'à l'échelle du canton et l'hétérogénéité des données sources représentent un réel défi de cartographie semi-automatisée.

Au sein de ce projet d'agglomération, huit zones de corridors biologiques seront analysées selon cette méthode de manière plus approfondie, afin d'aboutir à une cartographie au 1:10 000 de plus de 50 milieux naturels différents.

La méthode permettra en outre de délimiter des interfaces biologiques, tels que la limite forêt-prairie, prairie-cours d'eau, forêt-cours d'eau, etc. sous forme linéaire car non représentables à l'échelle de production de la carte. Ces interfaces sont des éléments du paysage fondamentaux pour les mouvements de la faune. Au final, ces produits, une carte précise des milieux naturels ainsi que des voies de communications pour la faune, vont permettre de mener une politique de développement de l'aménagement consciente, raisonnée et indiscutablement plus respectueuse de l'environnement.

Le dernier point clef de cette méthodologie de cartographie automatisée réside dans sa reproductibilité et dans sa capacité à intégrer les données acquises par les collectivités ou les instituts de conservation de la nature. Finalement, de nouvelles mises à jour des données sources, orthophotographies voire scènes Spot, planifiées actuellement tous les 4-5 ans, peut très rapidement permettre d'obtenir une cartographie mise à jour pour un coût raisonnable. ■

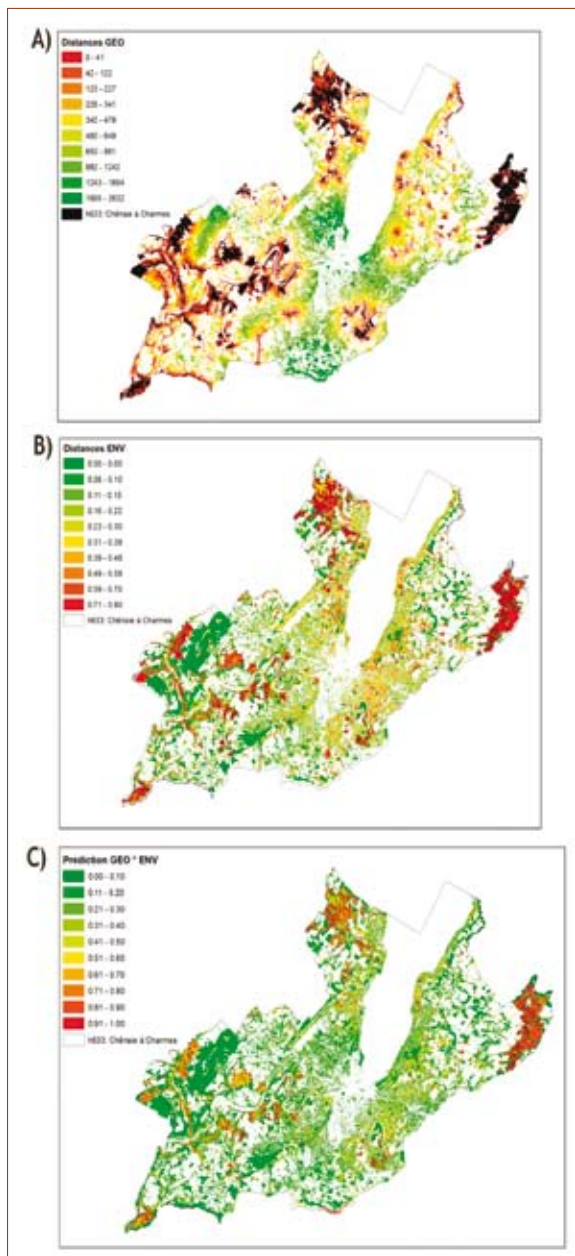


Figure 10 : Distances A) GEOgraphiques et B) ENVironnementales calculées pour chaque polygone identifié par la segmentation du départ pour le milieu 6.3.3 Chêne à Charmes sur l'ensemble du canton ; C) Combinaison des indices de distance GEOgraphiques et ENVironnementales pour calculer la distribution probable du milieu 6.3.3 Chêne à Charmes sur l'ensemble du canton.