

Ministère de l'Enseignement Supérieur et de la Recherche
Conservatoire National des Arts et Métiers
Ecole Supérieure des Géomètres et Topographes
1 Boulevard Pythagore - 72000 LE MANS



Mémoire de travail de fin
d'études présenté en vue de
l'obtention du Diplôme
d'Ingénieur de l'ESGT



Étude de l'utilisation des SIG dans le cadre d'un projet archéologique multi-échelles



Mémoire soutenu par :
Sébastien Merlet
le 10 juillet 2009

Membres du jury :
Président : M. Richard Thévenot
Maître de stage : Mme Anabel Ford
Professeur référent : Mme Joëlle Nicolas

Remerciements

Je tiens à remercier tout particulièrement :

Les jardiniers forestiers mayas modernes que j'ai pu côtoyer, et tout particulièrement **Narciso Torres**, pour avoir partagé leur savoir et leur conception du monde ainsi que pour leur engagement quotidien dans la protection de l'environnement et la sauvegarde du patrimoine culturel maya

Anabel Ford, archéologue responsable du projet, pour ce stage intéressant

Joëlle Nicolas, mon professeur référent, pour son suivi du projet

Ma binôme, Séverine Berné, pour son aide et ses conseils lors de la relecture de ce mémoire

Gary Raines, concepteur d'ArcSDM, pour ses conseils lors de l'application de la méthode des *Weights of Evidence*, et ses explications diverses sur la façon d'utiliser les outils

Jessica Everton et **Chrissie Bausch**, assistantes d'Anabel, pour leurs conseils en anglais et leur bonne humeur quotidienne

Constance Christensen, utilisatrice du SIG

Ma famille pour son soutien moral et financier

L'École Supérieure des Géomètres et Topographes pour avoir permis que ce stage ait lieu dans les meilleures conditions



Un des champs de maïs (milpa) de Narciso Torres, jardinier forestier maya moderne, exemple de diversité et de sauvegarde des méthodes traditionnelles mayas

Ce mémoire est dédié aux jardiniers forestiers mayas pour leur travail fabuleux, ainsi qu'à Jessica Everton, assistante du docteur Anabel Ford que j'ai eu l'occasion de former à l'usage d'ArcGis et aux notions de base de la géomatique. Je lui suis reconnaissant pour son aide précieuse dans l'exécution de certaines tâches répétitives prenant énormément de temps, et cela en gardant toujours le sourire. Merci Jessica, et bonne continuation pour tes études !

Table des matières

I. Introduction.....	6
II. Présentation du projet étudié.....	7
A. El Pilar : une ancienne cité Maya.....	7
B. Un parc binational dédié à la paix et à la protection de la faune et la flore.....	7
C. Étude du SIG existant.....	8
1. Les données.....	8
i. Quantité.....	8
ii. Qualité.....	9
2. L'organisation du SIG.....	10
3. L'utilisation du SIG.....	10
III. Mise en place d'une nouvelle organisation.....	11
A. Redéfinition du rôle du SIG.....	11
B. La géodatabase.....	11
1. Choix du format.....	11
2. Les jeux de classes d'entités.....	12
3. Outils de contrôle mis en place.....	13
i. Topologie de contrôle des sites archéologiques.....	13
ii. Topologie de contrôle des sols.....	13
4. Règles de fonctionnement.....	14
i. Insertion de nouvelles données.....	14
ii. Utilisation de données.....	15
iii. Modification de données.....	15
C. Les cartes.....	15
D. Le dossier de travail.....	15
IV. Méthode de conception d'un modèle prédictif.....	16
A. Introduction.....	16
B. Rappels de la méthode des Weights of evidence.....	16
1. Méthode de mise en œuvre.....	16
i. Sélection de points d'entraînement (sites).....	17
ii. Calcul des poids associés aux classes d'évidence.....	17
iii. Calcul du raster réponse.....	17
iv. Tests de qualité du résultat.....	17
2. Modèle utilisé au niveau de ArcGis.....	18
C. Réflexions sur l'implémentation.....	19
1. Choix des paramètres.....	19
2. Choix des points d'entraînement.....	19
3. Choix de l'aire de la cellule unitaire.....	19
4. Choix des masques à utiliser.....	19
5. Contrôle et choix des poids.....	20
6. Création de la carte des probabilités et analyse qualitative du résultat.....	21
V. Validation et amélioration du modèle local.....	22
A. Objectifs.....	22
B. Présentation de la zone et des paramètres.....	22
1. Rivières.....	23
2. Drainage et fertilité.....	24
3. Pente topographique.....	24
C. Mise en application de la méthode.....	25

1. Introduction.....	25
2. Présentation des sites et des masques.....	25
3. Étude qualitative des sites et des zones d'entraînement.....	26
4. Détail du calcul des poids : masque des transects.....	27
i. Drainage et fertilité.....	27
ii. Rivières.....	28
iii. Pente topographique.....	29
5. Calcul du résultat.....	29
D. Conclusions pour le modèle local.....	30
1. Fiabilité du modèle.....	30
2. Choix du résultat à prendre en considération.....	30
3. Comparaison avec le résultat de 2004.....	30
VI. Estimation de la population pour la zone d'étude.....	32
A. Étude des sites.....	32
B. Reclassification des classes prédictives.....	32
1. Analyse du résultat utilisé dans le passé.....	32
2. Nouvelle méthode mise en place.....	33
C. Propagation des sites et amélioration des probabilités.....	34
D. Analyse finale de la population de la zone étudiée.....	35
1. Étude de la densité et de la population.....	35
2. Analyse de la répartition des sites.....	37
E. Besoins en termes d'alimentation.....	39
1. L'agriculture Maya.....	39
i. De la forêt au milpa.....	39
ii. Du milpa au jardin forestier.....	39
iii. Du jardin forestier à la forêt et la haute canopée.....	40
iv. De la forêt à la haute canopée.....	40
2. Conception et calcul du modèle.....	40
i. Nombre de calories nécessaires et modèle d'alimentation.....	40
ii. Superficie nécessaire pour la culture du maïs.....	40
iii. Application du modèle du milpa.....	41
3. Analyse de valeurs possibles.....	43
i. Modification de la longueur du cycle du milpa.....	43
ii. Modification du nombre d'années d'utilisation du milpa.....	43
iii. Changement de la productivité des champs de maïs.....	43
F. Conclusion.....	44
VII. Conception d'un modèle régional.....	46
A. Introduction.....	46
B. Présentation de la zone et des paramètres.....	46
1. Points d'entraînement.....	46
2. Paramètres utilisables.....	47
i. SRTM.....	47
ii. Pente topographique.....	47
iii. Élévation au niveau d'une base régionale (ENBR).....	48
iv. Fertilité et drainage.....	48
C. Application avec les sites régionaux.....	49
1. Analyse des poids.....	49
i. ENBR (500 mètres).....	49
ii. SRTM.....	50
iii. Pente topographique.....	50

iv. Fertilité et drainage.....	50
2. Carte de probabilités.....	51
D. Application avec les points locaux.....	51
1. Analyse des poids.....	52
i. ENBR (500 mètres).....	52
ii. SRTM.....	52
iii. Pente topographique, drainage et fertilité.....	52
2. Cartes de probabilités.....	52
E. Choix d'une carte de probabilités à conserver.....	53
VIII. Modèle prédictif du monde maya et analyses supplémentaires.....	55
IX. Conclusion.....	56
X. Annexes.....	58

I. Introduction

Situés en Amérique Centrale, le Guatemala et son voisin le Belize sont des pays qui ne sont pas forcément très connus. C'est pourtant ici que durant plusieurs millénaires s'est développée une civilisation qui a laissé derrière elle des vestiges archéologiques incroyables. Un peuple qui passionne par la grandeur de ses constructions, par ses temples majestueux qu'il est toujours possible de voir dépasser de la canopée et par les multiples monts ayant recouvert ses habitations qui parsèment la jungle de ces pays. Alors que le Belize ne comportait en 2005 que 277 409 habitants pour 22 966 km² (Government of Belize 2006; SIB 2007), soit une densité de 12 hab./km², il est frappant de voir en se promenant dans la forêt à quel point les Mayas semblaient très présents.

Les densités avancées par certains archéologues vont dans ce sens, 280 hab./km² dans la région du Rio Bec (Fischbeck 2004) par exemple, ou encore 407 à 506 hab./km² pour la zone rurale de Copal Pocket (Webster & Freter 1990, cité dans Wheeler 2008). Comment une telle population pouvait-elle subvenir à ses besoins alors que les Mayas ne connaissaient pas le fer, n'utilisaient pas la roue ? Quels sont les secrets qui ont permis à cette civilisation d'asseoir sa splendeur dans le Yucatàn et d'y subsister pendant plus de 1 700 ans ?

Telles sont les questions auxquelles les archéologues essaient de répondre, et les systèmes d'information géographique prennent une place de plus en plus importante dans leurs recherches depuis les premières utilisations dans les années 1980 (Aldenderfer & Maschner 1996; Wheatley & Gillings 2002). De nombreuses publications ont été effectuées sur le sujet, notamment au niveau des possibilités de modélisation et de l'efficacité des modèles prédictifs (Mehrer 2006; Dermody 2005; Leusen 2002) pour tenter de définir des règles de bonne utilisation de ces outils en conciliant la connaissance et la mise en pratique des techniques associées avec le respect de l'aspect archéologique.

Le présent mémoire, effectué dans le cadre d'un programme d'étude de l'environnement maya, ne vise pas à être une énième analyse théorique des relations entre SIG et archéologie, mais à lier intimement la mise en œuvre d'un projet complet avec l'identification et la résolution des difficultés rencontrées. Comment concilier le travail d'un ingénieur en géomatique n'ayant pas de connaissances initiales en archéologie avec les besoins d'une équipe multidisciplinaire ne comportant pas de technicien en systèmes d'information géographique ? Quelles sont les phases à suivre pour la réalisation d'un tel projet, de l'organisation des données disponibles au sein d'un système utilisable par tous à l'analyse des résultats produits, en passant par les phases de concertation avec les membres de l'équipe ? La correction de modèles prédictifs biaisés est notamment clarifiée, mettant en lumière la complexité des choix à effectuer et les conclusions erronées auxquelles ils peuvent parfois conduire.

Cette étude insiste également sur les problèmes d'échelle lors de la mise en place d'une méthode prédictive : quels sont les critères à retenir pour le choix des données et l'application de la méthode des *Weights of Evidence* ici choisie ? Comment les choix de la taille des masques ou du type des sites utilisés influent-ils sur les résultats ? Est-il possible de généraliser à l'échelle du système du monde maya des conclusions obtenues à une échelle locale ?

Enfin, la question de l'application des résultats de modèles prédictifs et de leur traitement est présentée. Comment produire à partir de ces modèles de nouvelles informations sur la population étudiée, ses moyens de subsistance et son occupation des sols ? Est-il possible d'apporter grâce aux outils de la géomatique une nouvelle vision sur la façon dont un peuple pouvait vivre, et d'aborder la question des politiques actuelles en terme d'agriculture pour mettre en relation les enseignements du passé avec les défis du futur ?

Malgré la spécificité du programme étudié, ce mémoire est donc conçu de façon à être transposable à tout projet archéologique s'intéressant à une mise en œuvre d'une modélisation de population.

II. Présentation du projet étudié

A. El Pilar : une ancienne cité Maya

L'histoire des Mayas commence environ 2000 ans avant Jésus Christ, peuple vivant en tribus dispersées, cultivant le maïs et créant de la poterie (Von Hagen 1960). De 1000 à 300 ans avant Jésus Christ, cette culture se développe, établissant des centres indépendants, notamment dans la région du Petén, en Amérique Centrale, où se situe la zone d'étude du projet considéré. Les cités mayas ainsi que leurs majestueux temples perdurent de -500 à 1000 ap. JC. Après cette date, ce peuple ne crée plus de nouvelles constructions d'importance, et les grandioses cités sont peu à peu abandonnées à la jungle.

Au niveau d'El Pilar (se reporter à la carte ci-dessous), les premiers bâtiments publics sont établis aux alentours de -700, et les constructions continuent pendant plus de seize siècles. L'âge d'or de cette civilisation s'étale de 250 à 900 ap. JC, et alors que 869 ap. JC correspond au dernier monument daté construit à Tikal qui fut la plus importante des cités mayas, la maintenance d'El Pilar continue jusqu'à 1 000 ap. JC.

La région est occupée à nouveau à partir des années 1900 et le site redécouvert, jusqu'à l'établissement de la Réserve Archéologique d'El Pilar en 1990 (Gentle et al. 2000). Les champs de maïs au sein de celle-ci ont depuis laissé place à une jungle luxuriante, peu à peu repeuplée par les animaux y compris le fameux jaguar, dont la présence d'une femelle et de son petit a été confirmée en mai 2009.



Illustration 1: Réserve archéologique d'El Pilar

B. Un parc binational dédié à la paix et à la protection de la faune et de la flore

Bien plus qu'un simple parc archéologique, cette réserve de plus de 2 000 hectares est, de par sa localisation à cheval sur la frontière entre le Belize et le Guatemala, un véritable pari pour l'avenir puisque des tensions importantes existaient encore récemment entre ces deux pays.



En plus de sa vocation de parc binational dédié à la paix, El Pilar est également devenu un important lieu de protection de l'environnement, les visiteurs pouvant à loisir admirer la faune et la flore dans un cadre volontairement laissé sauvage.

Le type d'archéologie pratiqué ici est en lui-même très intéressant : plutôt que de mettre à nu les vestiges en les exposant, le principe de l'archéologie sous la canopée est utilisé. Protégés par la couverture de la jungle, les anciens temples s'offrent à l'imagination des personnes venant les découvrir, et semblent lancer un avertissement au monde moderne : après son heure de gloire, ils sont les dernières preuves de la grandeur d'une civilisation passée.

Parallèlement au parc en lui-même, diverses actions sont menées afin de comprendre et d'analyser les leçons du passé pour les appliquer aux problèmes du présent. *Exploring Solutions Past, the Maya Forest Alliance*¹ est ainsi un projet au sein duquel l'équipe du docteur Anabel Ford travaille en collaboration avec des jardiniers forestiers descendant des mayas pour sauvegarder et perpétuer leurs connaissances ainsi que leur savoir faire. Les populations du Belize et du Guatemala sont en pleine expansion, les habitations et les vastes champs adaptés à l'agriculture mécanisée grignotant toujours plus sur la jungle.

C'est dans le contexte de ce projet que l'étude présente se situe, le système d'information géographique de la forêt maya ayant pour but d'analyser aussi bien l'environnement passé et la façon de vivre des mayas que l'utilisation actuelle des sols et les projets de développement envisagés pour le futur, afin de répondre à diverses questions :

- Quel type d'utilisation des sols les mayas pratiquaient-ils ?
- Quels paramètres influaient sur l'emplacement de leurs habitations ?
- Peut-on modéliser ces emplacements afin d'estimer plus précisément les densités de populations au niveau de la zone d'étude ?
- Leur environnement pouvait-il supporter les nombres très importants d'habitants avancés par certains auteurs, ou cela pourrait-il être une des raisons de leur déclin ?
- En quoi les réponses à ces questions peuvent-elles influencer les politiques de développement actuelles au Guatemala et au Belize ?

C. Étude du SIG existant

Tout travail effectué dans le cadre d'un projet SIG se doit de suivre une certaine organisation afin d'assurer un bon déroulement de celui-ci. Ceci est d'autant plus vrai pour le projet El Pilar du fait que dans le cadre d'une application à l'archéologie, les systèmes d'information géographique sont parfois utilisés par des personnes n'ayant pas de connaissances approfondies dans ce domaine. Cela peut conduire à une idée approximative de ce qu'est un SIG et de ses possibilités, qui sont sur ou sous-évaluées. Dans le cas où une personne non archéologue gère la partie SIG du projet, le risque est au contraire que l'aspect archéologique soit sous estimé. Il convient donc d'utiliser les outils mis à disposition par les logiciels tout en gardant à l'esprit l'aspect et les besoins archéologiques (Gaffney et al. 1996).

La première tâche effectuée dans le cadre du projet El Pilar a donc été de regarder comment le SIG actuel est organisé. Cela a permis de faire ressortir les points faibles et les points forts de la situation existante afin d'effectuer un premier diagnostic.

1. Les données

i. Quantité

Le premier constat qui apparaît lors de la découverte du SIG de la forêt Maya est lié à l'abondance des données. Celles-ci sont en effet l'élément de base d'un SIG, et leur collecte peut être longue et coûteuse.

1 Site internet du projet : <http://espmaya.org>

Le SIG ici étudié bénéficie de données acquises depuis de nombreuses années sur le terrain par l'équipe du projet El Pilar, ainsi que d'autres provenant de sources diverses. Il est ainsi possible de trouver des données relatives à l'hydrologie, à la pédologie, à l'emplacement de sites archéologiques, ou encore concernant les actuels jardiniers forestiers et leur utilisation du sol.

L'importance du projet simplifie également l'acquisition d'éventuelles données manquantes, pouvant être fournies par diverses organisations travaillant au niveau de la région d'étude.

ii. Qualité

Le facteur quantitatif n'est cependant évidemment pas le plus important, la question de la qualité des données pour une certaine utilisation se posant également. Cette qualité peut être évaluée de deux manières :

- en se renseignant auprès des personnes s'occupant de ces données
- grâce aux informations fournies avec les données (métadonnées)

La source humaine, bien que parfois nécessaire, n'est pas vraiment satisfaisante. Dire qu'une donnée est de bonne qualité sans pouvoir quantitativement donner sa précision par exemple n'avance pas à grand chose, surtout quand le jugement ne provient pas d'une personne ayant des connaissances poussées sur le sujet étudié.

Les métadonnées (données à propos d'une donnée) sont donc primordiales pour une approche qualitative efficace. Depuis l'établissement du *Federal Geographic Data Committee* en 2004, et l'apparition de divers standards pour les métadonnées, de nombreux logiciels d'édition et de validation sont apparus (Benvenuti 2003). ArcCatalog permet notamment l'import, l'export et l'édition de celles-ci : une première vision globale de la qualité des métadonnées peut ainsi être obtenue en ouvrant certains fichiers du SIG avec ce logiciel.



Illustration 2: exemple de problèmes où l'absence de métadonnées s'avère très gênante

Dans le cadre du projet El Pilar, il apparaît hélas que la majorité des métadonnées sont de mauvaise qualité. Les informations renseignées sont la plupart du temps insuffisantes, que ce soit au niveau de la provenance des données ou de leur précision, et les champs requis par le profil utilisé (ESRI, 2003) ne sont notamment pas respectés. Il est dans ce cadre difficile de déterminer la qualité des données disponibles pour une utilisation particulière. Les droits d'accès et d'utilisation ne sont pas non plus mentionnés, ce qui peut amener d'éventuels problèmes.

L'illustration 2 présente ainsi deux jeux de mêmes sites archéologiques disponibles à l'échelle régionale. Certains d'entre eux se retrouvent distants de près de deux kilomètres, et aucune métadonnée ne permet de savoir la précision respective de ces jeux. Cela ne pose pas de problème dans le cadre d'une analyse suffisamment générale, mais dans le cas où une localisation plus fine d'un site serait demandée, rien ne permet de décider quel jeu devrait être utilisé. Cela participe également à la désorganisation du SIG puisque les données sont dupliquées : en connaissant le niveau de précision de la localisation des points, le meilleur jeu pourrait être gardé.

2. L'organisation du SIG

Le but d'un SIG est entre autres d'organiser les informations géographiques. Cela devient encore plus crucial dans le cas où un nombre très important de données sont utilisées. Il s'agit actuellement d'une faiblesse du SIG de la forêt Maya : il n'existe pas de vraie méthode de classement des données, certaines apparaissant dans des dossiers correspondant aux années d'acquisition, d'autres dans des dossiers créés par les personnes qui se sont servies de celles-ci, ou encore dans des dossiers thématiques. Aucun répertoire ne leur étant spécifiquement dédié, les cartes produites ainsi que tous les résultats des différentes analyses effectuées au moyen d'ArcGis se retrouvent dans ces mêmes dossiers.

De nombreux problèmes résultent de cette désorganisation, aggravés par le fait que les métadonnées ne sont pas à jour, ou inexistantes. Il y a ainsi une importante redondance due à la multiplication des dossiers, l'impossibilité de savoir si certains fichiers sont des données originales ou modifiées, ou encore la présence de fichiers apparemment identiques mais dont le contenu diffère sans qu'il ne soit possible de savoir quelle version est celle à utiliser.

3. L'utilisation du SIG

Le SIG de la forêt Maya est principalement utilisé pour des tâches ponctuelles telles que la conception d'une carte pour accompagner une présentation. Aucun protocole de classement des données, de mise à jour ou encore de veille technologique n'existe, et personne n'est réellement responsable de celui-ci.

Divers étudiants ont également eu l'occasion de travailler au *MesoAmerican Research Center* au niveau du Système d'Information Géographique pour leur mémoire ou leur thèse, mais la nature particulière de leurs projets participe à la création de dossiers annexes redondants plutôt qu'à une intégration des résultats et des améliorations au sein du système, ce qui pose un réel problème.

Il y a donc un besoin évident d'organisation et de création d'un système complet mais simple d'utilisation, doté de règles strictes quant à la gestion des données et à leur documentation.

III. Mise en place d'une nouvelle organisation

A. Redéfinition du rôle du SIG

Lors de la réorganisation, il est important d'être méticuleux puisque le travail effectué à ce niveau conditionnera toute la suite. Les phases de conduite d'un projet de système d'information géographique sont décrites dans de nombreux documents, notamment pour ArcGis (Booth & Mitchell 2001). Dans le cas présent, il convient de considérer ce qui existe déjà, tout en ne restant pas prisonniers de la situation actuelle. Les nouvelles propositions doivent en effet être correctement argumentées et documentées pour permettre leur mise en place effective.

Une définition intéressante des Systèmes d'Information Géographique est la suivante : « *A geographic information system (GIS) captures, stores, analyzes, manages, and presents data that is linked to location.* »² (Wikipedia 2009b). Il est bon de la rappeler puisqu'une des difficultés constatées du SIG étudié résultait de la profusion de données dont de nombreuses n'avaient pas de lien géographique justifiant cette présence (logiciels ou documents divers par exemple).

Le problème consiste à présenter efficacement les données géolocalisées pour les mettre à disposition de différents utilisateurs, ayant des besoins différents. Dans le cadre d'un tel projet archéologique, les utilisateurs et besoins identifiés sont résumés ci-dessous.

Rôle	Besoins
Responsables du projet	Accès aux différentes cartes présentant les résultats des analyses effectuées pour usage direct
Utilisateurs des données	Accès rapide aux « bonnes » versions des données, avec métadonnées
Analystes	Accès aux données et aux fichiers de travail utilisés dans le passé

Pour un projet aussi vaste et diversifié qu'El Pilar, de nombreuses personnes sont impliquées, une grande part d'entre elles n'ayant pas de formation en gestion ou utilisation de systèmes d'information géographique. Les résultats des analyses effectuées doivent donc être présentés dans un format directement exportable pour être inclus dans un article ou un site internet.

Les données contenues dans le SIG doivent être facilement récupérables par des personnes n'utilisant pas habituellement le système, et contenir les informations relatives à leur contenu, leur qualité, ou encore les questions légales. Les personnes travaillant plus particulièrement au sein du SIG (analystes, travaillant par exemple au niveau du modèle prédictif) ont également besoin d'un espace pour stocker les fichiers de travail et récupérer ceux d'analyses effectuées dans le passé pour reproduire ou améliorer un résultat.

Il apparaît également primordial de mettre en place des protocoles expliquant comment ces utilisations doivent être effectuées. Les besoins identifiés précédemment doivent être reflétés au niveau de la structure de fichiers et de dossiers mise en place. Celle ici retenue est divisée en :

- Une géodatabase³ pour le stockage et la gestion des données
- Un dossier contenant les fichiers de travail
- Un dossier contenant les résultats

² « Un SIG capture, stocke, analyse, gère et présente des données liées à une localisation »

³ Ce terme est celui utilisé par ESRI France pour désigner les bases de données géographiques

B. La géodatabase

1. Choix du format

Les géodatabases sont un outil très efficace pour stocker et gérer les données au sein d'un SIG. Elles permettent en effet de centraliser les données dans une structure adaptée, offrant des possibilités de gestion, stockage et contrôle très intéressantes. Dans le cas d'un projet archéologique, deux solutions sont principalement envisageables, résumées dans le tableau suivant (ESRI 2008).

Géodatabase fichier	Géodatabase personnelle
Chaque donnée est stockée dans un fichier séparé, la géodatabase étant constituée de dossiers contenant ces fichiers	Cette géodatabase est stockée dans un unique fichier Microsoft Access
Taille illimitée (la taille de chaque donnée étant limitée à 1 Tb par défaut)	Taille limitée à 2Gb, mais l'éditeur précise qu'au delà de 500Mb les performances sont dégradées
Multi-plateformes	Limité à Windows
Possibilité d'utiliser un format compressé en lecture seule pour économiser de l'espace disque	Les attributs peuvent directement être édités avec OpenOffice Base ou Microsoft Access

Afin de prendre en compte les perspectives d'évolution du projet, et notamment la taille des données gérées, le modèle choisi a été celui de la « *file geodatabase* ». Par tous ses aspects, ce format est plus intéressant que le second envisageable. Le format « ArcSDE geodatabase » qui s'appuie sur des serveurs Oracle, Microsoft SQL ou encore PostgreSQL n'est pas considéré ici, les capacités offertes par ces solutions dépassant les nécessités d'un projet archéologique tel que celui étudié.

2. Les jeux de classes d'entités

L'organisation interne de la géodatabase doit être pensée pour répondre aux besoins du projet après prise en compte des possibilités offertes, notamment la création de topologies permettant de contrôler les données (Booth & Mitchell 2001; Piroit & Saint Gerand 2005). Celle mise en place dans le cadre de cette étude pour le projet El Pilar est divisée en plusieurs jeux de classes d'entités :

- *Archaeology* : contient les classes d'entités correspondant aux sites archéologiques, aux limites des transects⁴
- *Bound_Admin* : contient diverses limites administratives, entre autres les limites de districts, les frontières
- *Forest_Gardens* : contient différentes informations sur les jardiniers forestiers collaborant au projet et leurs propriétés
- *Hydrology* : ce jeu de classes d'entités contient les données relatives à l'hydrologie : rivières, lacs
- *Names* : ce jeu est temporaire et contient l'emplacement de campements au niveau de la zone étudiée nécessitant d'être contrôlés
- *Pedology* : contient les données concernant les sols
- *Protected_areas* : contient les limites de diverses zones protégées

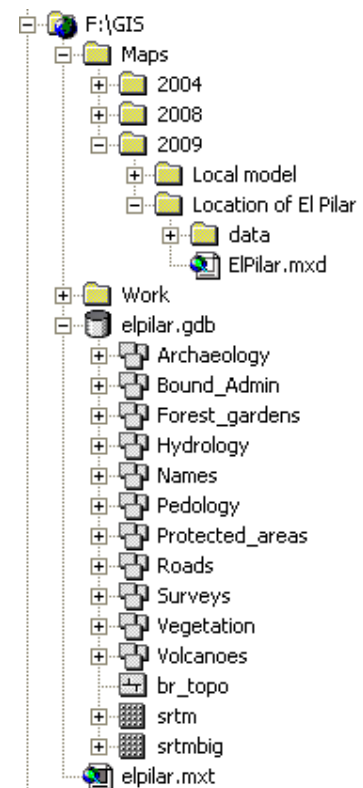


Illustration 3: Nouvelle structure du SIG

⁴ Transect : zone d'étude au niveau de laquelle les occurrences et les emplacements d'un phénomène sont relevés

- *Roads* : contient les routes
- *Surveys* : contient des données relatives à des levés ayant été effectués dans le passé, ainsi que des points de contrôle
- *Vegetation* : contient les informations concernant la végétation
- *Volcanoes* : contient les informations sur les volcans

Il est bon de noter que les images raster ne sont pas stockées directement dans des jeux de classes d'entités au sein de la géodatabase. Afin d'éviter que deux classes d'entités situées dans deux jeux différents n'aient le même nom, ceux-ci sont précédés par l'initiale du jeu concerné.

3. Outils de contrôle mis en place

Afin de s'assurer de la qualité des données, certains processus de contrôle peuvent être utilisés. Dans l'étude réalisée, les topologies servent ainsi à assurer divers aspects importants de la qualité au niveau des sites ou encore des sols.

i. Topologie de contrôle des sites archéologiques

Cet exemple de mise en place d'une topologie permet de vérifier que les points sont bien dans le périmètre des transects correspondants (classes de polygones). Il est en effet important pour le modèle prédictif de connaître l'étendue des levés effectués lors de l'entraînement par la méthode des *Weights of Evidence* (cf. IV. Méthode de conception d'un modèle prédictif).

Ce type de contrôle permet également de s'assurer qu'aucune manipulation malheureuse n'a amené à placer un site au mauvais endroit lors d'une séquence d'édition. Les règles définies dans cette topologie sont les suivantes :

Feature Class	Rule	Feature Class
a_El_Pilar	Must Be Properly Inside	a_El_Pilar_lim
a_sites	Must Be Properly Inside	a_sites_trans
a_local	Must Be Properly Inside	a_local_shape
a_local_shape	Must Be Covered By	a_local_trans

Dans le cas où un site serait en dehors de l'étendue d'un des levés, la validation de la topologie ferait apparaître ce problème qui pourrait alors être corrigé.

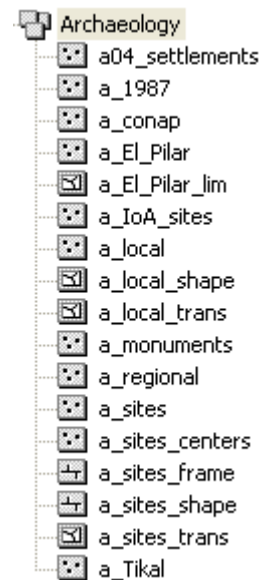


Illustration 4: Sites archéologiques

ii. Topologie de contrôle des sols

Les topologies sont également très utiles pour des opérations de contrôle lorsque des données provenant de sources différentes sont mises ensemble. Des problèmes n'apparaissant pas lors d'un examen visuel effectué avec ArcMap peuvent ainsi être identifiés et corrigés.



Illustration 5: contrôle de la fertilité des sols

L'illustration 5 montre le résultat d'une topologie appliquée à des données de fertilité des sols. Les règles utilisées dans cet exemple sont de deux types :

- *Must not overlap* : les polygones de la classe d'entités ne doivent pas se superposer
- *Must not have gaps* : il ne doit pas y avoir de vide entre les polygones

Cette carte des sols réalisée dans le passé (Sifuentes 2005) apparaît correcte visuellement dans ArcMap, mais la validation de la topologie montre que ce n'est pas réellement le cas. La délimitation du contour extérieur vient de la seconde règle et ne constitue pas un problème. Il en est de même pour le tracé de certaines rivières non incluses dans les polygones à l'ouest. En revanche, de nombreux problèmes sont visibles le long des frontières entre le Guatemala, le Belize et le Mexique. Lorsque les données de ces différents pays ont été regroupées, l'accrochage n'a pas été respecté et les côtés des polygones ne coïncident donc pas parfaitement, d'où la présence de vides ou de zones de superposition.

Ce genre de problème a son importance lors de l'application de la méthode des *Weights of Evidence* : les images rasters créées à partir de fichiers contenant de telles erreurs contiendront en effet des zones de *NoData* biaisant les probabilités calculées. Grâce à l'utilisation de topologies, ces problèmes sont identifiés et aisément corrigés en éditant manuellement les données dans ArcMap ou en utilisant par exemple l'outil « *integrate* » de *Spatial Analyst* pour supprimer les problèmes de faible dimension.

4. Règles de fonctionnement

Afin que le SIG reste organisé et efficace, il est important de mettre en place des règles d'utilisation, s'appliquant aussi bien aux utilisateurs ponctuels qu'aux analystes ou personnes gérant le système. Après toute modification du contenu de la géodatabase, la personne mettant la version mise à jour à disposition des autres utilisateurs indique dans les métadonnées générales son nom ainsi que la date et l'heure de la dernière modification.

D'autres règles importantes pour assurer la qualité des données sont données ci-après.

i. Insertion de nouvelles données

Tout jeu de données soumis pour insertion dans la géodatabase doit avoir des métadonnées correctement renseignées. Dans le cas contraire, la personne responsable du SIG se doit de le refuser en attendant une meilleure complétion des informations. Parmi les métadonnées les plus importantes pour justifier ou non de l'insertion figurent notamment :

- la description, devant clairement expliquer le contenu
- l'origine
- la période de validité (date de correspondance avec le terrain ou date de publication)
- l'état d'avancement (en cours, complet)
- la fréquence de maintenance
- les mots clés, associés au thésaurus du SIG de la Forêt Maya
- les contraintes d'accès et d'utilisation
- la qualité et la précision des données
- toute autre information nécessaire pour une utilisation efficace des données

Un modèle de métadonnées (fichier « .xml ») a été créé dans le cadre de cette étude pour les données produites par le *MesoAmerican Research Center*. Ce modèle peut être importé dans ArcCatalog et évite d'avoir à entrer les informations de distribution, de contraintes, ou encore de contact qui sont toujours les mêmes dans ce cas.

ii. Utilisation de données

L'accès et l'utilisation des données contenues dans la géodatabase sont conditionnés par le respect des contraintes explicitées dans les métadonnées. Lors d'une séquence de travail, l'utilisateur récupère les données qui lui sont nécessaires et en sauvegarde une copie servant pour son projet dans son répertoire de travail personnel, laissant la géodatabase disponible pour d'autres utilisateurs. Il ne travaille donc jamais directement avec les données stockées dans le SIG, seulement avec des copies.

iii. Modification de données

Dans le cas général, toute modification de données présentes dans la géodatabase doit se faire en concertation avec les autres personnes connaissant et utilisant celles-ci, le but étant d'éviter toute modification malheureuse (décision de supprimer un champ utile à d'autres utilisateurs par exemple). La personne effectuant les modifications travaille sur une copie, puis soumet celle-ci au responsable du SIG qui vérifie que les métadonnées ont bien été mises à jour pour refléter les changements opérés avant de remplacer l'ancienne version.

C. Les cartes

Le format de présentation des résultats ayant été retenu est le format de fichier « .mxd » de ArcGis. L'avantage de ce format réside dans le fait qu'il peut être aisément modifié pour changer la carte (date de publication par exemple). Un fichier image exporté au format « .png » (*Portable Network Graphics*) est également fourni pour pouvoir être directement utilisé. Ce format ouvert n'étant pas destructeur et offrant une bonne compression est en effet apparu comme le plus pertinent (Wikipedia 2009e) : pour une qualité égale, la taille du fichier est en effet considérablement réduite.

Les cartes se trouvent dans un dossier « *maps* », et sont classifiées par année de conception et nom du projet concerné : une classification par mois n'est pas nécessaire au vu du nombre limité de cartes produites chaque année.

Plusieurs règles régissent ces cartes :

- seules les couches nécessaires sont incluses (i.e. les couches visibles uniquement)
- toutes les données effectivement utilisées se trouvent dans un sous-dossier « *data* »
- les cartes sont enregistrées avec liens relatifs, aperçu, et leurs métadonnées sont correctement renseignées
- le cadre de données est configuré en « *fixed extent* » pour éviter tout changement accidentel de la mise en page
- le fichier image directement disponible est en qualité 200dpi, enregistré en « .png »
- la liste des rapports et articles utilisant cette carte est maintenue à jour dans la section « objectifs » des métadonnées

D. Le dossier de travail

Ce dossier est destiné à contenir les dossiers et documents de travail avec une organisation similaire. Cela concerne principalement les analyses effectuées au niveau du modèle prédictif : il est primordial de conserver les données utilisées pour pouvoir reproduire les résultats. Cela inclut des sélections aléatoires de points n'ayant pas de place dans la géodatabase par exemple, mais utiles pour les personnes travaillant à améliorer les résultats par la suite.

Tous les autres documents – courriels, documents OpenOffice, logiciels téléchargés par exemple – sont stockés ailleurs, en dehors du SIG.

La réorganisation du SIG étant effectuée, il est alors possible de passer à une phase d'analyse, à savoir la conception d'un modèle prédictif de sites archéologiques et d'un modèle de l'agriculture maya.

IV. Méthode de conception d'un modèle prédictif

A. Introduction

Depuis que les archéologues ont commencé à s'intéresser aux SIG au début des années 80, l'intérêt porté aux capacités de modélisation et de prédiction ne s'est pas démenti (Aldenderfer & Maschner 1996; Wheatley & Gillings 2002). Il s'agit d'ailleurs de la raison qui a poussé à la création d'un système d'information géographique au sein du projet étudié.

Afin de modéliser l'emplacement des sites archéologiques Maya au niveau régional, la méthode retenue est celle qui avait déjà été utilisée dans le passé pour la création du modèle à l'échelle locale : la méthode des *Weights of evidence* (Sirjean 2003; Monthus 2004; Clarke et al. 2009). Un important changement intervient dans la façon dont la méthode est mise en œuvre par rapport à ce qui a été fait précédemment puisque la nouvelle version de l'outil ArcSDM (*spatial data modeller*) parue en 2008 pour ArcGis 9.2 est utilisée, ainsi que celle pour ArcGis 9.3 (Sawatzky et al. n.d.).

Dans ces nouvelles versions, la mise en application ne se fait plus au moyen de menus comme dans le passé, mais en concevant un modèle au niveau d'ArcToolBox. Cet outil est donc plus adapté aux nouveautés qui ont été implémentées dans ArcGis par rapport à ArcView 3.2.

Il convient de réaliser deux études : l'une à l'échelle locale pour l'amélioration des résultats précédemment obtenus (Monthus 2004), et une autre à l'échelle régionale, avant une éventuelle généralisation au système du monde maya. Les paramètres pris en compte à des échelles différentes et les poids devant leur être attribués sont en effet différents. Ce problème bien connu des archéologues et qui a été soulevé dans de nombreux cas (Allen 2000) est très important dans le cadre de cette étude comme il est souligné par la suite.

B. Rappels de la méthode des *Weights of evidence*

Ces rappels sont nécessaires pour la conception du modèle conduisant au résultat dans ArcToolBox. Il convient de comprendre toutes les étapes et les éléments qui interviennent afin de produire un schéma logique et efficace. Cette phase est effectuée grâce à un modèle de démonstration mis à disposition par le concepteur d'ArcSDM. Le document communiqué décrit en détail la théorie avec l'étude d'un cas réel (Raines & Bonham-Carter 2006).

1. Méthode de mise en œuvre

ArcSDM est utilisable grâce à une boîte à outils qui s'intègre à ArcToolBox. Un certain nombre d'utilitaires sont mis à disposition, pour adapter la taille des rasters utilisés lors de l'analyse par exemple. Dans le cadre de cette étude, seuls les outils regroupés au sein du jeu « *Weights of Evidence* » seront considérés.

Il est rappelé que la méthode des *Weights of Evidence* permet de calculer la probabilité de trouver des sites à un endroit particulier en fonction de différents paramètres. Ces paramètres peuvent être liés à l'environnement naturel (pente, distance aux points d'eau), ou encore à des phénomènes sociaux, par exemple interactions entre différents groupes.

Le modèle prédictif est tout d'abord « entraîné » grâce à des zones où les emplacements de sites archéologiques sont connus (transects), produisant des poids associés à chaque classe d'évidence. Ces poids servent alors à calculer les probabilités a posteriori au niveau de la zone d'étude. Lors de la mise en œuvre de cette méthode, quatre étapes peuvent être identifiées, dont les outils associés sont séparés par couleurs dans le tableau page suivante.

L'outil « *Grand WOFÉ* » a un statut spécial, d'où sa couleur différente : il permet d'effectuer les trois premières étapes en une seule fois. N'offrant pas la même flexibilité d'analyse que les outils pris séparément (et notamment la possibilité d'effectuer des reclassifications au niveau des poids avant le calcul de la réponse), il n'a pas été utilisé lors de cette étude.

Outil	Paramètres	Effet
Training Sites Reduction	Points d'entraînement	Réduction du nombre de points d'entraînement afin de n'en conserver qu'un seul par cellule unitaire ou d'extraire une sélection aléatoire
Calculate Weights	Raster paramètre Points d'entraînement	Crée une table contenant les poids et les classes généralisées pour un certain paramètre
Calculate Response	Rasters paramètres Points d'entraînement Poids	Crée un raster contenant les probabilités a posteriori calculées pour une certaine zone à partir des poids utilisés
Logistic Regression	Rasters paramètres Points d'entraînement Poids	Calcule les probabilités par une méthode de régression logistique
Grand WOFE	Rasters paramètres Points d'entraînement	Exécute « <i>Calculate weights</i> », « <i>Calculate response</i> », et « <i>Logistic regression</i> »
Agterberg-Cheng CI Test	Post-probabilités Déviations standards Points d'entraînement	Test d'indépendance conditionnelle des rasters paramètres
Area Frequency Table	Post-probabilités Points d'entraînement	Crée une table permettant de tracer divers graphes faisant intervenir les classes d'évidence ainsi que le nombre de sites, permet d'évaluer la qualité de la classification et de la prédiction.

i. Sélection de points d'entraînement (sites)

Il est souhaitable pour obtenir un résultat correct de n'avoir qu'un site d'entraînement par cellule unitaire. Cette cellule unitaire dont la taille doit être déterminée est celle utilisée lors de l'analyse. En plus de cette réduction, l'outil associé permet d'effectuer des sélections aléatoires : cela permet de créer deux jeux de points, l'un servant pour entraîner le modèle et l'autre pour contrôler l'efficacité de la prédiction.

ii. Calcul des poids associés aux classes d'évidence

L'outil associé à cette étape analyse les classes d'évidence pour un certain paramètre au regard des points d'entraînement, et attribue à chacune d'elles un poids pouvant être soit négatif (la probabilité de trouver des sites au niveau de cette classe est diminuée), soit positif (la probabilité est augmentée). ArcSDM propose également à ce niveau une généralisation des classes.

iii. Calcul du raster réponse

Cette étape consiste à utiliser les poids calculés précédemment pour créer un raster des probabilités a posteriori au niveau d'une zone plus étendue. Deux outils sont disponibles pour cela. En effet, la méthode des *Weights of Evidence* ne donne des probabilités a posteriori valides que lorsqu'il y a indépendance conditionnelle des paramètres. Ce n'est cependant pas un problème si le résultat recherché consiste à savoir quel est le classement des zones. Dans le cas contraire, l'outil de régression logistique qui ne nécessite pas l'indépendance conditionnelle peut être utilisé pour obtenir des probabilités justes (Raines et al. 2009).

iv. Tests de qualité du résultat

La qualité du résultat obtenu peut être estimée de différentes façons. Dans le cadre d'un projet archéologique, il y a tout d'abord la vision des archéologues : « le résultat apparaît-il bon ? ». Cette appréciation – même si non fondée sur des méthodes mathématiques – a une grande valeur. Les arc-

héologues connaissent en effet le terrain et peuvent détecter rapidement une faille dans le modèle, pouvant par exemple être due à une mauvaise classification d'un paramètre.

L'outil « *Area Frequency Table* » utilisé avec des points d'entraînement ou de contrôle donne également un pourcentage de qualité de la classification ou de la prédiction. L'outil d'indépendance conditionnelle permet quant à lui de vérifier à partir du raster de probabilité et du raster de déviation standard l'hypothèse d'indépendance afin de choisir quelle méthode utiliser pour obtenir les bonnes probabilités a posteriori.

2. Modèle utilisé au niveau de ArcGis

À partir de ces considérations, le modèle utilisé au sein d'ArcGis est donné en Illustration 6. Ce modèle reprend les couleurs utilisées au sein de ModelBuilder 9.3, à savoir bleu pour les paramètres entrés, vert pour les résultats d'une exécution. En jaune sont représentées les différentes opérations effectuées.

Le modèle présenté ici est utile pour montrer l'enchaînement de la méthode, mais n'est pas réellement celui utilisé. En effet, il est plus simple de calculer les poids d'une part en faisant différents tests, puis de choisir lesquels utiliser pour calculer d'autre part les rasters de réponse. De la même façon, le test d'indépendance conditionnelle n'a pas besoin d'être lancé à chaque utilisation du modèle : une fois l'indépendance ou la dépendance établies, il n'est plus utile de l'exécuter lors de chaque test.

Il est bon de noter que dans notre cas nous utilisons comme paramètre « *Ignore missing data* » pour le calcul de la réponse : nous ne nous intéresserons donc pas au raster de variance ainsi qu'au raster de déviation standard générale, qui est dans ce cas le même que le raster de déviation standard associé aux probabilités a posteriori (Raines et al. 2009). Cela explique pourquoi « déviation standard » apparaît deux fois dans ce modèle : ces deux résultats sont en fait identiques.

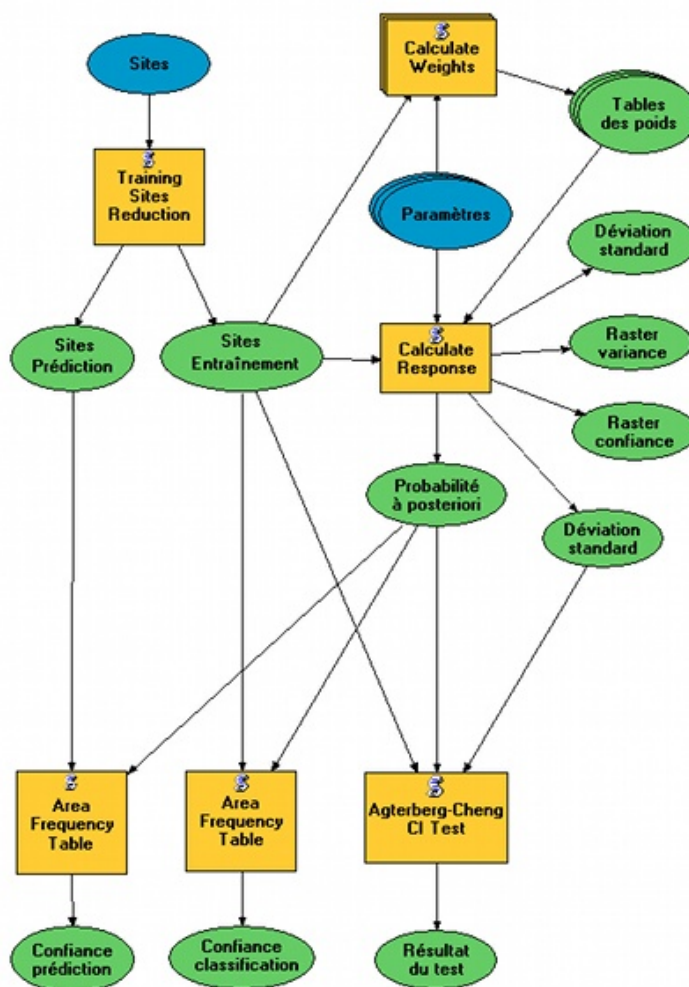


Illustration 6: Modèle prédictif utilisé

C. Réflexions sur l'implémentation

Les outils ayant été présentés, il est important de développer des réflexions insistant sur les aspects principaux pour l'implémentation de la méthode des *Weights of Evidence*, afin de donner des pistes pour la mise en place d'un modèle efficace. Ces considérations seront illustrées par la suite lors de l'application du modèle au projet archéologique considéré.

1. Choix des paramètres

Les paramètres utilisés lors de la mise en application de la méthode sont choisis en fonction de plusieurs raisons. Une étude préliminaire peut ainsi donner à penser qu'un paramètre aurait une influence sur la répartition des objets étudiés : une phase de récupération des données est alors nécessaire si celles-ci ne sont pas déjà disponibles.

Il est important que les classes d'évidence du paramètre étudié soient présentes de la même façon au niveau de la zone d'entraînement et de la zone d'étude (Wescott & Kuiper 2000) et la qualité des paramètres doit être considérée en terme d'adéquation à l'échelle du problème plutôt qu'en terme de précision absolue (Leusen 2002). Dans le cadre d'un projet archéologique, il est notamment important de prendre en compte le facteur temporel qui induit une part d'incertitude : le lit des rivières par exemple peut changer au cours du temps, ou encore la qualité des sols.

2. Choix des points d'entraînement

De même que les paramètres ont été choisis en fonction de leur pertinence au niveau de la zone étudiée, les points d'entraînement (qualifiés de « sites » dans le cadre d'un projet archéologique) se doivent d'avoir certaines qualités. Il convient de déterminer en premier lieu quelle est la définition d'un « site » par rapport à l'échelle de la zone d'étude, cela pouvant conduire à regrouper plusieurs unités en une seule plus importante (Leusen 2002).

La répartition des transects effectués joue également un rôle : il est très important de les choisir de telle façon qu'ils soient représentatifs de la zone étudiée. Leusen souligne ainsi le problème des biais apparaissant lors de recherches archéologiques (dus par exemple à l'influence de la végétation, la difficulté à détecter certains sites enfouis), ou encore le risque de transformer les zones à faible potentiel en zone inintéressantes. Dans ce cas, le modèle peut se transformer en un outil s'auto-prédissant : le résultat est influencé afin de correspondre à ce qui est attendu.

3. Choix de l'aire de la cellule unitaire

Il convient avant toute analyse de décider de la taille de la cellule unitaire. Cela implique de connaître approximativement les dimensions des sites dont la répartition est étudiée, une bonne cellule unitaire ne devant pas en contenir plus d'un, celui-ci devant être parfaitement contenu (Raines & Bonham-Carter 2006). Cependant, il peut arriver que malgré le choix d'une aire pertinente plusieurs points d'entraînement se retrouvent dans la même cellule. Pour éviter ce problème qui produit un biais dans le calcul des poids, il est important de toujours exécuter l'outil de réduction des sites au préalable.

Si le pourcentage de points retirés est trop important, cela signifie que l'aire de la cellule unitaire est probablement trop grande. Un examen de ces points permettra de décider de la décision à prendre : augmenter l'aire ou choisir de la conserver telle qu'elle est.

4. Choix des masques à utiliser

Lorsque les points d'entraînement peuvent être considérés comme ayant été uniformément recherchés au niveau de la zone d'étude (autrement dit, lorsque celle-ci correspond avec les transects), le masque servant lors du calcul des poids correspond tout simplement à l'étendue de la zone.

Dans le cas contraire, lors de la création de transects par exemple, il convient d'être prudent dans nos choix. Deux solutions sont envisageables :

- prendre comme masque d'entraînement l'étendue des transects
- prendre comme masque d'entraînement la zone d'étude

Le premier choix est celui qui devrait être employé (Raines et al. 2009). Pour peu que toutes les classes d'évidence présentes au sein de la zone d'étude soient représentées, cette méthode assure en effet les meilleurs résultats. Il peut hélas arriver que cela ne soit pas possible, pour des zones où les sites sont répertoriés mais pas les limites du lever effectué par exemple.

Dans ce cas, le masque d'entraînement retenu peut correspondre à la zone d'étude, ou à une zone plus restreinte englobant les sites. Ces choix ont deux conséquences :

- les valeurs des probabilités (a priori et a posteriori) ne sont pas significatives
- les poids affectés aux classes d'évidence sont susceptibles d'être faussés

Un modèle entraîné au niveau d'une zone ne peut être appliqué qu'à une zone constituée d'éléments similaires. Cela signifie qu'une classe de terrain absente du masque lors de la constitution des poids ne peut être prise en compte par la suite lors de la généralisation à la zone d'étude complète. Il est également nécessaire pour la qualité des poids dans le cadre d'un entraînement effectué à l'aide d'un masque couvrant une zone plus large que celle des transects d'avoir une proportion similaire des classes d'évidence dans les deux zones.

5. Contrôle et choix des poids

Il est rappelé que deux catégories sont disponibles pour chaque paramètre pour entraîner le modèle :

- *Categorical* : ce type de calcul est utilisé lorsque les classes d'évidence ne sont pas ordonnées
- *Ascending* : ce type correspond à des classes d'évidence ordonnées, par rapport un paramètre de distance à un cours d'eau

Le type le plus adapté doit être choisi, en prenant également en compte que :

- le type *categorical* peut faire apparaître des effets ignorés en utilisant l'autre type (par exemple un phénomène d'anneau : zone d'attraction entre 200 et 500 m)
- le type *ascending* est plus adapté à un paramètre ayant de nombreuses classes d'évidence contenant peu de sites

Pour les paramètres ordonnés, il sera donc toujours intéressant de considérer à la fois les résultats *categorical* et *ascending*.

Les paramètres utilisés au sein d'ArcSDM doivent être au format raster, avec des valeurs entières. Il convient donc de faire une reclassification éventuelle des données disponibles. Cette étape pourra également comporter une phase de généralisation pouvant avoir plusieurs raisons :

- adapter les données à l'échelle considérée
- un nombre plus limité de classes d'évidence et donc de plus grandes variations entre elles produira des résultats moins ambigus (Legg 2007)

Différents essais peuvent être nécessaires pour déterminer une bonne classification. Un nombre plus élevé de classes d'évidences ne produit donc pas forcément de meilleurs résultats : celles-ci comportent dans ce cas moins de points d'entraînement, ce qui peut mener à des erreurs d'interprétation.

Le premier calcul des poids ne se révélera pas être le plus judicieux dans la plupart des cas, et sera donc suivi d'une reclassification. ArcSDM donne dans la table des poids une proposition de reclassification pouvant servir de base. Il convient lors de cette étape de s'interroger sur les résultats apparaissant : une classe peut recevoir un mauvais poids en raison d'un manque de sites dû à l'emplacement des transects par exemple ; une reclassification en la couplant avec une autre classe peut alors s'avérer judicieuse.

Une valeur de confiance pour chaque classe d'évidence est calculée par ArcSDM : celle-ci est obtenue

nue en divisant le contraste par son écart type, et est utilisée par le logiciel pour la reclassification automatique. Le pourcentage de confiance peut être déduit de la valeur absolue de ce nombre en se référant à une table de Student (Wikipedia 2009f) avec une infinité de degrés de liberté (loi normale). Cette valeur est une approximation du fait de la conception d'ArcSDM, et est utilisée pour déterminer la généralisation (Raines et al. 2009, échanges par courriel avec Gary Raines) Une classe d'évidence sera qualifiée de « satisfaisant au niveau de confiance » lorsque sa confiance sera supérieure ou égale à la valeur choisie.

Afin d'obtenir des résultats cohérents, ces reclassifications doivent donc être le résultat d'une analyse menée en collaboration avec des personnes connaissant la réalité du terrain (archéologues). Cette étape est la plus importante, la conception de la carte des probabilités découlant directement de cela. Ces données seront également celles utilisées pour l'interprétation du résultat.

6. Création de la carte des probabilités et analyse qualitative du résultat

La production de la carte des probabilités se doit d'être accompagnée d'une analyse de la qualité. Cette question de savoir si un modèle est meilleur qu'un autre est difficile à résoudre, avec trois paramètres pouvant être pris en compte : les résultats, la méthodologie appliquée, ou les explications pouvant être fournies (Leusen 2002). La carte des probabilités établie peut être directement évaluée de différentes manières :

- **amplitude de l'intervalle des probabilités a posteriori** : il est possible de voir si la probabilité la plus importante est significativement supérieure à la probabilité a priori. De même, la probabilité la plus faible renseigne sur la capacité du modèle à définir des zones où peu de sites seront localisés
- **aspect visuel** : les personnes connaissant le terrain sont à même de dire si le résultat « semble juste » ou non. Cette estimation peut cependant conduire au phénomène déjà évoqué précédemment, dans lequel le résultat obtenu finit par correspondre après modifications à ce qui est attendu, et non pas à la réalité
- **spécificité** : une prédiction se doit d'être spécifique pour être utile, même si cela ne doit pas se faire au détriment de son exactitude (Leusen 2002). Cela doit bien sûr être considéré en fonction de l'échelle du modèle

L'analyse de la carte peut cependant contredire une analyse des poids par exemple : le choix de tel ou tel résultat aura donc une part de subjectivité. ArcSDM fournit plusieurs outils devant être utilisés à ce niveau. Le test d'indépendance conditionnelle renseigne ainsi sur la validité des probabilités a posteriori. L'outil *Area Frequency Table* fournit également :

- un pourcentage d'efficacité de classification lorsqu'il est utilisé avec les points d'entraînement (qualité de rétrodiction)
- un pourcentage d'efficacité de prédiction en utilisant des sites n'ayant pas servi pour l'entraînement

Dans ce second cas, il est bon de remarquer que si les points utilisés sont issus d'une sélection aléatoire visant à créer d'autre part les sites d'entraînement, les efficacités de classification et de prédiction seront de fait relativement liées puisque les points proviennent des mêmes transects. Il est donc plus intéressant de tester la qualité de prédiction au moyen de sites issus de levés indépendants. Ce fait est bien connu (Leusen 2002), mais n'est pas forcément suivi.

Lors de l'entraînement du modèle, il apparaît en effet naturel d'utiliser tous les transects disponibles au niveau de la zone d'étude. Choisir de réserver certains d'entre eux pour la prédiction uniquement insère d'ores et déjà un biais. Dans la pratique, l'efficacité de la prédiction peut être réévaluée lors de l'acquisition de nouveaux jeux de données, provenant dans le meilleur des cas de sources externes au projet. L'autre méthode consiste à effectuer de nouveaux transects, mais la question de l'indépendance de leur emplacement se pose alors : il peut apparaître naturel de les effectuer au niveau des zones de plus forte et de plus faible probabilités, alors que les zones de moyenne probabilité auront tendance à être plus négligées lors de ce type de contrôle, ce qui induit de fait un biais supplémentaire (Leusen 2002).

V. Validation et amélioration du modèle local

A. Objectifs

Avant de passer à la conception du modèle régional, une validation du modèle local conçu dans le passé (Monthus 2004, publié dans Clarke et al. 2009) s'impose, afin de contrôler en outre si la nouvelle version d'ArcSDM utilisée fournit bien les mêmes résultats. Des nouveautés sont de plus apparues, par exemple une proposition de reclassification des classes d'évidence des paramètres par le logiciel, alors que cela était fait manuellement dans le passé.

Cette validation double est d'une grande importance pour se familiariser avec les outils disponibles et aborder sereinement la modélisation à une échelle différente. Deux concepts de modèles peuvent être envisagés (Wheatley & Gillings, 2002, cités dans Dermody 2005) :

- les modèles déductifs, basés entièrement sur la théorie
- les modèles inductifs, basés entièrement sur l'observation

Celui conçu au sein du *MesoAmerican Research Center* est à caractère inductif. Le choix a été fait de se concentrer sur les paramètres environnementaux pour éviter que des présupposés théoriques n'induisent des erreurs. Les résultats obtenus servent alors à établir des explications archéologiques en prenant en compte les spécificités de la culture étudiée comme préconisé dans de nombreux ouvrages (par exemple Dalla Bona 1994). Il est ainsi important dans l'exemple traité de prendre en compte les usages en terme d'agriculture (spécificités du cycle du Milpa), de vie sociale.

B. Présentation de la zone et des paramètres

La zone qualifiée de « locale » est située à l'Est de la réserve archéologique d'El Pilar (Cf. Annexe 5 : localisation des zones locale et régionale, p.69). Cette zone de diagonale mesurant environ 65 km a été ainsi définie en fonction de plusieurs critères expliquant sa forme :

- existence de transects permettant d'entraîner le modèle
- présence de données intéressantes au niveau des sols (fertilité, drainage)

Les valeurs de drainage et de fertilité ont été établies à partir de cartes de classification des sols. La méthode appliquée pour cette classification est uniforme sur la zone d'étude (Wright et al. 1959), ce qui permet d'envisager la prédiction.

Les points d'entraînement ayant été utilisés dans le passé sont situés au niveau de trois transects nommés selon le plus important des centres qu'ils contiennent : le transect d'El Pilar, celui de Yaxox, et celui de Bacab Na. La validation a consisté à partir des images raster originales avant reclassification ou des versions plus précises quand celles-ci étaient disponibles, à recalculer les poids, à effectuer une reclassification, et enfin à voir le résultat final.

Les pixels des rasters paramètres ont tous une même taille de 20 m de côté.

Les paramètres retenus dans le passé (Sirjean 2003; Monthus 2004) sont ceux qui seront utilisés, aucun nouveau paramètre n'ayant été récupéré depuis :

Paramètre	Classes d'évidence
Distance aux rivières	Tampons tous les 100 m
Drainage	1 : bon ; 2 : moyen ; 3 : lent ; 4 : très lent
Fertilité	1 : élevée ; 2 : moyenne ; 3 : faible ; 4 : très faible
Pente topographique	En pourcentage, 100 classes

1. Rivières

Il apparaît intéressant d'utiliser un paramètre de distance aux rivières, l'hypothèse testée étant de savoir si les mayas privilégiaient la proximité aux cours d'eau pour leurs besoins quotidiens. Les tests effectués en 2004 et publiés récemment (Monthus 2004, publié dans Clarke et al. 2009) avaient amené à la conclusion que les mayas préféraient vivre dans une zone comprise entre 500 et 1000 m des rivières.

L'origine du jeu de données est cependant incertaine, les métadonnées étant manquantes. Il semblerait que diverses sources aient été regroupées, mais aucune information à propos de ces sources n'est disponible. Les différentes valeurs de bleu de l'illustration suivante représentent les tampons de distances aux rivières utilisés en 2004, du plus clair au plus foncé selon l'éloignement.

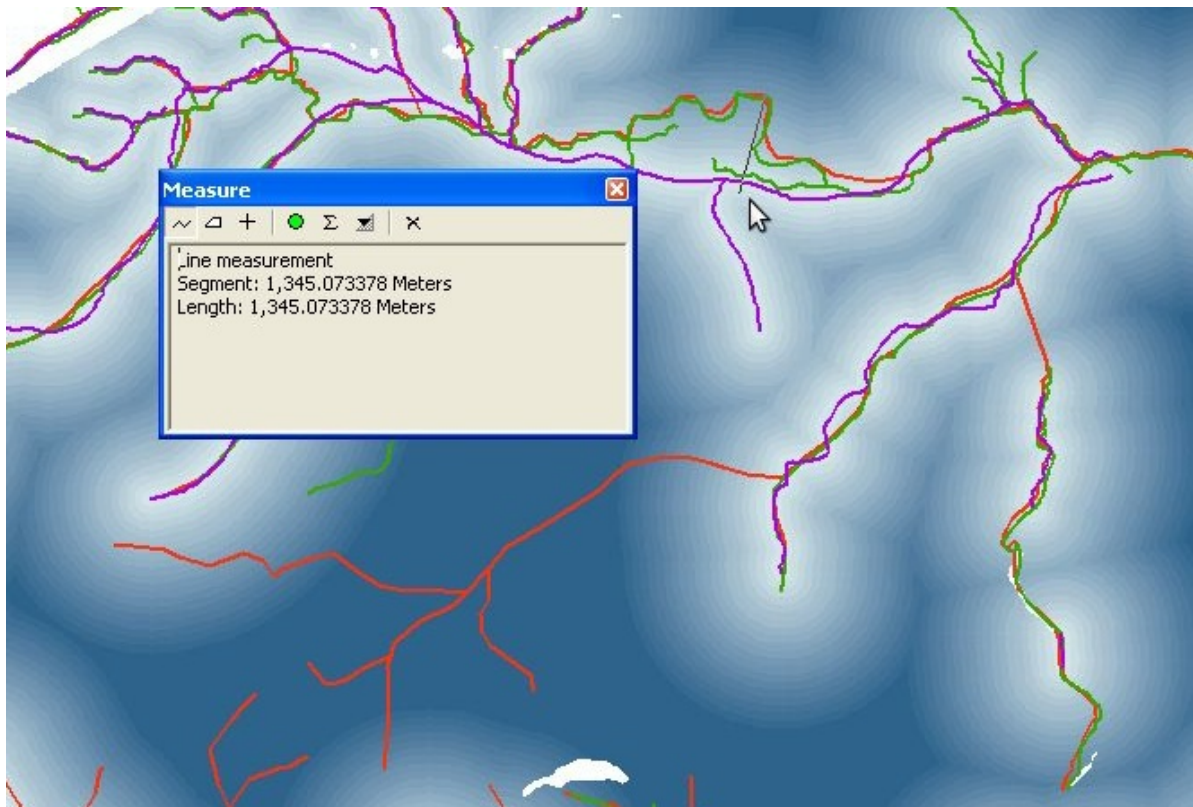


Illustration 7: Comparaison des jeux de rivières, et tampons utilisés dans le passé

Le site du BERDS (*Biodiversity and Environmental resource Data System of Belize*) met à disposition des fichiers shapefiles, contenant notamment les systèmes écologiques aquatiques (BERDS 2005). Ce fichier donne des rivières (en rouge) relativement différentes de celles utilisées dans le passé (en violet), tant au niveau de la forme que des ramifications (Illustration 7). Un troisième jeu de rivières (en vert) a également été trouvé dans un dossier du SIG, différent des deux précédents. Une fois de plus l'absence de métadonnées empêche de connaître sa source.

La question de la qualité de ce paramètre doit être considérée au vu de l'échelle. Les analyses précédemment effectuées portaient d'une classification par tampons successifs tous les cent mètres. À certains endroits, les données publiées par le BERDS en 2005 diffèrent du jeu utilisé lors de la conception du modèle de plus d'un kilomètre, ce qui influe évidemment sur les tampons apparaissant ici en arrière plan. Le modèle pourrait donc souffrir de ces insuffisances importantes.

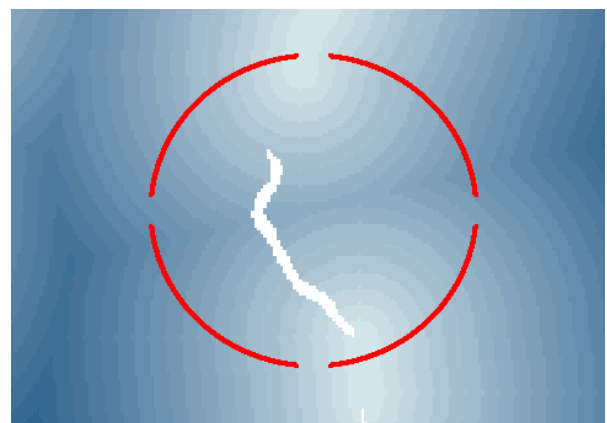


Illustration 8: Lac non pris en compte pour les tampons de distance aux rivières

Un autre problème apparaît : dans le passé les lacs avaient été considérés dans un jeu de données différent, qui avait été abandonné car ne fournissant pas de résultat probant. Le fait que le lit de certaines rivières passe par les lacs avait été occulté, ce qui crée donc des « trous » dans les tampons.

La largeur de la rivière Bélize, prise en compte lors de la création du masque d'application (les lacs ainsi que la rivière Bélize sont en *NoData*), n'est pas non plus considérée pour ces tampons (simple polyligne). Cela peut avoir une certaine importance puisque les transects utilisés lors de l'entraînement se trouvent près de cette rivière, et qu'elle atteint parfois 200 m de large.

Enfin, le lit des rivières est susceptible d'avoir changé puisque l'échelle de temps considérée concerne plus de mille ans, et les sites situés à proximité des rivières peuvent avoir disparu du fait des crues successives : ces évolutions avaient été ignorées.

Ce paramètre et les résultats associés doivent donc être analysés avec prudence. Le jeu de données produit par le BERDS peut ainsi servir de test, en incluant les lacs et en voyant les différences.

2. Drainage et fertilité

Les deux jeux de données de drainage et fertilité (présentés en Annexe 7 : carte locale du drainage et Annexe 8 : carte locale de la fertilité) sont relativement similaires, les valeurs ayant été établies à partir de la même carte des sols (Jenkin et al. 1976). L'indépendance conditionnelle risque de ne pas être respectée, mais utiliser un jeu de données comprenant les deux paramètres simultanément rendrait l'interprétation plus difficile, et en multipliant le nombre de classes réduirait la fiabilité. La conception d'un tel raster peut être effectuée de façon à ce que les paramètres soient reflétés par les valeurs, en considérant par exemple :

$$\text{Nouveau raster} = \text{Raster fertilité} + 10 * \text{Raster drainage}$$

Dans l'exemple précédent, la valeur 32 représente ainsi un drainage de classe 3 et une fertilité de classe 2. Il est en revanche toujours aussi difficile d'analyser la qualité des résultats, notamment au niveau des poids. Il est en effet raisonnable de penser que les poids pour la fertilité et le drainage seront meilleurs pour les classes de bonne qualité que pour les autres, et organisés de façon décroissante : un contrôle immédiat est donc possible pour valider leur pertinence ou identifier un éventuel problème.

Dans le cas d'une image raster regroupant les deux paramètres, un tel contrôle s'avère difficile : comment prédire en effet si un drainage de classe 3 et une fertilité de classe 2 s'avèrent meilleurs qu'un drainage de classe 2 et une fertilité de classe 3 ? Les intuitions a priori et les analyses a posteriori ne s'avèrent donc pas si aisées.

3. Pente topographique

Le jeu de données de pente topographique utilisé dans le passé était constitué d'une image raster ayant un pixel de 50 m. Un nouveau fichier créé à partir des données SRTM niveau 2 est désormais disponible, ayant un pixel de 30 m. Cette différence est importante à l'échelle considérée afin de rendre correctement le relief en respectant la taille des sites comme le montrent les illustrations 9 et 10 au niveau de la même zone (les fortes pentes sont en blanc, les faibles en noir).

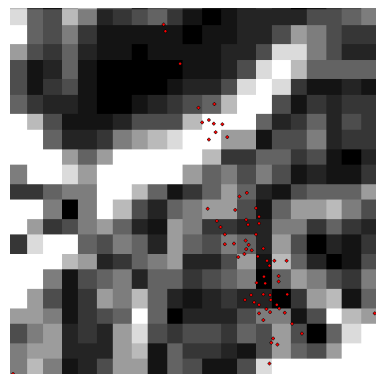


Illustration 9: Pixel de 50 m

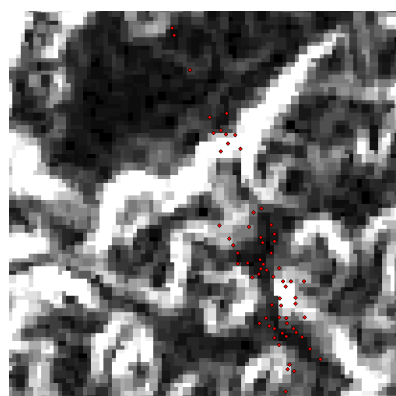


Illustration 10: Pixel de 30 m

La pente considérée sera donc cette fois celle ayant un pixel de 30 m, ce qui devrait améliorer le modèle grâce à des données plus adaptées à l'échelle étudiée. Il est en effet possible de constater qu'avec un pixel trop important certains sites se trouvent dans des classes différentes de la réalité : la taille de pixel d'un paramètre ne devrait pas être beaucoup plus grande que celle des objets étudiés, de même qu'il n'est pas souhaitable qu'elle soit beaucoup plus petite (les sites étant alors localisés sur plusieurs classes du même paramètre).

Le relief apparaissant très varié au niveau de la zone d'étude (pente de 0% à 101%), ce paramètre pourrait avoir une grande importance.

C. Mise en application de la méthode

1. Introduction

La démarche mise en place dans cette partie consiste à entraîner le modèle au moyen de divers jeux de points (dont celui utilisé dans le passé) et de plusieurs masques afin d'identifier les différences pouvant apparaître. Cet exemple réalisé avec le projet archéologique étudié vise à présenter des pistes pour augmenter la fiabilité des résultats, en s'appuyant sur les conseils donnés pour l'utilisation d'ArcSDM et en montrant en quoi la mise en application n'est pas si évidente et comment y remédier.

2. Présentation des sites et des masques

Trois masques présentés en Annexe 6 : masques utilisés à l'échelle locale vont être testés. Dans tous les cas, la zone finale de calcul des probabilités (zone d'étude) sera celle couverte par le masque local. Ces trois masques seront désignés de la façon suivante :

- **masque des transects** : correspond à la zone couverte par les transects utilisés en 2003 et 2004, à savoir les transects d'El Pilar, Bacab Na et Yaxox
- **masque « cover »** : celui-ci est constitué du masque des transects auquel ont été ajoutés les transects de Barton Ramie, Cadena Creek et Beaver Dam
- **masque local** : couvre toute la zone étudiée à l'échelle locale

Les sites utilisés au niveau du masque des transects sont ceux ayant servi dans le passé (Sirjean 2003; Monthus 2004) pour l'entraînement du modèle, mais le jeu de points d'entraînement n'est pas exactement le même : une sélection aléatoire avait été utilisée, mais non sauvegardée. Les points disponibles au niveau des autres transects sont ajoutés à ceux-ci, de même que des points dont l'étendue des levés n'est pas connue, ce qui donne un total de 1027 sites au niveau du masque local.

Lors des travaux passés, l'aire de la cellule unitaire considérée était de 0,01 km². La réduction du jeu de points à un site par cellule unitaire élimine plus de 40% des sites dans ce cas (1027 sites à l'origine, dont 615 seulement peuvent être utilisés), ce qui pourrait signifier un sur-dimensionnement. Or, il apparaît que les sites sont majoritairement d'une taille inférieure à 20 m, et séparés l'un de l'autre d'environ la même distance au minimum. Pour cette raison, il a été décidé de réduire la cellule unitaire à 0,0016 km² : la réduction retire dans ce cas 7% des sites, cette taille est donc plus pertinente.

Il est à noter que comme précédemment expliqué les sites sont séparés en deux jeux, l'un servant à contrôler l'efficacité de la prédiction, l'autre à entraîner le modèle. Les points servant à l'entraînement pour le masque des transects et le masque *cover* sont extraits de ce dernier. Les probabilités finales seront cependant calculées en utilisant l'ensemble des sites « réduits » afin qu'elles soient valides, et utilisables pour des calculs de densités. La Table 1 présente la liste des fichiers *shapefile* résultants de cela. Les sites servant à tester l'efficacité de la prédiction sont toujours les mêmes pour pouvoir réellement comparer : 25% des points sont réservés pour cela.

Description	Fichier associé	Nombre de sites
Tous les sites	sites_all.shp	1027 sites
Sites « réduits »	s_a_all.shp	954 sites
Sites d'entraînement local	s_a_tra.shp	716 sites
Sites de prédiction	s_a_pre.shp	238 sites (25% des sites)
Sites d'entraînement <i>cover</i>	s_c_tra.shp	452 sites
Sites d'entraînement transect	s_t_tra.shp	247 sites

Table 1: description des jeux de sites

3. Étude qualitative des sites et des zones d'entraînement

Dans le cas du masque des transects, les sites se trouvent au niveau de transects dont la zone est identifiée. L'entraînement du modèle effectué dans le passé ne prend cependant pas en compte celle-ci : celle retenue correspond en fait au masque local délimitant la zone d'étude totale. La Table 2 présente la répartition des différentes classes d'évidence obtenues pour le drainage, les résultats pour les autres paramètres pouvant être trouvés en Annexe 12 : ratios calculés avant reclassification.

Drainage	Masque local		Masque des transects			Masque cover		
	Pixels	Pourcentage	Pixels	Pourcentage	Ratio	Pixels	Pourcentage	Ratio
1	663816	20.61%	2373	18.93%	0.92	9316	41.36%	2.01
2	742333	23.04%	6965	55.57%	2.41	6965	30.92%	1.34
3	1053315	32.70%	1711	13.65%	0.42	2556	11.35%	0.35
4	761894	23.65%	1485	11.85%	0.5	3689	16.38%	0.69
	3221358	100.00%	12534	100.00%	1	22526	100.00%	1

Table 2: répartition des classes de drainage au niveau des masques

Il apparaît que la seconde classe de drainage est sur-représentée au niveau du masque des transects, alors que d'autres sont sous-représentées, le même problème apparaissant pour le masque *cover*. Les graphiques ci-dessous montrent en ordonnée les ratios définis précédemment pour les différentes classes des paramètres placées en abscisse. Aucun de ceux-ci n'est satisfaisant à ce niveau.

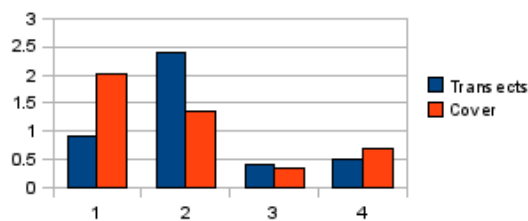


Illustration 11: ratios pour le drainage

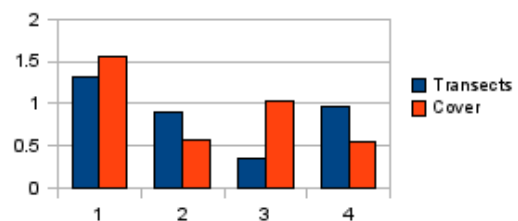


Illustration 13: ratios pour la fertilité

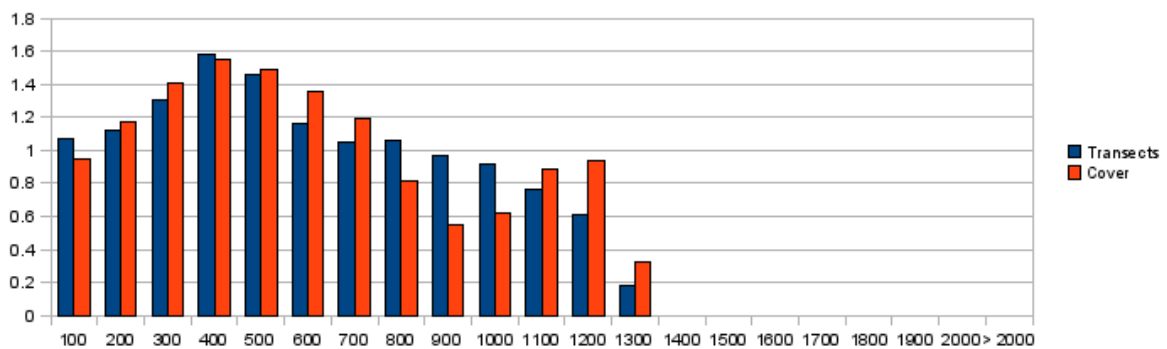


Illustration 12: ratios pour la distance aux rivières

La distance aux rivières souffre par exemple d'une grande faiblesse : à partir de 900 m, les tampons sont sous-représentés au niveau du masque des transects, voire même totalement absents pour les distances supérieures à 1400 m (pourtant présentes au niveau de la zone d'étude). Cela signifie qu'à partir de 900 m les poids seront sous-estimés s'ils sont calculés au moyen des sites correspondant à ces masques en utilisant le masque local. Si l'entraînement est fait à l'aide des bons masques, ces distorsions ne sont pas réellement problématiques, mais aucun poids ne sera calculé au delà de 1300 m, ce qui ne convient pas non plus.

Les résultats obtenus dans le passé (Monthus 2004, publié dans Clarke et al. 2009) signalant une zone à poids négatif au delà de 1100 m pourraient provenir de ces problèmes, hypothèse que cette validation peut faire apparaître. Entraîner le modèle au moyen du masque *cover* ou du masque des transects apparaît être la meilleure des solutions au vu des distorsions. Le jeu de sites d'entraînement local n'apparaît pas très fiable puisque les zones de lever ne sont pas connues : seul le masque local dans sa totalité peut être utilisé.

4. Détail du calcul des poids : masque des transects

Dans le passé, le fait que le type *ascending* soit destiné à un nombre élevé de classes d'évidence avait été ignoré. Le choix du type de classification effectuée dépendra donc cette fois de la pertinence des poids : lorsqu'ils sont correctement ordonnés avec *categorical* il n'est pas nécessaire de recourir à *ascending*. Les poids calculés sont donnés en Annexe 13 : entraînement au moyen du masque et des sites des transects.

i. Drainage et fertilité

La table 3 présente les résultats obtenus pour le drainage lors d'un entraînement au moyen du masque et des points des transects, W+ et W- étant les poids positifs ou négatifs calculés.

Drainage ascending							
Classe	Sites	W+	W-	Contraste	Confiance	Généralisation	Poids
1	113	1.01	-0.43	1.44	10.51	2	0.26
2	234	0.26	-1.64	1.91	6.62	2	0.26
3	247	0.14	-8.06	8.2	0.82	1	-1.64
4	247	0	2.46	-2.46	-0.17	1	-1.64

Table 3: poids pour un drainage de type ascending

Dans cette étude, la confiance qui a été choisie vaut deux (valeur proposée par défaut par ArcSDM), soit environ 97,5%. Dans le cas du drainage, il apparaît donc qu'en choisissant un type *ascending* deux classes généralisées apparaissent, séparant « bon et très bon » avec un poids positif, et « lent et très lent » avec un poids négatif.

Si maintenant une classification de type *categorical* est prise en compte, le résultat est le suivant :

Drainage categorical							
Classe	Sites	W+	W-	Contraste	Confiance	Généralisation	Poids
1	113	1.01	-0.43	1.44	10.51	1	1.01
2	121	-0.14	0.15	-0.29	-2.16	2	-0.14
3	13	-1	0.1	-1.11	-3.81	3	-1
4	0	0	0	0	0	99	-8.06

Table 4: poids pour un drainage de type categorical

Dans ce cas, les quatre classes d'évidence sont conservées lors de la généralisation, et il apparaît que les poids sont ordonnés correctement. Trois des classes satisfont au niveau de confiance. Le cas de la classe 4 est intéressant : la confiance nulle n'est en fait pas dérangeante puisqu'elle vient tout simplement du fait qu'aucun site d'entraînement ne se trouve dans une zone de drainage très lent.

L'analyse des poids en gardant à l'esprit la signification des classes permet donc dans ce cas de choisir une classification *categorical* qui fera plus ressortir la réalité, avec un poids très négatif pour

l'une des classes n'apparaissant pas aussi clairement dans l'autre cas. Le choix pour le drainage sera donc de prendre le type *categorical*, en conservant quatre classes.

Une analyse similaire conduit à choisir pour la fertilité le même type, avec cette fois trois classes.

ii. Rivières

Les classements des rivières en *ascending* ou en *categorical* ne sont pas satisfaisants (en *ascending* les niveaux de confiance sont trop faibles et ArcSDM ne peut pas calculer de généralisation valable, en *categorical* aucune organisation logique des poids n'apparaît), que ce soit pour celles utilisées en 2004 ou celles mises à disposition par le BERDS. Un entraînement utilisant les mêmes sites mais le masque local fait apparaître une reclassification avec le type *ascending*, la valeur de coupure se trouvant à 1100 m. Ainsi qu'il a été signalé lors de l'étude qualitative des paramètres, les classes d'évidences sont sous-représentées au niveau de la zone des transects à partir de cette valeur : il est donc préférable dans ce cas de ne pas utiliser ce paramètre. Les autres entraînements et essais de reclassifications effectués pour les distances aux rivières appuient cette décision : cette variable n'entre visiblement pas dans le choix de l'emplacement des sites.

Une reclassification satisfaisante consisterait à prendre d'une part le tampon de 100 m, et d'autre part les distances supérieures à 100 m. Dans ce cas, les valeurs de confiance sont bonnes, avec un poids négatif pour la première classe, positif pour la seconde. Cependant, il a été souligné que l'emplacement des rivières varie de plusieurs centaines de mètres selon les jeux de données. Cette zone de 100 m n'est donc pas significative et apparaît plus de nature à fausser le modèle qu'à l'améliorer.

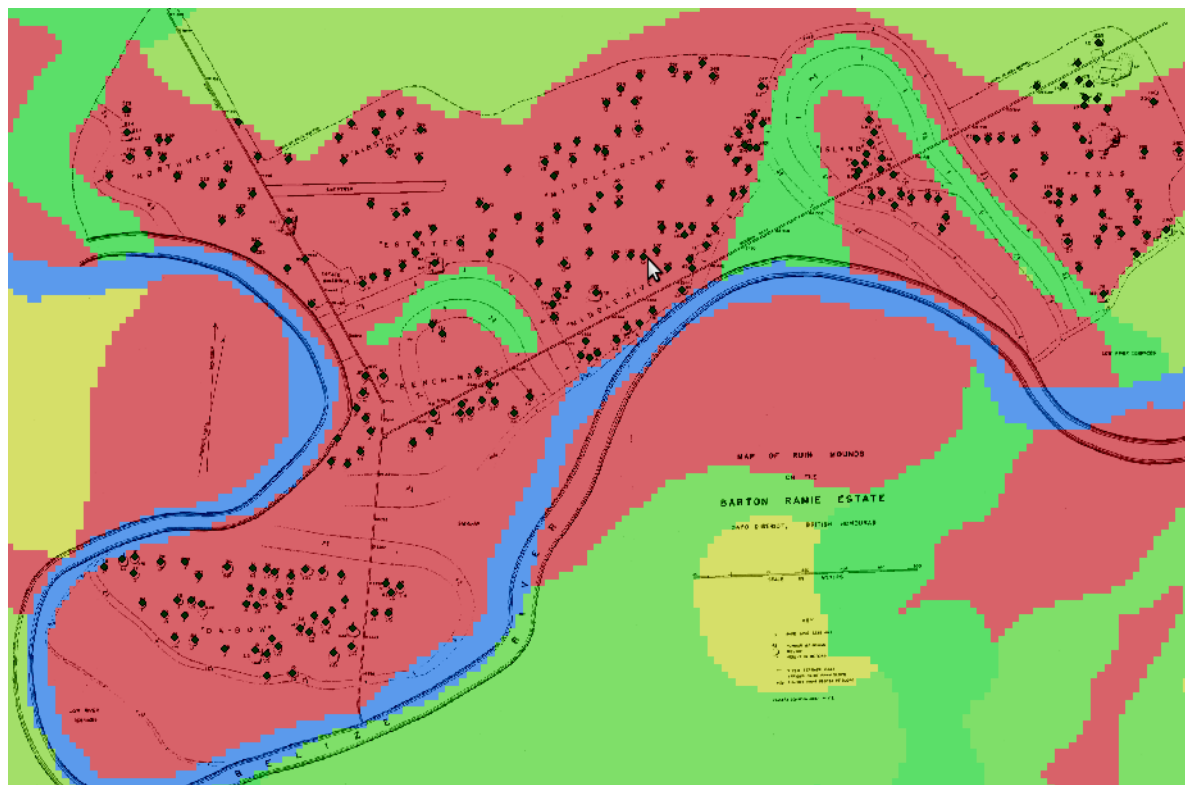


Illustration 14: difficultés inhérentes à la position des rivières, dans le présent ou le passé

L'illustration 14 donne un exemple avec la zone du transect de Barton Ramie et de ses sites, avec en fond un résultat du modèle prédictif ne faisant pas intervenir le paramètre de distance aux rivières (les zones à fortes probabilités sont en rouge, celles à faibles probabilités sont en vert). La rivière du plan ne correspond pas totalement avec celle du masque (en bleu), et il apparaît d'après la répartition des sites et du relief sur le plan que dans le passé le tracé de celle-ci était différent.

Au vu de toutes ces considérations, le paramètre de distance aux rivières ne sera donc pas utilisé pour le calcul des probabilités.

iii. Pente topographique

Le calcul avec le type *categorical* ne fait apparaître aucun résultat intéressant, ce qui est logique au vu du très grand nombre de classes. En revanche, un classement *ascending* fait apparaître une coupure au niveau de 33% de pente, cette valeur servira donc pour la reclassification.

5. Calcul du résultat

Les paramètres utilisés pour le calcul de la carte de probabilités avec le masque des transects sont donc les suivants :

Fertilité	Trois classes d'évidence
Drainage	Quatre classes d'évidence
Pente topographique	Deux classes d'évidence (< 33%, >= 33%)
Points de prédiction	s_a_pre.shp
Points et masques d'entraînement	transects, cover, local
Points de généralisation	s_a_all.shp

Le tableau suivant présente un récapitulatif des poids, valeurs de confiance, et résultats finaux obtenus au moyen des paramètres reclassifiés pour tous les masques et jeux d'entraînement.

Sites	Transects				Cover				Local	
Mas que	Transects		Local		Cover		Local		Local	
Paramètres	Poids	Confiance	Poids	Confiance	Poids	Confiance	Poids	Confiance	Poids	Confiance
Drainage										
Bon	1.01	10.51	0.8	9.23	0.56	11.63	1.2	21.02	1.22	26.8
Moyen	-0.14	-2.16	0.75	9.16	-0.15	-1.96	0.15	1.88	-0.19	-2.57
Lent	-1	-3.81	-1.83	-7.61	-0.94	-4.48	-1.95	-10.29	-1.06	-11.34
Très lent	-8.06	0	-8.67	0	-8.99	0	-9.28	0	-5.13	-5.4
Fertilité										
Élevée	0.48	8.56	0.7	11.37	0.42	11.32	0.82	16.83	0.77	20.7
Moyenne	-0.59	-5.07	-0.66	-5.99	-0.7	-5.57	-1.25	-11.1	-0.96	-12.57
Faible	-8.13	0	-8.55	0	-9.03	0	-9.15	0	-5.01	-5.23
Pente										
< 33%	0.03	2.35	0.01	1.18	0.02	2.37	0.01	2.04	0.02	2.7
>= 33%	-1.65	-2.35	-0.83	-1.18	-1.67	-2.37	-1.43	-2.04	-1.89	-2.7
Prob. à priori	0.001185		0.001185		0.001185		0.001185		0.001185	
Prob. à posteriori										
Minimum	2 10 ⁻¹¹		2 10 ⁻¹¹		3 10 ⁻¹²		2 10 ⁻¹²		7 10 ⁻⁹	
Maximum	0.005429		0.005366		0.003210		0.008985		0.008725	
Eff. de classification	76.40%		76.60%		82.10%		82.00%		80.40%	
Eff. de prédiction	79.60%		79.50%		79.60%		79.60%		79.60%	

Table 5: comparatif des résultats pour les différents masques et jeux de sites

Les valeurs apparaissant en bleu dans le tableau précédent sont celles ne satisfaisant pas au niveau de confiance de 2 choisi pour cette étude.

D. Conclusions pour le modèle local

1. Fiabilité du modèle

Il est remarquable de constater que les résultats fournis dans le tableau sont relativement similaires, ce qui prouve la fiabilité du modèle. Le niveau de confiance qui avait été décidé est satisfait dans la quasi-totalité des cas. Pour le problème surgissant pour la pente en entraînant au moyen des sites des transects et du masque local, il apparaît que les poids sont en effet différents pour la classe « pente supérieure à 33% ».

Pour ce qui est du drainage, les confiances sont proches du niveau fixé à l'origine. Il est bon de remarquer que les poids des couples (masque/site) transect/transect, *cover/cover* et local/local sont approximativement les mêmes (-0.14, -0.15, -0.19), alors que ceux obtenus grâce aux couples transects/local et *cover/local* sont différents (0.75, 0.15). Cela provient probablement du fait que la seconde classe de drainage est sur-représentée au niveau des zones des levers (cf. Illustration 11: ratios pour le drainage, p.26) : en utilisant le masque local avec ces points, les poids sont donc faussés. Pour la classe 2 du drainage, les poids acceptables sont donc ceux négatifs.

Le même phénomène apparaît pour la troisième classe, qui est cette fois sous-représentée et hérite d'un poids proche de -2 lors d'un entraînement au moyen du masque local pour les sites *cover* ou des transects, alors que celui-ci est environ égal à -1 pour les entraînements au moyen de masques adaptés aux sites.

2. Choix du résultat à prendre en considération

Au vu de ces considérations, le résultat le plus satisfaisant apparaît être celui obtenu avec le couple masque *cover* – sites *cover*. Le masque d'entraînement correspond à la zone des levers effectués sur le terrain, ce qui enlève les problèmes de sur-représentation de certaines classes, et il comporte une plus grande diversité et population de sites que pour le masque des transects. Alors que toutes les efficacités de prédiction sont quasiment les mêmes, ce résultat remporte la meilleure efficacité de classification.

Un test d'indépendance conditionnelle peut maintenant être effectué avec ce résultat pour déterminer si les probabilités obtenues sont valides ou si seules les zones doivent être considérées. Dans le cadre de cette étude, l'outil de régression logistique ne pourra pas être utilisé dans ce dernier cas puisqu'il est limité à une taille de cellule unitaire d'au moins 1 km². Le résultat du test apparaît en Annexe 15 : test d'indépendance conditionnelle (local) : dans ce cas, il est passé avec succès.

3. Comparaison avec les résultats de 2004

Une première comparaison avec les résultats de 2004 peut être faite au niveau des aspects des cartes de probabilités (voir illustrations 15 et 16). Ces cartes sont également disponibles en plus grandes tailles avec les valeurs de probabilités en Annexe 16 : pour le nouveau modèle, et Annexe 17 : pour l'ancien. Les valeurs de probabilités ne sont pas les mêmes puisqu'elles dépendent des sites utilisés pour le calcul du résultat, il convient donc de s'intéresser plutôt aux zones pour effectuer la comparaison.

Les fortes probabilités apparaissent ici en rouge, les probabilités moyennes en jaune, et les faibles probabilités en vert. Il apparaît qu'en 2004 les rivières avaient un effet relativement important sur le résultat. La zone de poids négatif pour ce paramètre (moins de 500 m et plus de 1000 m de distance aux rivières) est très visible, notamment au niveau des classes de faible probabilité.

Les zones de forte probabilité sont beaucoup plus marquées dans le nouveau modèle, notamment au niveau de la rivière Bélize. Les sites du transect de Barton Ramie (précédemment présentés en Illustration 14 p.28) étaient ainsi majoritairement classés en classe intermédiaire dans le passé au lieu d'apparaître en classe supérieure. Plus généralement, le bandeau de forte probabilité entourant la rivière Bélize et ses sols de bonne qualité disparaît en grande partie dans le modèle de 2004, du fait de la zone de poids négatif entourant les cours d'eau.

Il convient ici de rappeler que ces probabilités indiquent des zones de présence plus ou moins importante de sites mayas. Les couleurs donnent une première idée de la répartition, mais un calcul des densités associées aux zones définies est nécessaire pour une approche quantitative.

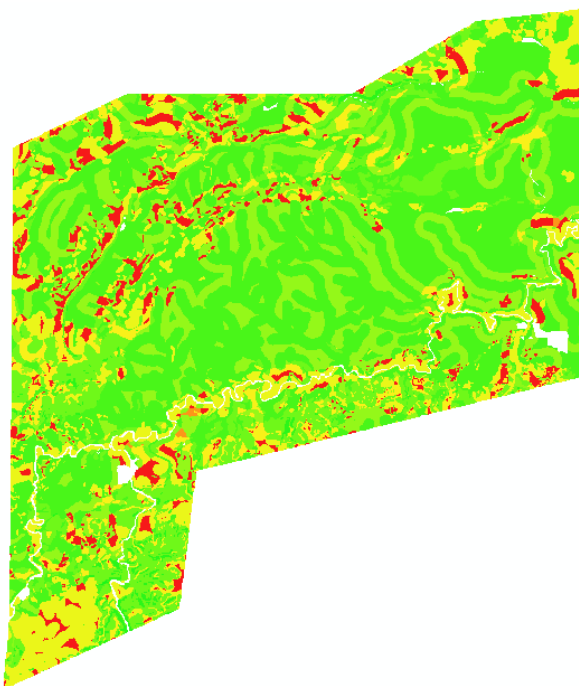


Illustration 15: modèle de 2004

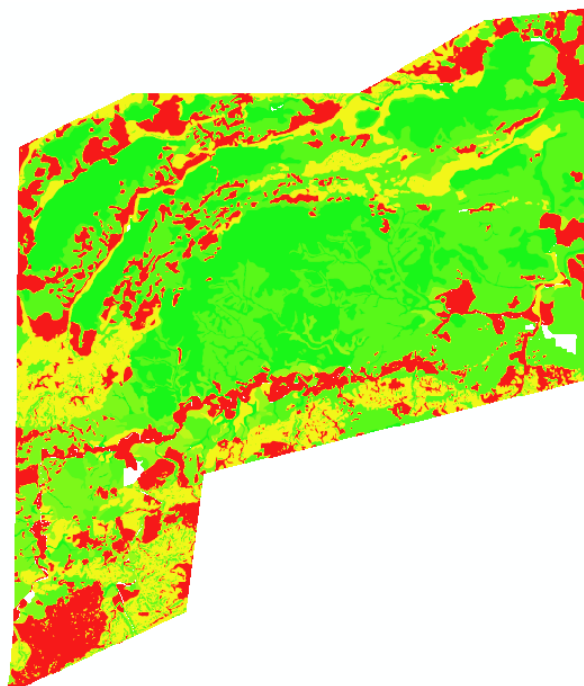


Illustration 16: nouveau modèle

Il est difficile de comparer plus précisément ces deux cartes à ce stade de l'étude. Celle de 2004 pourrait être considérée plus efficace en ce qu'elle est plus sélective, mais l'analyse de la reclassification de ses paramètres et de l'utilisation des rivières contrebalance ce jugement. La qualité du résultat sera donc évaluée en fonction de la répartition des sites et de la population qu'il permet d'effectuer après reclassification des probabilités.



Illustration 17: temples mayas dépassant de la canopée, Tikal

VI. Estimation de la population pour la zone d'étude

A. Étude des sites

Le modèle prédictif précédemment effectué permet de déterminer vingt probabilités. Celles-ci définissent des zones de présence de sites, mais ne permettent pas directement une analyse poussée. En effet, les sites utilisés lors de l'entraînement ont des natures très différentes. Dans le cadre de la réorganisation du SIG effectuée lors de ce travail de fin d'études, une activité importante a ainsi consisté à agréger et compléter la description des sites présents au niveau des transects disponibles (ensemble des sites dont la zone des levers était identifiée, précédemment désignés « sites du masque *cover* »). Cette description comporte notamment :

- l'identifiant du site
- son *labor investment* : nombre de jours de travail nécessaires à sa construction
- valeur de sa grande diagonale
- nombre de structures présentes
- distance à son plus proche voisin

Le *labor investment* est calculé au moyen d'une formule établie par Anabel Ford, faisant notamment intervenir la taille du site et le nombre de structures qu'il contient, et permettant de définir quatre rangs d'importance, allant de 1 à 4. Les sites de rang 1 (moins de 500 jours de travail) et ne comportant qu'une seule structure sont considérés comme des unités temporaires et reclassifiés en rang 0, tandis que les autres sont des unités résidentielles.

La première étape ne prend en compte que ces dernières, afin de répartir la population. La différenciation est importante puisque 41% des sites sont de rang 0 (voir Illustration 18). Ceux-ci seront propagés dans un second temps puisqu'ils jouent un rôle important dans l'occupation des sols.

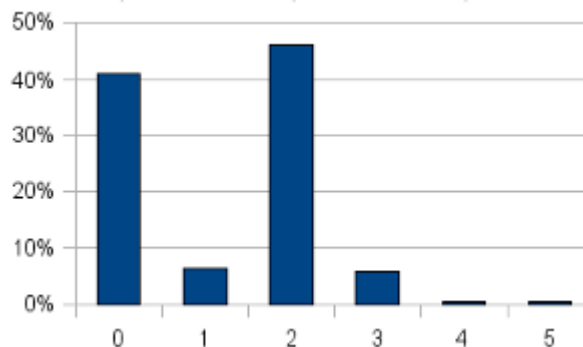


Illustration 18: répartition des sites par rang

Une estimation préalable de l'emplacement des habitations est effectuée sans prendre en compte les différents rangs, afin d'avoir une première vision globale de contrôle de la pertinence des résultats. Dans le cas où la reclassification des classes prédictives s'avère efficace, il est alors aisé de reproduire la même méthode en utilisant les rangs.

B. Reclassification des classes prédictives

1. Analyse du résultat utilisé dans le passé

Une estimation de la population de la zone locale étudiée avait été effectuée dans le passé. Cependant, cette estimation doit maintenant être refaite au vu des nouveaux éléments :

- elle s'appuyait sur le modèle de 2004 qui présente des incertitudes et dont les reclassifications (quatre classes de prédiction) ne paraissent plus pertinentes
- des reclassifications « manuelles » ont été effectuées après coup pour compenser les problèmes du modèle

Les illustrations ci-dessous présentent ainsi les classes de probabilités qui avaient été issues du modèle prédictif de 2004. La propagation de sites réalisée à partir des quatre classes (image de gauche) n'étant pas concluante, le parti pris avait été de créer une cinquième classe « nulle » dans laquelle au-

cun site n'était propagé, basée sur les sols : les catégories de sols nommées « *beaver dam* » et « *al-kache* » étaient simplement retirées.

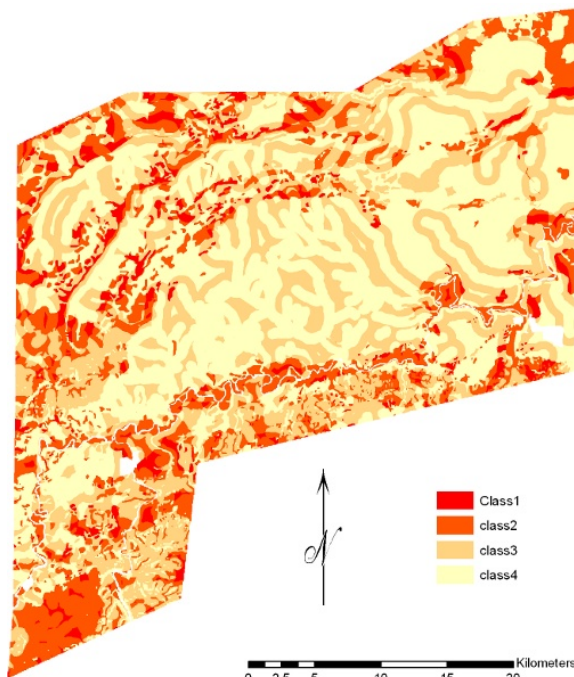


Illustration 19: Classes prédictives du modèle de 2004

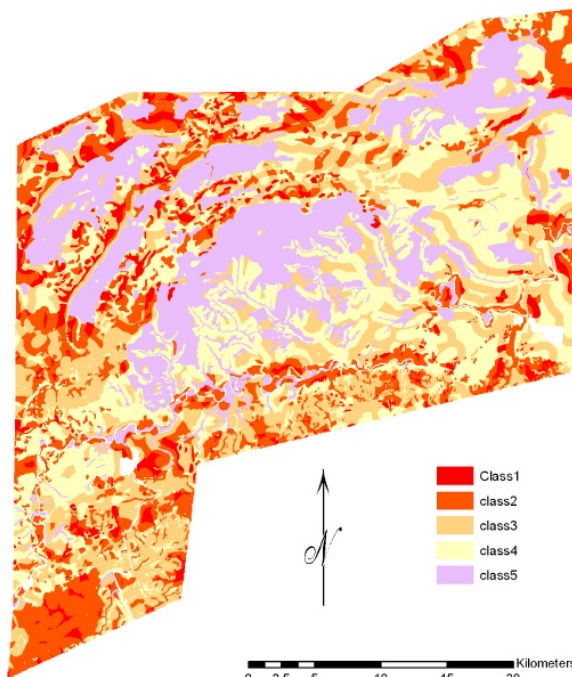


Illustration 20: application d'une cinquième classe pour compenser les problèmes

Cette méthode rejoint les problèmes évoqués précédemment (Leusen 2002) : le modèle prédictif est tout d'abord considéré comme valide, mais ses classes sont ensuite changées manuellement sur une base subjective pour obtenir le résultat escompté.

Le modèle utilisant la carte des sols pour le drainage et la fertilité, une telle reclassification a posteriori ne devrait pas être nécessaire. Le problème pourrait signifier qu'une autre classe issue des informations sur les sols intervient au point de fausser le résultat, et pourrait changer également les autres zones de celui-ci. Le fait de devoir « réutiliser » les sols pour créer artificiellement une nouvelle classe peut également montrer qu'un autre paramètre du modèle joue un rôle plus important qu'il ne le devrait, et atténue les effets dus aux classes d'évidence des sols. Lors de la reclassification en 4 classes (la carte de probabilités initiale en comporte vingt), cela ferait donc disparaître des nuances dans les paramètres, nuances étant importantes pour la classification. Cela rejoint les discussions précédentes (p.28) sur le problème des rivières et leur prédominance dans le modèle de 2004.

2. Nouvelle méthode mise en place

À partir des considérations précédentes, il apparaît que la reclassification est une phase cruciale. Afin de travailler efficacement, les valeurs du raster de probabilités sont multipliées par 100 000 afin d'obtenir, et la partie entière du résultat est conservée. Le modèle conçu dans le cadre de cette étude comporte alors treize classes, et les outils automatiques offerts par ArcGis pour reclassifier n'apparaissant pas être la meilleure solution, une méthodologie précise est donc appliquée.

La première étape consiste à s'intéresser à la répartition pour les transects. Bien que ceux-ci aient été utilisés lors de l'entraînement, ils ne comportent que dix des treize classes de probabilités. La reclassification se fait en prenant en compte la pertinence en terme de nombre de sites, de superficie, d'emplacement relativement aux autres classes sur la carte. Les trois zones manquantes sont intégrées en considérant les mêmes paramètres, le résultat complet étant communiqué en Annexe 18 : tables du modèle de population (première version), ainsi que sur la carte de l'annexe suivante.

Cinq classes résultent de cette opération, dont les quatre premières couvrent approximativement la même surface au niveau des transects. La dernière, correspondant à la plus importante des probabilités, couvre à elle seule 41% des transects, alors qu'elle ne représente que 20% de la zone totale. Cela

n'influe pas sur la validité du résultat du fait de la méthodologie appliquée (étude de densités) : les classes doivent toutes être représentées pour permettre le calcul, les proportions n'influent en revanche pas.

C. Propagation des sites et amélioration des probabilités

Les densités de population sont tout d'abord calculées au niveau des transects. La formule communément utilisée (Turner, 1990, cité dans Lutz et al. 2000) est la suivante :

- $Population(t) = StructuresOccupées(t) * HabitantsParStructure(t)$

avec :

- $StructuresOccupées(t) = Structures * Proportion(t) * Occupation(t) * Usage(t)$

Proportion(t) étant le nombre de structures datant de la période étudiée, Occupation(t) le pourcentage d'occupation de ces structures durant cette période, Usage(t) le pourcentage de structures occupées réellement utilisées.

Dans le cadre de cette étude, le facteur $Proportion(t) * Occupation(t)$ est estimé par le docteur Anabel Ford à 95% pour la fin de la période classique (800-900 ap. JC). L'illustration 18: répartition des sites par rang indique que le terme $Usage(t)$ vaut 59%, soit la proportion de sites de rang différent de zéro, en gardant à l'esprit que les unités prises en compte sont constituées d'ensembles de structures appartenant à une même famille.

Le nombre d'habitants par unité résidentielle occupée vaut 5,6, valeur communément utilisée (Lutz et al. 2000). Cela donne comme formule pour notre étude :

- $Population(late\ classic) = 95\% * NombreSites(rang > 0) * 5,6$

Les résultats apparaissent dans l'annexe précédemment citée (p.85), et la table 6 donne les valeurs pour l'intégralité de la zone pour les unités résidentielles occupées (URO).

Totalité de la zone d'étude						
Classe	URO late classic	Superficie (km ²)	Population	Densité	% population	% de la zone
1	0	311.8144	0	0	0%	24%
2	4,809	418.9492	26,928	64.28	15%	33%
3	1,813	74.5428	10,154	136.21	6%	6%
4	7,681	226.0332	43,012	190.29	24%	18%
5	17,924	257.2036	100,374	390.25	56%	20%
Total	32,226	1,288.5432	180,468	140.06	100%	100%

Table 6: propagation de la population à la zone étudiée

Avant d'étudier plus précisément ces chiffres, une représentation des sites permet de visualiser la répartition en terme d'occupation du terrain. Le modèle conçu pour la propagation des sites utilise la fonction « *Create Random Points* » d'ArcGis, et une distance minimale de 40 m entre deux structures est imposée, valeur choisie en prenant en compte les diamètres disponibles dans la table attributaire des sites connus afin d'éviter tout chevauchement. L'Annexe 20 : modèle de propagation des sites (première version) fait bien apparaître les formes d'occupation issues des classes de probabilités.

La question de l'uniformité de la répartition des unités résidentielles sur les grandes surfaces se pose alors, le modèle actuel ne prenant en compte que des critères purement environnementaux. Dans le cadre d'un projet archéologique, les paramètres sociaux sont également à considérer, et améliorer le modèle passe par cela (Gaffney et al. 1996). Le facteur d'attractivité des zones à forte population peut se déduire de tampons appliqués autour de celles-ci (voir Illustration 21, du rouge au vert selon l'éloigne-

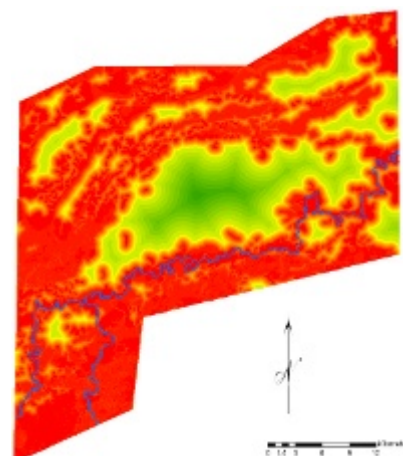


Illustration 21: attractivité sociale

ment aux zones les plus peuplées), le but étant de modéliser le fait qu'une zone de faible probabilité entourée de zones à fortes densités sera plus attractive qu'une autre de la même classe mais située loin des centres de vie sociale.

Les poids correspondant à ce nouveau paramètre sont calculés comme précédemment par la méthode des *Weights of Evidence*, et sont présentés ainsi que la nouvelle carte des probabilités en Annexe 21 : poids et probabilités au niveau local (seconde version). Les deux premières classes correspondent aux fortes densités, la troisième à un tampon de 500 m, et la dernière aux distances supérieures à cette valeur. La carte de probabilités résultante comporte plus de classes (27 au lieu de 20), ce qui permet de différencier les zones avec plus de précision pour une reclassification ultérieure.

D. Analyse finale de la population de la zone étudiée

1. Étude de la densité et de la population

En appliquant la méthodologie précédemment expliquée, une nouvelle population est obtenue pour la zone d'étude. Celle-ci ne varie que très peu, et cette différence n'est pas réellement quantifiable puisque la superficie globale change elle aussi du fait des passages de polygones à images raster.

Totalité de la zone d'étude						
Classe	URO <i>late classic</i>	Superficie (km ²)	Population	Densité	% population	% de la zone
1	0	484.6748	0	0	0%	38%
2	5,403	243.0280	30,255	124.49	17%	19%
3	1,753	76.2924	9,818	128.7	5%	6%
4	7,643	224.9168	42,800	190.29	23%	18%
5	17,808	255.5448	99,727	390.25	55%	20%
Total	32,607	1,284.4568	182,600	142.16	100%	100%

Table 7: population finale de la zone d'étude

Il apparaît que la densité de la seconde classe est celle qui est la plus changée : les unités résidentielles de cette zone de faible attractivité sont en effet concentrées à la périphérie des zones de meilleure qualité.

La densité moyenne de 142 hab./km² de la zone d'étude est qualifiée de dense (Boserup 1983). Elle s'apparente d'après la même source à celles rencontrées en Inde, Pakistan et Bangladesh en 1975. À titre de comparaison avec un pays moderne, notre zone d'étude d'une superficie deux fois moindre que celle du Luxembourg a une densité qui s'en approche (183 hab./km²) (Wikipedia 2009c).

Cette densité peut paraître élevée, mais est similaire à des valeurs de 150 – 200 hab./km² avancées par Culbert par exemple (cité dans Lutz et al. 2000). Des chiffres plus importants encore sont donnés dans certaines régions (280 hab./km² dans la région du Rio Bec), avec notamment l'argument que certaines techniques d'agriculture telles que l'utilisation de terrasses permettaient de subvenir aux besoins d'une telle population (Fischbeck 2004).

Au niveau de la zone étudiée, 78% des Mayas habitent une zone couvrant 38% du territoire, et à l'opposé 38% de celui-ci est inhabité, ce qui laisse une importante zone disponible pour la forêt et d'éventuelles cultures. La population élevée de 182 600 habitants peut être comparée avec la population actuelle du district de Cayo qui inclut la zone d'étude, et était de 53 715 personnes en 2000 (Wikipedia 2009a), soit une densité moyenne de 10,52 hab./km².

Une mise en relation des classes de densité avec des cartes de recommandations actuelles fait également apparaître des différences importantes : les sols privilégiés par les Mayas ne correspondent pas à ceux plébiscités dans le présent. La différence provient notamment en grande partie de la mécanisation et de l'extensibilité de l'agriculture : il est évident qu'un jardin forestier ne demande pas le même genre d'environnement que des champs exploités avec des tracteurs.

Une classification mécanique des sols telle que celle fournie en Illustration 22 (Jenkin et al. 1976) fait ainsi apparaître que de grandes zones non utilisées par les Mayas sont en fait considérées

comme intéressantes dans le cadre d'une agriculture mécanisée. La classe 1 correspond aux zones considérées fertiles avec peu ou pas de restrictions à une agriculture mécanisée, et la classe 4 correspond au contraire à des sols non propices à l'agriculture... alors que ces derniers correspondent aux zones privilégiées par les Mayas ! En prenant en compte le fait que le système maya soutenait une population bien plus importante, la pertinence de la classification et des techniques modernes est réellement discutable.

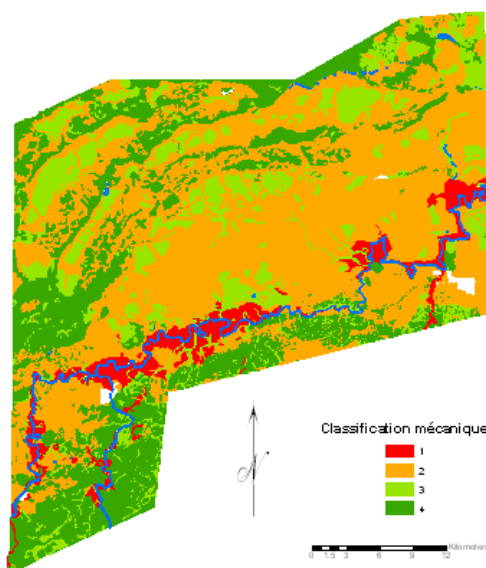


Illustration 22: classification mécanique des sols

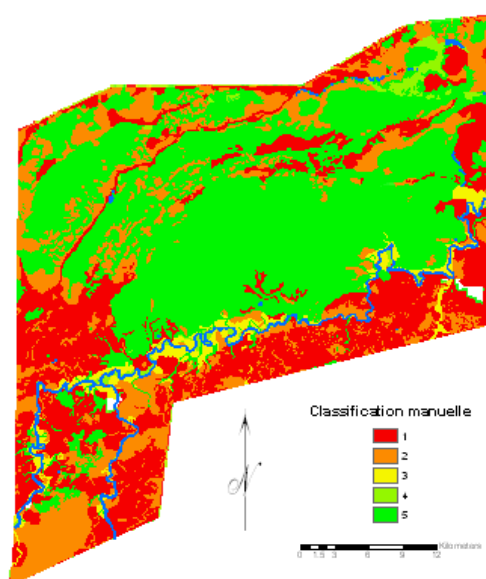


Illustration 23: classification manuelle des sols

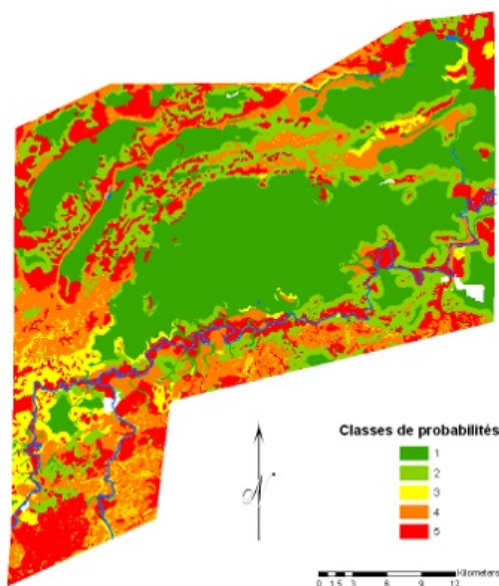


Illustration 24: carte des probabilités

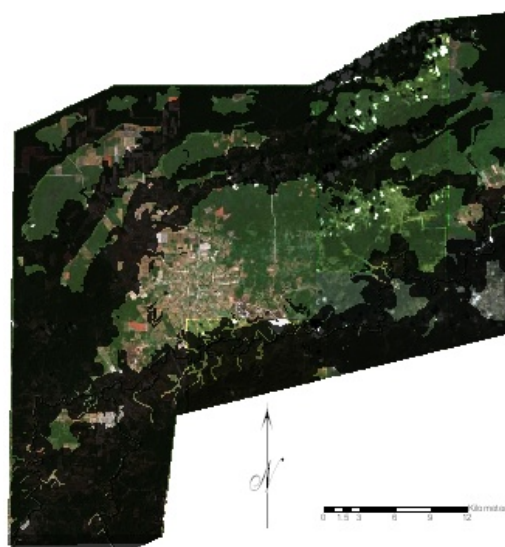


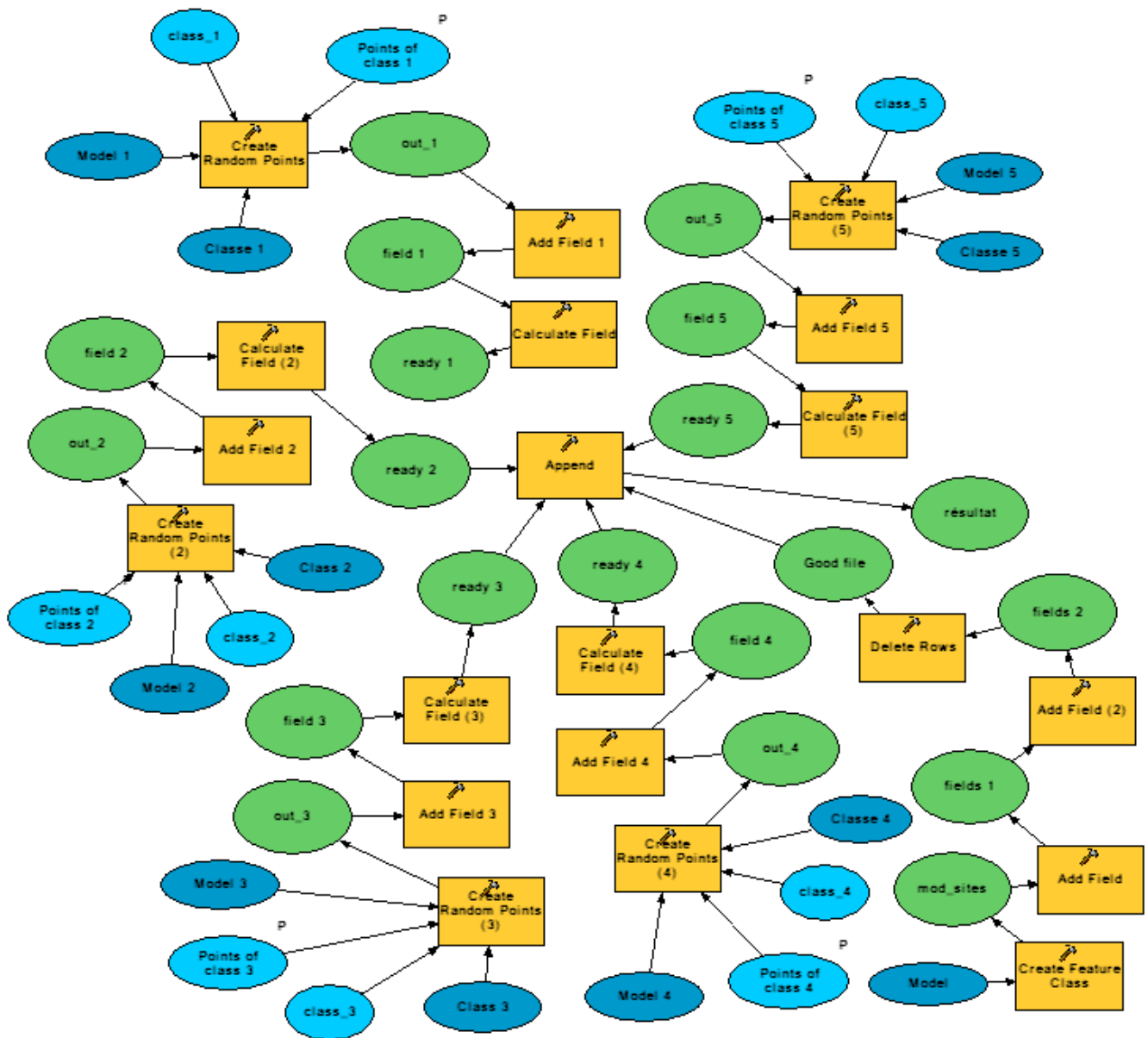
Illustration 25: vue aérienne

En revanche, une classification de la qualité des sols pour une agriculture manuelle effectuée dans le passé (Fedick 1988) apparaît au contraire plus proche du résultat du modèle prédictif (dans toutes ces cartes, les sols sont classés du rouge au vert par qualité décroissante).

L'analyse d'une photographie aérienne issue de *GoogleEarth* (reproduite en grande taille en Annexe 23 : p.90) sur laquelle est appliquée un masque couvrant les zones occupées du modèle fait clairement ressortir l'utilisation actuelle des sols. Il apparaît ainsi qu'une grande partie de la classe 1 de densité nulle pour les Mayas du fait de la pauvre qualité en terme de drainage et de fertilité est actuellement utilisée pour des champs, des pâturages, ou des habitations (zone de *Spanish Lookout*).

2. Analyse de la répartition des sites

Le modèle suivant est utilisé pour une génération aléatoire des sites sur la zone d'étude dans ArcGis avec la symbolisation et les couleurs par défaut :



Une distance minimale de quarante mètres entre deux sites est respectée dans chacun des cas, et les unités sont créées séparément pour chacune des classes. Une colonne « classe » comporte le numéro de celles-ci pour les identifier facilement après fusion des différentes tables obtenues. Un champ « rang » est alors ajouté, et les points sont classifiés grâce à la table de répartition.

Répartition des rangs de sites (zone d'étude)						
Classe	# Rang 0	# Rang 1	# Rang 2	# Rang 3	# Rang 4	# Rang 5
1	229	0	0	0	0	0
2	4,502	1,351	3,602	450	0	0
3	1,632	181	1,572	0	0	0
4	7,916	2,047	4,231	1,365	0	0
5	10,568	1,239	14,612	1,631	196	130
Total	24,847	4,819	24,017	3,446	196	130

Table 8: répartition des sites par rang et par classe

Les sites produits sont plus intéressants s'ils présentent la réelle occupation des sols plutôt qu'une simple représentation ponctuelle. Pour cela, un jeu de données composé de polygones est créé par l'application de tampons autour des sites propagés. Les valeurs des rayons de ceux-ci proviennent du tableau ci-dessous donnant le diamètre moyen des sites calculé en fonction de leur rang. Il faut souligner ici le diamètre important des sites de rang 5, valant près de 60 m. La distance minimale entre deux sites étant de 40 m, des problèmes de chevauchement peuvent apparaître. Un site de classe 4 doit par exemple se situer à au moins 53 mètres d'un site de classe 5 pour ne pas l'intersecter.

Rang des sites	0	1	2	3	4	5
Diamètre moyen (m)	11	20	23	38	47	59

Classe de probabilité	1	2	3	4	5
Distance entre URO	-	68	60	47	40

La solution apportée à ce problème intervient lors de l'attribution des rangs. Deux colonnes supplémentaires sont ajoutées pour cela :

- **distance au plus proche voisin** : valeur de la distance euclidienne au point le plus proche
- **identifiant du plus proche voisin** : cet attribut n'est pas utile pour l'instant mais peut avoir son utilité ultérieurement puisqu'il apparaît dans la table des sites originaux, il est donc intéressant de le calculer par la même occasion

Il est dès lors possible d'attribuer les rangs 4 et 5 aux sites ayant les plus grandes valeurs de distance au plus proche voisin, ce qui permet d'éviter tout chevauchement.

Cependant, un second problème apparaît : la propagation est effectuée séparément pour chacune des cinq classes de probabilités afin de satisfaire aux densités respectives. Cela signifie que la valeur minimale de 40 m entre deux unités est respectée au sein d'une même classe, mais pas forcément entre deux classes différentes.

Un site situé en classe 3 peut par exemple se retrouver à quelques mètres seulement d'un site de classe 2, et quelques soient leurs rangs ils se chevaucheront donc. Rien ne peut être fait dans le cadre de la méthode utilisée pour éviter cela, un traitement a posteriori est donc effectué : une topologie détectant les chevauchements est appliquée au jeu de données, et les douze problèmes identifiés sont corrigés manuellement et les distances recalculées.

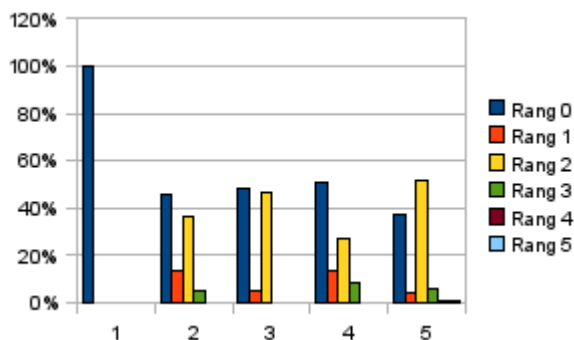


Illustration 26: répartition des rangs selon les classes de probabilités

Le résultat est donc un fichier contenant 57 455 sites complètement renseignés et répondant aux exigences du modèle. La carte de rendu est disponible en Annexe 24 : modèle de propagation des sites (seconde version). La lecture de celle-ci est plus simple que celle des cartes de probabilités ou de densités, et elle permet de visualiser l'occupation approximative du terrain. Il apparaît ainsi que dans les zones à très faible attractivité, seuls quelques sites peuvent être trouvés : ceux-ci sont exclusivement des structures temporaires, dont la grande diagonale mesure en moyenne onze mètres.

Il est intéressant de constater cependant que la répartition actuelle ne correspond pas encore à la réalité. Le transect de Barton Ramie comporte par exemple 259 sites dans la réalité, alors que la propagation en crée 296. Ces unités sont d'après le docteur Ford différentes de celles des autres transects : la distance moyenne à leur plus proche voisin valant 48 mètres est en effet plus importante que celle trouvable ailleurs au niveau de la classe 5 et valant 31 mètres.

Ces familles disposaient donc de plus de terrain, et la proportion de poteries peintes trouvées dans cette zone est également supérieure à la normale. Le fait que cet endroit soit apparemment propice à la culture du cacao qui était une monnaie utilisée par les Mayas pourrait expliquer ces différences, les habitants bénéficiant d'une sorte de traitement de faveur. Aucune des données disponibles ne permet d'intégrer ce paramètre à cette étude pour différencier ces sites.

E. Besoins en termes d'alimentation

La nouvelle population estimée va servir à recalculer les besoins en termes d'alimentation, en affinant la méthode utilisée. Celle utilisée dans le passé était basée sur un modèle de population considérant les besoins en terme de calories et de capacités de production au niveau de Tikal (Rose 2003), mais n'est plus satisfaisante. Le but de cette partie est d'essayer de déterminer si l'environnement des mayas leur permettait de subvenir à leurs besoins, une des raisons communément avancées pour leur déclin étant une surexploitation des ressources naturelles.

En prenant en compte les besoins d'une telle population et les connaissances disponibles sur certaines de leurs techniques d'agriculture, le modèle conçu devient alors un outil très intéressant pour ce type d'analyse. Si un modèle durable peut être établi pour la subsistance des Mayas et leur grand nombre d'habitants, ses enseignements seraient très intéressants pour des perspectives de développement durable au Guatemala et au Belize.

1. L'agriculture Maya

Il est important de rappeler à ce niveau les méthodes d'agriculture des Mayas. Il est suggéré que plutôt que d'utiliser des méthodes de déforestation extensives et permanentes, les Mayas vivaient dans un environnement constitué d'un patchwork de champs et de forêts (Ford 2008). Le système proposé est entièrement dynamique, une parcelle n'étant pas cantonnée à une utilisation particulière définitive, mais faisant au contraire partie d'un cycle dynamique.

Celui-ci, qualifié de cycle du Milpa (cf. Annexe 26 : cycle du milpa), évoque un passage progressif d'un état de défrichement effectué par la méthode du *slash and burn* (technique de brûlis) à un retour à la canopée, s'étalant sur une vingtaine d'années, et est encore utilisé de nos jours dans la région du Petén (Corzo Marquez & Schwartz 2008).

Quatre étapes principales peuvent être identifiées (ESP-Maya 2009).

i. De la forêt au milpa

Une parcelle de forêt est coupée et brûlée, le sol profitant de cet enrichissement temporaire. Pendant deux ou trois années, la parcelle sert à produire maïs, haricots et courgettes. À la différence des champs de maïs que l'on peut voir actuellement en France par exemple, il y a ici une véritable diversité. Le sol du milpa est en effet couvert de plantes, poussant au milieu du maïs. En plus des haricots et courgettes, il est ainsi possible de trouver des épices, des herbes, des tubercules, toute cette végétation participant à la qualité du milpa : certaines plantes servent à enrichir le sol, d'autres à éloigner ou attirer certains insectes pour qu'il ne s'attaquent pas à la nourriture, d'autres ont des vertus médicinales.

ii. Du milpa au jardin forestier

Des arbres fruitiers à croissance rapide (bananes, plantain, papaye) sont alors ajoutés au milpa et sont en mesure de produire des fruits en une année. Des arbres à croissance plus lente sont également plantés, pour apporter par exemple des avocats, mangues, citrons ou encore cherimoya. Alfonso Tzul, un jardinier forestier maya moderne, a également expliqué lors d'une discussion dans le cadre de ce projet comment le jardin se remplit alors d'oiseaux et d'animaux divers attirés par ces sources de nourriture.

Il est très intéressant de noter la philosophie que ces jardiniers appliquent à leur travail. Des oiseaux consommant les fruits ne seront pas chassés, mais des arbres fruitiers différents seront plantés pour qu'ils se détournent des fruits désirés. La même technique est d'ailleurs utilisée pour repousser certains insectes. Il apparaît également hors de questions pour ces jardiniers traditionnels d'utiliser des pesticides ou des désherbants chimiques tels que le Roundup qu'ils qualifient de « poison ».

Tout problème est traité naturellement, et les Mayas ont une relation spirituelle avec leur jardin, les plantes étant parfois considérées comme ayant leur propre personnalité, pouvant être humiliées ou corrigées verbalement si elles ne produisent pas suffisamment (Corzo Marquez & Schwartz 2008).

iii. Du jardin forestier à la forêt et la haute canopée

Les arbres fruitiers arrivent peu à peu à maturité, empêchant la lumière du soleil d'atteindre directement le sol. Le maïs, les haricots ou encore les courgettes ne sont plus viables, et de nouveaux arbres sont plantés pour le retour complet à la forêt, des cèdres ou des acajous par exemple.

iv. De la forêt à la haute canopée

Les nouveaux arbres dépassent alors les arbres fruitiers jusqu'à la reconstitution d'une haute canopée, et la forêt redevient ce qu'elle était avant le début du cycle. Les animaux retrouvent leur environnement d'origine, et le jardinier forestier peut vendre le bois ou l'utiliser pour sa consommation personnelle. Une vingtaine d'années après le brûlis initial, le cycle est complété et prêt à être renouvelé.

2. Conception et calcul du modèle

i. Nombre de calories nécessaires et modèle d'alimentation

Afin d'effectuer son analyse, Kathleen Rose considère le maïs comme l'alimentation de base des anciens mayas, et comme source nécessaire de carbohydrates. L'apport d'un kilogramme de maïs est estimé à 3 551 kilocalories (McHenry 1963, cité dans Rose 2003), et elle accepte la valeur de 34% de l'alimentation des Mayas constituée par celui-ci (Canto 2003; FAO 1998; cités dans Rose 2003). Les autres composantes de l'alimentation proviennent des différents éléments du cycle précédemment décrit : haricots, courgettes, tubercules, fruits divers ou encore viande provenant des animaux venant dans le jardin forestier.

La méthode appliquée consiste donc à s'intéresser à l'occupation des sols par les champs de maïs (milpas) nécessaires à nourrir la population totale. Une fois cette occupation déterminée, il est alors possible de calculer la superficie restante disponible pour la forêt et les autres étapes du cycle.

Kathleen Rose avance une valeur totale de 588 kg de grains de maïs produits par an et par hectare. Celle-ci est calculée en prenant en compte le fait que le milpa produit de moins en moins (de 1 603 kg la première année à 469 kg la troisième), puis elle applique un coefficient de 55% pour passer de l'épi aux grains (Cowgill 1962; Dickson 1980; cités dans Rose 2003).

Afin de déterminer le nombre moyen de calories nécessaires par jour et par personne, Rose considère les valeurs de 3 200 kilocalories pour les hommes et 2 300 kilocalories pour les femmes de 25 ans, ainsi que des coefficients liés à la température (FAO 1957). En faisant alors une moyenne considérant une proportion égale d'hommes et de femmes, elle arrive à une valeur de 2 544 kilocalories nécessaires par jour.

Or, les besoins en calories varient énormément avec l'âge, cette valeur basée sur une population de 25 ans est donc surévaluée. En considérant les chiffres du Ministère de la Santé américain, 3 200 kilocalories est le besoin maximal d'un homme actif entre 14-18 ans ; une femme de plus de 51 ans ne pratiquant pas d'activité physique n'a par exemple besoin que de 1 600 kilocalories (HHS & USDA 2005). Plutôt que de retenir la valeur clairement surestimée utilisée dans le passé, l'étude présente considère donc le besoin en calories moyen donné par le *World Food Programme of the United Nations* pour permettre une vie normale et saine, qui est de 2 100 kilocalories (WFP 2009).

ii. Superficie nécessaire pour la culture du maïs

L'Annexe 27 : *calculs de la superficie nécessaire pour la culture du maïs* présente les résultats des besoins en terme de milpa. Il apparaît ainsi d'après les calculs de calories que 227 km² de champs de maïs sont nécessaires au niveau de la zone d'étude pour nourrir la population. En considérant que les sols de classe 1 sont de très mauvaise qualité et donc non utilisés pour les milpas, celle-ci comporte donc environ 800 km² de terres cultivables sur un total de 1 284 km².

Cependant, ces terres sont déjà partiellement occupées par les structures des mayas. Il convient de retirer la superficie des unités résidentielles (aisément calculable puisque les rayons moyens ont été établis en fonction des rangs), ainsi que celle des infrastructures communes : places, bâtiments publics, voies de communication. Celles-ci se trouvant dans les zones de plus fortes densités, il a été décidé de retirer 10% des classes 4 et 5 à cet effet. Cela conduit à une valeur d'environ 66 km² de terres cultivables occupées.

Au final, il y a donc une superficie d'environ 730 km² de terres cultivables libres, dont 31% seront nécessaires pour la culture du maïs. Il convient dès lors de répartir ces milpas au niveau des différentes classes de densités.

L'illustration 27 présente un schéma d'occupation du sol d'une propriété moderne avec jardin forestier privé à Mexico (Lopez Morales 1993). En bas à droite apparaît l'habitation et le terrain l'entourant, en haut se situe le verger, jardin forestier fournissant fruits et ombrage à proximité directe, et en bas à gauche est représenté le milpa privatif. Une estimation graphique des superficies effectuée sur ce schéma

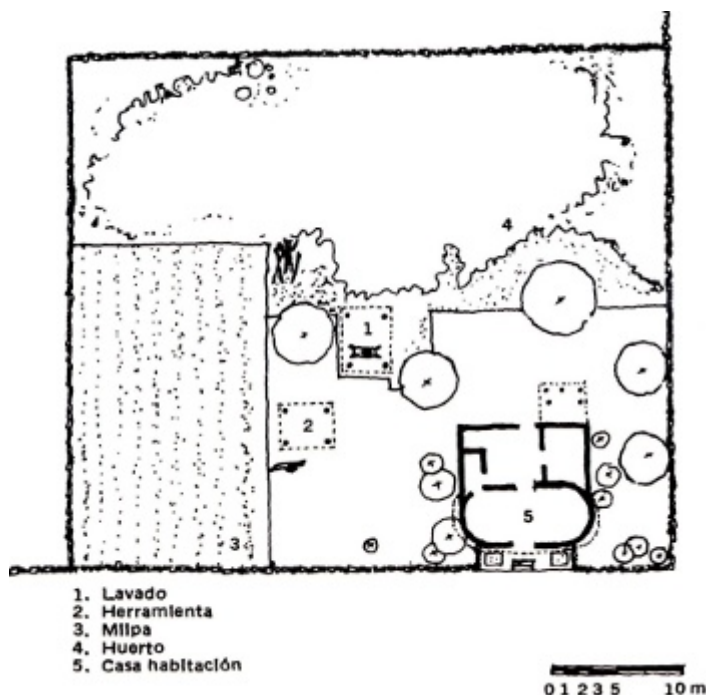


Illustration 27: schéma d'occupation du sol avec milpa et jardin forestier

montre que celui-là est environ 3 fois plus petit que le verger, chiffre également obtenu par Rose. À défaut d'autres valeurs disponibles, ces 30% pour le milpa et 70% pour le jardin forestier sont donc adoptés pour cette étude, Anabel Ford les considérant comme acceptables. Ces valeurs diffèrent des 50% choisis dans le passé pour cette répartition, et sont plus plausibles si l'on considère les raisons du cycle du milpa, ne permettant pas à celui-ci d'être trop important par rapport au jardin forestier.

iii. Application du modèle du milpa

La propagation est effectuée en prenant en compte les proportions déterminées ci-dessus. La première étape consiste à calculer la superficie de sol cultivable disponible pour chaque unité résidentielle occupée. La valeur de 22 506 m² obtenue doit ensuite être divisée en :

- **jardin privatif** : 30% de milpa et 70% de verger
- **terrains du cycle du milpa** : ceux correspondant aux quatre phases mentionnées

La superficie d'un jardin privatif est estimée entre 0,25 ha et 0,5 ha (Fedick 1992, cité dans Rose 2003), une valeur moyenne de 0,4 ha est donc retenue pour cette étude. Cela donne une première superficie de milpa que l'on peut qualifier d'interne, valant 1 200 m² par habitation.

Il faut alors déterminer la superficie nécessaire pour les terres du cycle du milpa externe. Celle-ci dépend de la longueur du cycle considéré. Certaines sources parlent d'approximativement vingt ans (ESP-Maya 2009), tandis que d'autres décrivent le cycle plus précisément. Ronald Nigh a ainsi étudié les jardiniers forestiers mayas lakandones des basses-terres de Chiapas, au Mexique (zone située à l'ouest du Guatemala). Après plusieurs années de culture du maïs, les lakandones identifient quatre étapes pour le retour à une forêt mature (Nigh 2008) :

- **Robir** : colonisation initiale, un à quatre ans
- **Jurupche** : forêt secondaire, quatre à dix ans
- **Nukuxche** : seconde étape de forêt secondaire, dix à vingt ans
- **Tamanche** : forêt mature, vingt ans et plus

Nigh souligne que la reconquête de la forêt secondaire sous la supervision des lakandones après culture du maïs peut ainsi se faire en douze à quinze années contre 25 à 200 ans nécessaires sans cet accompagnement (Chazdon 2008, cité dans Nigh 2008), notamment par le fait que des oiseaux et chauve-souris sont attirés par les arbres à croissance rapide propagés par les mayas, et amènent avec eux les graines des arbres composant la future canopée. La parcelle est alors prête à être reconvertie en milpa, même si dans la pratique les lakandones préfèrent laisser la forêt grandir encore pendant plusieurs années (Nations & Nigh 1980, cité dans Nigh 2008). Même si ces valeurs proviennent d'une région différente, elles sont prises en compte dans cette étude faute d'avoir d'autres sources. De plus, il est possible de supposer qu'étant donné que ce territoire est situé à la même latitude les conditions sont assez similaires, les différences ne dépassent sans doute donc pas un ou deux ans.

À partir de cette analyse, et pour avoir un cycle du milpa complet, la valeur moyenne de quatorze années va être retenue après culture du maïs pour revenir à la forêt secondaire initiale. Il serait possible de prendre une valeur un peu inférieure, mais il faut également envisager des cas de retour à une forêt mature, et le fait que les mayas préfèrent apparemment attendre plus que le temps minimal nécessaire (Nigh 2008).

Kathleen Rose décrit l'effet d'appauvrissement des sols du milpa au fil du temps (Rose 2003), ce qui pousserait à considérer trois ans de culture du maïs sur brûlis. La qualité des sols dépend également des compétences du jardinier forestier, la fertilité de la parcelle pouvant même augmenter après chaque nouveau cycle (Nigh 2008). Une valeur maximale de quatre ans de culture peut être envisagée, ceci se retrouvant également avec les jardiniers forestiers mayas modernes.

Cela donne donc quatre années de milpa pour un cycle total de 18 ans : 4,5 milpas sont donc plantées durant un cycle, ce qui nécessite une superficie de terrain non planté en milpa (en quête de reconquête par la forêt) 3,5 fois supérieure à la taille des milpas (voir le tableau ci-dessous).

Scénario 1	
Terrain disponible par habitation	
Superficie de terres cultivables libres (km ²)	733.84
Nombre d'unités résidentielles occupées	32607
Superficie de terrain disponible par URO (m ²)	22505.51
Jardins privés	
Jardin privé (m ²)	4000
Pourcentage de milpa	30%
Superficie verger (m ²)	2800
Superficie milpa interne (m ²)	1200
Cycle du milpa considéré	
Durée d'utilisation du milpa externe (années)	4
Durée du cycle du milpa (années)	18
Ratio entre verger/forêt et milpa	3.5
Superficies nécessaires pour le cycle du Milpa	
Superficie totale nécessaire pour les milpas (m ²)	6975.21
Superficie de milpa externe (m ²)	5775.21
Superficie pour le reste du cycle (m ²)	20213.24
Terrain disponible par habitation	
Superficie totale nécessaire par URO (m ²)	29988.45
Superficie restant libre par URO (m ²)	-7482.93
Sols restant à la nature	
Superficie totale de sols restant libre (km ²)	-244.00

La superficie totale de milpa nécessaire par habitation étant de 6 975 m², il apparaît en retranchant la

milpa interne qu'il faut 5 775 m² de milpa externe. En considérant que le cycle nécessite 20 213 m² supplémentaires pour pouvoir être complet, la superficie totale de terres nécessaire par unité résidentielle occupée est donc de 29 988 m², plus que les 22 505 m² réellement disponibles. En prenant en compte la totalité des sites, il y a donc un déficit de 244 km² pour arriver à subvenir aux besoins de la population de cette façon : cette modélisation ne convient pas.

3. Analyse de valeurs possibles

Il est possible de faire varier les paramètres pour essayer de trouver un point d'équilibre permettant de nourrir la population. Les tables de scénario sont communiquées en Annexe 28 : scénarios répondant aux besoins de la population. Les paramètres pouvant être modifiés sont :

- la longueur du cycle du milpa
- le nombre d'années d'utilisation du milpa
- la productivité d'un champ de maïs

i. Modification de la longueur du cycle du milpa

Un premier équilibre peut être atteint en réduisant la longueur du cycle du milpa de 18 à 12 ans et correspond au scénario 2. Dans ce cas, il y a un excès de terre valant 38 km² laissé à la nature. Cependant, cela signifie également que seulement huit années sont laissées pour un retour à la forêt, seule la première phase de forêt secondaire (appelée *jurupche* par les lakandones) pouvant être atteinte. Le résultat peut donc être une perte de diversité, Nigh recensant 11 variétés propres à la seconde phase de forêt secondaire (Nigh 2008). Le cycle du milpa n'étant pas complètement respecté, cela pourrait également conduire à un appauvrissement progressif des sols et une difficulté croissante à nourrir la population, avec les instabilités sociales que cela implique.

ii. Modification du nombre d'années d'utilisation du milpa

Ce paramètre ne peut pas être augmenté, quatre années d'utilisation du milpa étant considérées être un maximum pour conserver un sol suffisamment fertile. En revanche, il est possible d'envisager de le baisser à trois ans tout en diminuant la longueur globale du cycle. Ce n'est en fait pas satisfaisant, ramener l'utilisation du milpa à une valeur si basse impose un cycle de neuf ans au maximum pour pouvoir nourrir la population, soit six ans seulement pour la forêt (scénario 3).

iii. Changement de la productivité des champs de maïs

Anabel Ford avance l'idée que la productivité des champs de maïs pourrait être plus importante que la valeur utilisée ici, l'argument tenant dans la proportion importante de cendres volcaniques fertilisantes détectée pour cette époque (entretien avec Anabel Ford, 2009). Par manque d'informations plus précises de cet impact sur la productivité des champs, cela ne peut être testé ici.

En revanche, il est possible de connaître des chiffres récents de productivité au Guatemala. En 1961, ce pays avait ainsi des récoltes de maïs s'élevant à 828,56 kg/ha (FAO 2009), à comparer avec les 588 kg/ha qui sont considérés dans cette étude. Avec une telle valeur (scénario 4), le modèle est valide en considérant un cycle de 18 ans avec quatre années d'utilisation du milpa.

Cela signifie notamment qu'en utilisant le cycle du milpa, le Belize pourrait sans doute actuellement nourrir une population plus importante tout en sauvegardant la forêt. Le nombre d'habitants de ce pays en 2005 était en effet de 277 409 habitants (SIB 2007), soit moins du double de la population de la zone étudiée pour une superficie 18 fois plus importante.

En 2007, la productivité est passée à 1 666 kg/ha (FAO 2009), ce qui permettrait théoriquement un cycle du milpa de plus de 30 ans où le verger et la forêt représentent 6,5 fois la superficie cultivée au niveau de la zone étudiée. Les techniques pour parvenir à une telle productivité (agriculture extensive mécanisée) ne semblent cependant pas compatibles avec un cycle du milpa vertueux.

F. Conclusion

La validation et l'amélioration du modèle prédictif pré-existant (Monthus 2004 publié dans Clarke et al. 2009) imposaient pour cette étude de reconsidérer les résultats du modèle de population dont la validité n'était plus assurée. La nouvelle population obtenue au niveau de la zone (182 600 habitants) ne diffère que de 8 000 personnes avec celle calculée dans le passé, mais est maintenant expliquée par les données issues du modèle prédictif et non plus par des changements a posteriori sur critères subjectifs.

La méthode mise en place pour les questions d'alimentation n'était pas non plus satisfaisante, les paramètres étant simplement repris du travail de Kathleen Rose sans analyse critique, et le modèle étant bâti en terme de disponibilité plutôt qu'en terme de besoins.

Le modèle de propagation des sites ainsi que celui d'occupation des sols et de subsistance de la population conçus pour cette étude sont enregistrés dans un unique fichier tableur au format *OpenDocument Spreadsheet*, format ouvert et norme ISO internationale pour assurer l'inter-opérabilité (Wikipedia 2009d). Il est possible d'influer sur tous les paramètres considérés lors de la création du modèle, par exemple changer le pourcentage d'unités résidentielles occupées pour modéliser une autre période que le classique tardif, et d'obtenir instantanément tous les autres résultats tels que les besoins en maïs correspondant.

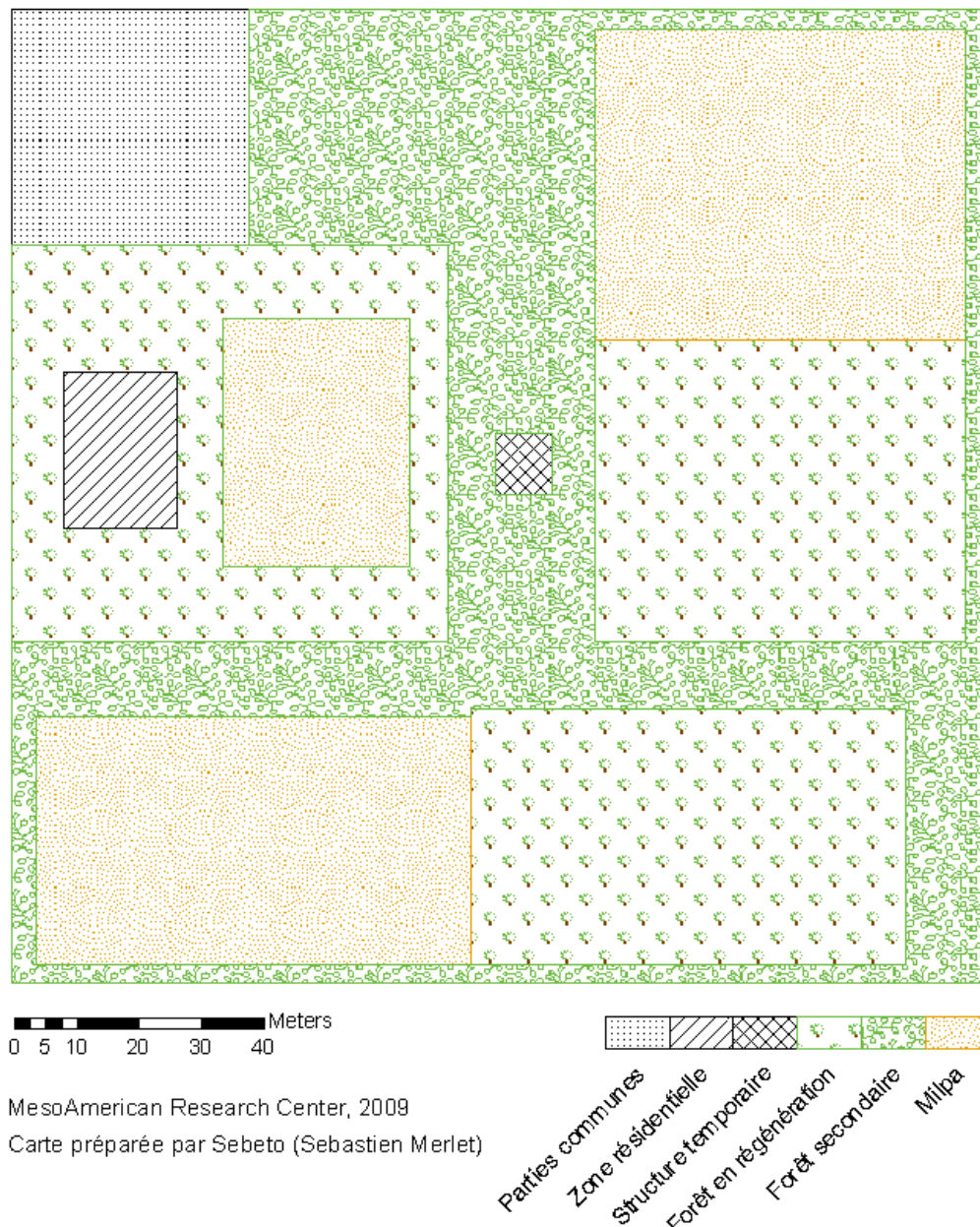


Illustration 28: occupation des sols pour le scénario n°2

La figure ci-dessus (Illustration 28) présente le résultat d'occupation des sols pour le scénario le plus satisfaisant (scénario 2). Ce schéma décrit la superficie de terres disponibles pour chaque unité résidentielle occupée des classes de sol cultivables du modèle. La zone résidentielle (qui peut être composée d'une grande structure ou de plusieurs plus modestes) est entourée d'une parcelle interne comportant un milpa et un jardin forestier doté d'une grande diversité de plantes et arbres ayant des usages variés, tels que nourriture, médecine, bois de construction ou combustible (Corzo Marquez & Schwartz 2008; Ford 2008; Gasco 2008; Nigh 2008).

La partie commune associée à cette habitation représente les structures publiques, cela incluant par exemple les temples, les places, ou encore les voies de communication principales. Ces parties communes sont évidemment regroupées pour former des parcelles importantes recevant ces bâtiments primordiaux pour la vie sociale.

Pour subvenir aux besoins de la famille comprenant environ 5,6 membres, des milpas externes sont également utilisés. Ceux-ci sont créés par une technique de brûlis qui enrichit les sols (Nigh 2008), et sont utilisés pendant quatre ans, un nouveau brûlis étant effectué ailleurs à la fin de cette période. L'ancien milpa n'est cependant pas abandonné à lui-même, mais au contraire accompagné de telle façon que la forêt secondaire peut réapparaître en 12 à 15 ans au lieu de 25 ans, ce qui permet également d'améliorer la qualité des sols d'un cycle au suivant (Nigh 2008). Du fait de la population importante de la zone étudiée, la durée du cycle est cependant réduite ici à douze ans, ce qui ne laisse que huit années à la forêt pour se régénérer, ne pouvant atteindre que la première phase de forêt secondaire décrite par Nigh. La forêt mature nécessitant 20 ans pour se reconstituer (Nigh 2008), ou la canopée fermée apparaissant 15 à 30 ans après le début du cycle (Ford & Nigh 2009) sont donc probablement quasiment absentes puisque cela imposerait de réduire d'autant le cycle du milpa à d'autres endroits.

Cette situation soulève la question de la viabilité de ce système sur la durée. Un cycle du milpa ainsi réduit permet-il de continuer à enrichir les terres ou tout au moins maintenir leur niveau de fertilité ? Le fait que la forêt mature se raréfie ou même disparaisse peut-elle avoir d'autres répercussions ? Cette étude est en effet basée sur la partie de l'alimentation constituée par le maïs, mais quels sont les effets sur la faune de cette situation, et donc sur les autres sources de nourriture des Mayas ?

Le système tel que modélisé apparaît être à la limite du déséquilibre, et nécessiterait pour arriver à des conclusions plus précises des études plus poussées sur la productivité des champs de maïs cultivés à l'époque des mayas, sur la viabilité d'un cycle du milpa restreint, ainsi que sur les effets de la diminution de la forêt mature au niveau des animaux. En étant à la limite de l'instabilité, il est tout à fait possible d'imaginer que toute survenue d'une période de stress climatique pourrait avoir de graves conséquences, tant au niveau de la population que de la fertilité des sols utilisés plus intensivement.

Les tensions sociales en résultant et l'affaiblissement des instances dirigeantes pourraient alors entraîner une baisse de la participation aux taxes, constituées notamment de culture des terres et constructions non rémunérées pour le compte des prêtres et des nobles, ou encore et surtout de constructions publiques et entretien des bâtiments (Von Hagen 1960). Sans soutien du peuple préoccupé à lutter pour sa subsistance, les classes dirigeantes seraient mises à mal, et sans entretien les bâtiments se dégraderaient rapidement. Ceci aurait pu participer au déclin de cette civilisation : des monuments archéologiques mayas révélés par les archéologues souffrent ainsi particulièrement de leur exposition sans entretien permanent et certaines de leurs merveilles sont en train de disparaître à jamais (Dunster 2004), montrant la complexité et l'importance du travail qui devait être effectué à ce niveau par les Mayas.



Illustration 29: aspect d'un milpa planté de maïs, haricots et plantes diverses par Narciso Torres, un jardinier forestier maya moderne

VII. Conception d'un modèle régional

A. Introduction

La conception d'une nouvelle version du modèle local amélioré a été l'occasion de rappeler le fonctionnement de la méthode des *Weights of Evidence*, et d'identifier les différents problèmes pouvant survenir et la façon d'y remédier.

Cette partie de l'étude vise à créer le même type de modèle prédictif à une échelle qualifiée de « régionale ». Alors que la zone locale représentait 1 284 km², celle régionale est largement plus importante : 44 902 km². Les méthodes mises en place doivent donc prendre en compte cet important changement d'échelle : choix des sites d'entraînement, de la taille de cellule unitaire, analyse de la qualité des paramètres.

B. Présentation de la zone et des paramètres

La zone d'étude est située à cheval sur la frontière entre le Belize et le Guatemala. Elle englobe notamment la zone locale, ainsi que la région du Petén. Le fait que deux pays soient représentés peut poser des problèmes aux frontières lorsque les paramètres proviennent de sources nationales : dans ce cas, il est nécessaire de trouver des jeux de données coïncidant et présentant le même niveau de détails.

1. Points d'entraînement

Plusieurs jeux de données de la géodatabase contiennent des sites archéologiques à l'échelle régionale, identifiés grâce à leurs métadonnées :

- *a_conap* : sites archéologiques référencés par l'*Instituto de Antropologia e Historia*
- *a_El_Pilar* : sites d'une partie du parc archéologique d'El Pilar
- *a_regional* : sites provenant de diverses sources, couvrant totalement la zone d'étude
- *a_sites* : sites des transects utilisés pour l'entraînement du modèle local
- *a_Tikal* : sites issus de levés effectués autour de Tikal
- *mod_population* : sites produits grâce au modèle local

Les sites du CONAP, de Tikal et d'El Pilar ne peuvent vraisemblablement pas être utilisés pour l'entraînement à l'échelle régionale : les zones de lever ne sont pas identifiées, et ces sites ne couvrant pas toute la zone d'étude, les poids ne seraient pas fiables. Pour les sites de la « croix » de Tikal, il apparaît même que leur zone de lever n'intersecte pas toutes les classes de drainage par exemple, empêchant le calcul.

Les sites des transects ayant servi pour le modèle local pourraient être considérés, mais la zone qu'ils couvrent est également trop limitée pour donner des résultats satisfaisants, de même que ceux du jeu *a_El_Pilar*.

a_regional et *mod_population* semblent donc être les candidats les plus intéressants pour cette partie de l'étude. Ces deux jeux sont cependant fondamentalement différents.

a_regional recense des sites sur une zone couvrant le Guatemala, le Belize et le Mexique. Ceux-ci ne sont pas de la même nature que ceux utilisés pré-

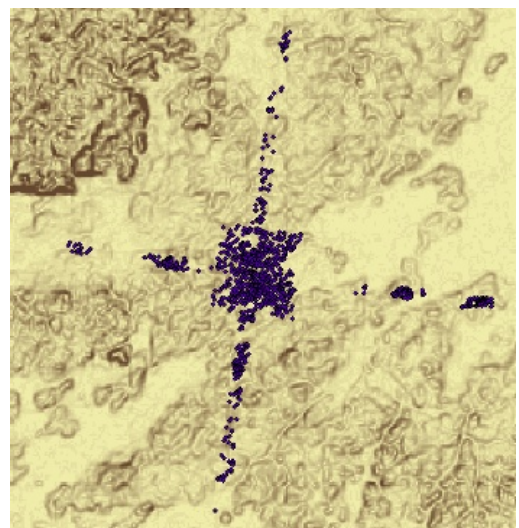


Illustration 30: sites du jeu *a_Tikal* sur fond de pente topographique

cédemment : seuls les grands centres sont identifiés. Tikal est ainsi représenté par un unique point, ainsi que par exemple El Mirador ou Corozal. 212 sites sont disponibles au niveau de la zone d'étude, et en considérant que ceux-ci ont été uniformément recherchés sur son étendue, l'entraînement au moyen de ce jeu en utilisant le masque régional est parfaitement valide. La taille maximale approximative de ces sites est estimée à quelques centaines de mètres, et l'utilisation de l'outil « *Calculate distance band from neighbor count* » d'ArcGis donne une distance minimale de 320 m entre deux sites, 300 m sera donc une bonne taille de cellule unitaire.

mod_population quant à lui comporte les sites créés à partir du modèle conçu précédemment. Celui-ci étant considéré comme satisfaisant, il est possible de les utiliser pour entraîner un modèle à une échelle différente. L'avantage de cette méthode est de dépasser le problème des transects qui ne couvrent pas une surface et une diversité de classes d'évidence de paramètres suffisantes. Il convient cependant de se demander si la taille des sites considérés est adaptée à l'échelle du problème : ainsi qu'il est ultérieurement expliqué, la taille de pixel des paramètres utilisés est de 100 m, et certains d'entre eux ont certainement une précision bien moindre, il faut donc être conscient de l'importante généralisation du résultat.

2. Paramètres utilisables

i. SRTM

Un modèle numérique d'élévation est disponible pour la totalité de la zone d'étude, aucune modification n'étant nécessaire. Le modèle du SRTM (*Shuttle Radar Topography Mission*) initial créé par la NASA comporte un pixel de 30 m environ à l'équateur, mais l'unique version disponible au public en dehors des États-Unis et leurs territoires a un pixel de 90 mètres (trois secondes d'arc). Conçu à partir d'une méthode interférométrique (*Interferometric Synthetic Aperture Radar*), la précision avancée est d'environ 20 m en planimétrie et 16 mètres en altimétrie (Farr et al. 2007). La version utilisée ici a été nouvellement incorporée dans le SIG dans le cadre de cette étude : elle est basée sur le SRTM V2, corrigé par la NASA afin de supprimer certains problèmes existants (artefacts dus à la technique utilisée), et amélioré par des méthodes d'interpolation afin de combler ses lacunes (CGIAR 2008).

Cette source de données apparaît être la plus pertinente pour le modèle numérique d'élévation, diverses études ayant prouvé que la précision du SRTM est globalement meilleure que des données telles que GTOPO30 par exemple, même s'il ne contient pas nécessairement plus de détails (Jarvis et al. 2004).

Il est ainsi possible de tester grâce à ce paramètre si les Mayas évitaient par exemple de s'installer au delà d'une certaine altitude, paramètre pouvant avoir une influence sur la température ou la végétation présente.

ii. Pente topographique

Ce paramètre est créé à partir du SRTM au moyen de l'outil « *surface analysis* » de *Spatial Analyst* dans ArcGis. Le résultat est une image raster exprimant la pente en pourcentage, valeurs dont seules les parties entières sont conservées pour les besoins d'ArcSDM. Ce paramètre était apparu pertinent à l'échelle locale, il est donc intéressant de pouvoir le comparer à l'échelle régionale.

Il est important de noter que la version de la pente topographique utilisée à l'échelle locale était plus précise : alors que le SRTM de pixel 30 m n'est pas disponible au public, certains de ses dérivés peuvent être téléchargés, dont la pente topographique couvrant le Belize. Cependant, cette donnée n'a pas pu être trouvée pour le Guatemala et le Mexique et ne peut donc pas servir ici.

Il faut également être conscient que cette pente peut comporter d'importantes erreurs : en tant que dérivée de premier ordre, des problèmes sont susceptibles de se produire au niveau des zones du SRTM comportant des données interpolées pour combler les zones de *NoData* des données originales publiées par la NASA (Jarvis et al. 2004).

iii. Élévation au niveau d'une base régionale (ENBR)

Ce paramètre résulte également d'un calcul effectué à partir du SRTM. À la différence d'une élévation absolue, l'Élévation au Niveau d'une Base Régionale prend en compte la position d'un point selon son environnement proche. La théorie consiste à définir un rayon d'analyse autour du site étudié (voir illustration 31), et à comparer l'élévation de celui-ci avec la moyenne de celle de la zone. Dans le cadre de cette étude, la formule choisie est la suivante :

$$\bullet \quad ENBR = ElevationPoint - ElevationMoyenne(rayon)$$

Une valeur positive signale que le point se situe au dessus de la moyenne, une valeur négative indique le contraire. Il est ainsi possible de tester si l'emplacement des sites était par exemple fait de manière à avoir une position dominante (sommet d'une colline offrant un large champ de vision). Testée à l'échelle locale, l'ENBR ne donnait aucun résultat concluant, mais pourrait être utile à une échelle différente, pour des types de sites considérés différents (Sifuentes 2005).

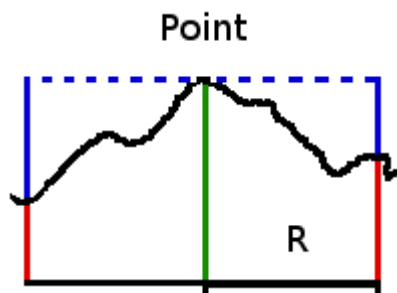


Illustration 31: ENBR

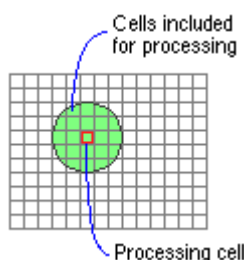


Illustration 32:
disque de calcul

Ce paramètre est calculé au moyen de l'outil « *Single Output Map Algebra* » de l'extension *Spatial Analyst* de ArcGis, aucune autre méthode directe n'ayant été trouvée.

La fonction mathématique utilisée se nomme *FocalMean*, et permet de calculer la moyenne des valeurs des cellules entourant un point choisi (ESRI 2008). Afin d'obtenir l'ENBR, la zone de moyenne est définie comme un disque (Illustration 32), le rayon étant exprimé en nombre de cellules : plusieurs valeurs de celui-ci ont été testées.

L'expression directement utilisée dans le logiciel est donc la suivante pour un rayon de 500 m (cinq cellules) :

$$\bullet \quad D:/GIS/data/srtm - FocalMean(D:/GIS/data/srtm, CIRCLE, 5, DATA)$$

Il apparaît que la palette de valeurs est relativement riche : celles-ci vont de -211 à 219 mètres pour l'exemple précédent, ce qui implique la nécessité de les reclassifier afin d'avoir une représentation significative utilisable lors de l'entraînement du modèle.

Une analyse des fréquences de distribution (Illustration 33) fait en effet apparaître le faible nombre de cellules pour de nombreuses valeurs, ce qui est d'autant plus gênant qu'un choix *ascending* pour l'entraînement n'a pas de raison d'être.

Dix classes sont donc créées : moins de -20 m, plus de 20 m, et paliers de 5 mètres entre ces deux valeurs. Si ce paramètre entre réellement en jeu dans le choix des emplacements, ce petit nombre devrait permettre d'obtenir des valeurs satisfaisantes en utilisant des entraînements prenant en compte un type *categorical*.

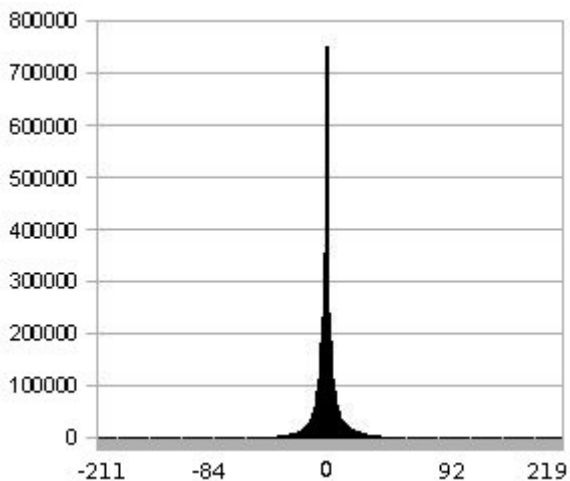


Illustration 33: fréquences de distribution

iv. Fertilité et drainage

Les classes d'évidence de ces paramètres sont issues de cartes numérisées dans le passé (Sifuentes 2005). Les valeurs s'échelonnant de 1 (très bon) à 4 (très mauvais) correspondent à celles présentées dans le travail au niveau local. Cependant, alors qu'à l'échelle précédente le drainage et la fertilité

étaient issus du même jeu de données, ce n'est plus le cas ici : les cartes originales ont en effet été créées par divers organismes pour le Belize, le Guatemala et le Mexique.

Le résultat de cette différence est que les niveaux de précision dans la délimitation des classes de sols ne sont pas les mêmes selon les pays. Le travail effectué par Sifuentes, puis plus tard par Connie a donc consisté à harmoniser les valeurs aux frontières et tenter d'obtenir un niveau de précision uniforme sur la zone. Ces jeux de données ont été considérablement améliorés dans le cadre de cette étude puisque de nombreux problèmes persistaient, notamment du fait que les possibilités d'accrochage entre polygones disponibles au sein d'ArcEditor n'avaient pas été utilisées. Les chevauchements ou trous apparaissant étaient donc problématiques pour la méthode des *Weights of Evidence*, et ont été corrigés grâce à la mise en place de topologies comme présentées précédemment en Illustration 5, p.13.

Des généralisations abusives avaient également été effectuées, ce qui a nécessité de revenir à de précédentes versions des cartes des sols. Les nouvelles décisions ont été prises en collaboration avec tous les utilisateurs du SIG afin d'arriver à un résultat plus satisfaisant.

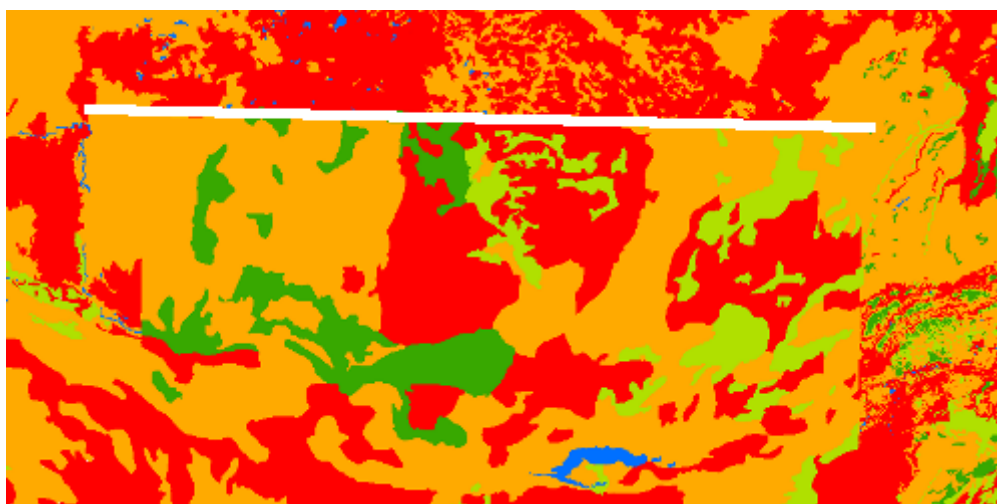


Illustration 34: changement de classes de fertilité aux frontières

L'image précédente (Illustration 34) présente un extrait de la fertilité telle qu'elle est au niveau de la zone, et fait apparaître en sa partie supérieure une ligne virtuelle de différence des classes (représentée sur cette image pour souligner la différence), correspondant à la frontière Guatemala - Mexique. Il est possible de voir les versions finales en Annexe 25 : cartes régionales du drainage et de la fertilité : alors que la frontière entre le Guatemala et le Belize est peu visible au niveau de la première, elle apparaît plus clairement dans la seconde. La zone d'étude n'intersecte pas le Mexique pour éviter la coupure encore plus nette visible en illustration 34, montrant un problème d'adéquation des données.

Au vu des incertitudes concernant ces paramètres, il est important de souligner que le résultat sera donc probablement très approximatif, même en considération de l'échelle (1:1 150 000).

C. Application avec les sites régionaux

La méthode localement utilisée est maintenant appliquée aux 212 points disponibles à l'échelle régionale pour les cinq paramètres présentés, avec une cellule unitaire de 300 m de côté, soit 0.09 km². Au vu du faible nombre de points en comparaison de la taille de la zone d'étude, il est nécessaire d'abaisser le niveau de confiance exigé pour accepter les poids, qui est ramené à un pour cette partie.

1. Analyse des poids

i. ENBR (500 m)

Un calcul des poids en considérant les dix classes d'évidence précédemment définies ne donne aucun

résultat satisfaisant, même avec une confiance acceptable de 1 (~85%) qui pourrait être plus adaptée à cette échelle. Il apparaît que les classes 1 à 4 (moins de -5 mètres) et 7 à 10 (plus de 5 m) couvrent des superficies beaucoup plus petites que les classes 5 et 6, elles sont donc groupées pour un nouvel essai. Les résultats n'étant pas plus concluants, un dernier essai est effectué avec seulement deux classes : élévation supérieure à la moyenne (codée « 1 »), et élévation inférieure à la moyenne (codée « -1 »).

Cette dernière reclassification donne un résultat satisfaisant avec un type *categorical*. Il apparaît qu'un poids négatif baissant les probabilités est donné à la classe d'évidence représentant une élévation inférieure à la moyenne de la base considérée, alors que l'autre hérite d'un poids positif.

ENBR <i>categorical</i>					
Classe	W+	W-	Contraste	Confiance	Poids
-1	-0.11	0.11	-0.23	-1.63	-0.11
1	0.11	-0.11	0.23	1.63	0.11

Un essai avec une ENBR calculée au moyen d'un rayon de base plus important (1 000 m) conduit à la même classification. Les poids sont cependant légèrement différents et la confiance est moins importante avec ce paramètre.

ii. SRTM

Le modèle d'élévation à disposition comporte de très nombreuses classes, avec des valeurs allant de - 11 mètres à 1 118 mètres. Cela exclut d'utiliser un type *categorical* sans reclassification préalable. Un entraînement *ascending* définit une valeur de coupure à 612 mètres, et un calcul en *categorical* après reclassification donne une valeur de confiance de 2,35.

SRTM					
Classe	W+	W-	Contraste	Confiance	Poids
0	0.04	-2.31	2.35	2.35	0.04
612	-2.31	0.04	-2.35	-2.35	-2.31

Les autres essais de reclassification ne sont pas concluants.

iii. Pente topographique

La pente topographique est testée en utilisant un type *ascending* du fait du nombre très important de classes (182) et de leur organisation. Une limite de classe apparaît à 28%, valeur intéressante puisque le modèle local faisait apparaître une coupure à 33% (Table 5, p.29).

Pente topographique					
Classe	W+	W-	Contraste	Confiance	Poids
0	0.02	-1.78	1.81	1.8	0.02
28	-1.78	0.02	-1.81	-1.8	-1.78

Après reclassification et calcul en mode *categorical*, la confiance finale vaut 1,8. Celle-ci est tout à fait acceptable au vu du faible nombre de points d'entraînement par rapport à l'étendue de la zone, d'autant plus que les poids correspondant sont quasiment les mêmes que ceux du modèle local, ce qui renforce la confiance en la qualité du résultat.

iv. Fertilité et drainage

Un entraînement utilisant un type *categorical* ne donne pas de résultat convenable pour les quatre classes d'évidence de ces paramètres : les poids ne sont pas ordonnés comme ils le devraient, les meilleures classes recevant des poids trop faibles par rapport aux secondes.

Pour la fertilité, le fait de regrouper les deux premières classes résout le problème : les poids sont correctement ordonnés, et une reclassification en les groupant donne un poids médian validant ces

valeurs.

Drainage					
Classe	W+	W-	Contraste	Confiance	Poids
1	0.05	-0.32	0.37	1.72	0.05
4	-0.32	0.05	-0.37	-1.72	-0.32

Fertilité					
Classe	W+	W-	Contraste	Confiance	Poids
1	0.09	-0.43	0.52	2.66	0.09
3	-0.36	0.05	-0.42	-1.85	-0.36
4	-0.59	0.03	-0.63	-1.74	-0.59

Pour le drainage, un entraînement *ascending* fait apparaître la nécessité de grouper les trois premières classes pour obtenir un résultat satisfaisant, cela venant peut-être également de la qualité très discutable du jeu de données utilisé.

2. Carte de probabilités

Les cinq paramètres précédents sont donc utilisés pour créer la carte de probabilités, présentée ci-dessous (Illustration 35).

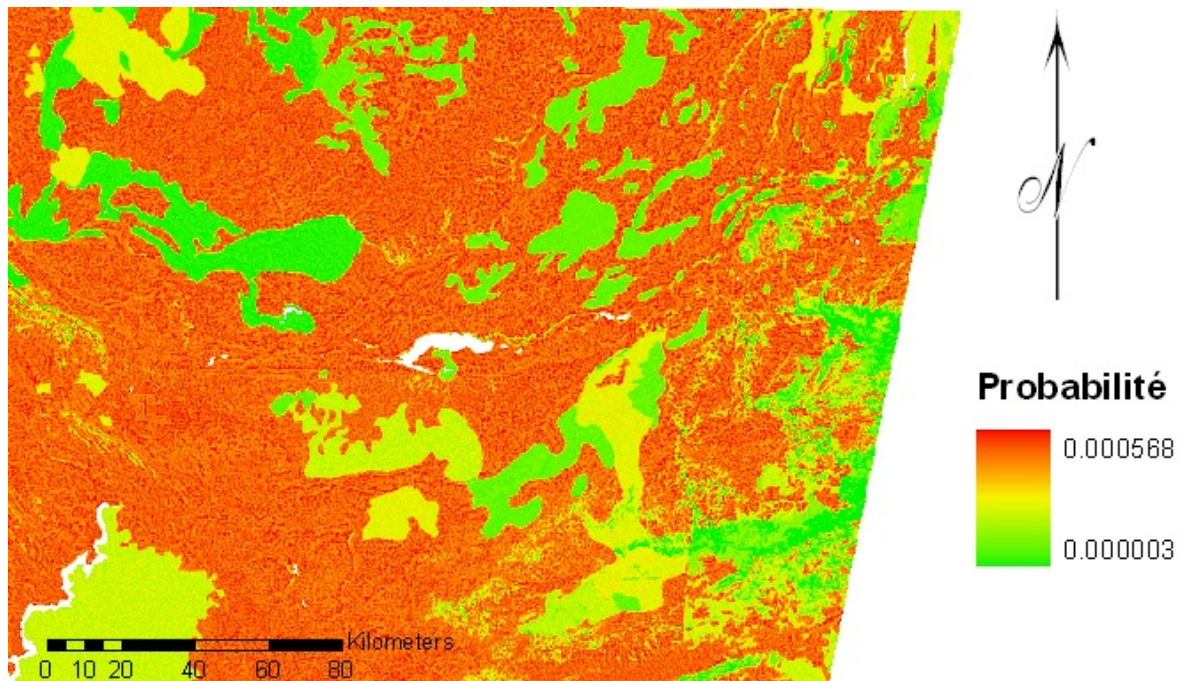


Illustration 35: probabilités pour la région obtenues avec *a_regional*

Une utilisation de l'outil *Area Frequency Tool* fait apparaître une efficacité de classification de 58% en utilisant le jeu d'entraînement, et une efficacité de prédiction de 63,5% en considérant les sites issus du modèle local. Ces valeurs apparaissent raisonnables au vu de la qualité des paramètres disponibles, et il est intéressant de voir qu'un entraînement au moyen de sites généralisés permet tout de même de prédire les unités résidentielles individuelles.

Le test d'Agterberg-Cheng donne une probabilité de 54,9% que le modèle ne soit pas conditionnellement indépendant, la dépendance n'apparaissant cependant pas être très forte.

D. Application avec les points locaux

La méthode mise en place dans cette partie consiste à utiliser pour entraîner le modèle les points générés au moyen du modèle prédictif local. Parmi les différences appliquées, il convient notamment

de changer la taille de cellule unitaire afin qu'elle soit adaptée à ces sites. La valeur choisie est de 0,0016 km², s'avérant pertinente d'après les calculs à l'échelle locale. L'outil de réduction des sites mis à disposition dans ArcSDM est utilisé : lors de la création des points, une distance minimale de 40 m avait été imposée au sein de chaque classe, mais des sites situés au niveau de deux classes différentes peuvent être plus proche que cette valeur.

Tous les paramètres sont restreints au masque local par extraction pour l'entraînement, il est donc important de prendre en compte le fait que toutes les classes d'évidences ne seront sans doute pas représentées, et d'envisager comment traiter ce problème.

1. Analyse des poids

i. ENBR (500 m)

Un premier entraînement est effectué avec les types *categorical* et *ascending*, mais ne donne aucun résultat satisfaisant. Une reclassification est effectuée, afin de tester une ENBR positive ou négative comme dans le cas précédent, mais sans succès, ce paramètre ne sera donc pas utilisé.

ii. SRTM

Une fois encore, les poids calculés ne donnent aucune information intéressante pour ce paramètre qui sera donc ignoré.

iii. Pente topographique, drainage et fertilité

Les paramètres de drainage et fertilité donnent de très bons résultats : ils ne nécessitent pas de reclassification, les poids sont correctement ordonnés et les niveaux de confiances très élevés du fait du nombre très important de points d'entraînement utilisés.

Drainage					
Classe	W+	W-	Contraste	Confiance	Poids
1	0.98	-0.5	1.48	128.09	0.98
2	0.17	-0.06	0.23	18.1	0.17
3	-0.49	0.18	-0.67	-47.5	-0.49
4	-2.81	0.27	-3.08	-66.65	-2.81

Fertilité					
Classe	W+	W-	Contraste	Confiance	Poids
1	0.68	-0.92	1.6	121.48	0.68
2	-0.56	0.26	-0.81	-60.81	-0.56
3	-2.47	0.11	-2.58	-44.79	-2.47
4	-4.7	0.1	-4.8	-25.37	-4.7

Pour la pente en revanche, le résultat n'est pas concluant. Il est cependant possible de prendre en compte les sites ayant servi à entraîner le masque local à l'origine plutôt que ceux propagés (ce qui se justifie par le fait que ceux-là sont à l'origine du modèle), et d'utiliser leurs poids : il est rappelé que deux classes sont considérées, selon que la pente est supérieure ou inférieure à 33%.

Pente topographique					
Classe	W+	W-	Contraste	Confiance	Poids
0	0.02	-1.67	1.69	2.37	0.02
33	-1.67	0.02	-1.69	-2.37	-1.67

2. Cartes de probabilités

Trois paramètres sont donc finalement retenus pour le calcul de la carte de probabilités : drainage, fertilité, et pente topographique.

Il est possible d'affiner cette carte en prenant en compte les poids calculés pour le SRTM et l'ENBR au moyen de la méthode précédente, auquel cas l'efficacité de prédiction est augmentée. La justification de cet ajout vient du fait que ces paramètres sont pertinents, les qualités du SRTM et de l'ENBR calculé à partir de celui-ci étant clairement identifiées et adaptées au problème à la différence de la fertilité et du drainage, il serait donc dommage de ne pas les utiliser.

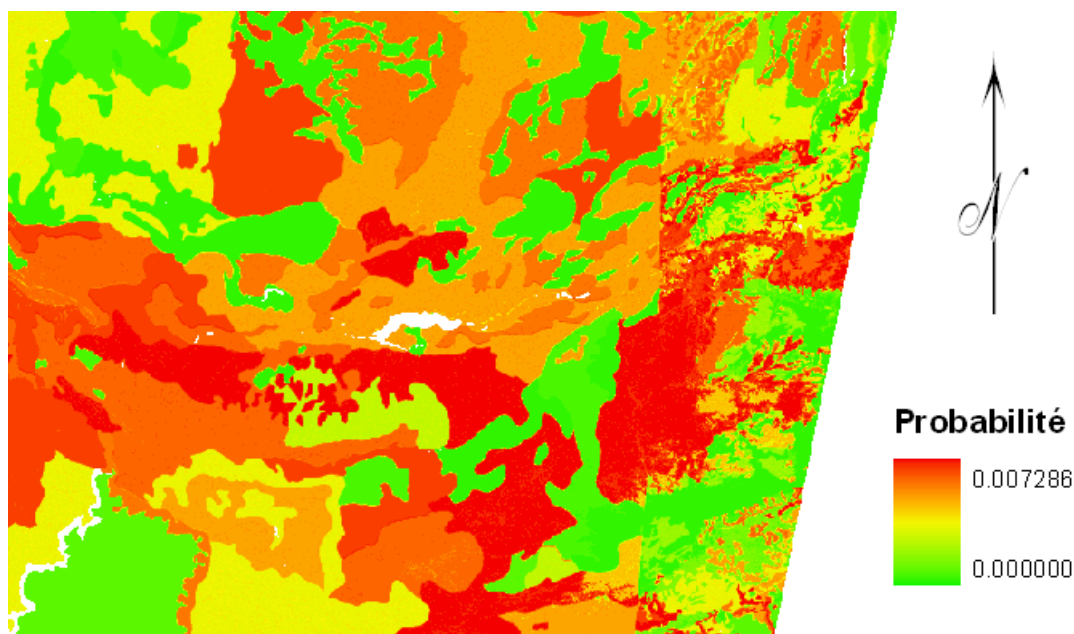


Illustration 36: probabilités pour la région obtenues avec mod_population

E. Choix d'une carte de probabilités à conserver

La carte de probabilités la plus efficace pourrait donc être celle calculée en premier lieu en utilisant les cinq paramètres disponibles. Produite en utilisant des sites très généralisés (seuls les grands centres mayas sont répertoriés) et en émettant l'hypothèse que leur répartition permet d'utiliser le masque régional comme zone d'entraînement, celle-ci permet de prédire 63,5% des sites créés au moyen du modèle local.

Ce chiffre est certes bien inférieur aux 80% d'efficacité à l'échelle précédente, mais est tout de même remarquable puisque ces sites sont d'une nature différente de ceux utilisés lors de l'entraînement. La piètre qualité des données de drainage et de fertilité empêche hélas d'avoir un résultat réellement exploitable, même à cette échelle.

Cependant, cette première approche se heurte au fait que cette carte de probabilités apparaît bien trop générale. Une très grande partie de la carte est ainsi couverte par les deux meilleures classes de probabilités, la valeur de 63,5% de prédiction s'en trouve donc largement contrebalancée. La zone centrale de la zone locale, qui ne comporte aucun site et est normalement de probabilité quasi-nulle se retrouve ainsi très bien classée...

Jeu de données	a_Tikal	a_regional	mod_population	a_sites	a04_settlements	a_El_Pilar
a_regional	60%	58%	64%	63%	60%	61%
mod_population	47%	52%	77%	83%	82%	56%

Table 9: efficacités de prédiction pour quelques jeux de sites, selon l'entraînement choisi

La table 9 présente les efficacités de prédiction (ou de classification) calculées pour un certain nombre de jeux de données afin de comparer les deux résultats. Ces pourcentages correspondent aux proportions de sites se trouvant dans une zone dont la valeur de probabilité a posteriori est plus importante que celle a priori. Cela n'est cependant pas tout à fait satisfaisant dans ce cas puisqu'il faut garder à l'esprit que l'indépendance conditionnelle n'est pas établie : les motifs sont bons, les probabilités en revanche peuvent être faussées.

La seconde version du modèle régional de probabilités produite à partir des points du modèle local apporte une efficacité de prédiction plus importante pour les jeux de données correspondant à l'échelle locale, ce qui était prévisible puisque les points d'entraînement sont similaires. Il apparaît également qu'elle est plus sélective que celle calculée en premier, ce qui est un paramètre très important. Il est ainsi très clair que les motifs correspondent plus à la réalité constatée au niveau local.

Il est possible de remarquer que la « croix » de lever des sites de Tikal est bien expliquée (Illustration 37), les sites étant clairement absents dans les branches au niveau des zones de faible probabilité. Le fait que certains « paquets » de points se retrouvent décalés par rapport aux limites des zones peut sans doute venir de l'imprécision de certains paramètres, mais le résultat général est extrêmement satisfaisant.

Il est difficile de faire des choix supplémentaires, le modèle étant clairement limité par la qualité des données de drainage et fertilité. Les classes où se trouvent les points de Tikal comportent visiblement de nombreux points régionaux, ce qui pourrait indiquer que leurs probabilités devraient être plus élevées. En même temps, la frontière entre le Belize et le Guatemala fait passer de celles-ci aux classes de fortes probabilités : une reclassification pourrait ramener à la même généralisation que pour l'entraînement avec les points régionaux.

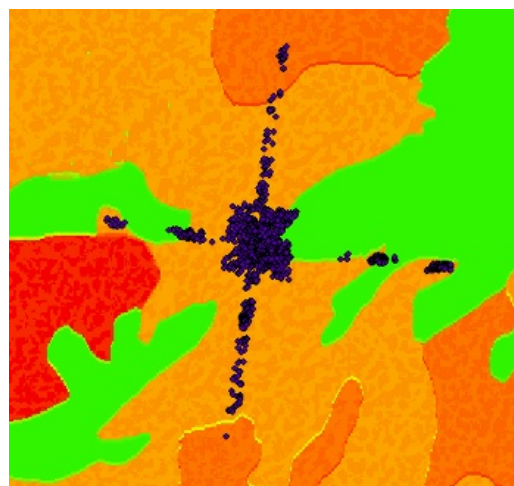


Illustration 37: "croix" de Tikal sur fond de probabilités

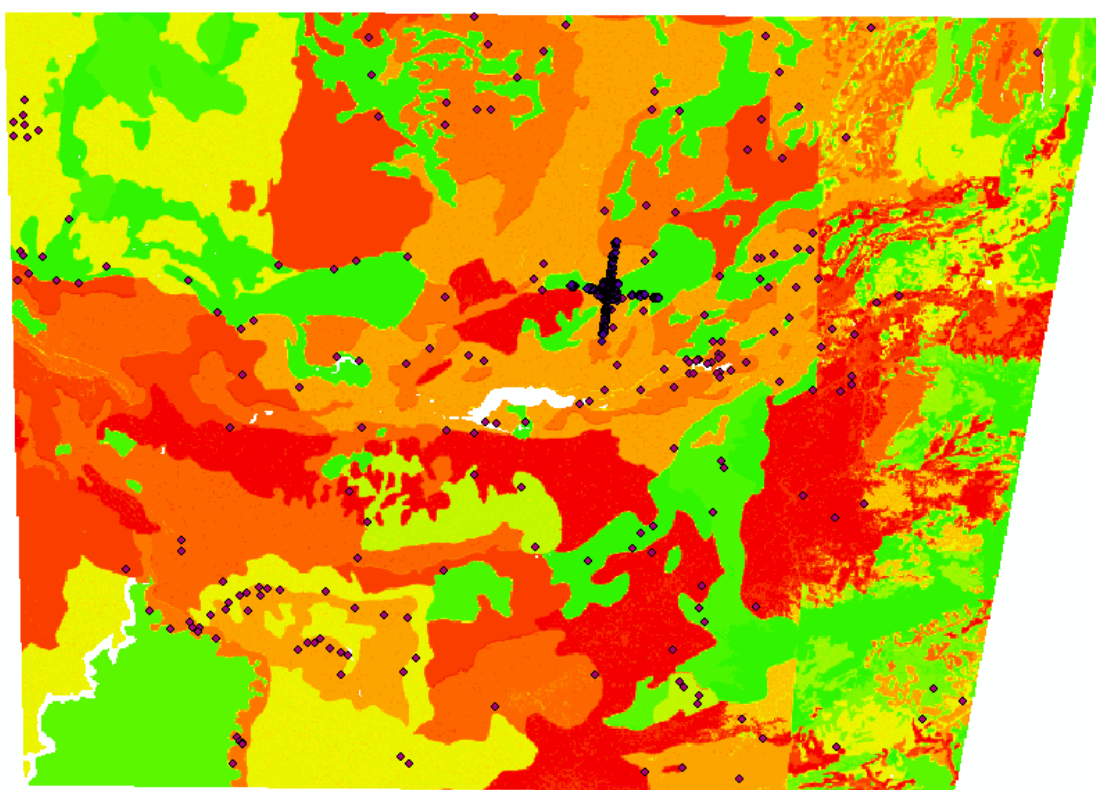
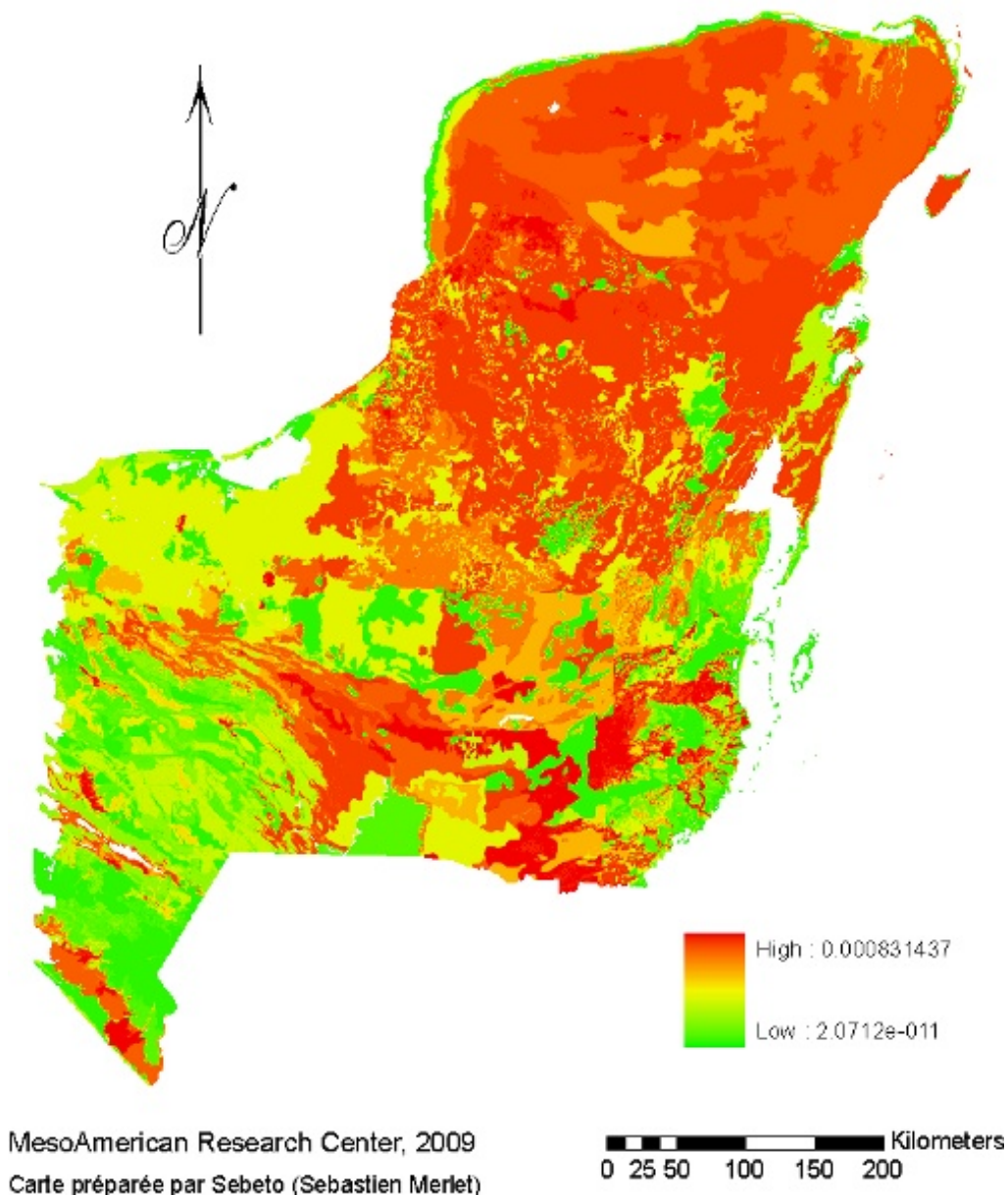


Illustration 38: points régionaux et "croix" de Tikal sur fond de probabilités

VIII. Modèle prédictif du monde maya et analyses supplémentaires

Malgré les défauts évidents du modèle régional, il est possible d'effectuer une propagation à l'échelle du système du monde maya. La carte résultante est présentée ci-dessous :



Cette généralisation est effectuée à titre informatif, les conclusions qu'elle permet de tirer étant très minces. Ainsi, toute la partie du Mexique située au nord du Guatemala et du Belize apparaît avoir de très bonnes probabilités, venant de bons drainage et fertilité.

Cependant, de nouveaux paramètres devraient être pris en compte à cette échelle. L'étendue de la zone d'étude entraîne par exemple des changements climatiques qui ne peuvent plus être ignorés, notamment un important gradient de pluviosité dû aux grandes différences de latitudes. Rien qu'au Belize par exemple, les précipitations varient de 4 445 mm au sud à 1 295 mm au nord, et les températures au niveau de la côte s'étendent de 10°C à 35,6°C, les différences étant encore plus importantes à l'intérieur des terres (Government of Belize 2006).

Dans l'état actuel des choses et en l'absence de données supplémentaires utilisables, cette étude ne peut donc aller plus loin. Un nombre plus important de sites aux échelles de la région du Petén et du système du monde maya serait également nécessaire pour entraîner le modèle efficacement.

IX. Conclusion

L'utilisation des systèmes d'information géographique au sein d'un projet archéologique multi-échelles nécessite énormément d'organisation. Les différents facteurs entrant en compte doivent être considérés par l'analyste qui, devant parfois travailler en l'absence d'autres personnes qualifiées dans le domaine des SIG, se doit de mettre en place et de maintenir une administration efficace des jeux de données disponibles.

Il apparaît primordial d'effectuer une veille technologique permanente pour utiliser les avancées en terme de stockage, normes, analyse et divulgation des informations. La présente étude révèle ainsi l'importance d'une bonne maintenance des métadonnées pour des raisons qualitatives et informatives, ainsi que de la définition de protocoles précis pour l'organisation, la modification, et l'accès aux jeux de données au sein de l'outil informatique. La solution privilégiée ici a consisté en la création d'une géodatabase ne comprenant que des données identifiées et documentées autant que possible.

Afin de prendre en compte le caractère éphémère de la présence de spécialistes en systèmes d'information géographique au sein du *MesoAmerican Research Center*, des règles très précises ont été établies et expliquées aux membres de l'équipe. Dans cette optique de simplification de l'utilisation tout en maintenant la qualité de la base de données géographiques sur la durée, des contrôles sont effectués au sein de celle-ci au moyen de topologies pour garantir des jeux adaptés aux usages auxquels ils sont destinés.

Un nouveau logiciel étant disponible pour la mise en place de la méthode des *Weights of evidence*, il a été décidé d'effectuer une validation d'une première version du modèle local conçue dans le passé (Monthus 2004, publié dans Clarke et al. 2009). L'analyse de la méthode mise en place a cependant amené à identifier certains problèmes, et certaines interprétations n'apparaissaient pas judicieuses. Un nouveau protocole de mise en œuvre de la méthode a donc été élaboré, et un modèle corrigé produit. Ceci a amené à recalculer la répartition de la population, dont les paramètres de base n'étaient plus valides et dont plusieurs paramètres n'apparaissaient plus pertinents.

La nouvelle technique mise en œuvre produit des résultats très satisfaisants, et leur analyse apporte une nouvelle vision de l'utilisation des sols par les Mayas habitant la zone d'étude locale. Les densités trouvées au niveau des différentes classes sont cohérentes avec celles établies au sein du monde maya par les travaux d'autres archéologues (Wheeler 2008). En considérant les besoins en maïs procurant les carbohydrates nécessaires à une diète équilibrée (Rose 2003), cette étude arrive à la conclusion que l'utilisation du cycle du milpa suffisait probablement à nourrir l'importante population (142 hab./km² en moyenne).

Ce cycle est cependant réduit à une durée de 12 ans, ce qui est moins que la durée nécessaire à un retour à une seconde phase de forêt secondaire ou à une forêt mature avec canopée fermée (Nigh 2008; Ford & Nigh 2009). Il serait intéressant d'étudier les effets de ce problème sur l'environnement : la faune s'en trouve-t-elle très impactée ? Le phénomène d'enrichissement des sols au fil de la succession des cycles est-il toujours possible ? Le modèle apparaît donc être à une limite de stabilité, et tout stress climatique diminuant la productivité des champs aurait alors pu avoir des conséquences très importantes et participer au déclin de cette civilisation.

Un autre constat s'impose : au vu de la population actuelle du Belize et de sa densité dix fois inférieure à celle de la zone locale étudiée (Government of Belize 2006; SIB 2007), une application du cycle du milpa pourrait être envisagée pour soutenir la production agricole tout en préservant l'environnement. Les Mayas ne connaissant pas le fer et n'utilisant pas la roue, la productivité actuelle pourrait sans doute être supérieure à celle obtenue il y a plus de 1 500 ans de cela, même en évitant de recourir à des méthodes trop modernes qui stresseraient les sols.

Ces résultats confortent les actions menées par les jardiniers forestiers modernes descendant des Mayas (EPFGN 2008), qui travaillent à sauvegarder et promouvoir les techniques ancestrales nécessaires à l'établissement de cycles durables, capables d'enrichir les sols au lieu de les appauvrir (Nigh 2008). En plus de pouvoir amener un moyen de subsistance supplémentaire aux populations étudiées (Corzo Marquez & Schwartz 2008; EPFGN 2008), ces méthodes pourraient être utilisées pour limiter la déforestation dans certaines zones sensibles, la forêt étant un élément extrêmement important

dans ce système.

Une réalisation du modèle prédictif à l'échelle régionale réalisée au moyen de deux jeux de points d'entraînements différents fait apparaître les difficultés à choisir les paramètres et le type de sites à prendre en compte. En utilisant des sites généralisés, l'efficacité de prédiction apparaît bonne, mais le résultat n'est pas assez sélectif pour être intéressant. Les sites produits à l'échelle locale peuvent servir dans ce cas, la carte produite étant satisfaisante.

Cependant, la qualité des données disponibles pose ici problème. Issues de sources diverses, celles-ci présentent des différences importantes tant au niveau de leur précision que de la continuité aux frontières. Les processus de reclassification et de réunification effectués dans le passé à partir des informations initiales n'ont pas été correctement menés, plusieurs versions de ces fichiers sont donc disponibles, se contredisant fortement pour plusieurs zones. Des corrections d'envergure ont été menées durant cette étude, notamment pour résoudre la question des chevauchements et trous dans la couverture de polygones.

Même si le modèle régional donne une bonne vision des motifs pouvant être observés, il est donc important de garder à l'esprit les aspects qualitatifs pour ne pas arriver à des conclusions qui se voudraient trop précises par rapport aux paramètres. C'est pour cette raison que la propagation des sites et le calcul d'une population régionale n'ont pas été faits : cela nécessite clairement d'obtenir au préalable des cartes de fertilité et de drainage plus adaptées pour éviter des résultats aberrants. Un modèle de subsistance tel que celui créé localement serait également un vrai défi, de nouvelles techniques d'agriculture pouvant intervenir à cette échelle, notamment l'utilisation de terrasses pour s'adapter à la pente (Fischbeck 2004).

La généralisation au Système du Monde Maya est très intéressante : le niveau de zoom découlant de la taille de cette carte permet d'avoir une vision plus globale, et a pour effet de « gommer » les imprécisions des paramètres. Toute la partie nord apparaît très attractive, alors qu'elle n'a été vraiment peuplée que tardivement. Il semble donc qu'à cette échelle de nouveaux paramètres doivent être pris en compte du fait de l'étendue de la zone. Rien qu'au Belize, la pluviosité est près de quatre fois plus importante au sud qu'au nord, et ce pays comporte également d'importants gradients de température (Government of Belize 2006).

Ces variables climatiques qui n'avaient peut-être pas d'effet de taille au niveau local (cette hypothèse n'a pas été vérifiée au cours de cette étude) jouent probablement un rôle bien plus grand à une telle échelle multinationale : influence sur la faune et la flore, ou sur la productivité des champs de maïs par exemple.

Dans l'état actuel des choses, les différents modèles semblent donc être arrivés à leurs limites : seule l'obtention de nouveaux jeux de données permettrait de pousser plus loin les analyses. Le travail effectué amène d'ores et déjà des informations extrêmement intéressantes sur l'occupation des sols et la situation limite des Mayas pouvant avoir contribué à leur déclin.

Une analyse réalisée en partenariat avec Marc Duquette (responsable de l'inventaire animalier pour le projet) est actuellement en cours pour étudier la répartition des animaux au sein des classes de végétation pour le modèle produit correspondant à la densité moyenne établie (Annexe 30 : représentation de la densité de population), afin de tenter de déterminer les effets que la réduction importante de la forêt mature pourrait entraîner. Cela ouvre notamment de nouvelles perspectives de recherche au niveau de l'alimentation des Mayas et des effets environnementaux sur le déclin de cette civilisation. Dans un pays composé à 72,5% de forêt, dont 37% est classifiée en forêt primaire porteuse d'une grande biodiversité (Butler 2009), la compréhension du mode de vie d'une civilisation passée peut apporter de nombreuses clés pour aller dans le sens d'un développement durable plus respectueux de l'environnement.

X. Annexes

Annexe 1 : résumé.....	59
Annexe 2 : contrat de non-plagiat.....	62
Annexe 3 : liste des illustrations et mentions légales.....	63
Annexe 4 : références.....	64
Annexe 5 : localisation des zones locale et régionale.....	69
Annexe 6 : masques utilisés à l'échelle locale.....	70
Annexe 7 : carte locale du drainage.....	71
Annexe 8 : carte locale de la fertilité.....	72
Annexe 9 : carte locale de la pente.....	73
Annexe 10 : carte des tampons de distance aux rivières (2004).....	74
Annexe 11 : carte des tampons de distance aux rivières (BERDS).....	75
Annexe 12 : ratios calculés avant reclassification.....	76
Annexe 13 : entraînement au moyen du masque et des sites des transects.....	77
Annexe 14 : poids pour les rivières, sites des transects avec masque local.....	81
Annexe 15 : test d'indépendance conditionnelle (local).....	82
Annexe 16 : carte du modèle prédictif local (première version).....	83
Annexe 17 : carte du modèle prédictif local de 2004.....	84
Annexe 18 : tables du modèle de population (première version).....	85
Annexe 19 : carte de reclassification des probabilités (première version).....	86
Annexe 20 : modèle de propagation des sites (première version).....	87
Annexe 21 : poids et probabilités au niveau local (seconde version).....	88
Annexe 22 : probabilités reclassifiées (seconde version).....	89
Annexe 23 : image aérienne d'utilisation des sols.....	90
Annexe 24 : modèle de propagation des sites (seconde version).....	91
Annexe 25 : cartes régionales du drainage et de la fertilité.....	92
Annexe 26 : cycle du milpa.....	93
Annexe 27 : calculs de la superficie nécessaire pour la culture du maïs.....	94
Annexe 28 : scénarios répondant aux besoins de la population.....	95
Annexe 29 : utilisation des sols au niveau de la zone d'étude.....	96
Annexe 30 : représentation de la densité de population.....	97

Annexe 1 : résumé

Les systèmes d'information géographique occupent une place de plus en plus importante au sein des projets archéologiques depuis leurs premières utilisations au début des années 1980 (Aldenderfer & Maschner 1996; Wheatley & Gillings 2002). Ils sont en effet un outil très efficace pour stocker, organiser, analyser et mettre à disposition les différents jeux de données disponibles, tout particulièrement grâce aux possibilités de modélisation et de conception de modèles prédictifs qu'ils offrent. Il est cependant important de garder à l'esprit le caractère particulier d'un contexte archéologique pour que les SIG remplissent pleinement leur rôle, que ce soit dans l'organisation ou l'analyse des résultats produits (Gaffney et al. 1996; Dalla Bona 1994)

L'objet de mon étude est le projet El Pilar, mené par le docteur Anabel Ford au sein du *MesoAmerican Research Center* (Université de Californie à Santa Barbara). El Pilar était une importante cité maya située dans la région du Petén, au niveau de ce qui est maintenant la frontière entre le Belize et le Guatemala, en Amérique Centrale. Ses premiers bâtiments publics sont datés aux environs de 700 ans avant Jésus Christ, et ce centre important perdure jusqu'à 1000 ap. JC.



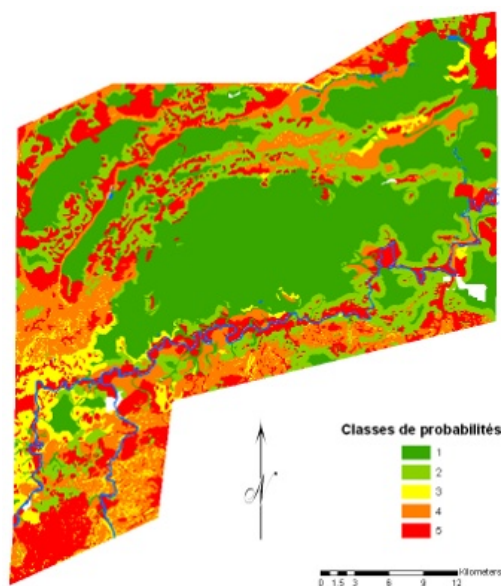
En collaboration avec l'organisation *Exploring Solutions Past: The Maya Forest Alliance*, l'équipe d'Anabel Ford étudie la civilisation maya du passé, mais également la manière de vivre de leurs descendants, les jardiniers forestiers mayas actuels, le but étant de trouver des modèles de développement durable dans cette région. Le territoire du Belize est en effet constitué de 72,5% de forêt, dont 37% est classifiée en forêt primaire, présentant une très grande biodiversité (Butler 2009), mais ce patrimoine environnemental exceptionnel pourrait à terme être menacé par l'expansion de l'agriculture moderne et de l'environnement urbain.

Une des raisons régulièrement avancées pour expliquer le déclin de la civilisation maya aux alentours de 900 ans après Jésus Christ serait une population trop élevée ayant conduit à une déforestation massive détruisant l'environnement. Plusieurs questions se posent alors : est-il possible d'estimer la population Maya à la fin de l'Ère Classique au niveau d'une zone d'étude suffisamment importante, et de modéliser en fonction de leurs besoins alimentaires leur occupation des sols ? Un tel modèle peut-il apporter de nouveaux éléments relatifs à l'effet de la présence des Mayas sur la fertilité des sols et la superficie laissée à la nature ? Leur modèle d'agriculture basé sur le cycle du milpa pourrait-il être utilisé pour protéger la forêt actuelle et comme moyen de subsistance pour une partie de la population ?

Le *MesoAmerican Research Center* dispose depuis 2004 d'un modèle prédictif conçu au niveau d'une zone de 1 285 km². Celui-ci présentant plusieurs problèmes, une nouvelle version a été créée

dans le cadre de ce travail de fin d'études. Au moyen d'ArcGis et de l'application ArcSDM (*Spatial Data Modeler*), une application de la méthode des Weights of Evidence (cf. Raines & Bonham-Carter 2006) a ainsi permis d'obtenir un modèle prédictif de l'emplacement des unités résidentielles maya à la fin de l'Ère Classique. Il apparaît que les principaux facteurs intervenant dans la sélection des zones d'habitation sont le drainage, la fertilité, et la pente topographique. ArcSDM calcule ainsi des poids pour chacune des classes d'évidence de ces paramètres au moyen de sites d'entraînements situés sur des transects levés dans le passé.

Il apparaît par exemple que les Mayas évitaient les pentes topographiques supérieures à 33% : un poids de -1,67 est donné à ces valeurs, alors que les valeurs inférieures ont un poids de 0,02. Le résultat est une carte de probabilités présentant les zones favorisées, ou au contraire celles à faibles probabilités. Afin de prendre en compte l'attractivité sociale des zones les plus peuplées, des tampons sont créés autour des fortes probabilités, et une nouvelle application de la méthode précitée fait ressortir de nouveaux motifs. Après reclassification, cinq classes de probabilités sont finalement identifiées, présentées ci-contre de la plus faible (classe 1 en vert) à la plus forte.



Une analyse des densités de sites au niveau des transects disponibles permet alors d'obtenir la répartition de la population à la fin de l'Ère Classique, et de calculer le nombre d'habitants au niveau de la zone d'étude. La densité moyenne est de 142 hab./km², et 38% du territoire ne comprend aucune habitation en raison d'un drainage et d'une fertilité très défavorables. 78% de la population est au contraire localisée sur 38% de la zone, la plus forte densité s'élevant à 390 hab./km², soit environ celle actuelle du Liban. Cette valeur peut sembler très élevée, mais elle est comparable aux valeurs de 407 à 506 hab./km² avancée par certains auteurs pour la zone rurale de Copal Pocket par exemple (Webster & Freter 1990).

L'étude des besoins de cette population en terme d'alimentation est effectuée en reprenant notamment des éléments d'un travail similaire effectué au niveau de Tikal (Rose 2003). La méthode consiste à considérer que les besoins en carbohydrates des Mayas provenaient principalement du maïs, composant 34% de leur alimentation (Canto 2003), ce qui définit leurs besoins en terme de sols cultivables. Un hectare de maïs produisait en moyenne 588 kg par an (Rose 2003), et l'apport d'un kilogramme de maïs est de 3 551 kilocalories (McHenry & Beaton 1963). En admettant que le besoin moyen journalier en calories pour mener une vie saine et normale est de 2 100 kilocalories (WFP 2009), il est donc possible d'obtenir la superficie de maïs devant être cultivée chaque année au niveau de la zone d'étude pour subvenir aux besoins de la population.

31% des sols cultivables sont ainsi nécessaires pour les milpas, mais il faut également prendre en compte les techniques d'agriculture des Mayas. Celles-ci sont basées sur le cycle du milpa, consistant en une culture sur brûlis suivie d'un accompagnement actif pour permettre le retour à une forêt mature. Proprement effectuée, cette technique permet d'enrichir les sols au fil des cycles, et entraîne une régénération de la forêt plus rapide (Nigh 2008). Plusieurs étapes sont identifiées dans ce cycle : culture du maïs, haricots, courgettes et plantes diverses (milpa) pendant une durée maximale de quatre années ; colonisation initiale par des arbres à croissance rapide (un à quatre ans) ; première phase de forêt secondaire (quatre à dix ans) ; seconde phase de forêt secondaire comportant de nouvelles essences (dix à vingt ans) ; et enfin un retour à la forêt mature en vingt ans et plus (Nigh 2008). D'après Nigh, douze à quinze ans apparaissent suffisant pour un retour à une forêt prête à être coupée et brûlée pour un nouveau cycle, même si dans les faits les Mayas préfèrent attendre plusieurs années supplémentaires.

Dans le cadre de cette étude, la durée maximale de quatre ans de culture du maïs est considérée. Le seul moyen d'arriver à un équilibre dans ces conditions – c'est-à-dire d'avoir assez de terres disponibles à la fois pour les milpas et pour les autres étapes du cycle – est d'appliquer une durée totale de douze années. À partir de ces résultats, il est possible de représenter pour chaque habitation maya la

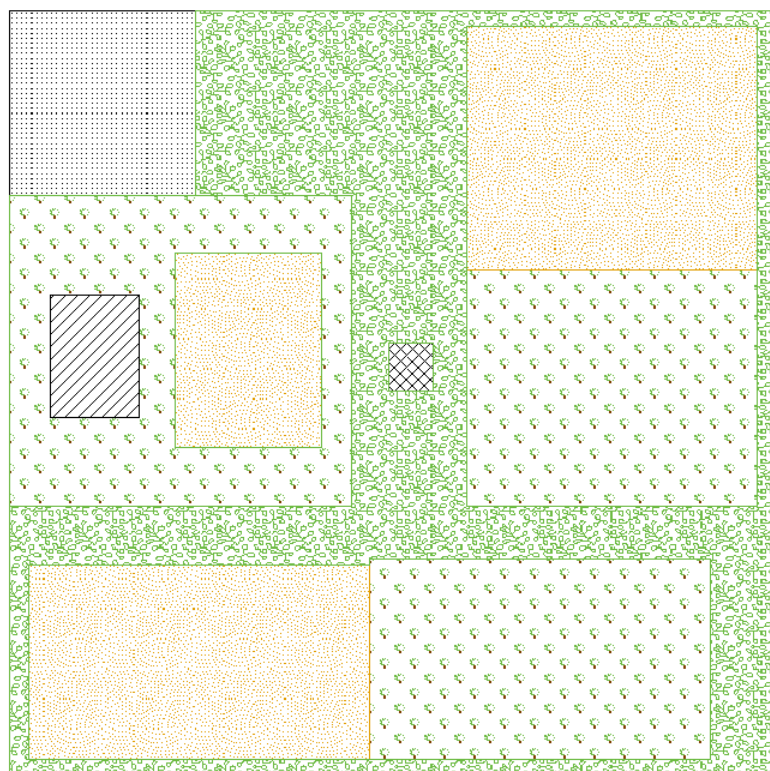
quantité de terrain disponible, et de la répartir selon les différents types d'utilisation.

Il apparaît que la forêt n'a dans ces conditions que huit années pour se régénérer, et n'arrive donc qu'à la première phase de forêt secondaire d'après la chronologie donnée par Nigh. Cela pose la question de la viabilité de ce modèle limite sur la durée : un cycle du milpa raccourci peut-il préserver la qualité des sols au fil du temps ? La raréfaction de la forêt mature et de la seconde phase de forêt secondaire a-t-elle des effets importants sur la vie animale, qui occupait également une grande place dans l'alimentation des Mayas ?

Les données pour répondre à ces questions ne sont pas encore établies, mais il apparaît que ce modèle d'agriculture pourrait être à sa limite à la fin de l'Ère Classique, et tout stress climatique supplémentaire aurait alors pu produire des problèmes d'alimentation et des tensions sociales participant au déclin de cette civilisation.

Il est cependant remarquable de constater par la même occasion que le système Maya était capable de subvenir aux besoins d'une population très importante avant la période de déclin. La densité moyenne de 142 hab./km² durant le Classique Tardif est à comparer avec celle dix fois inférieure du Belize actuel (Government of Belize 2006) : une utilisation du cycle du milpa pourrait être encouragée, ce qui participerait à la protection de la forêt de ce pays en limitant l'expansion de l'agriculture extensive. Les jardiniers forestiers mayas traditionnels modernes continuent de perpétuer les techniques et le savoir-faire (EPFGN 2008), et plusieurs études ont montré les avantages de ce système en terme de gestion des ressources, de sécurité alimentaire, et de conservation de l'environnement (Corzo Marquez & Schwartz 2008; Nigh 2008).

Au travers de l'utilisation d'un Système d'Information Géographique, la présente étude apporte donc de nouveaux éléments sur la répartition de la population, éléments qui pourront être pris en compte pour mieux comprendre l'environnement maya et les raisons du déclin de ce peuple. Elle met aussi en valeur les connaissances des Mayas du passé, et celles de leurs descendants qui continuent à protéger et perpétuer un système dans lequel chaque agriculteur participe activement à la protection et à la régénération de son environnement de vie et de travail.



0 5 10 20 30 40 Meters

Sébastien Merlet, 2009

Parties communes
Zone résidentielle
Structure temporaire
Forêt en régénération
Forêt secondaire
Milpa

Ci-contre : occupation des sols disponibles pour une famille maya de la zone d'étude lors du Classique Tardif

Annexe 2 : contrat de non-plagiat

Plagiat : action de plagier

Plagier : emprunter à un ouvrage et à son auteur, des éléments, des fragments dont on s'attribue abusivement la paternité en les reproduisant, avec plus ou moins de fidélité, dans une œuvre que l'on présente comme personnelle.

Je soussigné Sébastien Merlet,

déclare sur l'honneur connaître la définition du plagiat et ne pas m'y adonner. Je m'engage à ne pas copier, contrefaire ou falsifier mon mémoire de TFE, et à ne pas y utiliser, en tout ou en partie, l'œuvre d'autrui ou des passages appréciables tirés de celle-ci, sans les identifier expressément comme citations, et dans l'intention de les faire passer pour miens.

Fait à : San Ignacio, le 4 juin 2009

Signature de l'étudiant :

Sébastien Merlet

Annexe 3 : liste des illustrations et mentions légales

Les illustrations de ce mémoire ont été réalisées à partir des données mises à disposition par le *MesoAmerican Research Center* pour les besoins de cette étude. Certaines d'entre elles comportent cependant des éléments extérieurs dont les mentions légales n'ont pas pu être insérées directement pour des raisons de taille des images finales, et sont communiquées après l'index des illustrations.

Illustration 1: Réserve archéologique d'El Pilar.....	7
Illustration 2: exemple de problème où l'absence de métadonnées s'avère très gênante.....	9
Illustration 3: Nouvelle structure du SIG.....	12
Illustration 4: Sites archéologiques.....	13
Illustration 5: contrôle de la fertilité des sols.....	13
Illustration 6: Modèle prédictif utilisé.....	18
Illustration 7: Comparaison des jeux de rivières, et tampons utilisés dans le passé.....	23
Illustration 8: Lac non pris en compte pour les tampons de distance aux rivières.....	23
Illustration 9: Pixel de 50 mètres.....	24
Illustration 10: Pixel de 30 mètres.....	24
Illustration 11: ratios pour le drainage.....	26
Illustration 12: ratios pour la fertilité.....	26
Illustration 13: ratios pour la distance aux rivières.....	27
Illustration 14: difficultés inhérentes à la position des rivières, dans le présent ou le passé.....	28
Illustration 15: modèle de 2004.....	31
Illustration 16: nouveau modèle.....	31
Illustration 17: temples mayas dépassant de la canopée, Tikal.....	31
Illustration 18: répartition des sites par rang.....	32
Illustration 19: Classes prédictives du modèle de 2004.....	33
Illustration 20: application d'une cinquième classe pour compenser les problèmes.....	33
Illustration 21: attractivité sociale.....	34
Illustration 22: classification mécanique des sols.....	36
Illustration 23: classification manuelle des sols.....	36
Illustration 24: carte des probabilités.....	36
Illustration 25: vue aérienne.....	36
Illustration 26: répartition des rangs selon les classes de probabilités.....	38
Illustration 27: schéma d'occupation du sol avec milpa et jardin forestier	41
Illustration 28: occupation des sols pour le scénario n°2.....	44
Illustration 29: temple maya en cours de restauration à Tikal, Guatemala.....	45
Illustration 30: sites du jeu a_Tikal sur fond de pente topographique.....	46
Illustration 31: ENBR.....	48
Illustration 32: disque de calcul.....	48
Illustration 33: fréquences de distribution.....	48
Illustration 34: changement de classes de fertilité aux frontières.....	49
Illustration 35: probabilités pour la région obtenues avec a_regional.....	51
Illustration 36: probabilités pour la région obtenues avec mod_population.....	53
Illustration 37: "croix" de Tikal sur fond de probabilités.....	54
Illustration 38: points régionaux et "croix" de Tikal sur fond de probabilités.....	54

Mentions légales :

Illustration 1 : image de fond :

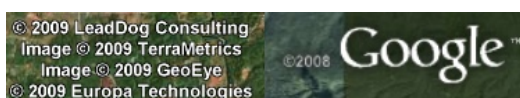
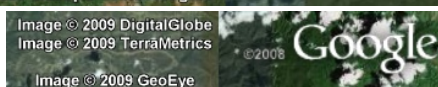


Illustration 25 : image de fond :



Annexe 4 : références

- Aldenderfer, M. & Maschner, H.D.G., 1996. *Anthropology, Space, and Geographic Information Systems*, Oxford University Press.
- Allen, S., 2000. Considerations of scale in modeling settlement patterns using GIS: An Iroquois example. In *Practical Applications of GIS for Archaeologists*. London: Taylor & Francis, pp. 101-112.
- Benvenuti, L., 2003. *Redlands Institute Manual for Creating Metadata in ArcCatalog*, The Redlands Institute, University of Redlands.
- BERDS, 2005. *Belize aquatic streams and rivers shapefile*, Biodiversity and Environmental Resource Data System of Belize. Available at: http://www.biodiversity.bz/bbms/downloads/BZE_Aquatic_Ecosystems.zip [Accessed February 2, 2009].
- Booth, B. & Mitchell, A., 2001. Getting started with ArcGis.
- Boserup, E., 1983. Regional differences in population and technology. In *Population and technological change, a study of long term trends*. The University of Chicago press, pp. 1-28.
- Butler, R., 2009. *Belize: forest cover*, Mongabay.com. Available at: <http://rainforests.mongabay.com/deforestation/2000/Belize.htm> [Accessed June 6, 2009].
- Canto, G., 2003. *Faculty of agriculture and natural resources visit*, edited by K. Rose, San Ignacio, Cayo.
- CGIAR, 2008. SRTM 90m Digital Elevation Data. Available at: <http://srtm.csi.cgiar.org/> [Accessed February 13, 2009].
- Chazdon, R.L., 2008. Chance and determinism in tropical forest succession. In *Tropical forest community ecology*. Oxford: Wiley-Blackwell Publishing, pp. 384-408.
- Clarke, K.C., Raines, G.L. & Ford, A., 2009. Modeling settlements patterns of the Late Classic Maya with Bayesian methods and GIS. *Annals of the Association of American Geographers*.
- Corzo Marquez, A.R. & Schwartz, N.B., 2008. Traditional home gardens of Petén, Guatemala: resource management, food security, and conservation. *Journal of ethnobiology*, 28, 305-318.
- Cowgill, U., 1962. An agricultural study of the Southern Maya lowlands. *American Anthropologist*, 64, 273-286.
- Dalla Bona, L., 1994. Methodological Considerations. In *Cultural Heritage Resource Predictive Modelling Project Vol. 4*. Centre for Archaeological Resource Prediction, Lakehead University, Ontario.

- Dermody, B.J., 2005. *An investigation into the performance of predictive GIS archaeological models in an Irish context*. Dublin Institute of Technology.
- Dickson, D., 1980. Ancient agriculture and population at Tikal, Guatemala: an application of linear programming to the simulation of an archaeological problem. *American Antiquity*, 45, 697-712.
- Dunster, C.H., 2004. About the artist. *Maya drawings*. Available at: <http://mayadrawings.com/page/2/> [Accessed May 23, 2009].
- EPFGN, 2008. *The Maya Forest as a garden*, The El Pilar Forest Garden Network. Available at: <http://www.mayaforestgardeners.org/marketpotential.php>.
- ESP-Maya, 2009. *The Milpa cycle*, Exploring Solutions Past, The Maya Forest Alliance. Available at: <http://espmaya.org/milpa.html> [Accessed May 18, 2009].
- ESRI, 2008. ArcGis Desktop 9.3 help. Available at: <http://webhelp.esri.com/arcgisdesktop/9.3> [Accessed March 23, 2009].
- ESRI, 2003. ESRI Profile of the Content Standard for Digital Geospatial Metadata. Available at: <http://www.esri.com/metadata/esriprof80.html> [Accessed March 23, 2009].
- FAO, 1998. *Balance sheets*, Food and Agriculture Organization of the United Nations.
- FAO, 1957. *Calorie requirements; report of the second Committee on Calorie Requirements*, Rome: Food and Agriculture Organization of the United Nations.
- FAO, 2009. *FAOSTAT*, Food and Agriculture Organization of the United Nations. Available at: <http://faostat.fao.org/> [Accessed May 21, 2009].
- Farr, T.G. et al., 2007. The Shuttle Radar Topography Mission. *Reviews of Geophysics*, 45. Available at: http://www2.jpl.nasa.gov/srtm/SRTM_paper.pdf [Accessed May 20, 2009].
- Fedick, S.L., 1992. An agricultural perspective on prehistoric Maya household location and settlement density. In *Memorias Del Primer Congreso Internacional De Mayistas*. México: Universidad Nacional Autónoma de México.
- Fedick, S.L., 1988. *Prehistoric Maya settlement and land use patterns in the upper Belize river area, Belize, Central America*. Arizona State University.
- Fischbeck, S.L., 2004. Agricultural terrace productivity in the Maya lowlands of Belize. In *Journal of undergraduate research*. University of Wisconsin-La Crosse, pp. 105-112. Available at: <http://murphylibrary.uwlax.edu/digital/jur/2001/fischbeck.pdf> [Accessed May 4, 2009].
- Ford, A., 2008. Dominant plants of the Maya forest and gardens of El Pilar: implications for paleoenviron-

mental reconstructions. *Journal of ethnobiology*, 28, 179-199.

Ford, A. & Nigh, R., 2009. Origins of the Maya forest garden: a resource management system.

Gaffney, V., Stancic, Z. & Watson, H., 1996. Moving from Catchments to Cognition: tentative Steps Toward a Larger Archaeological Context for GIS. In *Anthropology, Space, and Geographic Information Systems*. Oxford University Press, pp. 132-154.

Gasco, J., 2008. 'Le da alegria tener flores' homegardens in the Soconusco region of Chiapas, Mexico. *Journal of ethnobiology*, 28, 259-277.

Gentle, L. et al., 2000. *El Pilar Community Creek tail, life returns to the forest*, Friends for Conservation and Development (FCD).

Government of Belize, 2006. *About Belize*, Government of Belize. Available at: http://www.governmentofbelize.gov.bz/about_belize.html [Accessed May 25, 2009].

HHS & USDA, 2005. Dietary guidelines for Americans, 2005. Available at: <http://www.health.gov/dietary-guidelines/dga2005/document/pdf/DGA2005.pdf> [Accessed May 18, 2009].

Jarvis, A. et al., 2004. Practical use of SRTM data in the tropics - Comparisons with digital elevation models generated from cartographic data.

Jenkin, R.N. et al., 1976. *The agricultural development potential of the Belize Valley, vol. 1*, Tolworth Tower, England: Land Resources Division, Ministry of Overseas Development.

Lamy, G., 2008. *Utilisation de l'outil sig dans la cartographie de sites archéologiques mayas, application pour la modélisation de la population*. École Supérieure des Géomètres et Topographes.

Legg, R.J., 2007. Site catchment analysis. In *Encyclopedia of Archaeology*. Elsevier.

Leusen, P.M., 2002. *Pattern to process: methodological investigations into the formation and interpretation of spatial patterns in archaeological landscapes*. Available at: <http://dissertations.ub.rug.nl/faculties/arts/2002/p.m.van.leusen/> [Accessed April 1, 2009].

Lopez Morales, F.J., 1993. *Arquitectura vernácula en México*, México: Fondo Internacional para la Promoción de la Cultura, Unesco: Editorial Trillas.

Lutz, W., Prieto, L. & Sanderson, W., 2000. *Population, development, and environment on the Yucatan Peninsula: From Ancient Maya to 2030*, International Institute for Applied Systems Analysis.

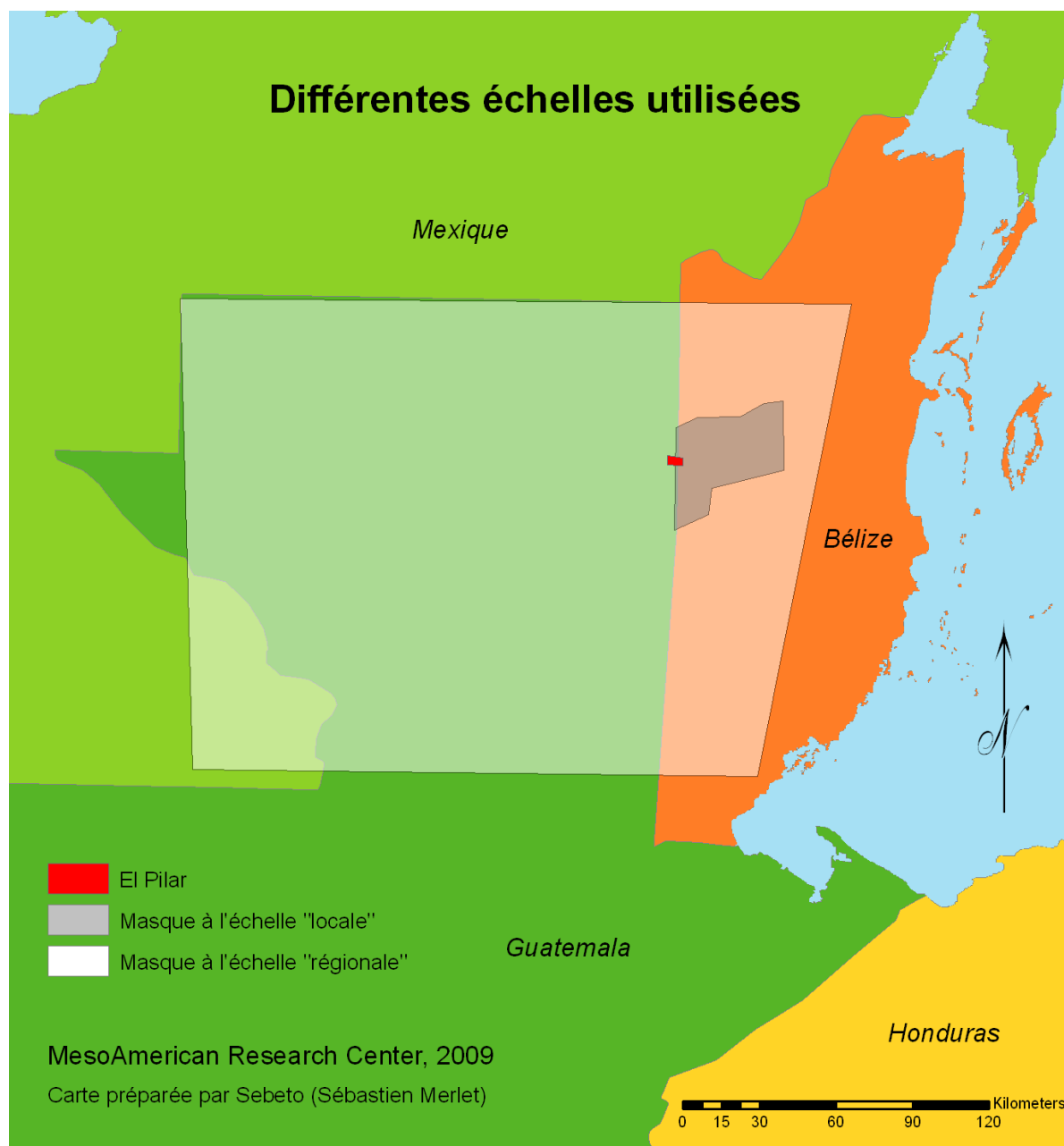
McHenry, E.W. & Beaton, G.H., 1963. *Basic Nutrition* Rev. ed., Philadelphia: Lippincott.

- Mehrer, M., 2006. *GIS and Archaeological Site Location Modelling*, Boca Raton, Fla: CRC.
- Monthus, F., 2004. *Mémoire en vue d'obtenir le diplôme ingénieur ESGT*. École Supérieure des Géomètres et Topographes.
- Nations, J.D. & Nigh, R., 1980. The evolutionary potential of Lacandon Maya sustained-yield tropical forest agriculture. *Journal of Anthropological research*, 36(1).
- Nigh, R., 2008. Trees, Fire And Farmers: Making Woods And Soil In The Maya Forest. *Journal of ethnobiology*, 28, 231-243.
- Pirot, F. & Saint Gerand, T., 2005. La Géodatabase sous ArcGIS, des fondements conceptuels à l'implémentation logicielle. *Géomatique Expert*, 41/42, 62-66.
- Raines, G.L. & Bonham-Carter, G.F., 2006. Exploratory Spatial Modelling; Demonstration for Carlin-type deposits, Central Nevada, USA, using Arc-SDM.
- Raines, G.L., Sawatzky, D. & Bonham-Carter, G.F., 2009. ArcSDM 9.3 help.
- Rose, K., 2003. *A population estimation model based on caloric requirements and land production capabilities at Tikal, Guatemala*. University of California at Santa Barbara.
- Sawatzky, D., Raines, G.L. & Bonham-Carter, G.F., *Spatial Data Modeller (SDM): ArcMAP 9.3 geoprocessing tools for spatial data modelling using weights of evidence, logistic regression, fuzzy logic and neural networks*, Available at: <http://arcscrips.esri.com/details.asp?dbid=15341> [Accessed February 15, 2009].
- SIB, 2007. *Belize: total population estimates and projections (1980-2050)*, Belmopan City: The Statistical Institute of Belize. Available at: http://www.cso.gov.bz/dms20uc/dynamicdata/docs/20071101112912_2.pdf [Accessed May 25, 2009].
- Sifuentes, J.M., 2005. *Geographical Information Systems data implementation for analyses of settlement patterns of an early agricultural society*. University of California, Santa Barbara.
- Sirjean, É., 2003. *Creation and validation of a predictive model of ancient Maya settlements with a statistical method*. École Supérieure des Géomètres et Topographes.
- Von Hagen, V.W., 1960. *World of the Maya*, Mentor: Ancient civilization.
- Webster, D. & Freter, A., 1990. The demography of Late Classic Copan. In *Precolombian population history in the Maya lowland*. Albuquerque: University of New Mexico Press, pp. 37-61.
- Wescott, K. & Kuiper, J., 2000. Using a GIS to model prehistoric site distributions in the upper Chesapeake Bay. In *Practical Applications of GIS for Archaeologists*. pp. 59-72.

- WFP, 2009. *What is hunger?*, World Food Programme of the United Nations. Available at: http://wfp.org/aboutwfp/introduction/hunger_what.asp [Accessed May 20, 2009].
- Wheatley, D. & Gillings, M., 2002. Location models and prediction. In *Spatial Technology and Archaeology, The Archaeological Applications of GIS*. London: Taylor & Francis, pp. 165-180.
- Wheeler, D.H., 2008. *The organization of agricultural production on the southwest periphery of the Maya lowlands: a settlement patterns study in the upper grijalva basin, chiapas, Mexico*. University of Pittsburgh. Available at: http://etd.library.pitt.edu/ETD/available/etd-07032008-131858/unrestricted/Wheeler_DH_ETD_2008.pdf [Accessed May 21, 2009].
- Wikipedia, 2009a. Cayo district. In *Online Encyclopedia Wikipedia*. Available at: http://en.wikipedia.org/wiki/Cayo_District [Accessed May 6, 2009].
- Wikipedia, 2009b. Geographic Information System. In *Online Encyclopedia Wikipedia*. Available at: http://en.wikipedia.org/wiki/Geographic_information_system [Accessed March 20, 2009].
- Wikipedia, 2009c. Luxembourg (pays). In *Online Encyclopedia Wikipedia*. Available at: [http://fr.wikipedia.org/wiki/Luxembourg_\(pays\)](http://fr.wikipedia.org/wiki/Luxembourg_(pays)) [Accessed May 5, 2009].
- Wikipedia, 2009d. OpenDocument Spreadsheet. In *Online Encyclopedia Wikipedia*. Available at: <http://en.wikipedia.org/wiki/OpenDocument> [Accessed April 20, 2009].
- Wikipedia, 2009e. Portable Network Graphics. In *Online Encyclopedia Wikipedia*. Available at: http://fr.wikipedia.org/wiki/Portable_Network_Graphics [Accessed April 1, 2009].
- Wikipedia, 2009f. Student's t distribution. In *Online Encyclopedia Wikipedia*. Available at: http://en.wikipedia.org/wiki/Student%27s_t-distribution [Accessed March 30, 2009].
- Wright, A.C.S. et al., 1959. *Land in British Honduras, report of the British Honduras land use survey team*, Her Majesty's Stationery Office.

Annexe 5 : localisation des zones locale et régionale

La carte ci-dessous présente les différentes échelles utilisées pour la méthode des *Weights of evidence*. Celles-ci sont définies par la taille des masques servant lors de l'application des poids.



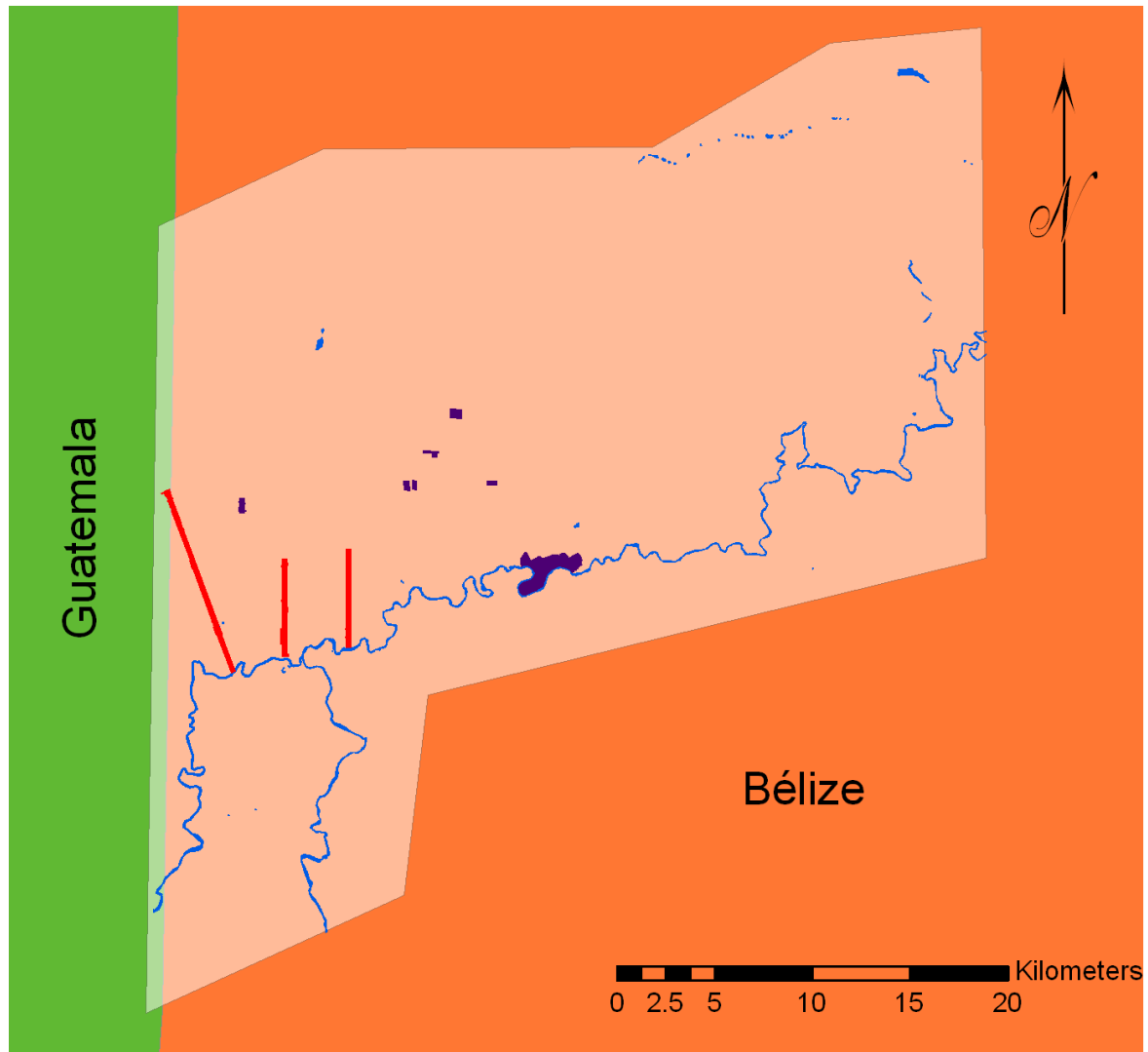
Dans le cadre de cette étude, nous parlons par souci de concision d'échelles locale, régionale, ou d'El Pilar, cela se référant en fait à la taille de ces masques. Le choix de celles-ci est expliqué au niveau de la mise en application de la méthode. La grande diagonale du masque régional mesure environ 260 km, alors que celle du masque local vaut environ 65 km et celle d'El Pilar 6 km.

Annexe 6 : masques utilisés à l'échelle locale





La carte suivante présente les composantes des masques utilisés lors des entraînements :

- **masque des transects** : composante rouge
- **masque *cover*** : composantes rouge et violette
- **masque local** : en blanc

Masques utilisés

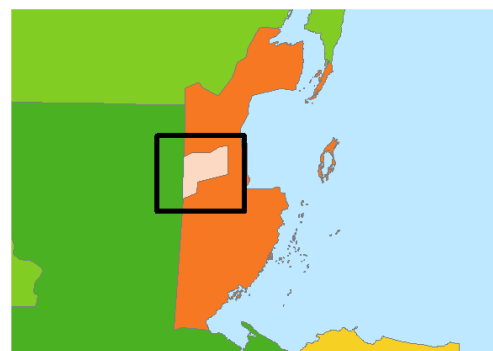


Composantes des masques

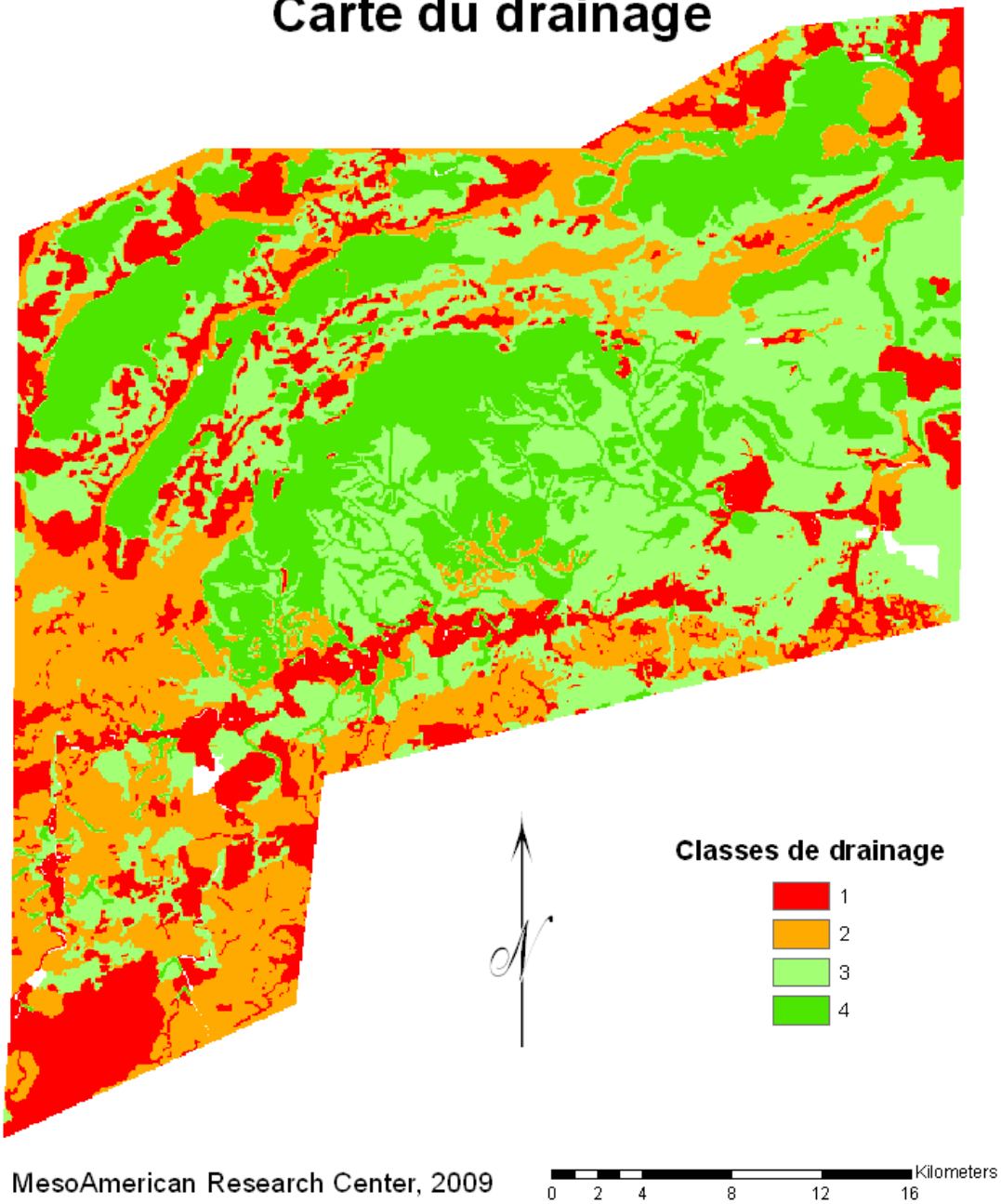
-  Rivière Bélize et lacs
-  Bacab Na; El Pilar; Yaxox
-  Barton Ramie; Beaver Dam; Cadena Creek
-  Masque local

MesoAmerican Research Center, 2009

Carte préparée par Jessica Everton et Sebeto (Sébastien Merlet)



Carte du drainage

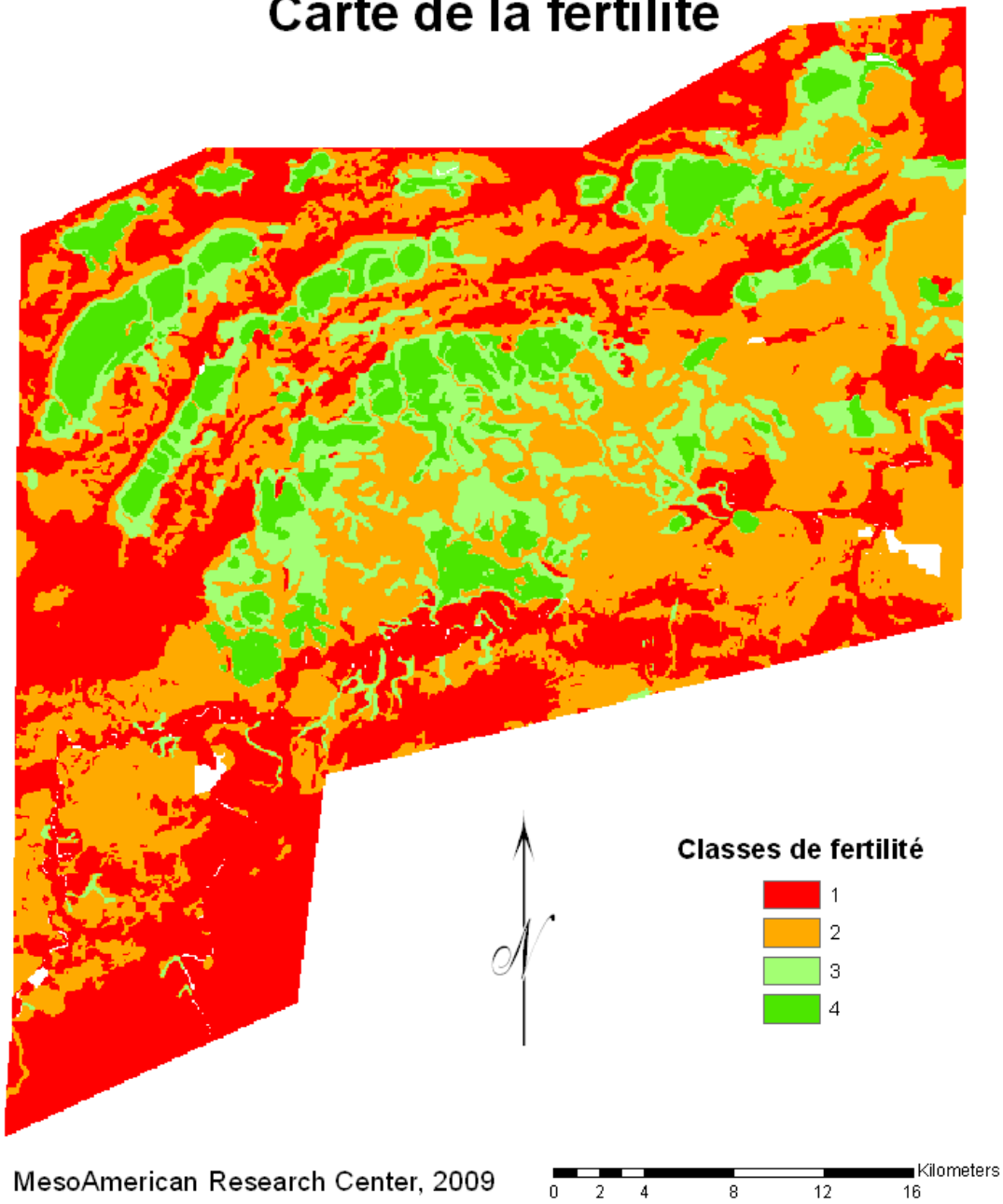


MesoAmerican Research Center, 2009

Map prepared by Sebeto (Sebastien Merlet)

0 2 4 8 12 16 Kilometers

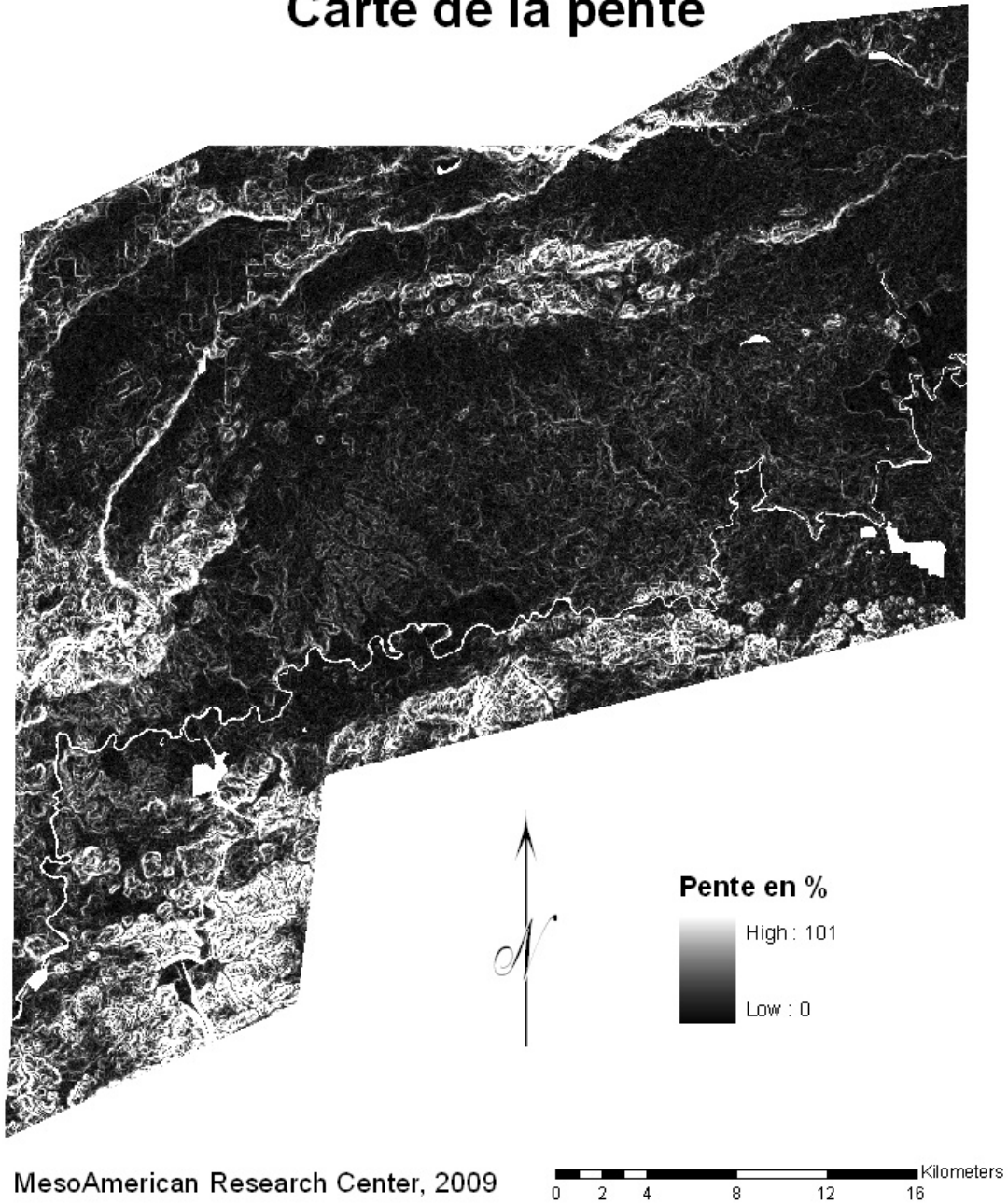
Carte de la fertilité



MesoAmerican Research Center, 2009

Map prepared by Sebeto (Sebastien Merlet)

Carte de la pente

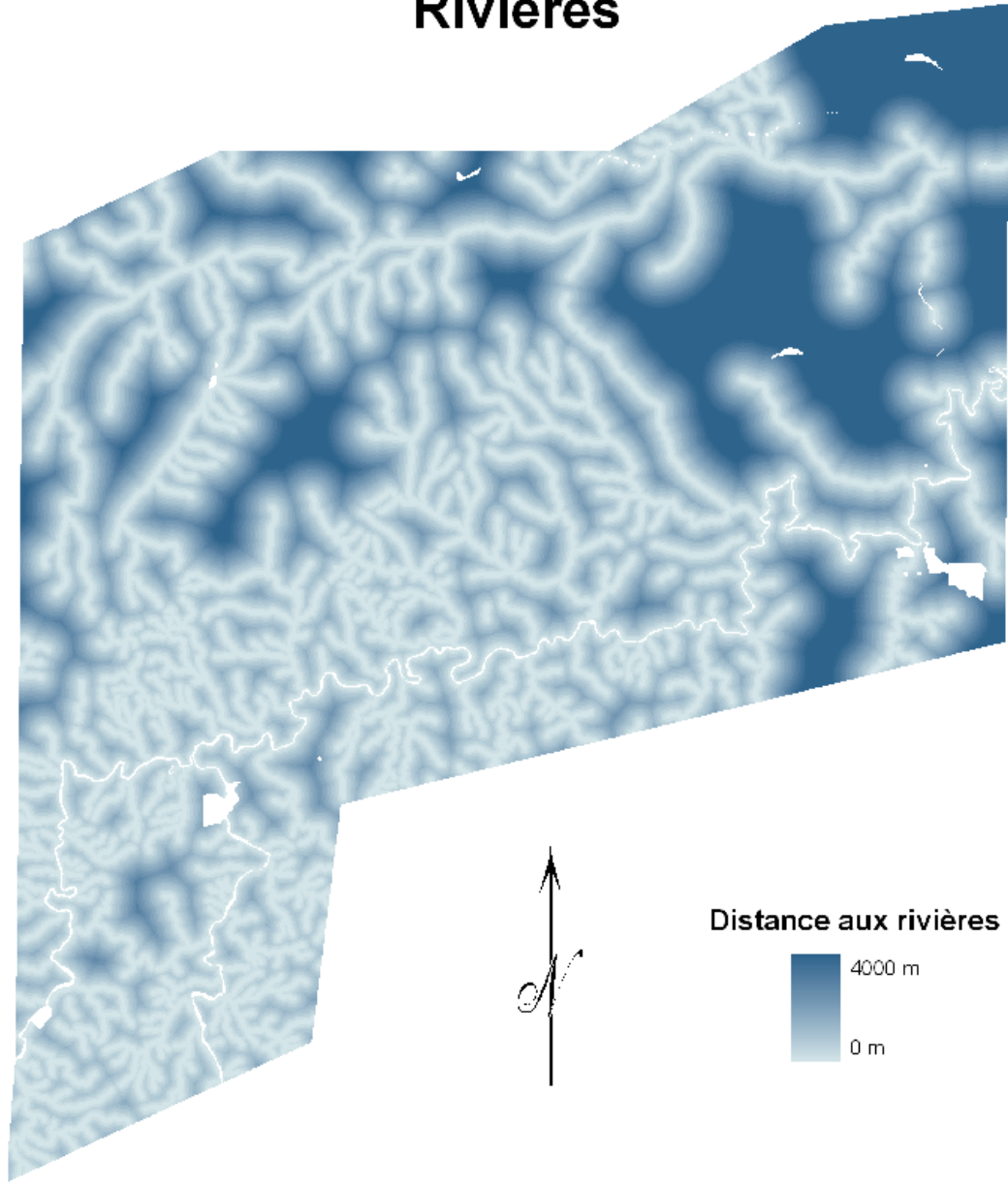


MesoAmerican Research Center, 2009

Map prepared by Sebeto (Sebastien Merlet)

Annexe 10 : carte des tampons de distance aux rivières (2004)

Rivières

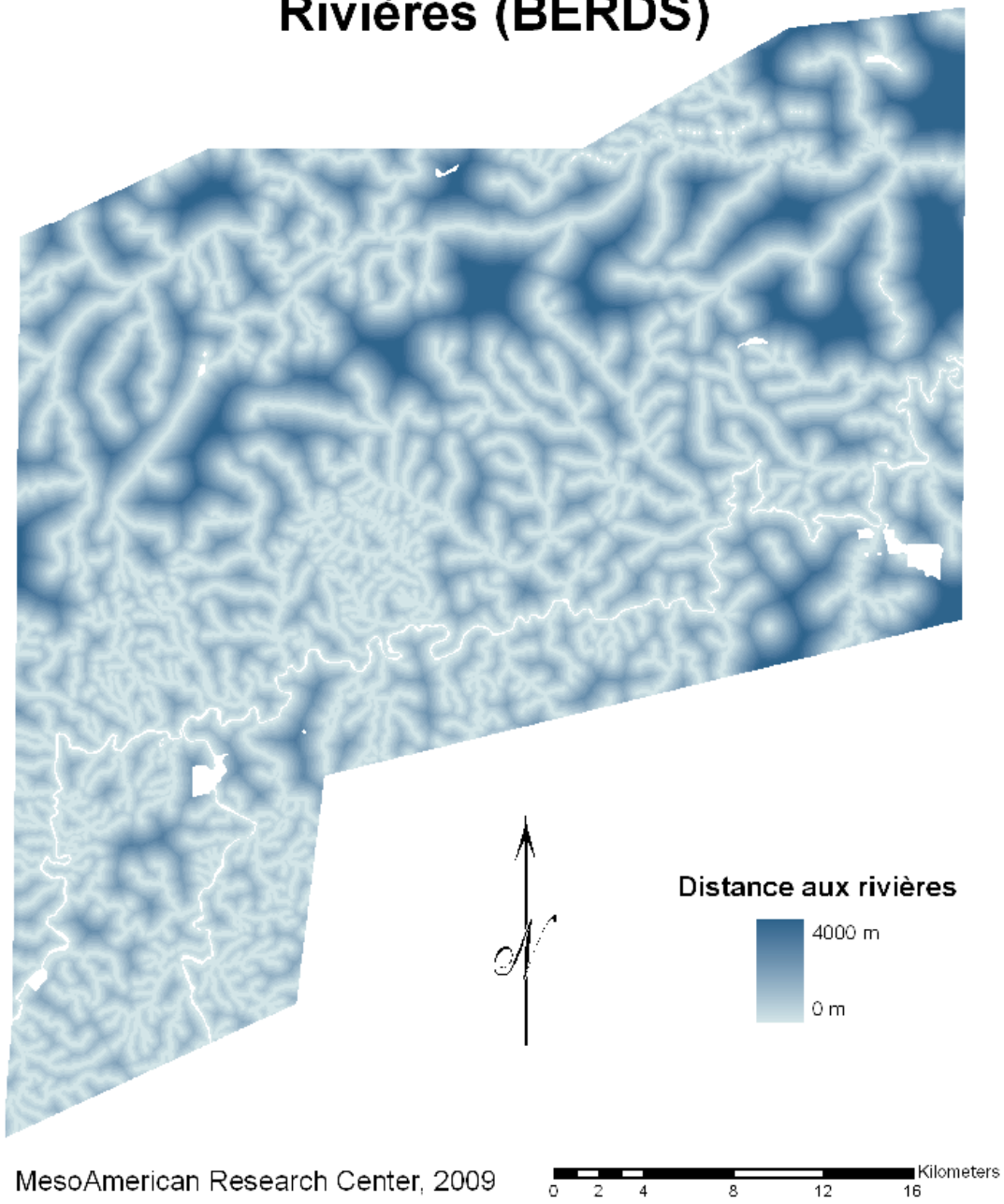


MesoAmerican Research Center, 2009
Map prepared by Sebeto (Sebastien Merlet)

0 2 4 8 12 16 Kilometers

Annexe 11 : carte des tampons de distance aux rivières (BERDS)

Rivières (BERDS)



MesoAmerican Research Center, 2009
Map prepared by Sebeto (Sebastien Merlet)

Annexe 12 : ratios calculés avant reclassification

Les tableaux suivants présentent les ratios calculés pour les différents masques. Les résultats pour la pente ne sont pas donnés ici en raison du nombre très important de classes (97), en revanche ils apparaissent pour ce paramètre reclassifié dans une autre annexe. Les distances aux rivières présentées sont celles provenant des tampons utilisés en 2004.

Drainage	Masque local		Masque des transects			Masque cover		
	Pixels	Pourcentage	Pixels	Pourcentage	Ratio	Pixels	Pourcentage	Ratio
1	663816	20.61%	2373	18.93%	0.92	9316	41.36%	2.01
2	742333	23.04%	6965	55.57%	2.41	6965	30.92%	1.34
3	1053315	32.70%	1711	13.65%	0.42	2556	11.35%	0.35
4	761894	23.65%	1485	11.85%	0.5	3689	16.38%	0.69
	3221358	100.00%	12534	100.00%	1	22526	100.00%	1

Fertilité	Masque local		Masque des transects			Masque cover		
	Pixels	Pourcentage	Pixels	Pourcentage	Ratio	Pixels	Pourcentage	Ratio
1	1261671	39.17%	6452	51.48%	1.31	13690	60.77%	1.55
2	1289128	40.02%	4494	35.85%	0.9	5016	22.27%	0.56
3	376649	11.69%	497	3.97%	0.34	2701	11.99%	1.03
4	293910	9.12%	1091	8.70%	0.95	1119	4.97%	0.54
	3221358	100.00%	12534	100.00%	1	22526	100.00%	1

Rivières	Masque local		Masque des transects			Masque cover		
	Pixels	Pourcentage	Pixels	Pourcentage	Ratio	Pixels	Pourcentage	Ratio
100	494552	15.35%	2069	16.51%	1.08	3297	14.64%	0.95
200	431747	13.40%	1877	14.98%	1.12	3558	15.80%	1.18
300	380829	11.82%	1933	15.42%	1.3	3751	16.65%	1.41
400	313649	9.74%	1929	15.39%	1.58	3397	15.08%	1.55
500	253214	7.86%	1434	11.44%	1.46	2636	11.70%	1.49
600	202360	6.28%	919	7.33%	1.17	1923	8.54%	1.36
700	161210	5.00%	656	5.23%	1.05	1347	5.98%	1.19
800	130249	4.04%	540	4.31%	1.07	739	3.28%	0.81
900	110475	3.43%	417	3.33%	0.97	427	1.90%	0.55
1000	91997	2.86%	328	2.62%	0.92	398	1.77%	0.62
1100	77793	2.41%	231	1.84%	0.76	480	2.13%	0.88
1200	67471	2.09%	161	1.28%	0.61	443	1.97%	0.94
1300	56941	1.77%	40	0.32%	0.18	130	0.58%	0.33
1400	48835	1.52%	0	0.00%	0	0	0.00%	0
1500	45431	1.41%	0	0.00%	0	0	0.00%	0
1600	40063	1.24%	0	0.00%	0	0	0.00%	0
1700	35461	1.10%	0	0.00%	0	0	0.00%	0
1800	30665	0.95%	0	0.00%	0	0	0.00%	0
1900	26815	0.83%	0	0.00%	0	0	0.00%	0
2000	23565	0.73%	0	0.00%	0	0	0.00%	0
> 2000	198036	6.15%	0	0.00%	0	0	0.00%	0
	3221358	100.00%	12534	100.00%	1	22526	100.00%	1

Annexe 13 : entraînement au moyen du masque et des sites des transects

Drainage ascending							
Classe	Sites	W+	W-	Contraste	Confiance	Généralisation	Poids
1	113	1.01	-0.43	1.44	10.51	2	0.26
2	234	0.26	-1.64	1.91	6.62	2	0.26
3	247	0.14	-8.06	8.2	0.82	1	-1.64
4	247	0	2.46	-2.46	-0.17	1	-1.64

Drainage categorical							
Classe	Sites	W+	W-	Contraste	Confiance	Généralisation	Poids
1	113	1.01	-0.43	1.44	10.51	1	1.01
2	121	-0.14	0.15	-0.29	-2.16	2	-0.14
3	13	-1	0.1	-1.11	-3.81	3	-1
4	0	0	0	0	0	99	-8.06

Fertilité ascending							
Classe	Sites	W+	W-	Contraste	Confiance	Généralisation	Poids
1	196	1.18	-1.3	2.48	11.55	2	1.18
2	247	0.29	-8.73	9.01	0.9	1	-1.3
3	247	0.19	-8.35	8.54	0.85	1	-1.3
4	247	0	0.03	-0.03	0	1	-1.3

Fertilité categorical							
Classe	Sites	W+	W-	Contraste	Confiance	Généralisation	Poids
1	196	1.18	-1.3	2.48	11.55	1	1.18
2	51	-0.9	0.47	-1.37	-6.82	2	-0.9
3	0	0	0	0	0	99	-8.73
4	0	0	0	0	0	99	-8.73

Rivières BERDS ascending							
Classe	Sites	W+	W-	Contraste	Confiance	Généralisation	Poids
100	52	0.04	-0.01	0.05	0.32	0	0
200	93	-0.03	0.02	-0.05	-0.37	0	0
300	129	-0.07	0.08	-0.14	-1.08	0	0
400	148	-0.14	0.26	-0.4	-2.94	0	0
500	165	-0.13	0.35	-0.48	-3.38	0	0
600	178	-0.1	0.34	-0.45	-2.97	0	0
700	192	-0.07	0.32	-0.4	-2.46	0	0
800	207	-0.03	0.2	-0.23	-1.26	0	0
900	212	-0.04	0.3	-0.34	-1.75	0	0
1000	223	-0.02	0.23	-0.25	-1.1	0	0
1100	238	0.01	-0.26	0.27	0.74	0	0
1200	246	0.01	-5.9	5.91	0.59	0	0
1300	246	0	-4.01	4.01	0.4	0	0
1400	246	0	-3.05	3.06	0.3	0	0
1500	246	0	2.46	-2.46	-0.17	0	0

Rivières BERDS categorical							
Classe	Sites	W+	W-	Contraste	Confiance	Généralisation	Poids
100	52	0.04	-0.01	0.05	0.32	1299	-0.01
200	41	-0.12	0.03	-0.14	-0.8	1299	-0.01
300	36	-0.15	0.03	-0.18	-0.95	1299	-0.01
400	19	-0.54	0.06	-0.6	-2.46	400	-0.54
500	17	-0.08	0.01	-0.08	-0.32	1299	-0.01
600	13	0.38	-0.02	0.4	1.31	1299	-0.01
700	14	0.41	-0.02	0.43	1.48	1299	-0.01
800	15	0.76	-0.03	0.79	2.73	800	0.76
900	5	-0.31	0.01	-0.32	-0.68	1299	-0.01
1000	11	0.46	-0.02	0.47	1.44	1299	-0.01
1100	15	0.65	-0.03	0.68	2.38	1100	0.65
1200	8	0.17	-0.01	0.17	0.46	1299	-0.01
1300	0	0	0	0	0	1299	-0.01
1400	0	0	0	0	0	1299	-0.01
1500	0	0	0	0	0	1299	-0.01

Rivières 2004 ascending							
Classe	Sites	W+	W-	Contraste	Confiance	Généralisation	Poids
100	27	-0.44	0.07	-0.51	-2.44	0	0
200	62	-0.24	0.1	-0.34	-2.24	0	0
300	98	-0.18	0.14	-0.32	-2.36	0	0
400	130	-0.18	0.25	-0.43	-3.25	0	0
500	163	-0.12	0.28	-0.4	-2.87	0	0
600	190	-0.06	0.22	-0.27	-1.73	0	0
700	202	-0.06	0.31	-0.37	-2.14	0	0
800	222	-0.01	0.08	-0.09	-0.41	0	0
900	232	0	0	0	-0.01	0	0
1000	243	0.02	-0.8	0.82	1.6	0	0
1100	246	0.01	-1.44	1.45	1.43	0	0
1200	247	0	-4.45	4.45	0.44	0	0
1300	247	0	2.46	-2.46	-0.17	0	0

Rivières 2004 categorical							
Classe	Sites	W+	W-	Contraste	Confiance	Généralisation	Poids
100	27	-0.44	0.07	-0.51	-2.44	100	-0.44
200	35	-0.06	0.01	-0.07	-0.37	1099	-0.03
300	36	-0.06	0.01	-0.07	-0.38	1099	-0.03
400	32	-0.19	0.03	-0.22	-1.1	1099	-0.03
500	33	0.17	-0.02	0.19	0.99	1099	-0.03
600	27	0.44	-0.04	0.49	2.24	600	0.44
700	12	-0.08	0	-0.08	-0.28	1099	-0.03
800	20	0.71	-0.04	0.75	2.99	800	0.71
900	10	0.22	-0.01	0.22	0.66	1099	-0.03
1000	11	0.59	-0.02	0.61	1.86	1099	-0.03
1100	3	-0.45	0.01	-0.45	-0.76	1099	-0.03
1200	1	-1.21	0.01	-1.22	-1.2	1099	-0.03
1300	0	0	0	0	0	1099	-0.03

Pente ascending							
Classe	Sites	W+	W-	Contraste	Confiance	Généralisation	Poids
0	2	-1.07	0.02	-1.08	-1.5	2	0.03
1	18	-0.28	0.03	-0.3	-1.2	2	0.03
2	42	-0.05	0.01	-0.07	-0.38	2	0.03
3	67	-0.01	0	-0.02	-0.11	2	0.03
4	85	-0.03	0.01	-0.04	-0.28	2	0.03
5	113	0.08	-0.06	0.14	1.03	2	0.03
6	128	0.07	-0.07	0.14	1.04	2	0.03
7	141	0.06	-0.08	0.14	1.05	2	0.03
8	158	0.09	-0.14	0.22	1.62	2	0.03
9	167	0.08	-0.15	0.23	1.63	2	0.03
10	173	0.06	-0.13	0.2	1.37	2	0.03
11	183	0.08	-0.2	0.28	1.87	2	0.03
12	191	0.09	-0.25	0.34	2.18	2	0.03
13	198	0.09	-0.31	0.4	2.42	2	0.03
14	202	0.08	-0.3	0.39	2.26	2	0.03
15	213	0.11	-0.5	0.61	3.23	2	0.03
16	217	0.11	-0.56	0.67	3.37	2	0.03
17	220	0.1	-0.58	0.68	3.25	2	0.03
18	221	0.09	-0.53	0.62	2.89	2	0.03
19	224	0.09	-0.58	0.67	2.98	2	0.03
20	227	0.08	-0.63	0.72	2.99	2	0.03
21	231	0.09	-0.78	0.86	3.27	2	0.03
22	233	0.08	-0.84	0.92	3.27	2	0.03
23	235	0.08	-0.89	0.96	3.2	2	0.03
24	235	0.07	-0.8	0.87	2.88	2	0.03
25	236	0.06	-0.78	0.84	2.66	2	0.03
26	236	0.04	-0.65	0.7	2.2	2	0.03
27	238	0.04	-0.69	0.73	2.09	2	0.03
28	241	0.04	-1	1.04	2.48	2	0.03
29	243	0.04	-1.27	1.31	2.57	2	0.03
30	244	0.04	-1.43	1.47	2.51	2	0.03
31	244	0.03	-1.33	1.36	2.32	2	0.03
32	245	0.03	-1.65	1.68	2.35	2	0.03
33	245	0.03	-1.49	1.52	2.12	1	-1.65
34	245	0.03	-1.41	1.44	2	1	-1.65
35	246	0.03	-1.97	2	1.98	1	-1.65
36	246	0.02	-1.87	1.89	1.88	1	-1.65
37	246	0.02	-1.78	1.8	1.79	1	-1.65
38	246	0.02	-1.68	1.7	1.68	1	-1.65
39	246	0.02	-1.55	1.56	1.54	1	-1.65
40	246	0.01	-1.39	1.41	1.39	1	-1.65
41	246	0.01	-1.31	1.32	1.3	1	-1.65
42	247	0.01	-5.87	5.89	0.99	1	-1.65
43	247	0.01	-5.77	5.78	0.98	1	-1.65
44	247	0.01	-5.63	5.64	0.96	1	-1.65
45	247	0.01	-5.57	5.58	0.96	1	-1.65
46	247	0.01	-5.51	5.52	0.95	1	-1.65
47	247	0.01	-5.46	5.47	0.95	1	-1.65
48	247	0.01	-5.41	5.41	0.94	1	-1.65
49	247	0.01	-5.29	5.3	0.93	1	-1.65
50	247	0.01	-5.18	5.19	0.92	1	-1.65
51	247	0.01	-5.08	5.08	0.91	1	-1.65
52	247	0.01	-4.92	4.93	0.89	1	-1.65
53	247	0.01	-4.89	4.89	0.89	1	-1.65
54	247	0	-4.73	4.74	0.87	1	-1.65
55	247	0	-4.61	4.61	0.86	1	-1.65
56	247	0	-4.26	4.26	0.83	1	-1.65
57	247	0	-4.13	4.13	0.81	1	-1.65
58	247	0	-4.09	4.09	0.81	1	-1.65
60	247	0	-3.85	3.85	0.78	1	-1.65
61	247	0	-3.47	3.47	0.73	1	-1.65
62	247	0	-2.55	2.55	0.25	1	-1.65
63	247	0	-2.36	2.36	0.24	1	-1.65
64	247	0	-1.85	1.85	0.18	1	-1.65
66	247	0	-0.72	0.72	0.07	1	-1.65
67	247	0	2.46	-2.46	-0.17	1	-1.65

Pente categorial							
Classe	Sites	W+	W-	Contraste	Confiance	Généralisation	Poids
0	2	-1.07	0.02	-1.08	-1.5	99	-0.08
1	16	-0.12	0.01	-0.13	-0.47	99	-0.08
2	24	0.15	-0.02	0.17	0.75	99	-0.08
3	25	0.06	-0.01	0.07	0.32	99	-0.08
4	18	-0.07	0.01	-0.08	-0.31	99	-0.08
5	28	0.47	-0.05	0.52	2.44	5	0.47
6	15	0.01	0	0.01	0.03	99	-0.08
7	13	0.01	0	0.01	0.03	99	-0.08
8	17	0.3	-0.02	0.32	1.22	99	-0.08
9	9	-0.01	0	-0.01	-0.04	99	-0.08
10	6	-0.32	0.01	-0.33	-0.78	99	-0.08
11	10	0.44	-0.01	0.45	1.32	99	-0.08
12	8	0.29	-0.01	0.3	0.79	99	-0.08
13	7	0.21	-0.01	0.22	0.53	99	-0.08
14	4	-0.33	0.01	-0.34	-0.66	99	-0.08
15	11	0.84	-0.03	0.87	2.58	15	0.84
16	4	0.13	0	0.13	0.25	99	-0.08
17	3	-0.42	0.01	-0.43	-0.71	99	-0.08
18	1	-1.39	0.01	-1.4	-1.38	99	-0.08
19	3	0.04	0	0.05	0.07	99	-0.08
20	3	-0.18	0	-0.18	-0.3	99	-0.08
21	4	0.34	0	0.34	0.65	99	-0.08
22	2	-0.26	0	-0.26	-0.36	99	-0.08
23	2	-0.44	0	-0.44	-0.61	99	-0.08
24	0	0	0	0	0	99	-0.08
25	1	-1.03	0.01	-1.04	-1.02	99	-0.08
26	0	0	0	0	0	99	-0.08
27	2	-0.47	0	-0.47	-0.65	99	-0.08
28	3	0.59	-0.01	0.59	0.95	99	-0.08
29	2	-0.03	0	-0.03	-0.04	99	-0.08
30	1	-0.49	0	-0.49	-0.48	99	-0.08
31	0	0	0	0	0	99	-0.08
32	1	0.08	0	0.08	0.08	99	-0.08
33	0	0	0	0	0	99	-0.08
34	0	0	0	0	0	99	-0.08
35	1	-0.01	0	-0.01	-0.01	99	-0.08
36	0	0	0	0	0	99	-0.08
37	0	0	0	0	0	99	-0.08
38	0	0	0	0	0	99	-0.08
39	0	0	0	0	0	99	-0.08
40	0	0	0	0	0	99	-0.08
41	0	0	0	0	0	99	-0.08
42	1	1.9	0	1.9	1.51	99	-0.08
43	0	0	0	0	0	99	-0.08
44	0	0	0	0	0	99	-0.08
45	0	0	0	0	0	99	-0.08
46	0	0	0	0	0	99	-0.08
47	0	0	0	0	0	99	-0.08
48	0	0	0	0	0	99	-0.08
49	0	0	0	0	0	99	-0.08
50	0	0	0	0	0	99	-0.08
51	0	0	0	0	0	99	-0.08
52	0	0	0	0	0	99	-0.08
53	0	0	0	0	0	99	-0.08
54	0	0	0	0	0	99	-0.08
55	0	0	0	0	0	99	-0.08
56	0	0	0	0	0	99	-0.08
57	0	0	0	0	0	99	-0.08
58	0	0	0	0	0	99	-0.08
60	0	0	0	0	0	99	-0.08
61	0	0	0	0	0	99	-0.08
62	0	0	0	0	0	99	-0.08
63	0	0	0	0	0	99	-0.08
64	0	0	0	0	0	99	-0.08
66	0	0	0	0	0	99	-0.08
67	0	0	0	0	0	99	-0.08

Annexe 14 : poids pour les rivières, sites des transects avec masque local

Rivières BERDS ascending							
Classe	Sites	W+	W-	Contraste	Confiance	Généralisation	Poids
100	53	0.18	-0.04	0.22	1.41	2	0.04
200	94	0.1	-0.06	0.16	1.25	2	0.04
300	130	0.09	-0.09	0.18	1.38	2	0.04
400	149	0.01	-0.02	0.03	0.26	2	0.04
500	166	-0.02	0.04	-0.06	-0.47	2	0.04
600	179	-0.04	0.12	-0.17	-1.16	2	0.04
700	193	-0.04	0.14	-0.18	-1.15	2	0.04
800	208	-0.01	0.07	-0.08	-0.45	2	0.04
900	213	-0.03	0.18	-0.21	-1.14	2	0.04
1000	224	0	0.05	-0.05	-0.23	2	0.04
1100	239	0.04	-0.76	0.8	2.22	2	0.04
1200	247	0.05	-7.18	7.24	0.72	1	-0.76
1300	247	0.04	-6.91	6.95	0.69	1	-0.76
1400	247	0.03	-6.61	6.64	0.66	1	-0.76
1500	247	0.02	-6.3	6.33	0.63	1	-0.76
1600	247	0.02	-6	6.01	0.6	1	-0.76
1700	247	0.01	-5.7	5.72	0.57	1	-0.76
4000	247	0	8.09	-8.09	-0.57	1	-0.76

Rivières 2004 ascending							
Classe	Sites	W+	W-	Contraste	Confiance	Généralisation	Poids
100	27	-0.34	0.05	-0.39	-1.92	2	0.19
200	62	-0.14	0.05	-0.19	-1.27	2	0.19
300	98	-0.02	0.02	-0.04	-0.29	2	0.19
400	130	0.05	-0.05	0.09	0.73	2	0.19
500	163	0.13	-0.21	0.33	2.48	2	0.19
600	190	0.18	-0.43	0.61	4.03	2	0.19
700	202	0.16	-0.52	0.68	4.13	2	0.19
800	222	0.2	-0.96	1.16	5.52	2	0.19
900	232	0.2	-1.33	1.53	5.76	2	0.19
1000	243	0.21	-2.52	2.73	5.42	2	0.19
1100	246	0.19	-3.78	3.98	3.97	2	0.19
1200	247	0.17	-8.26	8.43	0.84	1	-3.78
1300	247	0.15	-8.14	8.29	0.83	1	-3.78
1400	247	0.13	-8.03	8.16	0.82	1	-3.78
1500	247	0.12	-7.91	8.02	0.8	1	-3.78
1600	247	0.1	-7.79	7.89	0.79	1	-3.78
1700	247	0.09	-7.67	7.76	0.78	1	-3.78
1800	247	0.08	-7.55	7.63	0.76	1	-3.78
1900	247	0.07	-7.44	7.51	0.75	1	-3.78
2000	247	0.06	-7.33	7.39	0.74	1	-3.78
2001	247	0	8.09	-8.09	-0.57	1	-3.78

Annexe 15 : test d'indépendance conditionnelle (local)

Overall CI: 1.8%

Conditional Independence Test: w

Observed No. training pts, n = 954

Expected No. of training points, T = 812.6

Difference, T-n = -141.4

Standard Deviation of T = 59.597

Conditional Independence Ratio: 1.17 <simply the ratio n/T>

Values below 1.00 may indicate conditional dependence

among two or more of your data sets. <Bonham-Carter(1994,ch.9)
suggest that values <0.85 may indicate a problem>

Agterberg & Cheng Conditional Independence Test

<See Agterberg and Cheng, Natural Resources Research 11(4), 249-255, 2002>

This is a one-tailed test of the null hypothesis that T-n=0. The test statistic is (T-n)/standard deviation of T. Probability values greater than 95% or 99% indicate that the hypothesis of CI should be rejected, but any value greater than 50% indicates that some conditional dependence occurs>

Probability that this model is not conditionally independent with
(T-n)/Tstd = -2.373191 is 0.9%

Input Data:

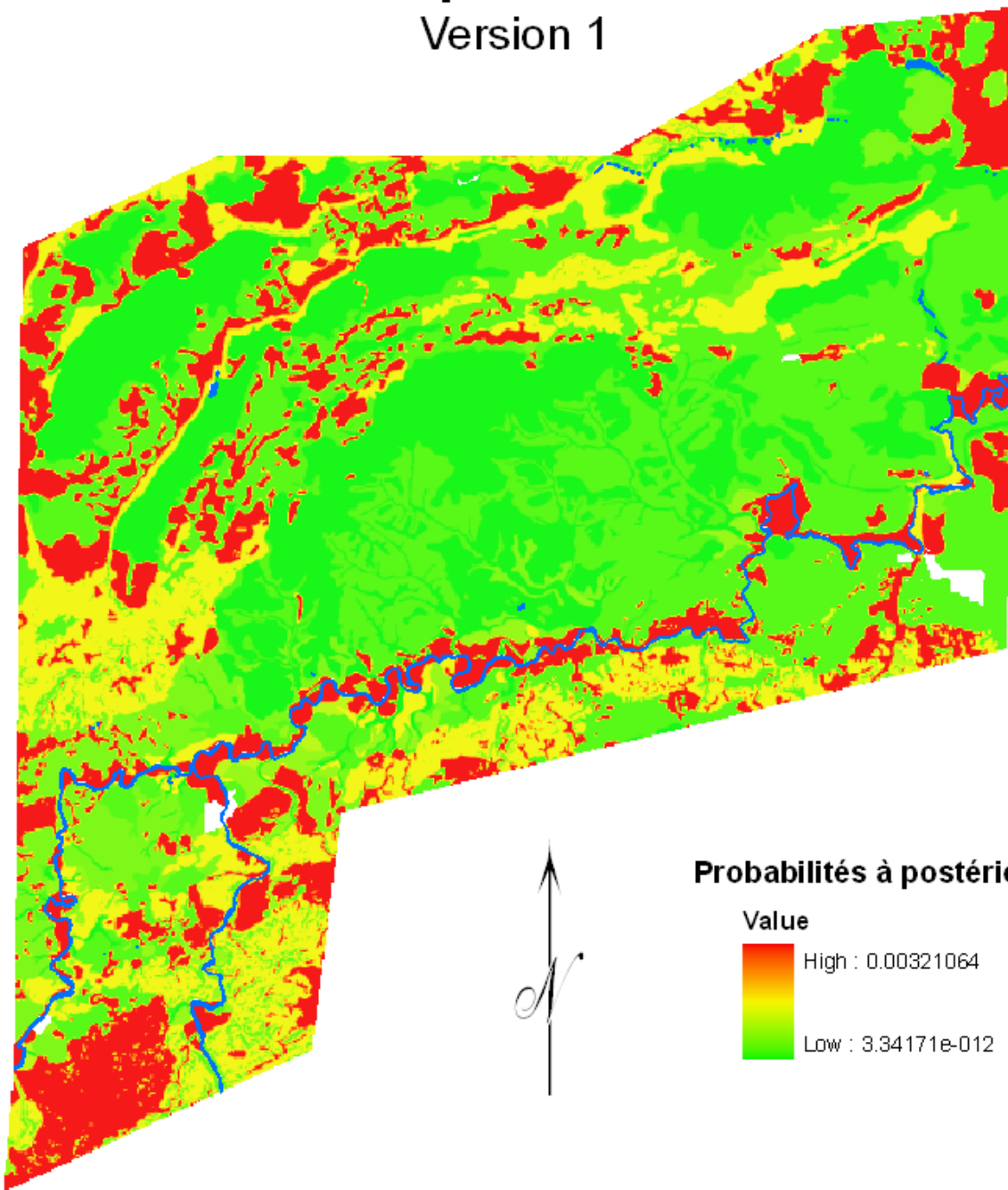
Post Probability: w_pprb

Post Probability Std Deviation: w_std

Training Sites: s_a_all

Modèle prédictif 2009

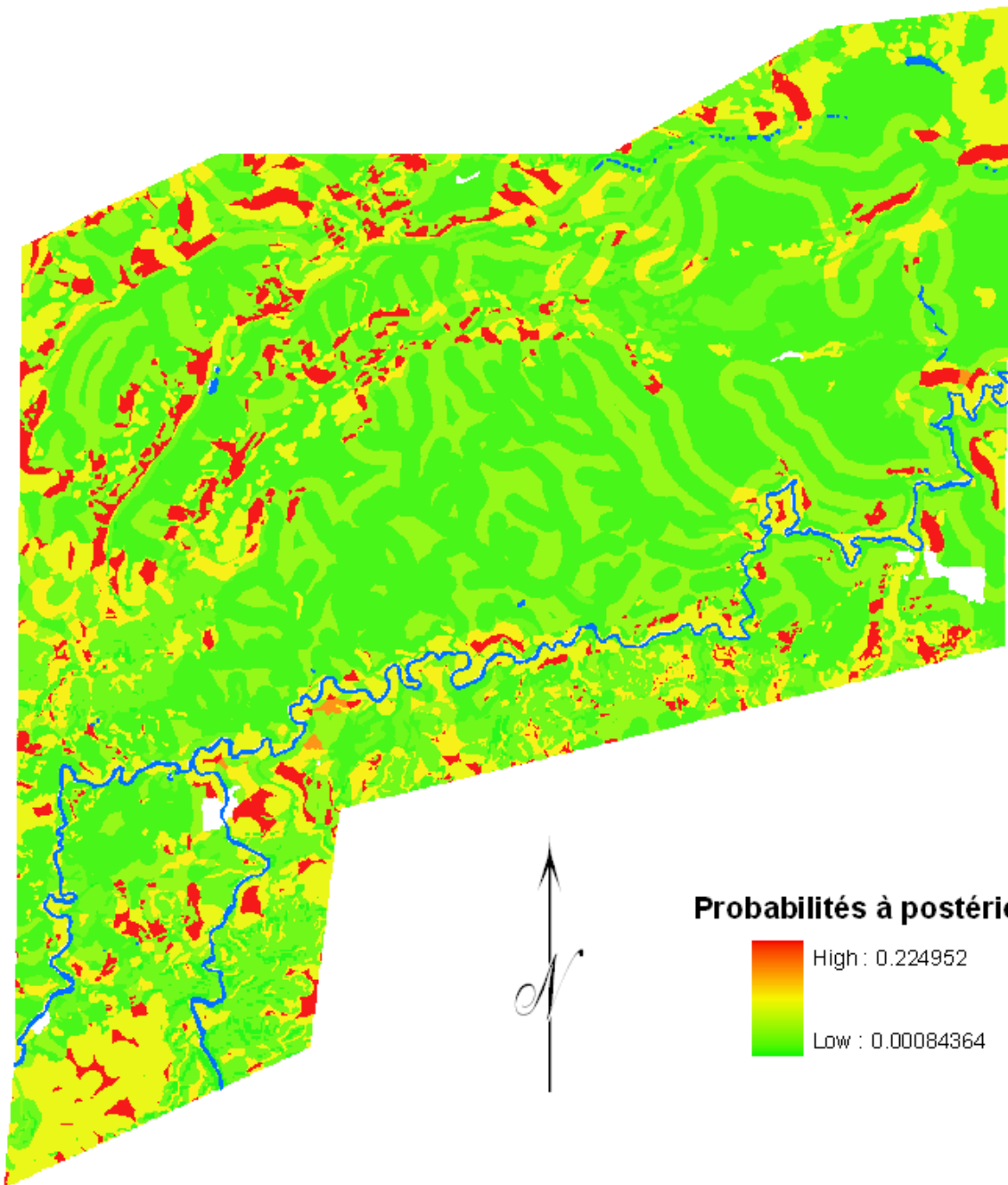
Version 1



MesoAmerican Research Center, 2009
Map prepared by Sebeto (Sebastien Merlet)

0 1.5 3 6 9 12 Kilometers

Modèle prédictif 2004



MesoAmerican Research Center, 2009

Map prepared by Sebeto (Sebastien Merlet)

0 1.5 3 6 9 12 Kilometers

Annexe 18 : tables du modèle de population (première version)

% d'URO late classic	95%
----------------------	-----

Habitants par URO	5.6
-------------------	-----

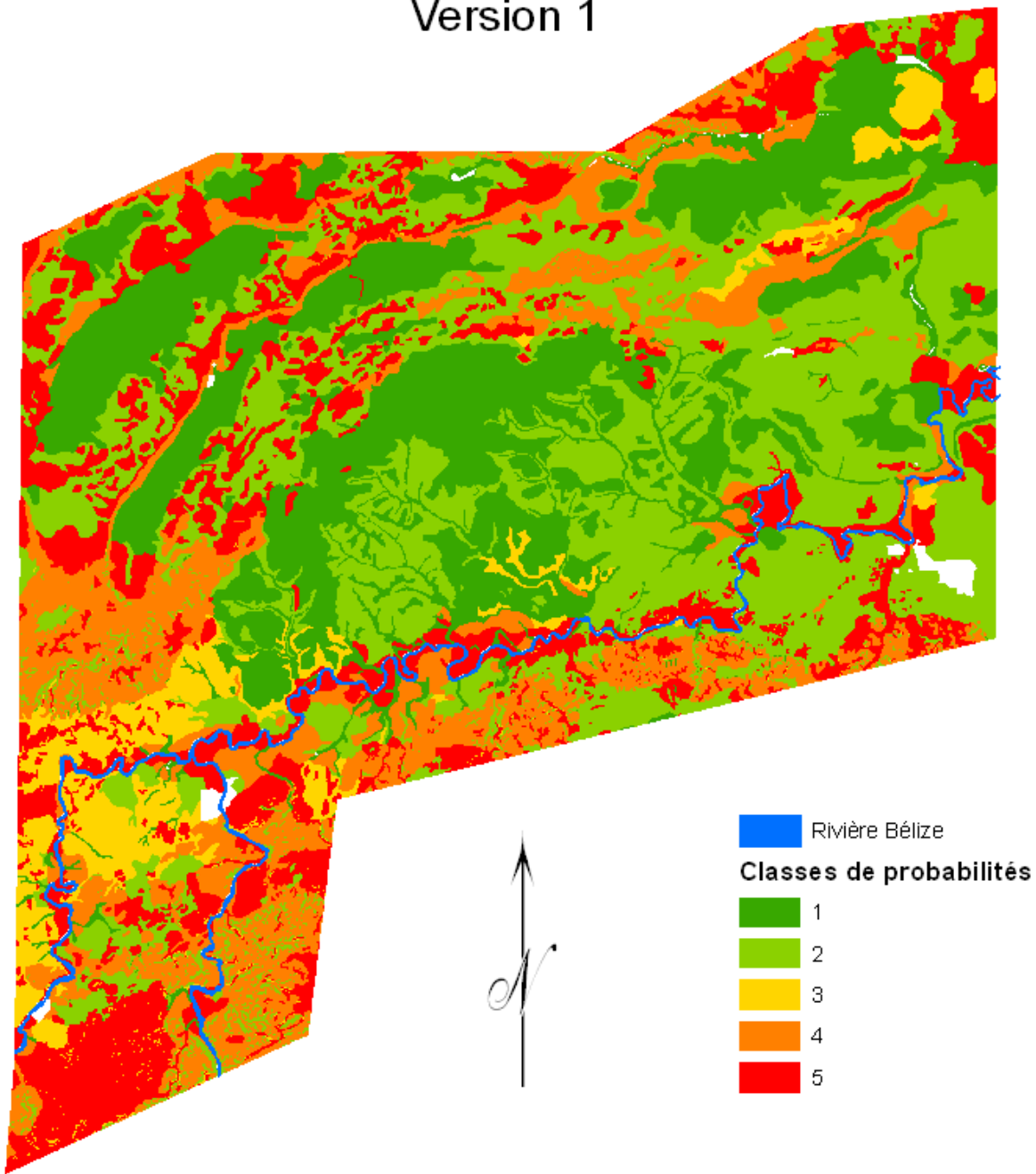
Répartition des sites (transects)					Rangs des sites					
Proba	URO	Superficie (km ²)	Population	Densité	0	1	2	3	4	5
0	0	1.5484	0	0	0	0	0	0	0	0
4	0	0.0012	0	0	0	0	0	0	0	0
9	0	0.0040	0	0	0	0	0	0	0	0
13										
19										
23	12	0.8920	64	72	11	3	8	1	0	0
28	1	0.1840	5	29	4	1	0	0	0	0
51	28	1.0888	149	137	23	2	26	0	0	0
59	0	0.0048	0	0	0	0	0	0	0	0
71	5	0.1180	27	225	4	1	3	1	0	0
105										
156	51	1.4476	271	187	54	14	28	9	0	0
321	273	3.7216	1,452	390	162	19	224	25	3	2
	370	9.0104	1,968	218	258	40	289	36	3	2

Totalité des transects						
Classe	URO <i>late classic</i>	Superficie (km ²)	Population	Densité	% population	% de la zone
1	0	1.5536	0	0	0%	17%
2	12	1.0760	69	64.28	4%	12%
3	27	1.0936	149	136.21	8%	12%
4	53	1.5656	298	190.29	15%	17%
5	259	3.7216	1,452	390.25	74%	41%
Total	352	9.0104	1,968	218.46	100%	100%

Totalité de la zone d'étude						
Classe	URO <i>late classic</i>	Superficie (km ²)	Population	Densité	% population	% de la zone
1	0	311.8144	0	0	0%	24%
2	4,809	418.9492	26,928	64.28	15%	33%
3	1,813	74.5428	10,154	136.21	6%	6%
4	7,681	226.0332	43,012	190.29	24%	18%
5	17,924	257.2036	100,374	390.25	56%	20%
Total	32,226	1,288.5432	180,468	140.06	100%	100%

Probabilités reclassifiées

Version 1



MesoAmerican Research Center, 2009

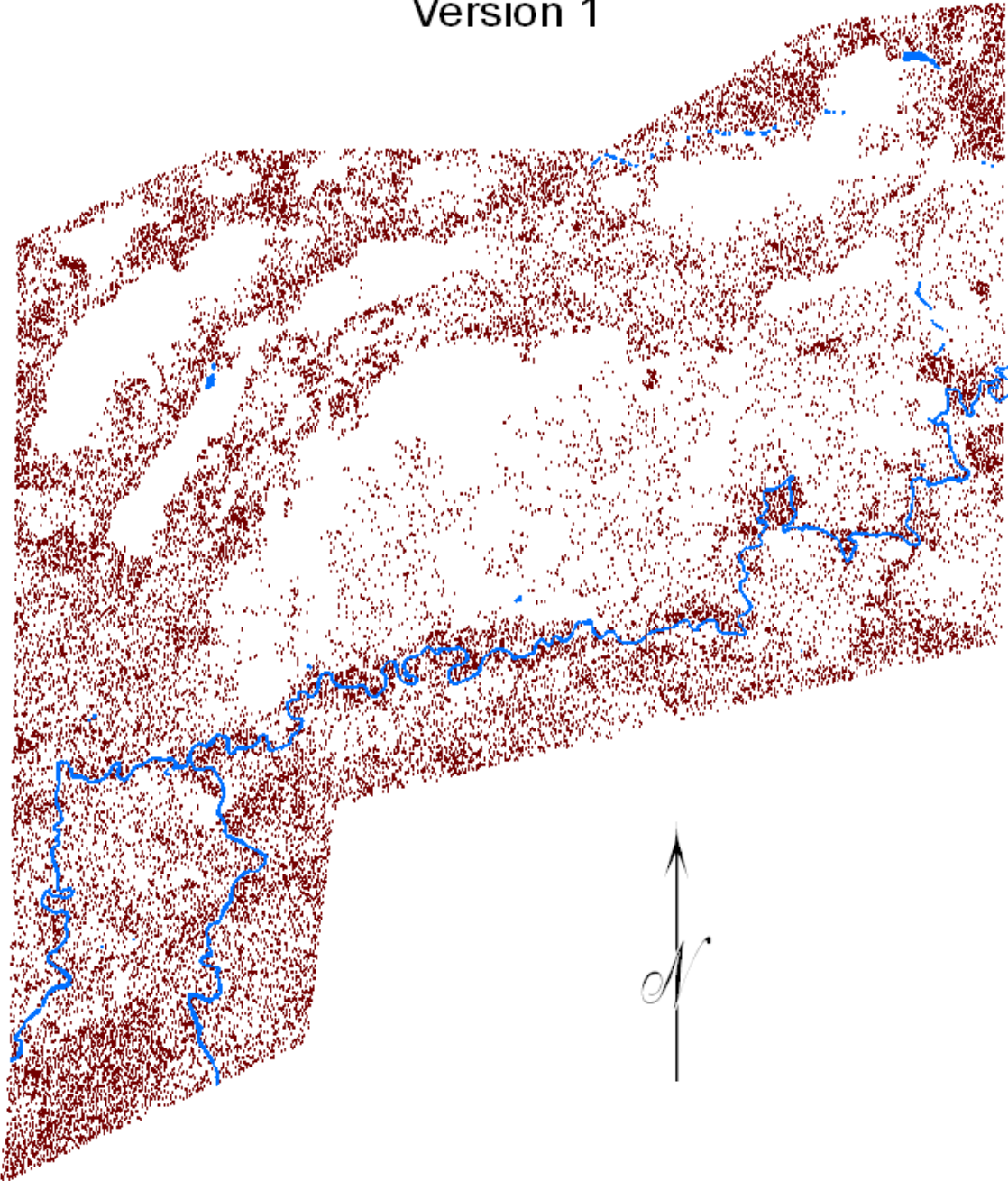
Map prepared by Sebeto (Sebastien Merlet)

0 1.5 3 6 9 12 Kilometers

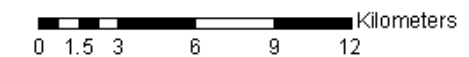
Annexe 20 : modèle de propagation des sites (première version)

Chaque point correspond à une unité résidentielle maya propagée au moyen des densités.

Répartition des sites Version 1



MesoAmerican Research Center, 2009
Map prepared by Sebeto (Sebastien Merlet)

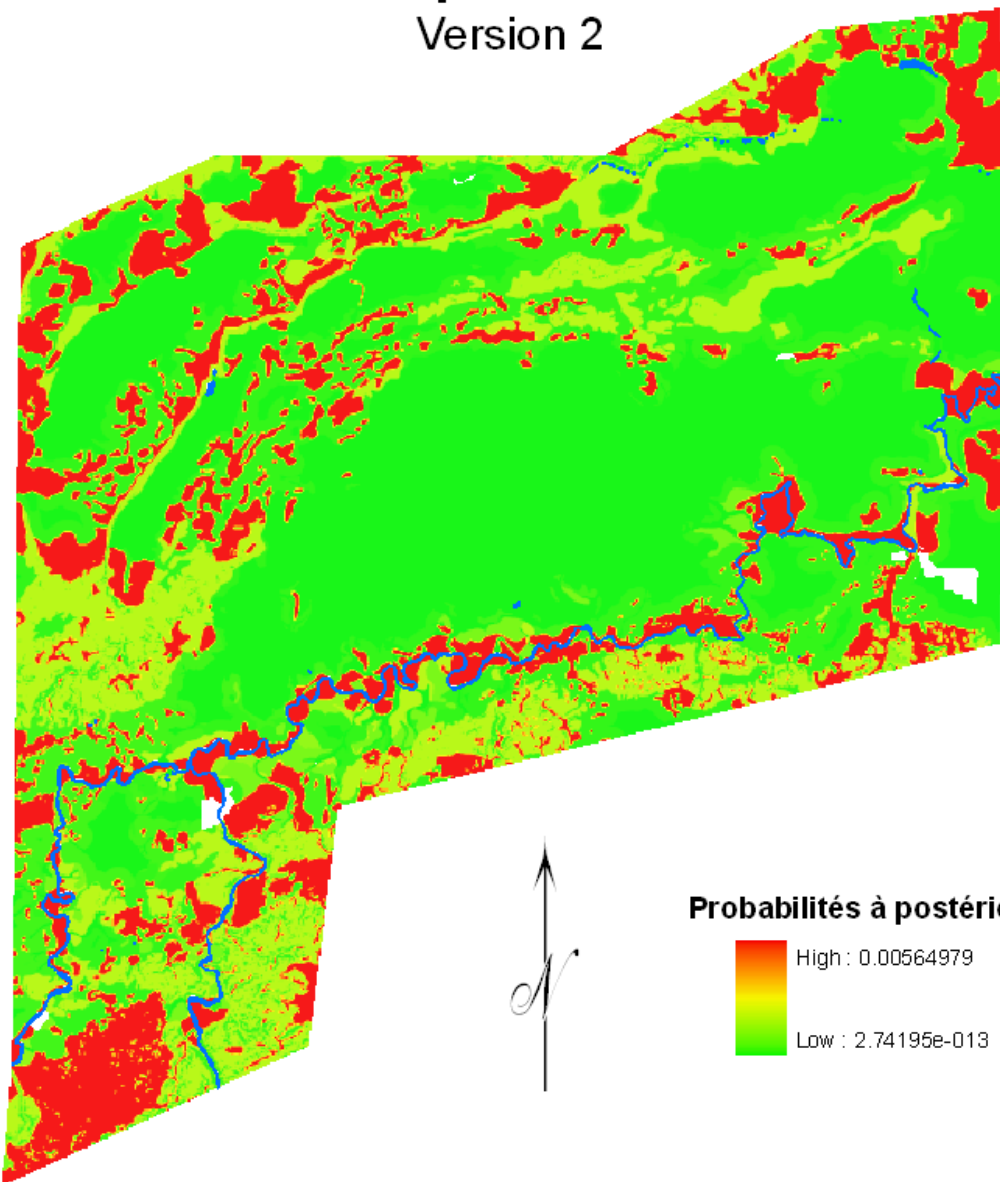


Annexe 21 : poids et probabilités au niveau local (seconde version)

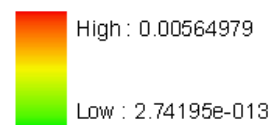
Attractivité des zones de forte densité (categorical)							
Classe	Sites	W+	W-	Contraste	Confiance	Poids	
1	307	0.57	-0.66	1.23	11.66	0.57	
2	88	0.14	-0.03	0.17	1.35	0.14	
3	52	-0.93	0.22	-1.15	-7.62	-0.93	
4	1	-4.19	0.16	-4.35	-4.34	-4.19	

Modèle prédictif 2009

Version 2



Probabilités à postériori



MesoAmerican Research Center, 2009

Map prepared by Sebeto (Sebastien Merlet)

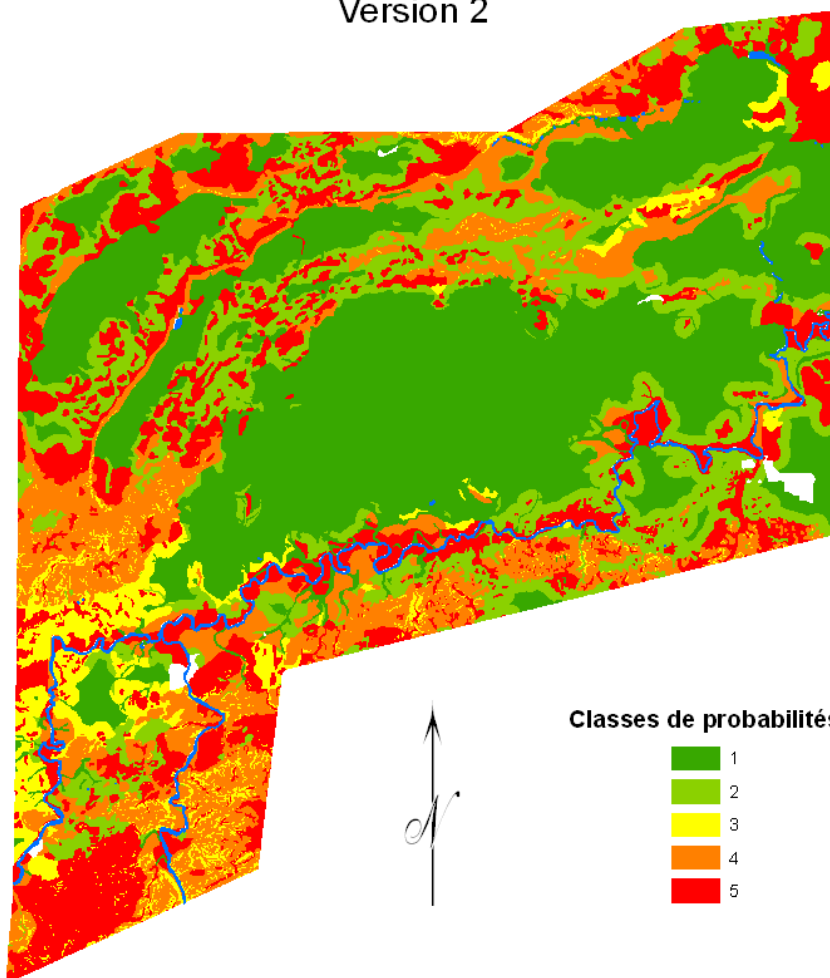
0 1.5 3 6 9 12 Kilometers

Annexe 22 : probabilités reclassifiées (seconde version)

Totalité des transects						
Classe	URO <i>late classic</i>	Superficie (km ²)	Population	Densité	% population	% de la zone
1	0	2.0116	0	0	0%	22%
2	11	0.5128	64	124.49	3%	6%
3	28	1.1988	154	128.7	8%	13%
4	53	1.5656	298	190.29	15%	17%
5	259	3.7216	1,452	390.25	74%	41%
Total	352	9.0104	1,968	218.46	100%	100%

Totalité de la zone d'étude						
Classe	URO <i>late classic</i>	Superficie (km ²)	Population	Densité	% population	% de la zone
1	0	484.6748	0	0	0%	38%
2	5,403	243.0280	30,255	124.49	17%	19%
3	1,753	76.2924	9,818	128.7	5%	6%
4	7,643	224.9168	42,800	190.29	23%	18%
5	17,808	255.5448	99,727	390.25	55%	20%
Total	32,607	1,284.4568	182,600	142.16	100%	100%

Probabilités reclassifiées Version 2



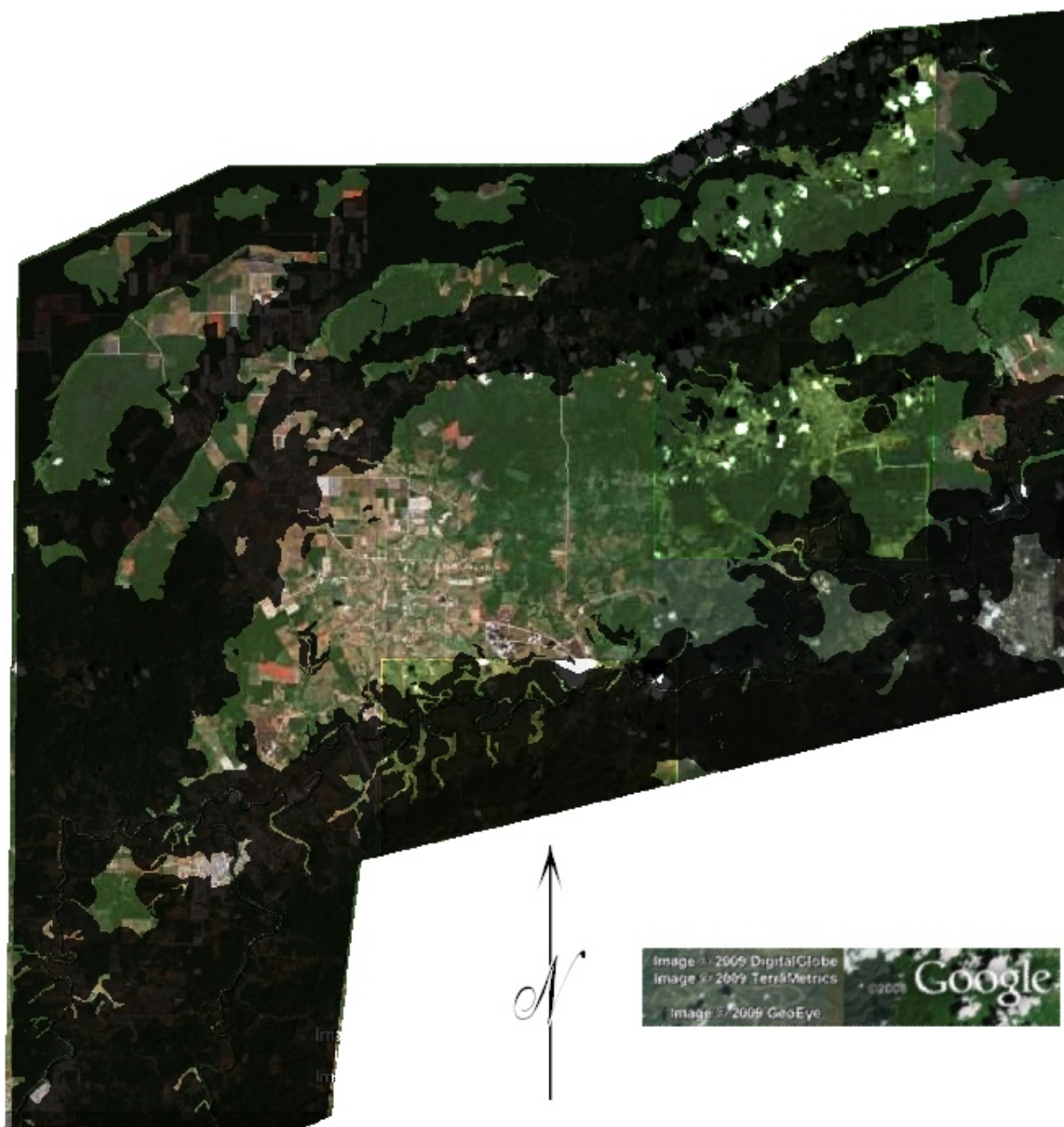
MesoAmerican Research Center, 2009
Map prepared by Sebeto (Sebastien Merlet)

0 1.5 3 6 9 12 Kilometers

Annexe 23 : image aérienne d'utilisation des sols

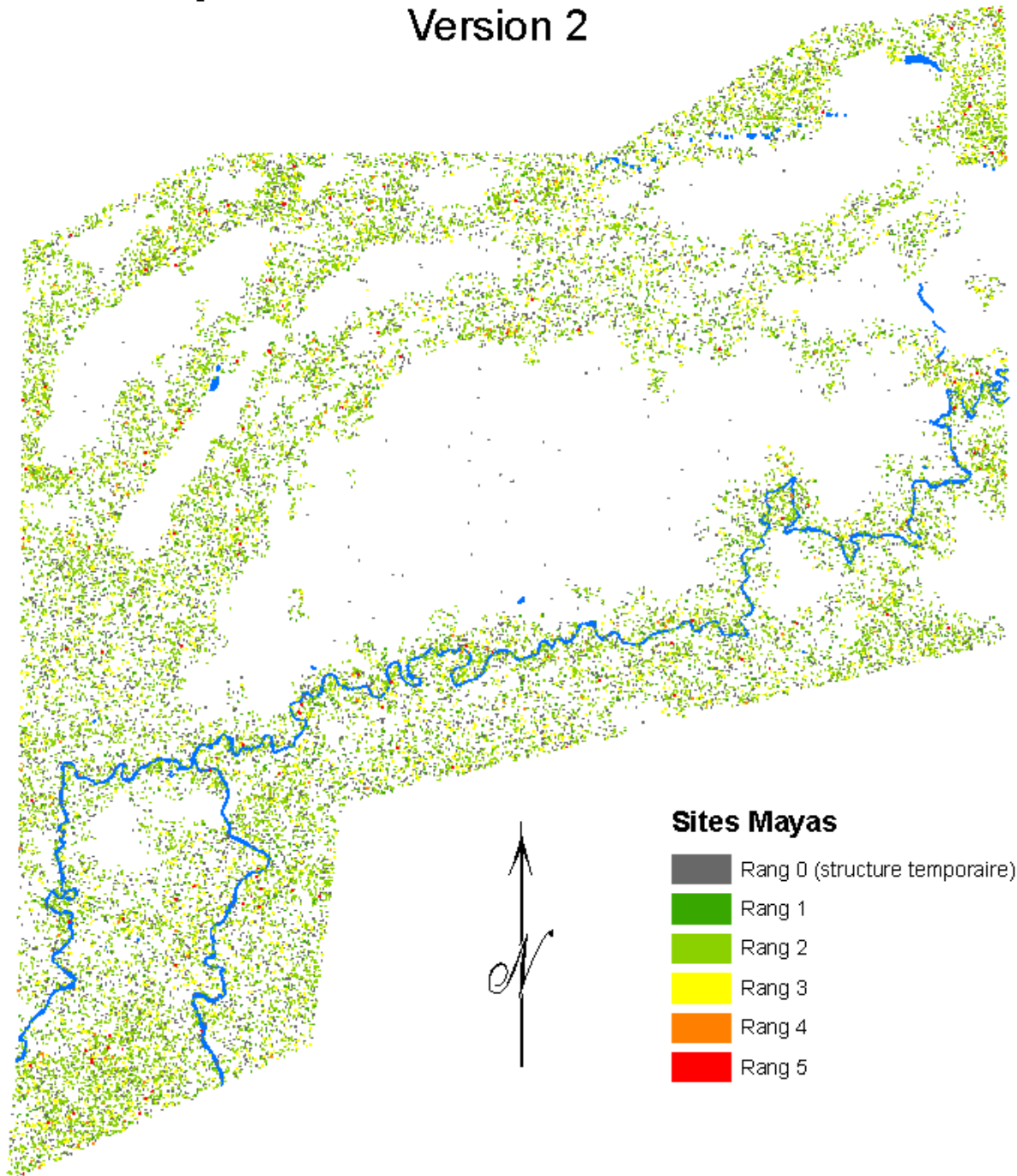
Un masque représentant les zones d'occupation résidentielle des Mayas (classes de probabilités 2 à 5) est superposée à une image aérienne issue du logiciel GoogleEarth et fait apparaître des différences importantes par rapport au présent : les zones apparentes des photos aériennes n'étaient pas utilisées par les mayas à part quelques structures temporaires, mais servent actuellement à l'agriculture.

Utilisation actuelle des sols



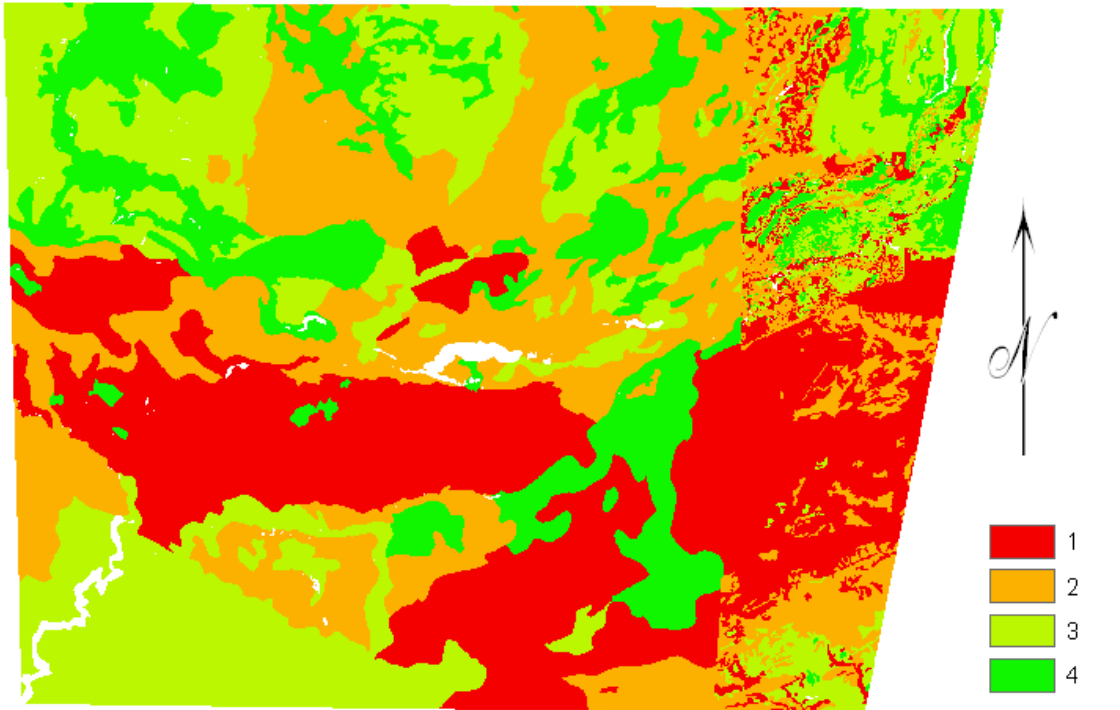
MesoAmerican Research Center, 2009
Map prepared by Sebeto (Sebastien Merlet)

Répartition aléatoire des sites Version 2



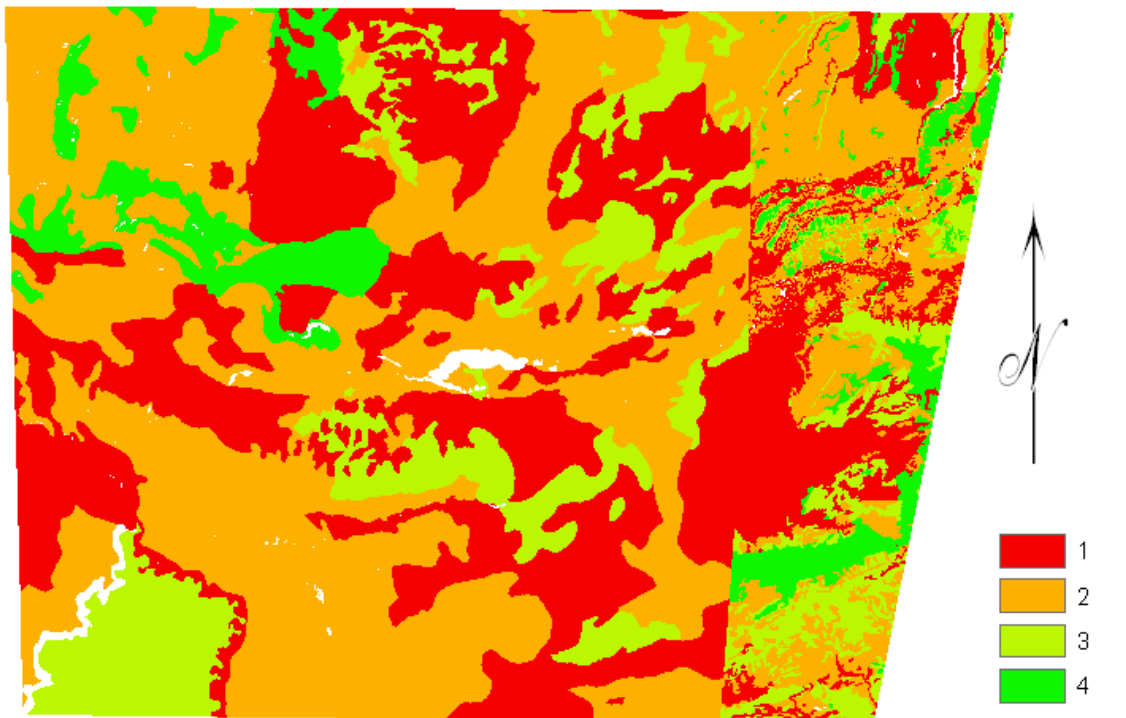
MesoAmerican Research Center, 2009
Carte préparée par Sebeto (Sebastien Merlet)

Drainage pour la zone régionale



MesoAmerican Research Center, 2009
Map prepared by Sebeto (Sebastien Merlet)

Fertilité pour la zone régionale

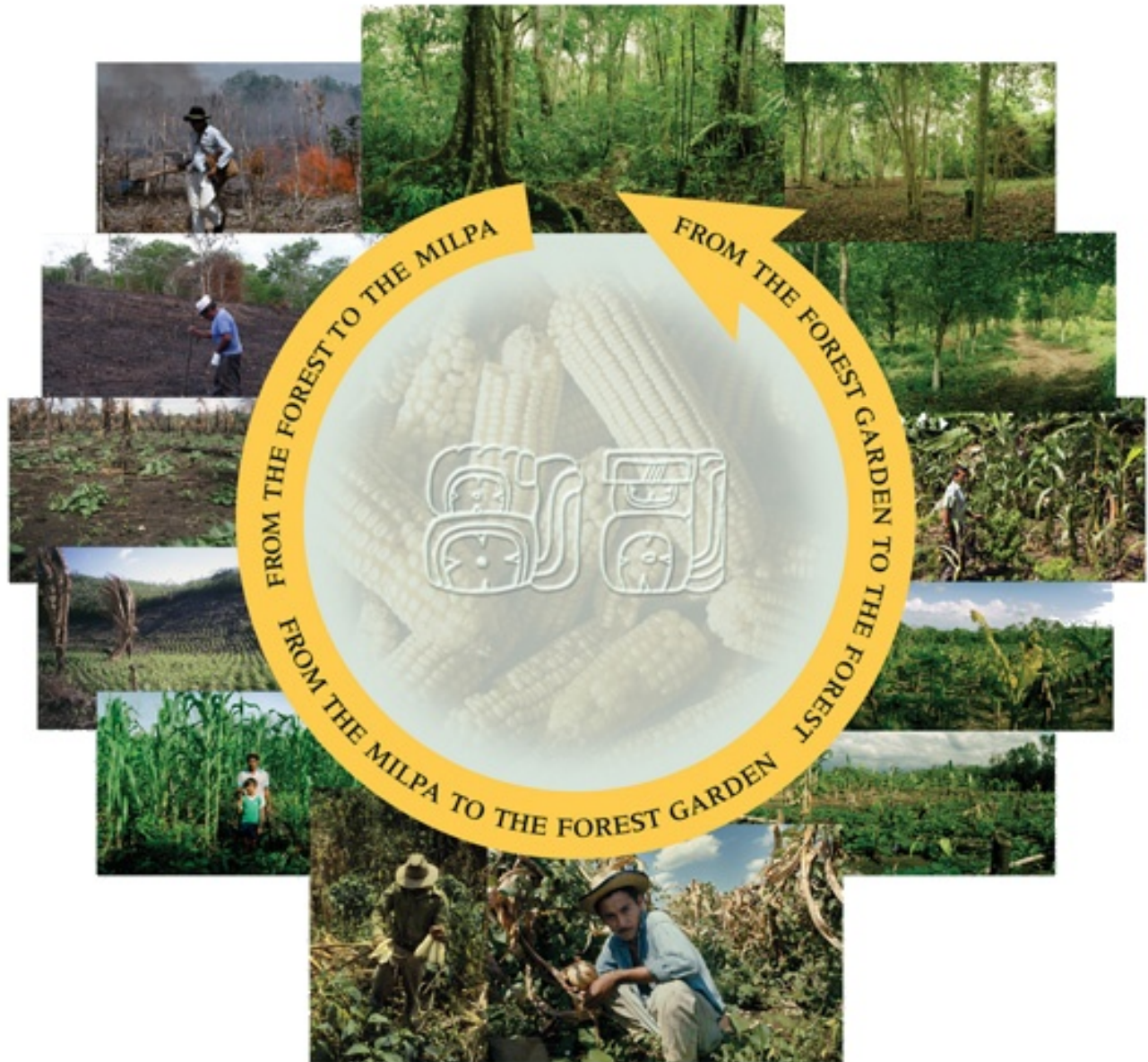


MesoAmerican Research Center, 2009
Map prepared by Sebeto (Sebastien Merlet)

Annexe 26 : cycle du milpa

Image provenant de <http://espmaya.org/>

THE MILPA CYCLE



Annexe 27 : calculs de la superficie nécessaire pour la culture du maïs

Paramètres de calcul	
Population totale de la zone	182600
Calories par jour par personne	2100
Calories par an par personne	764925
Total de calories pour la population par an	139675305000
Pourcentage provenant du maïs	34%
Total de calories (maïs) nécessaire par an	47489603700
Calories d'un kilogramme de maïs	3551
Kilogrammes de maïs nécessaires par an	13373585
Kilogrammes de maïs produits par an par km ²	58800
Superficie nécessaire pour la culture du maïs (km ²)	227

Terres cultivables	
Superficie totale de la zone (km ²)	1284.46
Classe 1 non cultivable (km ²)	484.67
Total de terres cultivables (km ²)	799.78

Sols cultivables occupés	
Sites de rang 0	2.16
Sites de rang 1	1.45
Sites de rang 2	9.71
Sites de rang 3	3.87
Sites de rang 4	0.33
Sites de rang 5	0.36
Structures communes (10% des classes 4 et 5)	48.05
Total de terres cultivables occupées (km ²)	65.94

Sols cultivables nécessaires pour le maïs	
Terres cultivables libres (km ²)	733.85
Terres nécessaires pour le maïs (km ²)	227
Pourcentage de terres cultivables nécessaires	31%

Annexe 28 : scénarios répondant aux besoins de la population

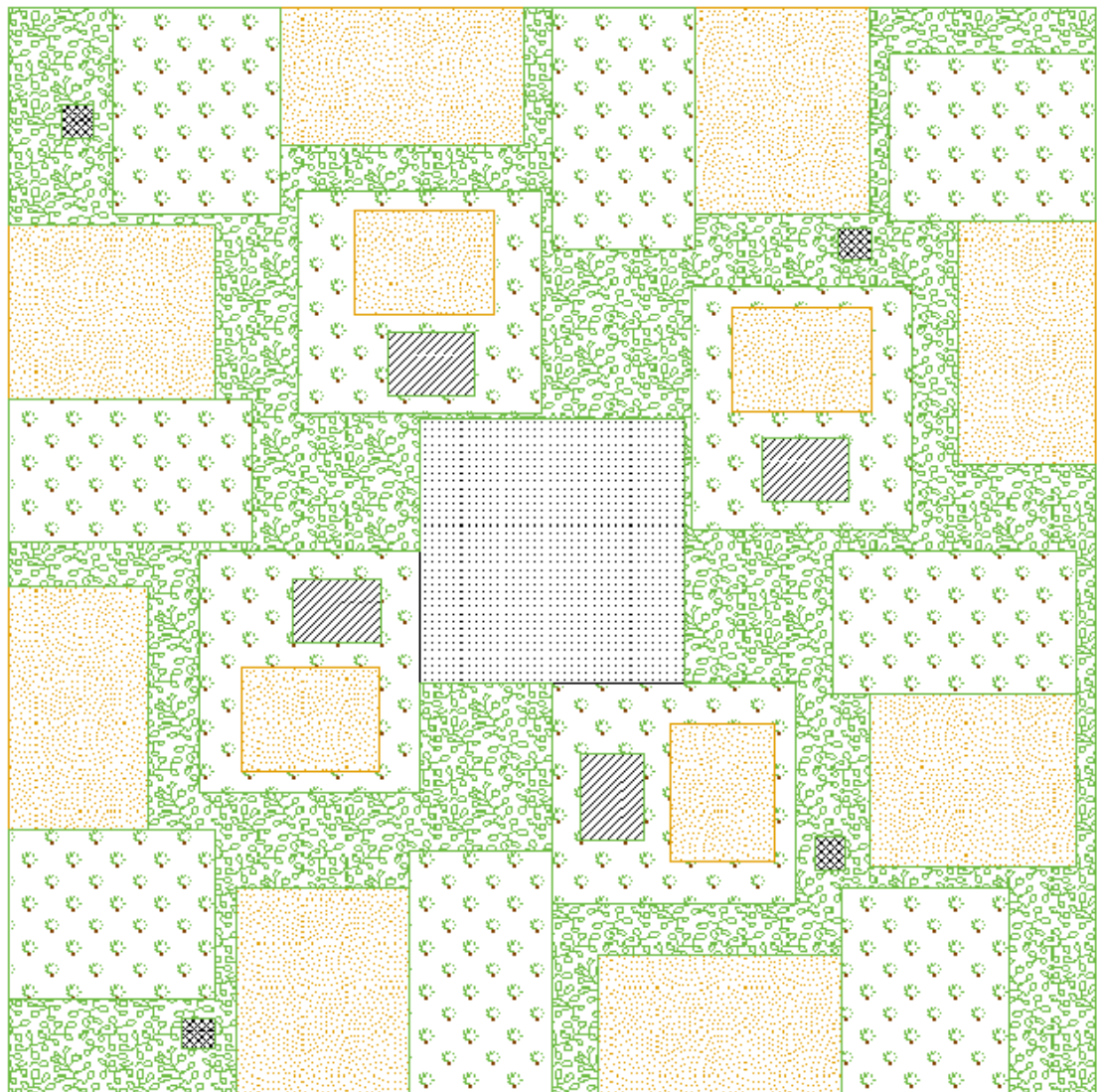
Les paramètres utilisés pour ces scénarios sont décrits en page 43 de ce mémoire. Les scénarios 2 et 3 conduisent aux mêmes valeurs, mais avec des durées de cycle du milpa différentes (réciproquement 12 ans et 9 ans).

Scénario 2	
Cycle du milpa considéré	
Durée d'utilisation du milpa externe (années)	4
Durée du cycle du milpa (années)	12
Ratio entre verger/forêt et milpa	2
Superficies nécessaires pour le cycle du Milpa	
Superficie totale nécessaire pour les milpas (m ²)	6975.21
Superficie de milpa externe (m ²)	5775.21
Superficie pour le reste du cycle (m ²)	11550.42
Terrain disponible par habitation	
Superficie totale nécessaire par URO (m ²)	21325.63
Superficie restant libre par URO (m ²)	1179.88
Sols restant à la nature	
Superficie totale de sols restant libre (km ²)	38.47
Scénario 3	
Cycle du milpa considéré	
Durée d'utilisation du milpa externe (années)	3
Durée du cycle du milpa (années)	9
Ratio entre verger/forêt et milpa	2
Superficies nécessaires pour le cycle du Milpa	
Superficie totale nécessaire pour les milpas (m ²)	6975.21
Superficie de milpa externe (m ²)	5775.21
Superficie pour le reste du cycle (m ²)	11550.42
Terrain disponible par habitation	
Superficie totale nécessaire par URO (m ²)	21325.63
Superficie restant libre par URO (m ²)	1179.88
Sols restant à la nature	
Superficie totale de sols restant libre (km ²)	38.47
Scénario 4	
Cycle du milpa considéré	
Durée d'utilisation du milpa externe (années)	4
Durée du cycle du milpa (années)	18
Ratio entre verger/forêt et milpa	3.5
Superficies nécessaires pour le cycle du Milpa	
Superficie totale nécessaire pour les milpas (m ²)	4950.05
Superficie de milpa externe (m ²)	3750.05
Superficie pour le reste du cycle (m ²)	13125.17
Terrain disponible par habitation	
Superficie totale nécessaire par URO (m ²)	20875.22
Superficie restant libre par URO (m ²)	1630.30
Sols restant à la nature	
Superficie totale de sols restant libre (km ²)	53.16

Annexe 29 : utilisation des sols au niveau de la zone d'étude

Cette carte donne une idée de ce à quoi pouvait ressembler le paysage au niveau de la zone d'étude en terme d'occupation des sols. Quatre unités résidentielles occupées sont ici représentées, avec la quantité de terres qu'elles avaient effectivement à leur disposition en utilisant un cycle du milpa de douze ans au total, dont quatre ans de culture du maïs. Le cycle s'arrête au niveau de la forêt secondaire puisque les huit ans de régénération sont insuffisantes à un retour à la forêt mature. Les parties communes ont été regroupées, il est possible d'imaginer une place comportant un petit temple par exemple.

Exemple de regroupement Maya



0 12.5 25 50 75 100 Meters

MesoAmerican Research Center, 2009

Carte préparée par Sebeto (Sebastien Merlet)

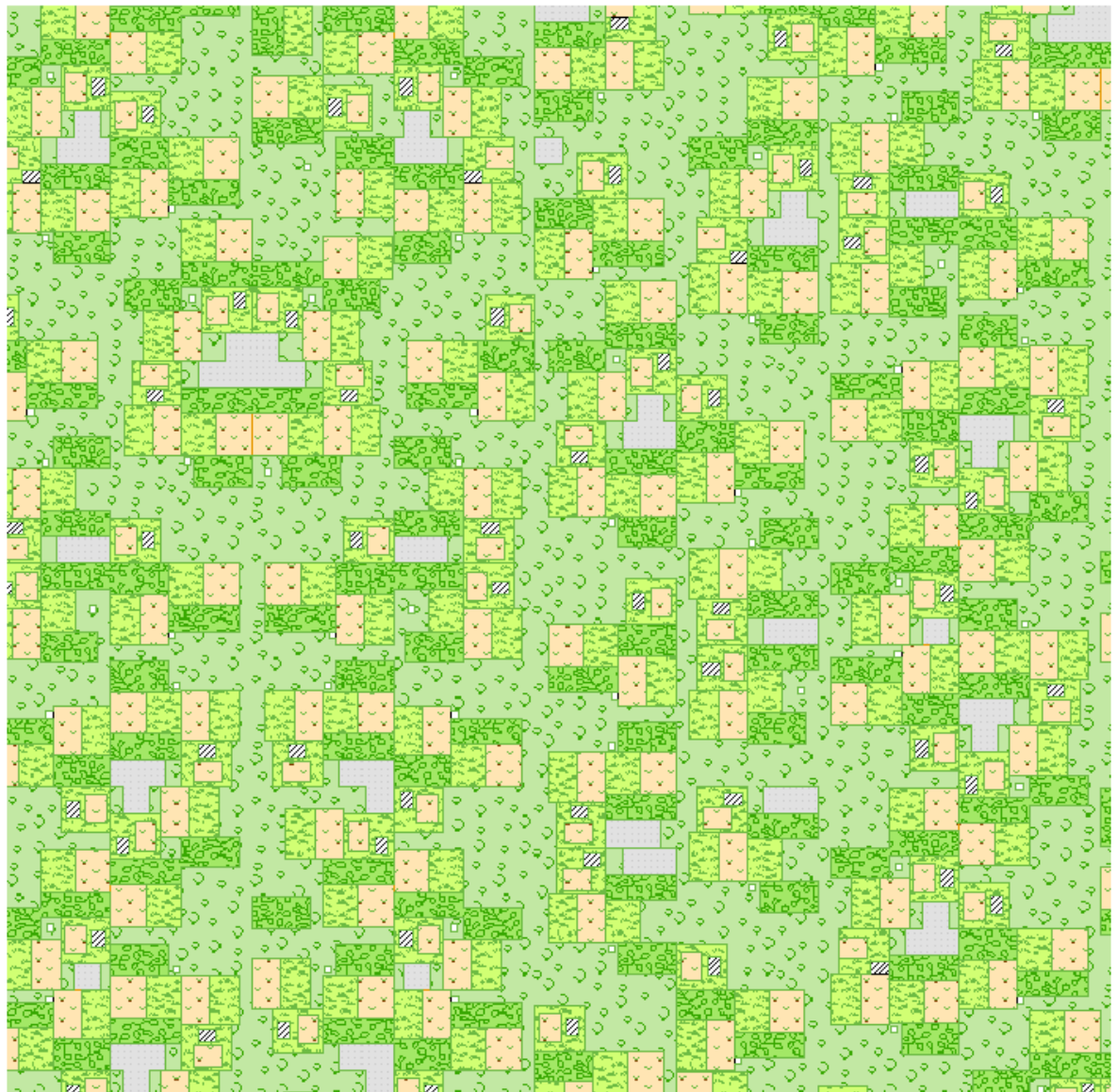
Parties communes
Zone résidentielle
Structure temporaire
Forêt en régénération
Forêt secondaire
Milpa

Annexe 30 : représentation de la densité de population

La carte présentée ici n'a pas vraiment de réalité physique au niveau de notre zone d'étude. Elle vise simplement à donner une idée de ce que représente la densité moyenne de 142 hab./km². Les proportions de sols correspondent aux moyennes rencontrées, la zone inutilisée schématisant la classe 1 du modèle. Ce type d'occupation du territoire pourrait être rencontrée à la limites entre les bons sols et les mauvais, où l'attrait pour les zones densément peuplées se fait encore sentir.

Modèle d'occupation des sols

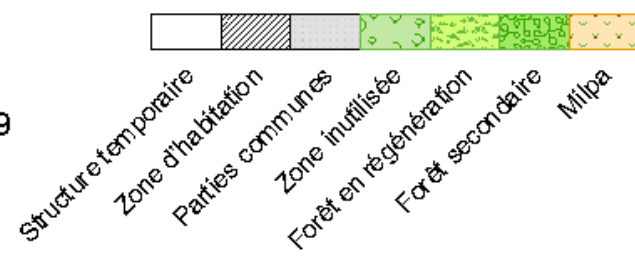
Densité de 142 hab./km²



0 75 150 300 450 600 Meters

MesoAmerican Research Center, 2009

Carte par Sebeto (Sebastien Merlet)



Étude de l'utilisation des Systèmes d'Information Géographique dans le cadre d'un projet archéologique multi-échelles

Mémoire d'ingénieur ESGT
Présenté par Sébastien Merlet

RÉSUMÉ

Les Systèmes d'Information Géographique sont devenus des outils couramment utilisés dans le cadre de projets archéologiques en raison des possibilités qu'ils offrent en terme de stockage, d'organisation, d'analyse et de partage des données. Cette étude présente un exemple complet de mise en place d'un SIG pour le *MesoAmerican Research Center*, de la création d'une géodatabase et de ses protocoles d'utilisation à la réalisation de cartes de répartition de sites mayas pour la période du Classique Tardif. La méthode des *Weights of Evidence* analysée dans cette étude sert à produire un modèle prédictif permettant de déterminer des densités par classe de probabilités ainsi que la population totale d'une zone proche de la Réserve Archéologique d'El Pilar, au Belize. Le modèle d'occupation des sols construit à partir de ce modèle en prenant en compte les méthodes d'agriculture et le cycle du milpa apporte des éléments nouveaux sur la relation des anciens Mayas avec leur environnement, ainsi que sur l'importance de la sauvegarde de ce patrimoine culturel par les jardiniers forestiers mayas modernes. Diverses modélisations de la répartition des Mayas à différentes échelles pour la région du Petén et le Système du Monde Maya sont également présentées, offrant de nouvelles possibilités de recherches.

Mots clés : Mayas, Modèle prédictif, SIG, Modèle de population, Modèle d'occupation des sols, Cycle du Milpa, Jardiniers forestiers, Classique Tardif, Projet Archéologique, Weights of Evidence

SUMMARY

Geographic Information Systems have become commonly used tools in archaeological projects due to the possibilities they offer to store, organize, analyze and share datasets. This study presents a complete example of the implementation of a GIS for the *MesoAmerican Research Center*, from the conception of a geodatabase and its protocols of use, to the creation of maps showing the Maya settlement patterns for the Late Classic Period. The method of the "*Weights of Evidence*" analyzed in this study is used to produce a predictive model allowing one to determine densities for the different probability classes, as well as the total population of an area near the Archaeological Reserve of El Pilar, in Belize. The patterns of land use, built from this model by taking into account the agricultural methods and the Milpa Cycle, bring new elements about the relationship between Ancient Maya and their environment, and the importance of modern Maya forest gardeners safeguarding this cultural heritage. Various models of the distribution of the Maya at different scales for the region of Petén and the Maya World System are also presented, offering new research possibilities.

Keywords: Mayas, Predictive model, GIS, Population Model, Soils Use Model, Milpa Cycle, Forest Gardeners, Late Classic, Archaeological Project, Weights of Evidence