



Universiteit
Leiden
The Netherlands

SMALL Savannah : an information system for the integrated analysis of land use change in the Far North of Cameroon

Fotsing, E.

Citation

Fotsing, E. (2009, December 8). *SMALL Savannah : an information system for the integrated analysis of land use change in the Far North of Cameroon*. Retrieved from <https://hdl.handle.net/1887/14619>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/14619>

Note: To cite this publication please use the final published version (if applicable).



Photo 3. Le déplacement d'un troupeau de bœufs en saison sèche, à la recherche de pâturages dans la zone de Mindif, région de l'Extrême Nord du Cameroun.

Chapitre 3. Outils pour l'analyse et la modélisation des changements d'utilisation de l'espace

Résumé

L'analyse et la modélisation intégrée des changements d'utilisation de l'espace nécessitent une gamme variée de techniques et d'outils logiciels. L'application de ces outils à des problèmes concrets dépend le plus souvent de plusieurs critères comme les objectifs visés par l'étude, la discipline scientifique ou l'échelle d'analyse. Par exemple, les changements d'utilisation de l'espace à l'échelle régionale sont étudiés le plus souvent en combinant les observations directes, la télédétection et les SIG pour mieux comprendre l'organisation des paysages et suivre leur évolution. Les outils de SIG et d'analyse spatiale sont utilisés pour représenter et expliquer les phénomènes géographiques par rapport à leur localisation. Les outils d'analyse statistique sont utiles pour expliquer et prédire les évolutions en s'appuyant sur les tendances actuelles et la comparaison des séries de données. L'analyse des facteurs qui déterminent les changements d'utilisation de l'espace est en général effectuée à l'échelle locale en utilisant des outils tels que la cartographie participative et les interviews afin d'identifier les motivations et les stratégies des acteurs impliqués dans les processus de changement. A différentes échelles, les modèles dynamiques peuvent être utilisés pour explorer les trajectoires des changements d'utilisation de l'espace et fournir ainsi des informations utiles à la prise de décision. Ce chapitre a pour objectif de présenter ces outils et décrire comment ils ont été combinés et utilisés pour mettre en œuvre SMALL Savannah, le SIE conçu pour l'analyse intégrée des changements d'utilisation de l'espace dans la région des savanes de l'Extrême Nord du Cameroun.

Mots clés : Changement d'utilisation de l'espace, télédétection, morphologie mathématique, SIG, analyse statistique, modélisation dynamique.

Abstract

The integrated analysis and modelling of land use change requires the use of a wide range of techniques and software tools. The application of these tools often depends on many criteria like the objectives of the study, the scientific discipline or the scale of analysis. For example land use change at regional scale are often analysed by combining direct observations, remote sensing and GIS to better understand landscape organisation and monitor it evolution. GIS and spatial analysis tools are used to represent and explain geographic phenomena according to their location. Statistical tools are useful to explain and predict the evolution based on actual trend and comparison of data series. The analysis of land use change drivers is often undertook at local scale using tools like participatory mapping and interviews in order to identify the motivation and strategies of actors involved in the process of change. At different scales, dynamic simulation models can be used to explore land use change trajectories and provide key information for decision making. The objective of this chapter is to present these tools and describe how they were combined and used to implement Savannah, the EIS designed for the integrated analysis of land use change in the Far North region of Cameroon.

Key words: land use change, remote sensing, mathematical morphology, GIS, statistical analysis, dynamic modelling.

3.1. Introduction

Les systèmes d'utilisation de l'espace comme la plupart des systèmes environnementaux sont très complexes. Le développement d'un Système d'Information pour l'analyse des changements d'utilisation de l'espace nécessite donc une approche et des outils appropriés. Au chapitre 2, nous avons montré que la démarche suivie dans cette thèse pour analyser les dynamiques d'utilisation de l'espace comprend les quatre principales étapes suivantes : 1) identifier les structures d'occupation du sol, quantifier et localiser les changements qui surviennent au cours du temps ; 2) identifier les facteurs biophysiques et socio-économiques qui déterminent ces changements aux différentes échelles spatiales ; 3) quantifier les relations entre les structures d'utilisation de l'espace et les facteurs déterminants ; 4) élaborer un modèle de simulation des changements et tester des scénarios d'évolution afin d'anticiper les réponses aux différentes options de gestion. La mise en œuvre de cette démarche exige l'utilisation et la combinaison d'outils appropriés à chaque étape. De nombreuses approches et outils d'analyse et de modélisation des changements d'utilisation de l'espace ont été développés. Leur application à des problèmes concrets dépend le plus souvent de plusieurs critères comme les objectifs visés par l'étude, la discipline scientifique, l'échelle d'analyse (Levin, 1992). En géographie et en écologie par exemple, les changements d'utilisation de l'espace à l'échelle régionale sont étudiés le plus souvent en combinant les observations directes, la télédétection, les SIG pour mieux comprendre l'organisation des paysages et suivre son évolution. Les outils de SIG et d'analyse spatiale sont utilisés pour représenter et expliquer les phénomènes géographiques par rapport à leur localisation. Les outils d'analyse statistique sont utiles pour prédire les évolutions en s'appuyant sur les tendances et la comparaison des séries de données. Dans le domaine des sciences sociales par contre, l'analyse des facteurs qui déterminent les changements d'utilisation de l'espace est en général effectuée à l'échelle locale en utilisant des outils tels que la cartographie participative et les interviews afin d'identifier les motivations et les stratégies des acteurs impliqués dans le processus de changement. A différentes échelles, les modèles de simulation peuvent être utilisés pour explorer les scénarios d'évolution future et fournir ainsi des informations utiles à la prise de décision. Ce chapitre a pour objectif de présenter ces outils et décrire comment ils ont été utilisés et combinés pour mettre en œuvre SMALL Savannah, le SIE conçu pour l'analyse intégrée des changements d'utilisation de l'espace dans la région des savanes de l'Extrême Nord du Cameroun. La section 2 décrit comment les outils de la télédétection et de traitement d'image sont utilisés pour identifier les structures d'occupation du sol et les changements dans le contexte des zones de savane. La méthode d'analyse d'image proposée combine itérativement traitement numérique, cartographie participative et photo-interprétation pour extraire des structures d'occupation du sol pertinentes. La section 3 met en exergue la contribution des outils de SIG d'analyse spatiale à la mise en œuvre du SIE. La section 4 présente le modèle statistique et la méthode, utilisés pour explorer les facteurs déterminants et quantifier les relations avec les structures et les changements d'utilisation de l'espace. La section 5 fait un aperçu des modèles de simulation des dynamiques d'utilisation de l'espace existant et justifie le choix de la plate-forme de modélisation retenue pour construire le modèle développé dans cette thèse.

3.2. Outils de télédétection et de traitement d'images

Le point de départ de la démarche suivie pour analyser les dynamiques d'utilisation de l'espace est la cartographie de l'occupation du sol. Dans cette section, nous présentons une méthode qui s'appuie sur les

outils de la télédétection et des SIG pour identifier et extraire les structures d'occupation du sol et les changements dans le contexte des zones de savane d'Afrique centrale (Fotsing et al., 2006). En effet, la télédétection est en évolution constante grâce au développement de nouveaux capteurs et méthodes de traitement numérique des images en réponse aux divers besoins de gestion de l'environnement tels que l'observation, l'inventaire des ressources naturelles, la modélisation des dynamiques et la prévision. Malgré les importants développements de la télédétection, on s'interroge toujours sur l'intégration effective de cette technique à des programmes opérationnels destinés à la gestion des problèmes environnementaux et de développement (Regis et al., 1999). L'une des raisons de cette situation est liée à la difficulté de généraliser les méthodes de traitement testées au laboratoire ou dans un contexte donné. En effet, de nombreuses méthodes ont été développées pour extraire des informations spécifiques sur des territoires bien connus (Bonn et Rochon, 1996). Dans les zones de savanes, les paysages sont très hétérogènes et les dynamiques d'utilisation de l'espace sont loin d'être connues dans l'espace et le temps, l'application exclusive d'une méthode de traitement automatique n'est pas toujours praticable. De plus, les données spatiales détaillées comme les photographies aériennes ou les images satellites nécessaires à cette connaissance sont peu abondantes, souvent anciennes et de qualité approximative (Poncet, 1986). Dans ce contexte, il est important de mettre en œuvre des procédures de traitement et d'analyse d'image appropriées pour arriver à extraire des informations utiles à la compréhension des processus et à la prise de décision. La méthode développée dans cette étude combine itérativement traitement numérique d'image, vérité de terrain, cartographie participative et photo-interprétation. Les traitements numériques appliqués aux images pour extraire les structures d'occupation du sol sont décrits et quelques résultats et éléments de validation de la méthode sont discutés.

3.2.1. Principe de la méthode d'analyse d'images

La procédure d'analyse d'image utilisée pour dériver les structures d'occupation du sol à partir des images satellite comprend les quatre principales étapes suivantes telles que le montre la figure 3.1 : les traitements préliminaires, l'extraction de l'information, la généralisation cartographique, la validation et la correction des classes d'occupation du sol. Chacune de ces étapes nécessite la combinaison de plusieurs traitements d'image de base parmi lesquels on peut citer : l'analyse spectrale, les indices de végétation, les classifications multispectrales, la morphologie mathématique et la photo-interprétation.

On distingue principalement deux types d'approche d'analyse numérique des images spatiales : l'analyse ponctuelle et l'analyse spatiale. L'analyse ponctuelle ou radiométrique est basée sur les lois de comportement spectral des objets et permet d'identifier les objets de façon assez précise. Toutefois, le principal inconvénient c'est qu'il est difficile de généraliser les modèles de comportement spectral testés dans un contexte donné à d'autres situations. En particulier pour les zones de savanes d'Afrique où les paysages sont très hétérogènes, l'analyse spatiale est une approche complémentaire à l'analyse radiométrique. Elle est basée sur la caractérisation des objets par leur position géographique et par des connaissances sur leurs positions relatives dans l'espace (Pumain et Saint-Julien, 1997). Cette approche d'analyse dite géographique est fondée sur l'hypothèse selon laquelle les unités cartographiques définies ne sont pas réparties de façon aléatoire dans l'espace mais présentent une certaine logique dans la distribution spatiale qu'il faut rechercher. Les objets sont reconnus à partir d'études sur le terrain et en relation avec des informations déjà acquises (cartes existantes, photo aériennes, etc.). Une technique complémentaire est la cartographie participative au cours de laquelle les cartes et les images satellites en composition colorée

sont utilisées comme support de base des discussions sur les changements environnementaux, l'appropriation des terres ou tout autre processus qui gouverne les changements d'occupation du sol. L'approche participative a été également choisie comme un moyen d'explorer les perceptions des populations locales sur les changements d'occupation du sol à travers les informations issues des discussions en groupe (Bocco and Toledo, 1997). Lors des interviews semi-structurés, nous avons utilisé une série de questions ouvertes concernant les changements d'occupation du sol et l'utilisation des ressources naturelles pour guider la discussion. Lors de la description des zones utilisées pour une activité donnée, les populations peuvent identifier et tracer les limites des zones utilisées. Ces informations spatiales sont numérisées, importées et superposées avec les images géoréférencées sous le SIG, ce qui facilite la phase de photo-interprétation.

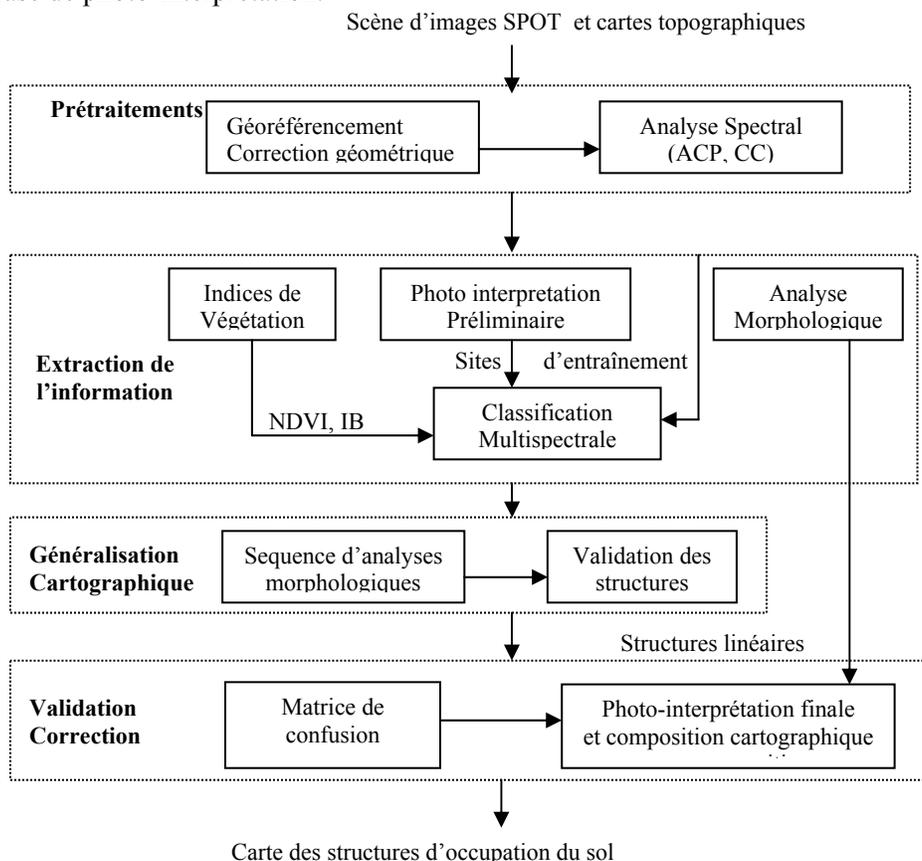


Figure 3.1 : Procédure de traitement et d'analyse des images

3.2.2. Traitements préliminaires

Les traitements préliminaires comprennent la sélection et le géoréférencement des images, et les analyses spectrales préliminaires tels que les compositions colorées et les analyses en composantes principales.

a) Sélection des images et géoréférencement

Le choix des images utilisées pour cette étude a été guidé par le besoin d'une analyse de l'extension récente de la culture du sorgho de contre saison et ses conséquences sur les autres formes d'utilisation de

l'espace, notamment les défrichements de la savane boisée et la dégradation des pâturages. En se référant au calendrier agricole de la région, les images acquises entre le 15 janvier et le 15 février sont apparues les plus appropriées pour atteindre cet objectif car cette période correspond aux dates de croissance maximale et de récolte du sorgho de contre saison. Les données de base acquises comprennent un ensemble de trois scènes d'images satellites (1 scène SPOT de janvier 1987 et 2 scènes SPOT de février 1999), une base de données de points GPS contenant les vérités de terrain et une carte du zonage agroécologique. Les deux scènes d'images SPOT récentes ayant les mêmes paramètres d'acquisition (SPOT XI 4 bandes, 28/02/1999, niveau de traitement 2A) ont été assemblées en une image pour couvrir la zone d'étude. La scène d'image d'archive (SPOT XS 3 bandes, 15/01/1987) a été corrigé géométriquement pour être facilement superposée et comparée à l'image de 1999. Les images étant de niveau de traitement 1B et 2A, les corrections atmosphériques ainsi que les corrections géométriques permettant de compenser les distorsions internes de l'image dues aux conditions de prise de vue sont déjà effectuées (variation d'altitude, effet panoramique, courbure de la terre et rotation de la terre pendant la prise de vue, etc.). Les deux scènes ont été ensuite géoreférencées dans un système de projection défini pour la zone d'étude et dont les caractéristiques sont les suivantes : UTM, Ellipsoïde Clarke 1880, zone 33 hémisphère nord, datum Andinda. Les points de contrôle utilisés ont été collectés avec un récepteur GPS sur des sites identifiables facilement sur le terrain et sur les images disponibles (pont, radier, carrefour, etc.).

b) Analyses spectrales préliminaires

Les compositions colorées de type standard obtenues à partir des canaux XS permettent de localiser certaines entités du paysage qui peuvent déjà être sélectionnés comme zone d'entraînement pour la classification multispectrale. Toutefois, plusieurs autres informations restent cachées. L'analyse en composantes principales (ACP), encore appelée transformée de Hotelling est l'une des nombreuses techniques utilisées pour améliorer la qualité des images en vue de leur interprétation. Etant donnée la difficulté de visualiser simultanément plus de trois bandes spectrales d'une image, on cherche à se ramener à trois composantes sans perdre beaucoup d'information. C'est une transformation orthogonale qui suit la même équation que celle de Fourier (Bonn et Rochon, 1996; Girard and Girard, 1999).

3.2.3. Extraction de l'information sur l'occupation du sol

Ce paragraphe décrit les traitements de base qui ont été combinés et implémentés pour extraire les structures et les thèmes de l'occupation du sol à partir des images satellites.

a) Indices de végétation

Un indice de végétation est une combinaison linéaire des canaux qui est utilisée principalement pour estimer la densité du couvert végétal, différencier les grandes unités naturelles ou culturelles et entreprendre dans certains cas la prévision des récoltes. Les courbes de réflectance d'un végétal vert et d'un sol nu montrent que les plus grandes différences de réflectance entre deux bandes s'observent dans les longueurs d'onde correspondant au rouge et à l'infrarouge. De plus Baret et Guyot (1991) ont montré que ces canaux contiennent 90% de l'information sur le couvert végétal. Ces résultats expliquent le choix de ces canaux pour la formulation de nombreux indices de végétation. Plus de quarante indices de végétation ont été développés dans divers domaines par les chercheurs en télédétection pour différents besoins (Bannari et al., 1998; Bariou et al., 1985; Baret et Guyot, 1991). Ces indices peuvent être regroupés en fonction de plusieurs critères parmi lesquels : l'utilisation qui en est faite, le nombre de

bandes combinées et les opérations de bases appliquées aux canaux. Pour cette étude, l'indice de végétation normalisé et l'indice de brillance généralisé sont les deux indices qui ont été calculés et exploités pour extraire spécifiquement les informations sur le couvert végétal.

L'indice de végétation normalisé (NDVI, Normalised Difference Vegetation Index), qui est l'un des indices les plus utilisés, fournit des bons résultats pour la détection des végétations chlorophylliennes. Il est défini par : $NDVI = (PIR - R) / (PIR + R)$. L'indice de brillance généralisé permet de mieux distinguer les sols nus. En présence de recouvrement de végétaux verts inférieur à 25%, cet indice fournit une meilleure corrélation que l'indice de végétation car la réponse spectrale fournie par les données des satellites dans cette région dépend essentiellement du sol (Belghith, 1990 ; Bonn et Rochon, 1996). Il est

défini par : $IB = \sqrt{(IR)^2 + (R)^2}$

b) Classifications multispectrales

La classification ou la segmentation d'une image consiste à la diviser en zones ayant une occupation identique comme la forêt, les habitations ou les cultures. Il existe plusieurs algorithmes de classification et chacune a ses avantages et ses inconvénients. Plusieurs classifications ont été effectuées en utilisant les méthodes de classification supervisée ou non supervisée appliquées aux canaux de base (XS), aux composantes principales (ACP) et aux indices de végétation (NDVI et IB). L'algorithme de classification par les nuées dynamiques utilisée pour l'exemple présenté dans cette section est une méthode non supervisée qui a été appliquée avec succès dans plusieurs domaines tels que la reconnaissance des formes et la neurologie pour dissocier des objets en se basant sur un ensemble de paramètres caractéristiques (Diday, 1978 ; Diday, 1971). Les images obtenues par classification multispectrale présentent le plus souvent une très forte hétérogénéité locale des structures qui ne facilite pas l'interprétation visuelle ou l'intégration avec d'autres données sous un SIG. Pour résoudre ce problème, une méthode de généralisation cartographique basée sur une combinaison convenable des opérateurs de morphologie mathématique est présentée dans la section suivante (Simonneaux, 1995).

c) Analyses morphologiques

Les opérations de morphologie mathématique sont un ensemble d'opérations de traitement d'images basées sur les formes et qui permettent de filtrer des images binaires ou en teintes de gris (Schowengerdt, 1983 ; Coster et Chermant, 1985 ; Serra, 1988). Le principe repose sur la théorie des ensembles et consiste à comparer les structures d'une image avec un objet de référence qu'on appelle élément structurant. Les opérateurs de base de la morphologie mathématique sont l'érosion et la dilatation. Ils permettent de définir des opérateurs plus complexes tel que l'ouverture, la fermeture, la reconstruction géodésique et les Chapeaux Haut de forme (Debaine et al., 1988 ; Legeley et Mering, 1997). Dans ce travail, ces opérations ont été utilisées respectivement pour délimiter des formes d'occupation du sol, extraire des structures linéaires sur des images spatiales et mettre en œuvre une procédure de généralisation automatique d'images classées.

La reconstruction géodésique est une opération qui vise à éliminer les petites composantes connexes sans modifier la taille ou la forme des objets. L'image originale I est appelée masque et on utilise en général comme marqueur, une érosion de taille n donné. Le marqueur est successivement dilaté et on effectue l'intersection avec le masque. L'opération s'arrête lorsque deux intersections sont égales (figure 3.2). On obtient ainsi une image R où toutes les petites zones ont été éliminées. La figure 3.3 illustre les résultats de

chaque étape de cet algorithme. On remarque sur la figure 3.3d que seuls les pixels connexes à un point de l'intersection du marqueur et de l'image initiale sont reconstruits.

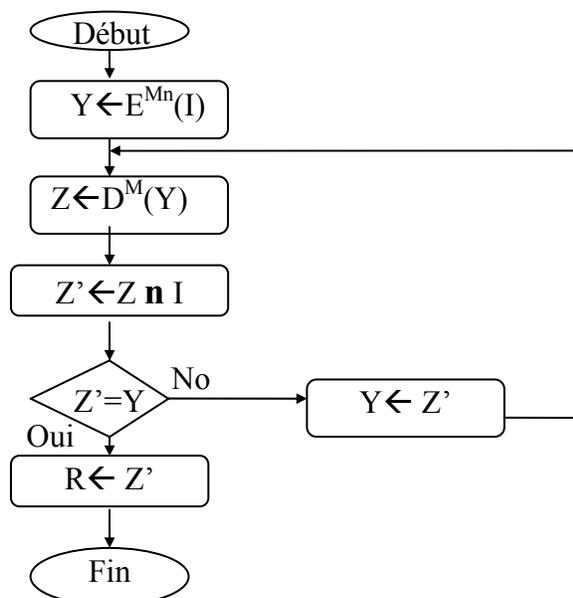


Figure 3.2 : Algorithme de la reconstruction géodésique

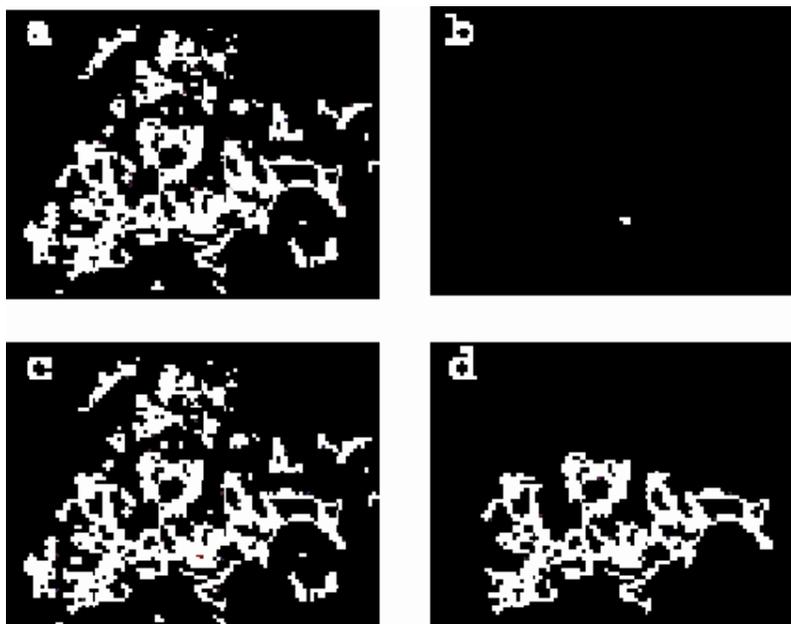


Figure 3.3 : Reconstruction géodésique (d) d'une image binaire (a) Utilisant une érosion de taille 2 comme marqueur (b), masque + marqueur (c).

La reconstruction géodésique est exploitée pour « boucher » de façon efficace les petits trous qui apparaissent dans une zone homogène (figure 3.4). Il suffit d'inverser l'image, d'appliquer une reconstruction géodésique et d'inverser à nouveau l'image. Lorsque l'image est inversée (figure 3.4b), les

petits trous sont considérés comme du « bruit ». En appliquant une reconstruction géodésique, on élimine ce bruit (figure 3.4c). On inverse ensuite l'image résultante pour obtenir une image où les trous sont « bouchés » (figure 3.4d).

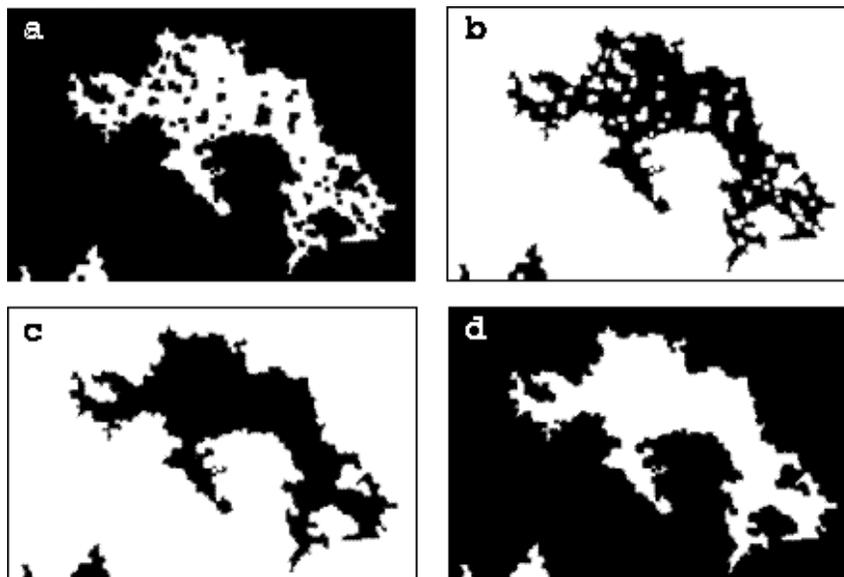


Figure 3.4 : Le « bouchage » des trous sur une image (a) par inversion, reconstruction de l'image inversée (c) et une nouvelle inversion (d).

Les Chapeaux Haut de forme blanc et noir sont des filtres dits morphologiques et sont adaptés pour l'extraction d'objets linéaires fins (sombres ou clairs), d'une épaisseur donnée. Ils ont été exploités par Legeley et al. (1997) pour extraire les failles à partir d'images SPOT panchromatiques. Pour extraire le réseau hydrographique, un chapeau haut de forme blanc de taille 1 avec un élément structurant 5 x 5 (figure 3.5a) est calculé sur la composante 1 de l'ACP et on obtient une image en teintes de gris où le réseau hydrographique est représenté par les plus hautes valeurs de gris. Sur le Chapeau Haut de Forme blanc, nous effectuons deux seuillages avec un choix convenable d'un seuil bas (SB = 10) et d'un seuil haut (SH = 40). On effectue ensuite une reconstruction géodésique du seuil bas en utilisant le seuil haut comme marqueur. Les deux principales rivières de l'image ressortent clairement comme le montre la figure 3.5d malgré quelques bruits qui persistent. On procède ensuite au nettoyage en effectuant un étiquetage automatique des entités connexes (en 8 connexité puis en 4 connexité) suivie d'une extraction des composantes correspondant au réseau.

3.2.4. Procédure de généralisation cartographique

La généralisation cartographique a été mise en œuvre pour produire des unités cartographiques plus homogènes à partir des cartes issues des classifications automatiques soit en éliminant les taches couvrant des petites surface ou en remplissant les trous dans une composante connexe. La procédure comprend quatre étapes principales : extraction et codage de chaque classe, traitement séparé de chaque classe, fusion des classes traitées et généralisation ultime de la carte obtenue.

Une fonction de transfert d'histogramme b_i est appliquée à l'image classifiée et permet de coder chaque classe d'occupation du sol cli en puissance de deux (2^{i-1}). Pour une image de k classes, la fonction b_i est définie par :

$$b_1(x) = \begin{cases} 1 & \text{si } x = 0 \\ 0 & \text{si } x \in [1, k-1] \end{cases} \quad b_i(x) = \begin{cases} 2^{i-1} & \text{si } x = i-1 \\ 0 & \text{si } x \in [0, i-2] \cup [i, k-1] \end{cases} \quad b_k(x) = \begin{cases} 2^{k-1} & \text{si } x = k-1 \\ 0 & \text{si } x \in [0, k-2] \end{cases}$$

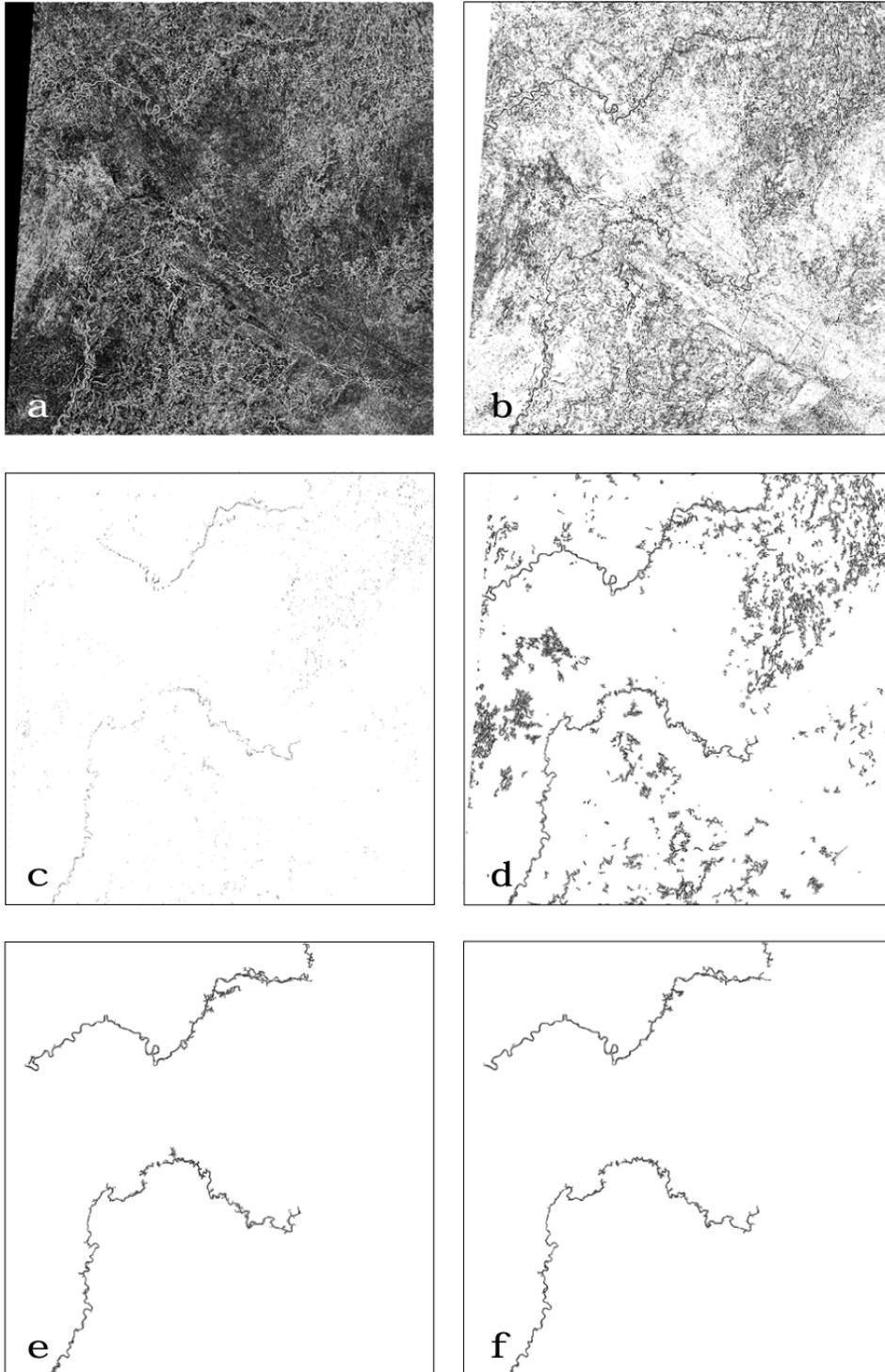


Figure 3.5 : Extraction du réseau hydrographique (f) à partir du Chapeau Haut de forme morphologique blanc WTH (a), par reconstruction géodésique (d) du seuil bas, SB = 10 (b) avec le seuil haut, SH = 40 (c) utilisé comme marqueur. Etiquetage en 8 connexité (e), puis 4 connexité et nettoyage (f).

Le traitement d'une classe consiste à éliminer les entités connexes de petite taille qui sont considérées comme du bruit en appliquant une séquence d'opérateurs de Morphologie Mathématique. Par exemple, une reconstruction géodésique est utilisée pour nettoyer les éléments des petites surfaces et remplir les trous des composantes connexes (figure 3.6b). Puisque des trous persistent, une fermeture peut être utilisée pour joindre les éléments voisins, mais ce traitement induit de larges imperfections (figure 3.6c). On utilise une technique de bouchage de trous qui consiste à inverser l'image, appliquer une reconstruction géodésique et inverser à nouveau le résultat. Pour une image donnée, la même suite de traitement est appliquée à chaque classe en utilisant les mêmes paramètres (nombre d'érosion ou de fermeture et taille de l'élément structurant) de manière à conserver l'importance relative de chaque classe. Le choix de ces paramètres dépend de l'échelle de généralisation qui est la taille approximative des plus grandes entités à éliminer. En théorie, la fusion des classes traitées produit une carte ayant un nombre de classes compris entre k et 2^k-1 . Ce nombre dans la pratique dépend du niveau d'homogénéité de l'image classée et de la qualité du traitement des classes. L'apparition des nouvelles classes correspond à des zones n'appartenant à aucune classe soit à zones appartenant à plusieurs classes après la fusion. La généralisation ultime permet d'affecter chacune des unités de ces nouvelles classes à une des classes initiales. Cette opération consiste en une combinaison de traitement automatique et une interprétation visuelle avec la classification initiale comme référence.

3.2.5. Résultats et Validation

Pour chacune des deux images (1987 et 1999), huit principales classes d'occupation du sol ont été identifiées : habitation, maraîchers/vergers, cultures pluviales, sols nus, sorgho de contre saison, feu, savane boisées, montagnes. Le croisement des deux cartes sous le SIG fournit des informations sur les grandes évolutions et conversions de l'occupation du sol. L'analyse à plusieurs niveaux permet d'avoir des connaissances plus pertinentes sur la diversité des situations et de mieux comprendre les stratégies des acteurs face aux différentes mutations. Une analyse plus détaillée et substantielle de ces résultats est donnée au Chapitre 8.

Validation des thèmes d'occupation du sol

La validation de la qualité sémantique des unités d'occupation du sol est effectuée sur la carte la plus récente. La source de référence utilisée est un ensemble de données relevées au GPS sur des sites où la nature de l'occupation du sol est identifiée et caractérisée sur un rayon de 100 m autour du point. Un effectif de 525 points a été utilisé et la répartition entre les classes d'occupation du sol dépend de l'importance des superficies et des difficultés de détection de chaque classe. Une jointure spatiale a été effectuée entre la carte des points GPS et la carte d'occupation du sol de 1999 pour produire une matrice de confusion. La précision globale mesurée par le coefficient Kappa est de 0.733. Les erreurs (excédent ou déficit) de chaque classe donne des indications détaillées sur les limites de réalisation ou d'utilisation de la carte obtenue (Girard et Girard, 1999). Ces informations sur les erreurs de réalisation ont été localisées et des corrections ont été effectuées afin d'améliorer la précision des cartes obtenues.

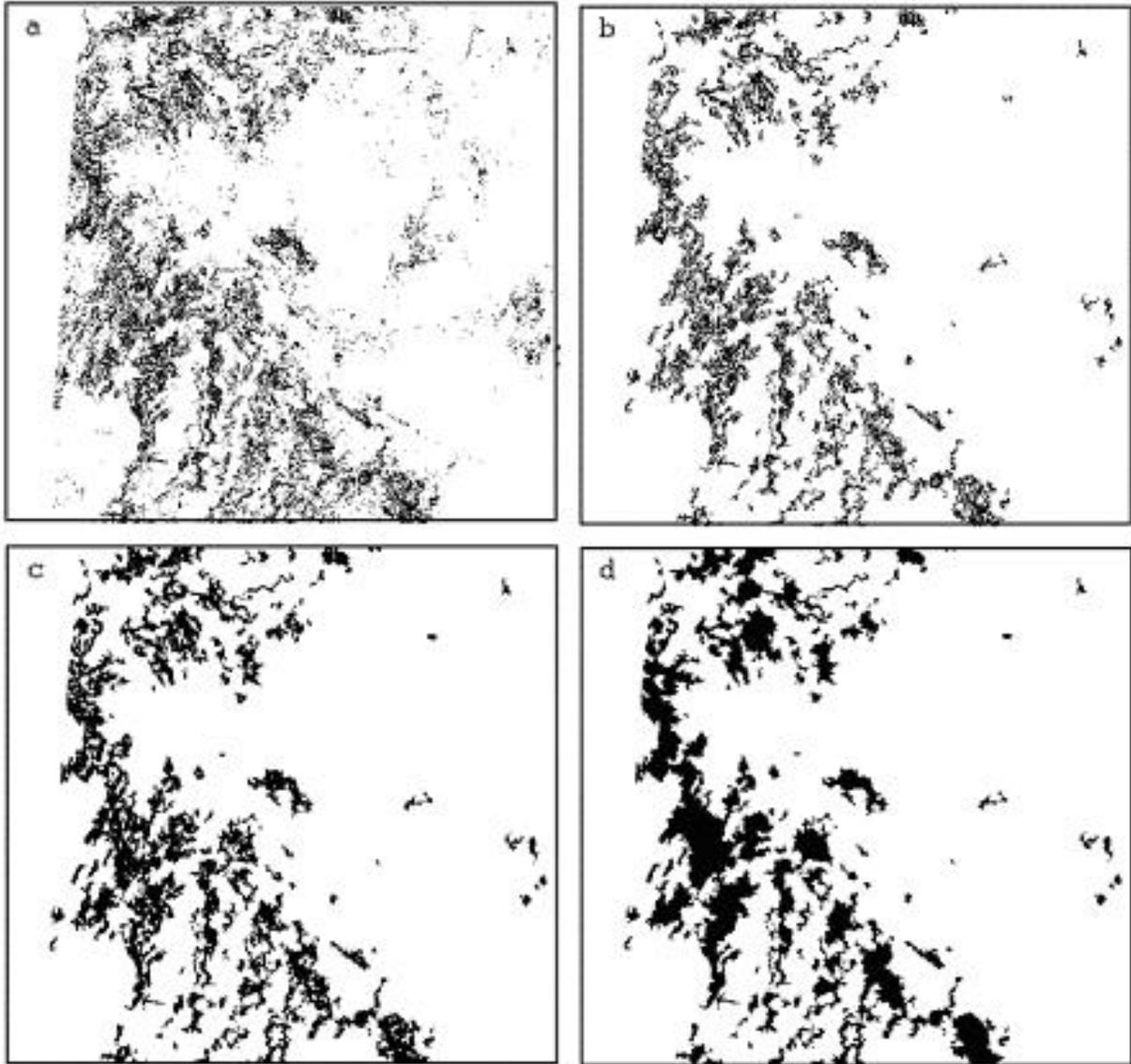


Figure 3.6 : Exemple de traitement d'une classe. a - classe extraite ; b - reconstruction géodésique avec érosion de taille 2 et un élément structurant 3x3 ; c - fermeture de taille 1 avec élément structurant 5x5 ; d - bouchage des trous.

Validation de la procédure de généralisation cartographique

Une première méthode simple de validation de la procédure de généralisation cartographique consiste à calculer les histogrammes de l'image classifiée et de l'image issue de la généralisation et comparer les proportions de chaque type d'occupation du sol. On constate une quasi-conservation de l'importance de chaque type d'occupation du sol après la généralisation. Toutefois, ce résultat ne donne aucune indication sur les changements de la structure de l'occupation du sol au cours de la procédure de généralisation. La différence entre les structures de la carte brute obtenue par classification multispectrale et la carte généralisée peut être appréciée visuellement (figure 3.7).

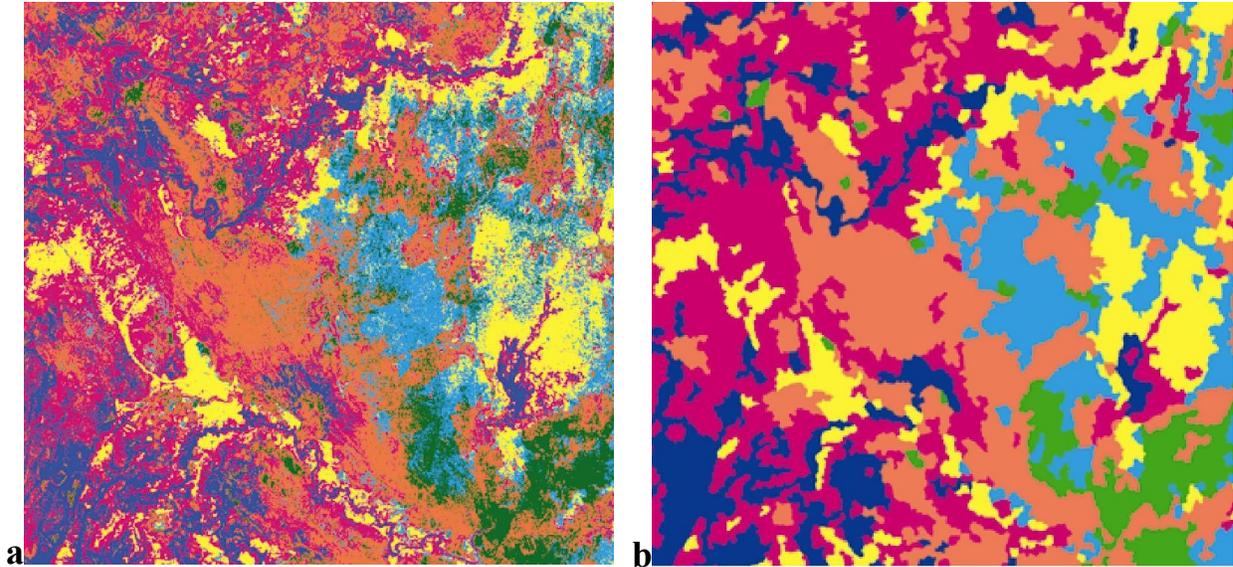


Figure 3.7. Illustration du résultat de la procédure de généralisation cartographique. a. Image classifiée à partir de IB et NDVI. b. Image issue de la généralisation cartographique

La comparaison visuelle donne une idée mais reste une méthode subjective. L’algorithme multi-résolution proposé par Costanza (1989) a été utilisé pour calculer des coefficients de similarité entre les structures d’occupation des deux cartes. Cet algorithme est également utilisé au chapitre 9 pour effectuer un test qui permet de déterminer une plage de résolution appropriée pour représenter l’ensemble des données. Il est ensuite utilisé au chapitre 10 pour évaluer les performances du modèle d’allocation des changements d’utilisation de l’espace en comparant la carte produite par le modèle avec une carte de référence (réalité ou résultat d’un autre modèle). Le principe de l’algorithme de Costanza (1989) repose sur l’idée selon laquelle la mesure à une seule résolution est insuffisante pour décrire et comparer des structures complexes. Ainsi, la comparaison est effectuée en mesurant la similarité des structures sur plusieurs fenêtres de taille variable. Pour une fenêtre de taille w centrée sur une cellule, le coefficient d’ajustement ou qualité de la correspondance entre les deux scènes comparées est défini par la formule suivante:

$$F_w = \frac{\sum_{s=1}^{tw} \left[1 - \frac{\sum_{i=1}^p |a_{1i} - a_{2i}|}{2w^2} \right]^s}{t_w}, \text{ où } a_{ki} \text{ est le nombre de cellules de type } i \text{ dans la scène } k \text{ pour la fenêtre de}$$

calcul, p est le nombre de type différents et t_w est le nombre total de fenêtre de la scène pour une taille w donnée. Le coefficient d’ajustement global entre les deux scènes noté F_t est une moyenne pondérée des coefficients d’ajustement pour différentes tailles de fenêtre où un poids plus important est accordé aux fenêtres de petite taille sans toutefois ignorer les fenêtres plus grandes. F_t est calculé en utilisant la formule suivante :

$$F_t = \frac{\sum_{w=1}^n F_w e^{-k(w-1)}}{\sum_{w=1}^n e^{-k(w-1)}}, \text{ où } k \text{ est une constante indiquant le poids donné aux fenêtres de faible taille par}$$

opposition aux fenêtres de grande taille.

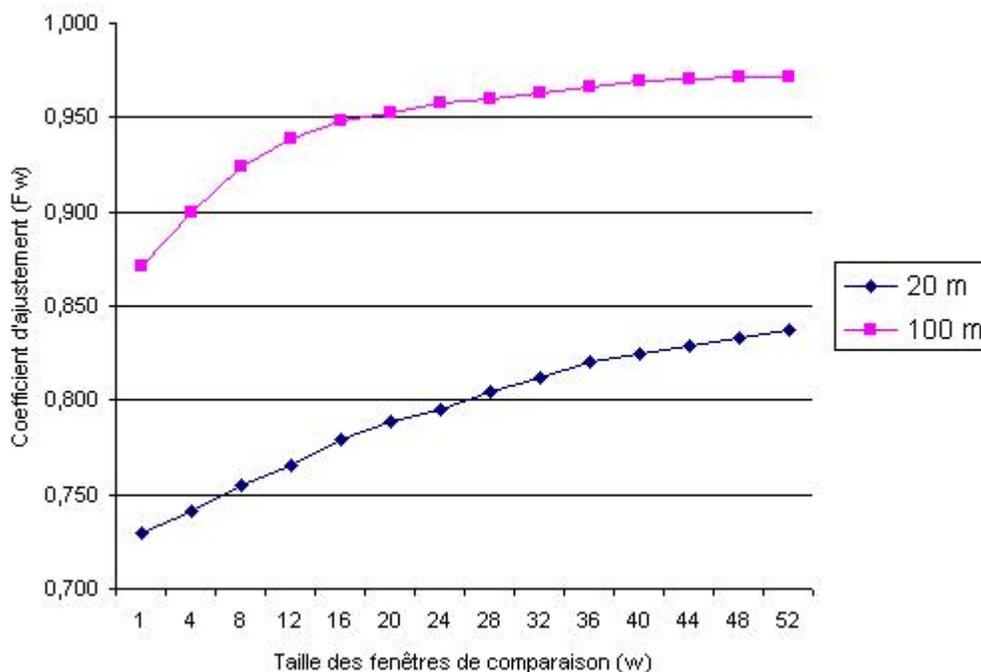


Figure 3.8 : Validation de la procédure de généralisation par comparaison des structures de la classification brute et de la classification généralisée.

Les deux cartes comparées ayant une résolution de 20 m, les résultats montrent que le coefficient de similarité n'est pas assez élevé (0,76) et la courbe des coefficients est sensiblement une droite indiquant des différences perceptibles entre les structures des deux images. Ceci s'explique par le fait que les unités de faibles tailles ont été éliminées au cours de la généralisation cartographique modifiant ainsi les microstructures de l'image. Toutefois, en ramenant les deux cartes à 100 m de résolution, on obtient un coefficient de similarité global assez élevé 0.92 et la courbe des coefficients de similarité pour les fenêtres successives de comparaison a une forme qui traduit une meilleure correspondance des structures entre les deux scènes (figure 3.8). Ce résultat amène à conclure qu'il serait donc plus pertinent d'utiliser la carte obtenue à une résolution supérieure à 100 m. Cette résolution correspond d'ailleurs à la taille de l'unité d'observation des vérités de terrain utilisées pour l'analyse d'image et renseigne sur le niveau de précision des résultats cartographiques.

La validation de la qualité sémantique des unités d'occupation du sol effectuée dans ce cas n'est pas uniquement un moyen d'évaluation de la méthode, mais une étape de la méthode dans la mesure où elle permet d'affiner la photo-interprétation et fournit à l'utilisateur des informations sur les limites de réalisation et d'utilisation de la carte. Les résultats montrent que cette approche combinant traitement

numérique d'image, cartographie participative et photo-interprétation est un moyen réaliste pour dériver les structures d'occupation du sol pertinentes et réunir les connaissances sur les facteurs déterminants ces structures et leurs changements. La cartographie participative assistée avec les images satellites révèle de nombreux détails concernant le repérage des changements dans l'espace et le temps ainsi que sur les causes possibles. Dans le contexte spécifique des zones des savanes d'Afrique, cette approche représente donc une perspective intéressante dans le développement des techniques d'interview pour l'exploration des changements d'occupation du sol et d'utilisation de l'espace.

3.3. Outils de SIG et d'analyse spatiale

Le terme SIG se définit de différentes manières. La première perception du terme qui est la plus courante considère le SIG comme un outil logiciel permettant l'acquisition, le stockage, l'analyse et l'affichage des données qui ont une référence spatiale (Laurini et Thompson, 1996). Une seconde perception du terme SIG le considère comme un modèle, c'est-à-dire une représentation simplifiée du monde réel. Le troisième point de vue qui est étymologiquement le plus correct considère le SIG comme une entité organisée qui comprend un ensemble de données, de ressources humaines, matérielles et logicielles permettant de gérer les données à référence spatiale (Gayte et al, 1997). Dans ce chapitre c'est la première conception qui est considérée. Toutefois nous pourrions faire référence à la deuxième pour indiquer le modèle de données spatial de l'application développée. Dans cette thèse nous utilisons le terme SIE pour faire référence à la troisième perception. Les données à gérer dans un SIE sont presque toujours localisées et il serait impensable de concevoir un SIE sans accorder une place importante aux SIG. C'est ce qui justifie la confusion qui est le plus souvent faite entre SIG et SIE. Compte tenu de notre démarche d'analyse des dynamiques d'utilisation de l'espace qui prend explicitement en compte les références spatiales des processus étudiés, les SIG jouent un rôle important dans les différentes étapes de l'analyse et de la mise en œuvre des applications. Il assure deux principales fonctions à savoir l'acquisition et la gestion des données géographiques, l'analyse spatiale et la présentation des informations géographiques. La figure 3.9 montre comment chacune de ces fonctions contribue à la mise en œuvre de chacun des principaux modules du SIE SMALL Savannah notamment le diagnostic et la caractérisation, l'identification des facteurs déterminants et la simulation des dynamiques spatiales de l'utilisation de l'espace.

3.3.1. Acquisition et gestion des données géographiques

Le SIG permet d'abord d'importer et de géoréférencer les données issues de la télédétection (images satellites et relevées GPS) et les cartes existantes. Les traitements dépendent du format de la source de données. Les données au format vecteur sont directement intégrées dans la base de données géographique avec d'autres sources de données. Les données au format raster sont digitalisées et les données attributaires y sont associées. Les fonctions de superposition des couches d'information du SIG sont utilisées pour faciliter le diagnostic et la caractérisation du système d'utilisation de l'espace en combinant les données et connaissances issues de sources diverses : revue de la littérature, statistiques agricoles, recensement (population, cheptel), analyse d'images satellites ou de cartes existantes (Bocco et Toledo, 1997).

3.3.2. Analyse spatiale et présentation des résultats de simulation

Les opérations d'analyse spatiale du SIG sont utilisées sous deux formes : la première consiste à combiner les données spatiales multi-dates et multi-thèmes afin de détecter les changements et formuler des hypothèses sur les facteurs potentiellement déterminants ; la deuxième consiste à construire des variables spatiales qui traduisent l'inscription spatiale des phénomènes étudiés comme par exemple la pression humaine sur l'espace et l'influence de l'accessibilité ou de l'aptitude des sols sur les changements d'utilisation de l'espace. Les fonctions de présentation graphique des SIG sont utilisées pour visualiser les résultats de la simulation qui peuvent être superposés avec d'autres couches d'information pour faciliter la localisation et l'interprétation des dynamiques d'utilisation de l'espace.

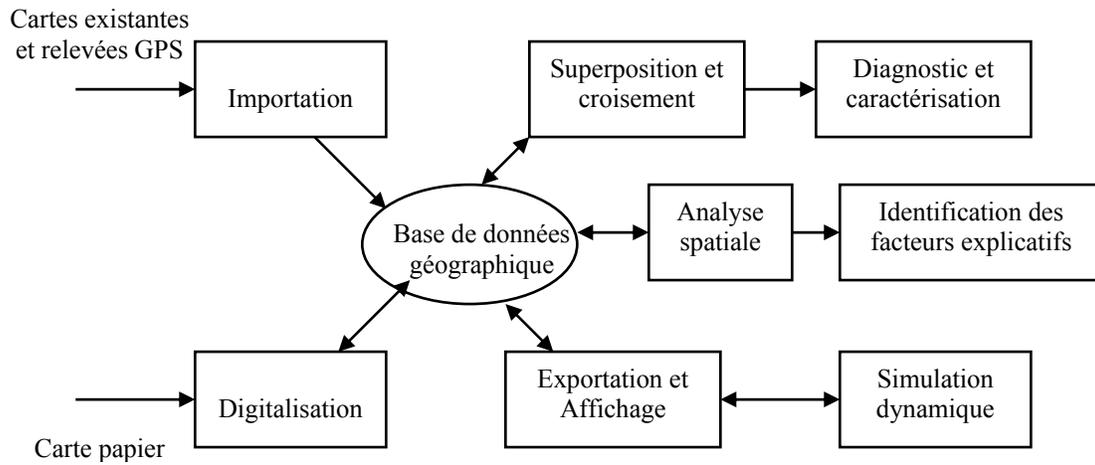


Figure 3.9 : Rôle des outils de SIG dans l'analyse intégrée des dynamiques d'utilisation de l'espace.

3.4. Outils et modèles d'analyse statistique

3.4.1. Modèle d'analyse statistique

L'analyse quantitative de la relation entre l'utilisation de l'espace et les facteurs déterminants est effectuée afin d'identifier les facteurs qui se combinent pour influencer significativement la distribution spatiale de chaque utilisation et mesurer l'importance de chaque facteur dans cette interaction. Le modèle de régression logistique est adapté pour ce type de problème où la variable dépendante est de type binaire (Agresti and Finlay, 1997). En effet la carte de chaque utilisation de l'espace LU est de type binaire c'est-à-dire $LU(x, y) = 0$ pour l'absence ou $LU=1$ pour la présence de cette utilisation à la position i de coordonné (x, y) . Pour une seule variable indépendante X représentant un facteur déterminant potentiel, le modèle de régression simple est défini par :

$$P_i(LU = 1) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{(\beta_0 + \beta_1 X)}}, \text{ qui peut encore s'écrire } P_i(LU = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

où P_i est la probabilité de l'évènement « LU=1 dans la portion i de l'espace ». Plus généralement et de façon analogue, pour plusieurs variables dépendantes X_1, X_2, \dots, X_p , représentant les facteurs déterminants potentiels, le modèle de régression multiple est défini par :

$$P_i(LU = 1) = \frac{1}{1 + e^{-z}} \quad , \text{ où } z \text{ est la combinaison linéaire } Z = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

Des deux équations précédentes on déduit que :
$$\text{Log}\left(\frac{P_i}{1 - P_i}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

La relation entre les variables indépendantes et cette probabilité est non linéaire et la probabilité estimée est toujours entre 0 et 1. La valeur Z peut être interprétée comme une mesure de la préférence à affecter l'utilisation LU à la portion i de l'espace. De cette façon on arrive à calculer pour toutes les portions de l'espace, la probabilité d'affectation de chaque utilisation de l'espace qui pourrait être utilisée pour définir les préférences en cas de compétition entre plusieurs utilisations de l'espace. Les régressions multiples de type logistique sont ainsi utilisées pour construire des modèles de prédiction de la distribution spatiale des différentes utilisations de l'espace sous l'influence des différents facteurs déterminants.

3.4.2 Procédure d'exploration des facteurs et de construction des modèles de prédiction

La construction des modèles de type empirique est difficile à cause du nombre important de facteurs qui interagissent pour influencer la structure et la dynamique de l'utilisation de l'espace. De plus, ces facteurs peuvent changer de nature ou d'importance au cours du temps. Une des difficultés qui se posent est de sélectionner les variables les plus pertinentes à inclure dans le modèle. La procédure d'exploration et d'identification des facteurs déterminants proposée ici est principalement de type inductive. Toutefois, elle intègre les aspects de raisonnement déductif. Elle commence par des observations de la réalité et cherche des explications à partir de l'analyse des liens ou des régularités dans les données. A l'opposé, le raisonnement déductif part d'une connaissance théorique des processus, formule des hypothèses qui sont testées dans le monde réel pour fournir des explications aux phénomènes (Overmars et al., 2006). L'idée principale de la méthode est de tirer profit à la fois profit des connaissances sur les processus et des possibilités offertes par les données spatiales explicites disponibles. Lorsqu'on ne dispose d'aucune théorie ni connaissance sur un processus, on explore les données à la recherche des corrélations et structure explicatives. Lorsqu'on a une connaissance même approximative ou théorique des facteurs pertinents pour une explication, celle-ci sert à formuler des hypothèses qui sont testées avec l'exploration des données disponibles. On arrive ainsi à construire des modèles de prédictions robustes c'est à dire qui sont assez représentatifs des processus en cours et à venir.

On peut distinguer plusieurs situations qui justifient l'intérêt de l'approche d'exploration des facteurs utilisée. Le pire cas est celui où un processus est observé, une bonne connaissance sur celui-ci est disponible mais, les moyens d'acquérir les données spatiales explicites pour le mettre en évidence font défaut. Dans certains cas, on dispose des données spatiales explicites, mais l'analyse des données ne permet pas de mettre en évidence le processus pourtant observé du fait que les données ne sont pas collectées à l'échelle appropriée. Inversement, dans une autre situation, la relation peut être trouvée entre les données mais ne présente pas d'intérêt soit par manque de connaissances sur le processus, soit parce

qu'on ne s'y intéresse pas. Ce dernier cas est caractéristique des approches réductionnistes et disciplinaires où on s'intéresse à étudier l'influence de facteurs prédéfinis. Une autre situation, intermédiaire, est celle où l'on n'a pas de certitude sur la pertinence du processus et l'on ne dispose pas de données explicatives sur la relation observée.

Les étapes suivies pour la construction des modèles de prédiction de chaque utilisation de l'espace sont illustrées à la figure 3.10. Le diagnostic et la caractérisation du système d'utilisation de l'espace de la zone d'étude permettent de formuler les hypothèses sur les processus et la manière dont ils opèrent dans l'espace. Ces informations sont utilisées pour calculer des variables représentatives des phénomènes étudiés en utilisant les données spatiales disponibles. La première étape de l'algorithme consiste à construire des régressions simples entre l'utilisation de l'espace et toutes les variables. Pour limiter les effets de multi-colinéarité, certaines variables sont exclues du modèle lorsqu'elles sont à priori redondantes avec d'autres variables. La deuxième étape consiste à sélectionner les variables qui contribuent plus significativement à l'explication de la structure et les changements d'utilisation de l'espace. Les informations utilisées sont les paramètres statistiques des régressions simples et les connaissances disponibles sur les processus modélisés. Les trois paramètres statistiques sur lesquels s'appuie la sélection sont : le sens de la corrélation déterminé par le signe du coefficient β_i , la signification statistique de la corrélation et le pouvoir explicatif de la variable représenté par la valeur ROC ou Relative Operating Characteristics (Pearce et Ferrier, 2000; Pontius et Schneider, 2001).

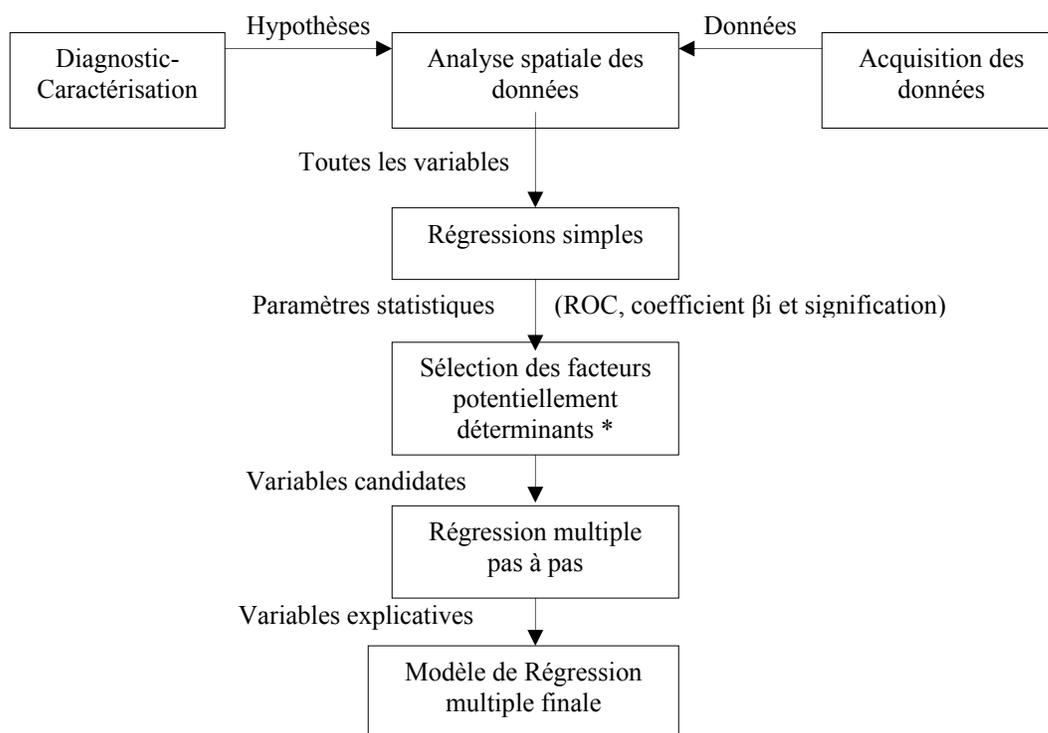


Figure 3.10 : Algorithme de construction des modèles de prédiction de chaque utilisation de l'espace

Pour toutes les régressions simples, la variable est rejetée si le sens de la corrélation est contraire aux connaissances qu'on dispose sur le processus. Le niveau de signification statistique de la corrélation est une évaluation de la contribution de la variable indépendante à l'explication de la variable dépendante.

Dans un test statistique, c'est en général à l'utilisateur de déterminer la valeur qui convient à sa situation. Une valeur inférieure à 0,05 est le plus souvent considérée comme significative, soit un coefficient de confiance de 95%. Dans cette application la plupart des variables retenues sont significatives avec un niveau inférieur à 0.01. Toutefois, quelques variables ont été incluses dans les modèles avec un niveau compris entre 0.01 et 0.9. La valeur ROC est une mesure de la qualité de la régression logistique tout comme le coefficient de détermination R^2 de la régression linéaire. La valeur de la variable dépendante peut être prédite à « présente » ou « absente » si la probabilité prédite est supérieure ou inférieure à une probabilité seuil dont le choix dépend des besoins de précision du modèle. Les cellules peuvent être réparties en fonction de la conformité entre la réalité et le modèle comme le montre le tableau de la figure 3.11. La sensibilité définie par $\frac{A}{A+C}$ est la fraction de cellules de valeur « présente » qui ont été prédites à « présente ». La spécificité définie par $\frac{D}{B+D}$ représente la fraction de cellule de valeur « absente » qui ont été prédites à « absente ». La valeur ROC représente la superficie de la courbe décrite par les valeurs de la sensibilité et de la spécificité pour un ensemble de probabilité seuil. Un exemple de courbe ROC d'un modèle de prédiction est donné à la figure 3.11. Un modèle aléatoire donne une valeur de 0.5 alors qu'une correspondance parfaite donne une valeur ROC de 1.

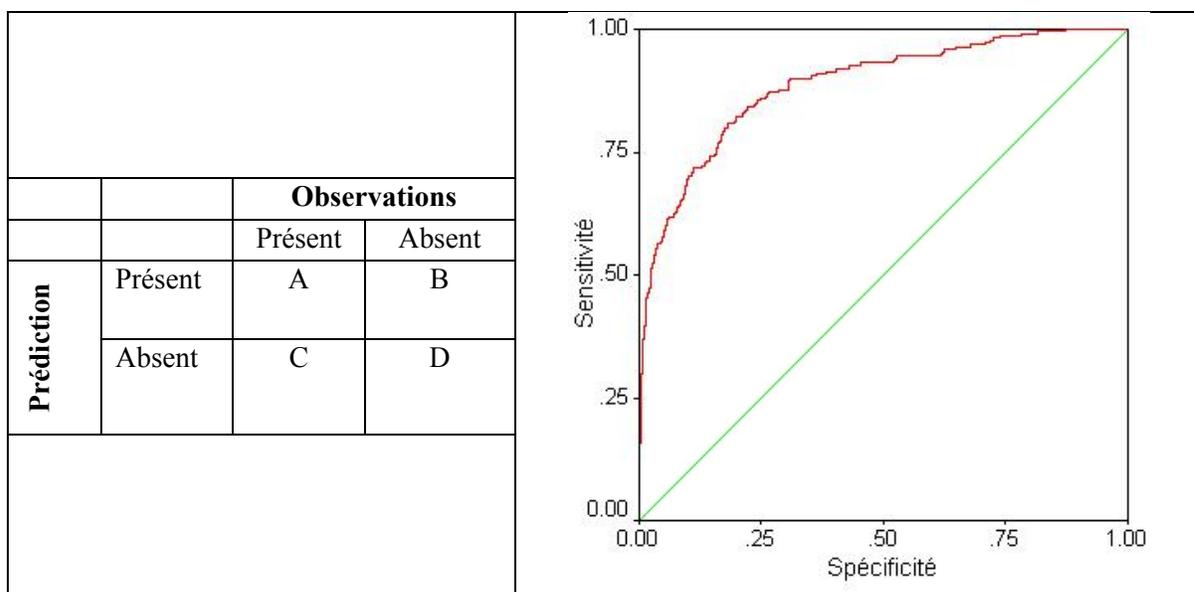


Figure 3.11 : Exemple de courbe ROC du modèle de prédiction de l'utilisation de l'espace « Habitation ». ROC = 0,887 représente la valeur de la surface en dessous de la courbe.

Deux valeurs seuils ROCs1 et ROCs2 telles que $ROCs1 < ROCs2$ sont fixées comme performances minimales exigées respectivement pour les variables candidates ou déterminantes par hypothèse et les variables non-candidates ou non déterminantes par hypothèse. Ces valeurs seuils peuvent varier d'une utilisation de l'espace à une autre en fonction des précisions de prédictions attendues. Une variable déterminante par hypothèse est sélectionnée lorsque son pouvoir explicatif est supérieur à ROCs1. Une variable non déterminante par hypothèse est sélectionnée si le pouvoir explicatif est supérieur à ROCs2 et si elle traduit un processus hypothétique même si l'on ne dispose pas assez d'information sur ce dernier.

La troisième étape consiste à effectuer une régression pas à pas avec l'ensemble des variables sélectionnées. Cette étape permet également de trancher sur certains cas de multi-colinéarité non résolus à l'étape 1 en éliminant du modèle final certaines variables. La dernière étape consiste à calculer le modèle de régression multiple finale qui permet ensuite de calculer la probabilité pour qu'une cellule soit affectée à une utilisation de l'espace. Les performances de ces modèles sont également évaluées en utilisant la méthode ROC (Pearse, et Ferrier, 2000; Pontius et Schneider, 2001). Un bon modèle de prédiction doit être sûr et avoir une bonne capacité de discrimination. La sûreté est considérée dans le sens où chaque prédiction est une estimation précise de la probabilité de rencontrer une utilisation sur une cellule donnée. La capacité de discrimination est différente et nécessite une approche de mesure différente. Dans ce dernier cas, le modèle doit être capable de distinguer correctement les sites occupés de ceux non occupés par une utilisation, indépendamment de la sûreté des prédictions. (Pearce et Ferrier, 2000). Les résultats de cette procédure d'exploration des facteurs déterminants sont présentés au chapitre 9 pour un ensemble de données sur une petite région du Nord Cameroun. Les modèles obtenus sont ensuite utilisés au chapitre 10 pour prédire la distribution de l'utilisation de l'espace. La comparaison de l'évaluation des performances de prédiction de plusieurs modèles indique que le modèle le plus pertinent n'est pas nécessairement celui correspondant aux hypothèses, ni celui correspondant aux meilleures corrélations fournies par les données mais un compromis, tendant à refléter la réalité des processus qui opèrent dans le système d'utilisation de l'espace.

3.5. Outils et modèles de simulation

3.5.1. Typologie des modèles de changement d'utilisation de l'espace

Les modèles de changement d'utilisation de l'espace en particulier sont conçus principalement pour répondre à quatre types de questions fondamentales : pourquoi, quand et où les processus opèrent-ils, quels sont les effets et les conséquences de ces processus dans l'espace et dans le temps ? Une très large gamme de modèles appliqués à l'utilisation de l'espace a été développée pour répondre à ces questions et des revues assez complètes ont été réalisées par Baker (1989) et Lambin (1994). Des revues plus récentes sur les modèles d'utilisation de l'espace ont été effectuées par Briassoulis (2000), Irwin et Geoghegan (2001), Sanders (2001), Lambin et al. (2003) et Parker et al. (2003). Dans cette étude, nous nous intéressons aux modèles de simulation qui sont développés pour prédire les changements des structures d'utilisation de l'espace. La classification de ces modèles peut s'appuyer sur plusieurs critères : le type de questions auxquelles ils permettent de répondre ou les domaines d'applications privilégiés, les possibilités techniques et la base théorique ou la manière dont les processus sont appréhendés. On peut distinguer trois principales classes de modèles qui dépendent principalement de la manière dont les processus sont appréhendés : les modèles empiriques, les modèles mécanistes et les modèles de type système (figure 3.12).

Les approches de modélisation spatiale de types empiriques basées sur les transitions de probabilité, sont nées de la combinaison de la télédétection, des SIG et des modèles mathématiques multi-variés ou multi-temporels. La construction de ces modèles qui sont en général de type spatial et statistique s'appuie sur une hypothèse de stationnarité du processus : on suppose que les relations observées entre les variables restent valables au cours de la période de simulation. Les techniques statistiques sont utilisées à cet effet

afin de sélectionner les facteurs déterminants les plus importants et de quantifier leurs relations avec l'utilisation de l'espace. L'implémentation dans les modèles de simulation se fait par une translation directe de l'analyse empirique en probabilité de transition (Bell et Hinojosa, 1977 ; Veldkamp et Fresco, 1996).

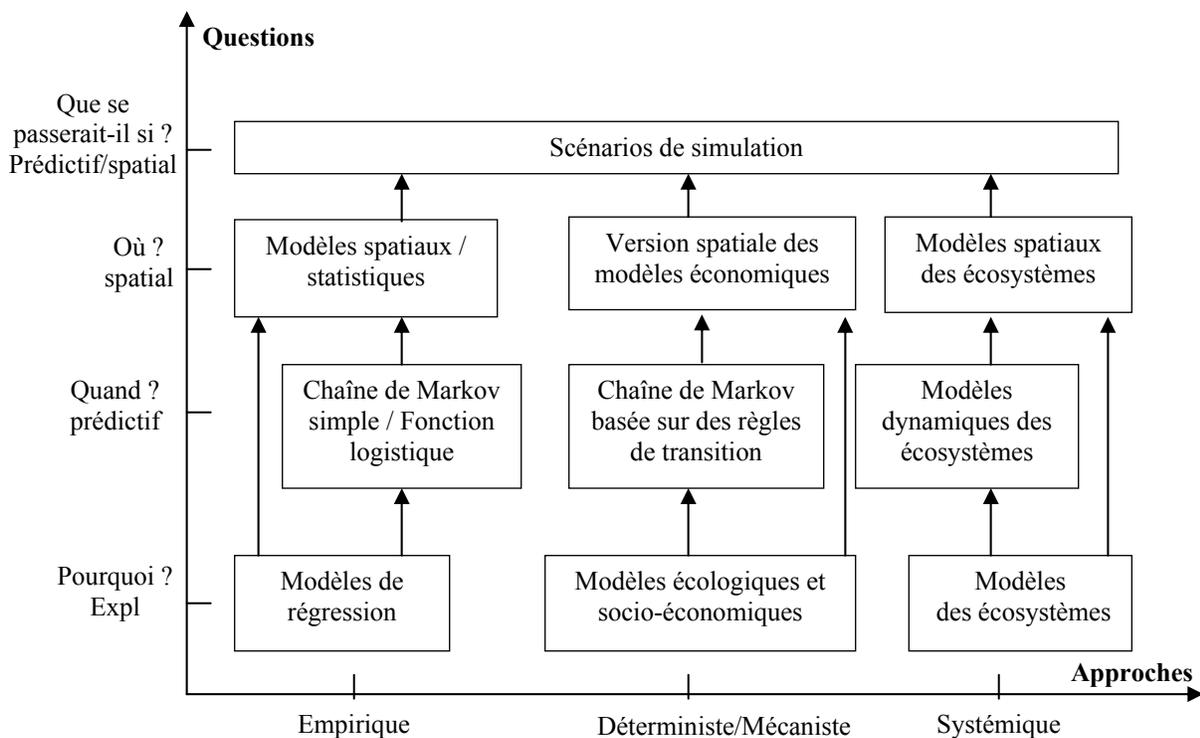


Figure 3.12 : Typologie des modèles de simulation des dynamiques d'utilisation de l'espace.

Les modèles mécanistes qui sont le plus souvent de type déterministe permettent de représenter les interactions entre les types d'utilisation de l'espace et les rétroactions dans le système. Ceci suppose une connaissance des processus qui gouvernent le fonctionnement du système. Les processus individuels sont modélisés en utilisant des lois ou règles décrites par des équations simples. Les modèles d'optimisation de type équation différentielle ou de type programmation linéaire (Njiti et Sharpe, 1994 ; Roetter et al., 2005) appartiennent à cette classe. Dans ces modèles mathématiques, la relation entre l'utilisation de l'espace et ses facteurs déterminants est supposée connue. Les modèles de type système décrivent plusieurs processus complexes en interaction en insistant sur les relations, les interactions entre tous les composants. Ils permettent ainsi de mieux simuler les interactions entre les systèmes sociaux et environnementaux. Les modèles de type multi-agents permettent de représenter explicitement les interactions entre agents et fournissent ainsi un cadre qui permet de prendre en compte les dynamiques de systèmes et les logiques individuelles ou collectives de prise de décision (Bousquet et al., 1998 ; Parker et al., 2003). Les modèles basés sur les automates cellulaires permettent d'implémenter les interactions entre unités spatiales en prenant en compte les relations de voisinages entre elles. Ces types de modèles sont appropriés pour la simulation des systèmes complexes mais, leur mise en œuvre nécessite également une bonne connaissance des interactions élémentaires.

Chacune des classes de modèles présente ses avantages et ses inconvénients. Même au sein de chaque classe, les modèles sont différents en fonction des objectifs, des disciplines ou des échelles d'analyse, mais chaque modèle a une contribution significative pour représenter les processus et prédire leur évolution. Toutefois, l'utilisation exclusive de plusieurs de ces modèles ne permet pas toujours de comprendre le comportement du système de façon assez complète à cause des contraintes liées à l'approche, à l'échelle ou à une discipline. C'est ce constat qui a stimulé très récemment un besoin de développement de modèles intégrés au sein de la communauté scientifique intéressé par la modélisation des changements d'utilisation de l'espace (Lambin, 1994 ; Rielsame et al., 1994 ; Lambin et al., 2003).

3.5.2. Démarche et choix du type de modèle de simulation de changement d'utilisation de l'espace

La démarche suivie dans le cadre de cette thèse avait pour objectif de développer un modèle intégré qui tire profit des articulations ou combinaisons possibles entre les différentes classes de modèles existants. Le plus souvent, les comparaisons faites aux sujets des modèles ne portent ni sur la même région, ni sur les mêmes processus, ni dans les mêmes conditions (Pontius et al, 2007). Théoriquement, il aurait été intéressant de réaliser un modèle de chacune des trois classes sur la même zone d'étude et de comparer les résultats pour dégager les contributions respectives de chaque type de modèle avant d'étudier les possibilités d'intégration. Ce travail est difficile à cause des contraintes de temps et surtout du fait de l'indisponibilité des données et des connaissances assez limitées qu'on dispose sur le fonctionnement interactif et dynamique entre les systèmes socio environnementaux dans la zone d'étude. Une deuxième possibilité serait de concevoir directement un modèle intégrant les aspects complémentaires des différents types de modèle. Cette approche semble plus logique mais très difficile à mettre à œuvre aussi bien sur le plan théorique que pratique. La troisième possibilité qui a été suivie qui semble plus réaliste était de partir d'un modèle et de le faire évoluer en intégrant les aspects des deux autres types de modèle. Les travaux antérieurs ont montré que lorsqu'un système devient très complexe pour être décrit analytiquement, une solution empirique doit nécessairement être recherchée (Easterling, 2000), les modèles hypothétiques de type déterministe étant limités compte tenu du fait qu'on a une compréhension incomplète du système (Kok, 2001). Le point fort des approches de modélisation spatiale de type empirique réside dans la distribution spatiale des éléments du paysage et des changements de sa structure dans la mesure où il permet de projeter et afficher dans un format cartographique la structure du paysage futur qui résulterait de la continuation des pratiques actuelles de gestion de l'espace. Les approches de type empirique constituent donc un bon point de départ de notre démarche d'analyse du système agricole qui vise à identifier les problèmes, décrire le système et concevoir les trajectoires d'évolution (chapitre 1).

Les modèles de type transition de probabilité ont été étudiés et se trouvent appropriés pour l'étude de la transformation des paysages ainsi que l'analyse des processus de diffusion spatiale. Les chaînes de Markov appartiennent à ces approches stochastiques permettant de décrire des processus qui se déplacent dans une séquence à travers un ensemble d'états. Le modèle des chaînes de Markov est défini par une fonction qui lie l'utilisation de l'espace à chaque instant avec l'utilisation la plus récente et les probabilités de transition d'un état à un autre au cours d'un intervalle de temps (Bell et Hinojosa, 1977). La principale difficulté de la mise en œuvre de ce modèle réside dans la définition de la fonction de transition. Les probabilités de transition peuvent être estimées statistiquement à partir d'un échantillon de transitions survenues au cours d'un intervalle de temps donné. Ceci nécessite d'importantes quantités de données qui ne sont pas toujours

disponibles dans le contexte de la zone d'étude. En l'absence de données détaillées sur les transitions, les probabilités de transition peuvent être définies en formulant des règles de comportement des éléments du paysage (Lambin, 1994). Un autre moyen consiste à les relier à un ensemble de variables exogènes et dans ce cas, les probabilités de transition sont définies par les modèles de régression. CLUE (Conversion of Land Use and its Effects) est un exemple de cadre de modélisation dynamique basée sur cette approche empirique qui a été conçu pour explorer les effets des facteurs déterminant l'utilisation de l'espace et de simuler les changements en réponse à la croissance démographique. Le modèle dynamique prend en compte le système d'interaction complexe entre l'utilisation passé et présente de l'espace, les conditions socio-économiques et les contraintes biophysiques pour allouer les changements (Verburg et al., 1999).

3.5.3. La plateforme de modélisation CLUE

Structure de la plate-forme et modèle d'allocation des changements

La plateforme de modélisation CLUE (Conversion of Land use and its Effects) a été développée pour simuler les changements d'utilisation de l'espace en s'appuyant sur une analyse des relations quantitatives entre l'utilisation de l'espace et ses facteurs déterminants, combinée avec une procédure de modélisation dynamique qui permet de simuler la compétition entre plusieurs formes d'utilisation de l'espace. Depuis sa première version (Veldkamp et Fresco, 1996), le cadre de modélisation CLUE a connu plusieurs développements ultérieurs mais le principe fondamental reste le même et repose sur les étapes suivantes: 1) la collecte et le traitement préalable de grand ensemble de données ; 2) l'analyse statistique des facteurs déterminants l'utilisation de l'espace à différents niveaux ; 3) la modélisation dynamique et spatiale explicite des changements l'utilisation de l'espace et 4) la possibilité d'exécution de scénarios d'évolution future.

Les données sur l'utilisation de l'espace sont enregistrées avec une résolution choisie par le modélisateur en fonction des objectifs, de l'étude, de la nature du processus étudiée et de la précision recherchée. Les cellules ne sont pas nécessairement homogènes et contiennent un pourcentage de chaque type d'utilisation de l'espace. Une analyse quantitative de la relation entre la distribution spatiale actuelle et les facteurs potentiels permet de dériver des équations de régressions multiples qui sont utilisés pour le calcul de l'utilisation de l'espace à une date t donnée (De Koning et al., 1998). En comparant l'utilisation actuelle à celle calculée par régression, on identifie facilement les zones dont le pourcentage de l'utilisation de l'espace a une tendance à la hausse (inférieure à celle calculée par régression) ou à la baisse (supérieure à celle calculée par régression). On peut produire une carte montrant les zones où une augmentation ou une diminution de l'utilisation de l'espace peut être escomptée.

Le modèle dynamique d'allocation des changements est composée de deux principaux modules : un module de calcul de la demande et un module d'allocation spatiale des changements. Le modèle repose sur une hypothèse Boserupienne qui suppose que la production agricole se développe en réponse à la demande alimentaire. Ainsi, le module de calcul de la demande évalue à l'échelle nationale, les changements de la demande totale en produit sur la base de la taille de la population et de la productivité des terres. Cette évaluation s'appuie sur des méthodes d'extrapolation simples ou sur des modèles économiques plus ou moins complexes qui considèrent les mutations en cours les plus importantes et les scénarios formulés (De Koning et al., 1999a; Veldkamp et Fresco, 1997). La procédure d'allocation spatiale des changements

d'utilisation de l'espace tient compte des facteurs biophysiques et socio-économiques déterminants (fertilité des sols, structure et importance de la population, relief, etc.) à différentes échelles (De Koning et al., 1999b, Verburg et al., 1999). La tendance générale des changements est d'abord calculée à grande échelle afin de capturer l'influence des facteurs qui agissent sur une très grande distance. En cas d'augmentation de la demande pour une utilisation considérée, on sélectionne toutes les cellules dont l'utilisation considérée a une tendance à augmenter. La différence est calculée pour toutes les cellules sélectionnées et une fraction de cette différence est additionnée à l'utilisation actuelle. Cette fraction est ajustée itérativement de telle manière que l'utilisation agrégée de toutes les cellules soit égale à la demande pour ce type d'utilisation à l'échelle nationale. Cette première allocation permet d'identifier les zones avec un changement relatif plus ou moins grand en comparant les changements absolus des cellules individuelles avec la moyenne des changements pour toutes les cellules. Une deuxième allocation est ensuite effectuée à fine échelle sur la base des résultats de l'allocation à grande échelle en prenant en compte les contraintes locales et la compétition entre les types d'utilisation de l'espace. A fine échelle, on sélectionne également les cellules dont le l'utilisation a tendance à augmenter et on calcule la différence. A cette échelle, la fraction d'utilisation allouée dépend du changement relatif calculé à grande échelle (Verburg et al., 1999). Les scénarios effectués fournissent des résultats différents suivant que l'allocation est faite uniquement à grande échelle, uniquement à fine échelle ou par une allocation par l'approche multi-échelle.

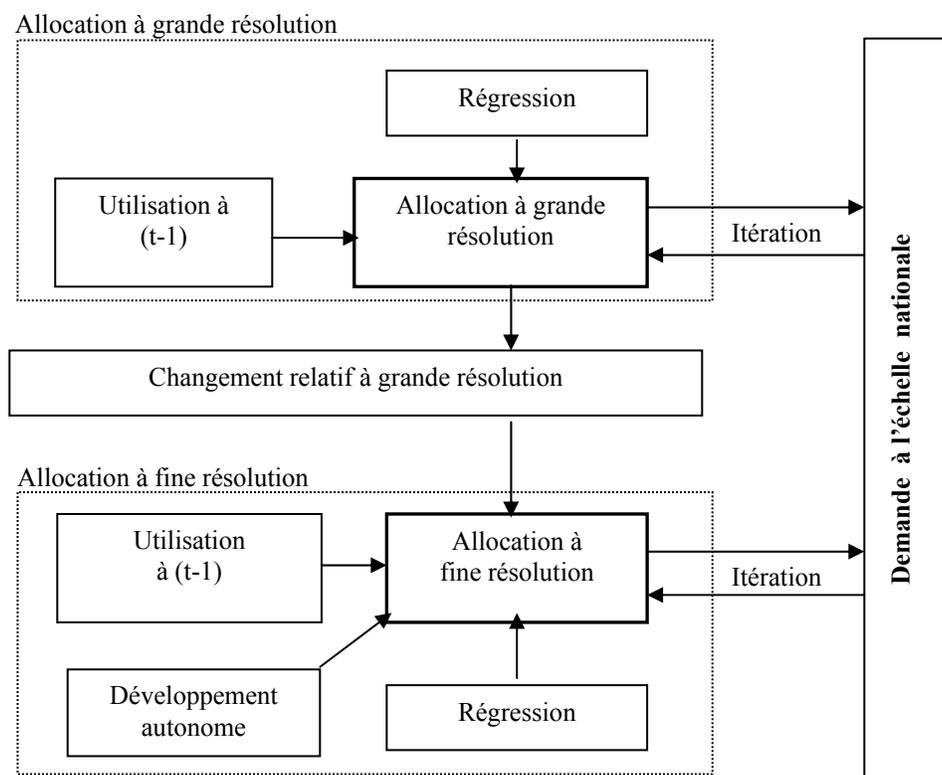


Figure 3.13 : Procédure d'allocation des changements dans la plateforme de modélisation CLUE (extraite et adaptée de Verburg et al., 1999).

Une limite majeure de la procédure d'allocation qui est proposée vient du fait que les changements ne sont effectués que pour les cellules dont la tendance du couvert est du même sens que celui de la demande au niveau national. Afin de prendre en compte les effets ascendants les développements autonomes sont mis

en oeuvre pour prendre en compte les changements exclusivement basés sur les conditions locales (Verburg et al., 1999 ; De Koning et al., 1999a). Ainsi les cellules qui n'ont pas été sélectionnées pour les changements dans la procédure précédente subissent un changement proportionnel à la différence spécifique entre le couvert actuel et celui calculé par régression. Dans cette situation, la fraction est déterminée par l'utilisateur du modèle et devra en général être plus petit comparé au changement dans la direction de la demande. Les résultats des scénarios montrent que l'implémentation des développements autonomes produit une plus grande dynamique du système. Enfin, la compétition entre les différents types d'utilisation (total des utilisations >100%) nécessite une modification de l'allocation des changements. La modification des changements des différents types d'utilisation est fonction de la force de compétition qui dépend de la différence entre l'utilisation actuelle et celui calculé par régression et du changement dans la demande.

Application de CLUE aux petites régions : CLUE-S

La première version de la plate-forme CLUE a été principalement appliquée à de très vastes régions à l'échelle nationale et continentale où les données sont le plus souvent issues de recensement et disponibles à très petite échelle, induisant une représentation à faible résolution spatiale ou la taille du grain est grande. Les exemples d'application ont été développées pour le Costa Rica, l'Equateur et la Chine (De Koning et al., 1998 ; De Koning et al., 1999 ; Verburg, 2000 ; Kok et Winograd, 2001). Une représentation pertinente de l'utilisation de l'espace dans ces cas est donnée par la proportion de chaque type d'utilisation dans une cellule. Lorsqu'on s'intéresse à l'utilisation de l'espace à l'échelle d'une petite région, cette forme de représentation n'est très adaptée dans la mesure où on ne dispose que des données à grande échelle où l'information disponible sur une unité (polygone ou cellule) représente l'utilisation de l'espace dominante. C'est le cas lorsqu'on utilise des cartes existantes ou les cartes produites par analyse d'image de télédétection à haute résolution. En plus du problème de représentation des données, les applications régionales présentent des spécificités qui ne sont pas prises en compte dans cette première version de la plateforme. Les développements récents de la plateforme ont par exemple porté sur l'application du modèle aux espaces de taille relativement petite en utilisant des données à très haute résolution. La version CLUE-S (CLUE for Small regions), a été développée pour des petites régions et intègre les spécificités et propriétés des systèmes d'utilisation de l'espace à cette échelle (Verburg et al., 2002).

La procédure d'allocation du modèle traduit ces demandes en changement dans l'espace en s'appuyant sur les probabilités de chaque cellule à être affectée aux différentes utilisations. En s'appuyant sur les résultats de régression, une carte de probabilité peut être calculée pour chaque type d'utilisation de l'espace. Une nouvelle carte de probabilité est calculée chaque année avec les valeurs mises à jour des facteurs déterminants qui changent dans le temps tel que la distribution de la population ou l'accessibilité. La probabilité P_i ($\sum_j P_{ij} = 1$) pour que la cellule i de la carte d'utilisation de l'espace LU soit affectée à l'utilisation j est déterminée par le modèle de régression logistique défini à la section précédente. L'allocation des changements est faite suivant une procédure itérative qui utilise les données suivantes : la carte actuelle de l'utilisation de l'espace, les cartes de probabilités et la demande de chaque utilisation de l'espace (figure 3.14). L'utilisateur de la plateforme peut spécifier des règles de décision pour indiquer des restrictions de changement ou une politique de gestion de l'espace donnée. Ce sont des règles qui déterminent les conditions sous lesquelles une utilisation de l'espace peut changer à chaque pas de temps.

L'implémentation de certaines règles dans le système consiste à affecter une résistance au changement à certaine utilisation de l'espace de façon à générer la stabilité dans la structure du système. Une élasticité relative ($ELAS_u$) comprise entre 0 et 1 est définie pour l'utilisation de l'espace considérée sur la base des connaissances du système agraire ou lors de la calibration du modèle. Trois possibilités sont envisageables : $ELAS_u = 1$ lorsque les transitions sont difficiles ou impossibles après la première conversion, $ELAS_u = 0$ lorsqu'il n'existe aucune restriction au changement et $0 < ELAS_u < 1$ pour le cas des utilisations de l'espace opérant entre les deux extrêmes. Plus l'élasticité sera grande, et plus il sera difficile à convertir cette utilisation. L'algorithme d'allocation à chaque pas de temps se résume dans les étapes suivantes :

1. Sélection des cellules candidates au changement,
2. Pour chaque cellule i , la probabilité totale est calculée suivant la formule $TPROP_{i,u} = P_{i,u} + ELAS_u + ITER_u$ où $ITER_u$ est une variable d'itération propre à l'utilisation de l'espace u et $ELAS_u$ son élasticité.
3. Une allocation préliminaire est faite avec une valeur égale de la variable $ITER_u$ pour toutes les utilisations de l'espace. Chaque cellule est allouée à l'utilisation de l'espace qui a la plus grande probabilité totale. Cette opération produit un changement d'utilisation pour un nombre de cellules.
4. La surface totale allouée à chaque utilisation est comparée à la demande. Pour les utilisations où la surface allouée est plus petite que la demande, la variable d'itération est incrémentée et dans le cas contraire, la variable d'itération est décrétementée.
5. Les étapes de 2 à 4 sont exécutées jusqu'à ce que les allocations soient égales aux demandes. La carte d'utilisation de l'espace de l'année en cours est sauvegardée et l'algorithme continue au pas de temps suivant.

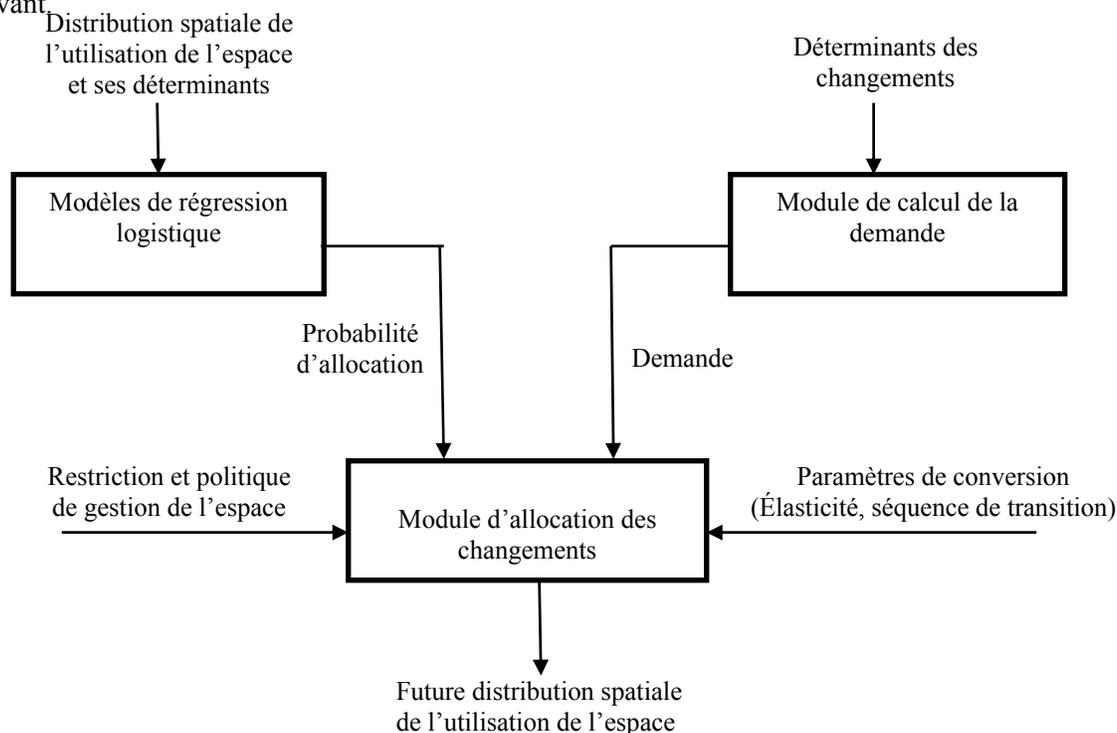


Figure 3.14: Echange d'information avec et entre les modules de la plateforme CLUE-S.

En assurant un lien continu entre la demande régionale et les potentialités locales pour chaque type d'utilisation de l'espace, cette procédure d'allocation prend ainsi en compte l'interaction entre différentes échelles qui est requise dans les modèles de changement d'utilisation de l'espace. Une autre façon de prendre en compte la propriété multi-échelle des systèmes d'utilisation de l'espace consiste à inclure dans les modèles de régression, des variables représentant des facteurs déterminants qui opèrent sur de grandes distances comme l'accessibilité au centre urbain ou la distance à la route. Une première version du modèle développé pour la petite région autour de Maroua dans l'Extrême Nord du Cameroun est présentée au chapitre 10.

3.6. Conclusion

Dans ce chapitre, nous avons décrit les méthodes et outils logiciels qui sont utilisés pour la mise en oeuvre du SIE SMALL Savannah. Une première remarque qui se dégage de cette présentation est le fait que chaque outil apporte une contribution spécifique aux objectifs et à la finalité globale du SIE. Il s'agit notamment d'observer l'environnement, gérer l'information environnementale, analyser et prévoir les phénomènes en vue de supporter le processus de prise de décision pour une gestion rationnelle de l'environnement. Les outils de traitement d'images et de la télédétection (Cartographie, GPS, etc.) contribuent à l'observation et au suivi de l'environnement. Les SIG et outils d'analyse spatiale contribuent à la gestion de l'information environnementale et à l'analyse des phénomènes environnementaux dans un contexte spatial. Les outils d'analyse statistique contribuent à explorer les grandes bases de données et aident à expliquer ou prévoir les phénomènes environnementaux. Les modules de simulation aident à mieux comprendre le fonctionnement des systèmes environnementaux et explorer les scénarios d'évolution, fournissant ainsi un support d'aide à la décision. Une deuxième remarque qui se dégage de cette présentation et qui caractérise les SIE concerne la grande diversité de ces outils. Une raison majeure qui explique cette diversité est l'hétérogénéité des données manipulées par les SIE qui exige de recourir à des techniques de stockage, des méthodes de traitement et d'analyse qui ne sont pas toujours toutes disponibles dans un seul logiciel. De plus, plusieurs tâches du SIE ne sont pas informatisables contrairement aux SI classiques. Les outils développés restent des outils d'analyse isolés où très peu d'accent est mis sur les interfaces Homme-Machine. Les chercheurs et experts dans le domaine de l'environnement ont donc hérité d'une longue tradition d'utilisation d'outils spécifiques pour des traitements donnés. Le cloisonnement du traitement de l'information environnementale a contribué à la génération d'un nombre croissant d'applications logicielles qui sont très faiblement intégrées. Ceci ne permet pas toujours une gestion cohérente et la restitution de l'information environnementale aux utilisateurs. Les méthodes de développement des SIE devraient prendre en compte cette spécificité en proposant des techniques pour faciliter l'intégration des applications.