HAL
open science

Interprétation du LIDAR : les défis dans l'identification des anciens sites mayas

Hugo Bihr

To cite this version:

Hugo Bihr. Interprétation du LIDAR : les défis dans l'identification des anciens sites mayas. Sciences de l'ingénieur [physics]. 2013. dumas-00919369

HAL Id: dumas-00919369

https://dumas.ccsd.cnrs.fr/dumas-00919369v1

Submitted on 16 Dec 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CONSERVATOIRE NATIONAL DES ARTS ET METIERS

ÉCOLE SUPÉRIEURE DES GÉOMÈTRES ET TOPOGRAPHES

_____

MÉMOIRE

Présenté en vue d'obtenir

le DIPLÔME D'INGÉNIEUR CNAM

Spécialité : Géomètre et Topographe

par

Hugo BIHR

_____

Interprétation du LiDAR: les défis dans l'identification des anciens sites mayas

Soutenu le 23 septembre 2013

_____

JURY

PRÉSIDENT : **M. Christophe PROUDHOM**

MEMBRES : **M. Laurent POLIDORI, professeur référent**
M. Éric LABERGERIE

_____

Maître de stage : **Dr Anabel FORD**

Remerciements

Ce stage de six mois passés en Californie, au Belize et au Guatemala aura été l'occasion de faire de belles rencontres. Je tiens à remercier toutes les personnes ayant participé de près ou de loin à la réalisation de mon travail.

Tout d'abord j'adresse ma profonde reconnaissance à l'égard de Dr Anabel Ford. Je lui suis reconnaissante pour la passion qu'elle porte au projet BRASS/El Pilar et qu'elle transmet au quotidien toujours avec beaucoup de dynamisme et d'énergie. Ce fut un honneur que de côtoyer une personnalité s'investissant corps et âme dans une mission d'une aussi belle envergure.

Mes remerciements vont à Anne Girardin, la première étudiante à avoir scellé le lien entre BRASS/El Pilar et l'ESGT. Grâce à Anne, j'ai pu caresser l'espoir de me lancer dans cette mission. J'ai également apprécié l'aide de membres de l'UCSB lors de mes recherches : Dr Keith Clarke, Thomas Jensen, Thomas Pingel, Shane Grigsby et Michael Alonzo.

J'ai une pensée toute particulière aux collègues de bureau et de terrain, de Santa Barbara et de Santa Elena. Je remercie l'équipe du MARC pour son accueil chaleureux, son soutien dans les moments de doutes : Addison, Brandon, Connie, Hannah, Kaylah, Khahn et Lisete. Mes salutations s'envolent aussi jusqu'à la base de BRASS, à destination d'Andres, Carmen, Maria, Narciso, Nehanda. Merci pour le partage de la douceur de vivre bélizienne. Je pense bien sûr à Tina, Grecia, Paulino qui ont participé aux expéditions sous la canopée.

Bien sûr, un projet à l'étranger nécessite plus d'investissements financiers. Pour cela j'ai pu compter sur l'aide précieuse de l'ESP (Exploring Solutions Past) de la région Pays-de-la-Loire, de la commune de Brette-les-Pins via leurs subventions (Bourse Envoléo et Bourse sur Critères Sociaux).

Enfin, je remercie ma famille et plus particulièrement mes parents pour leurs soutiens moral et financier.

Table des matières

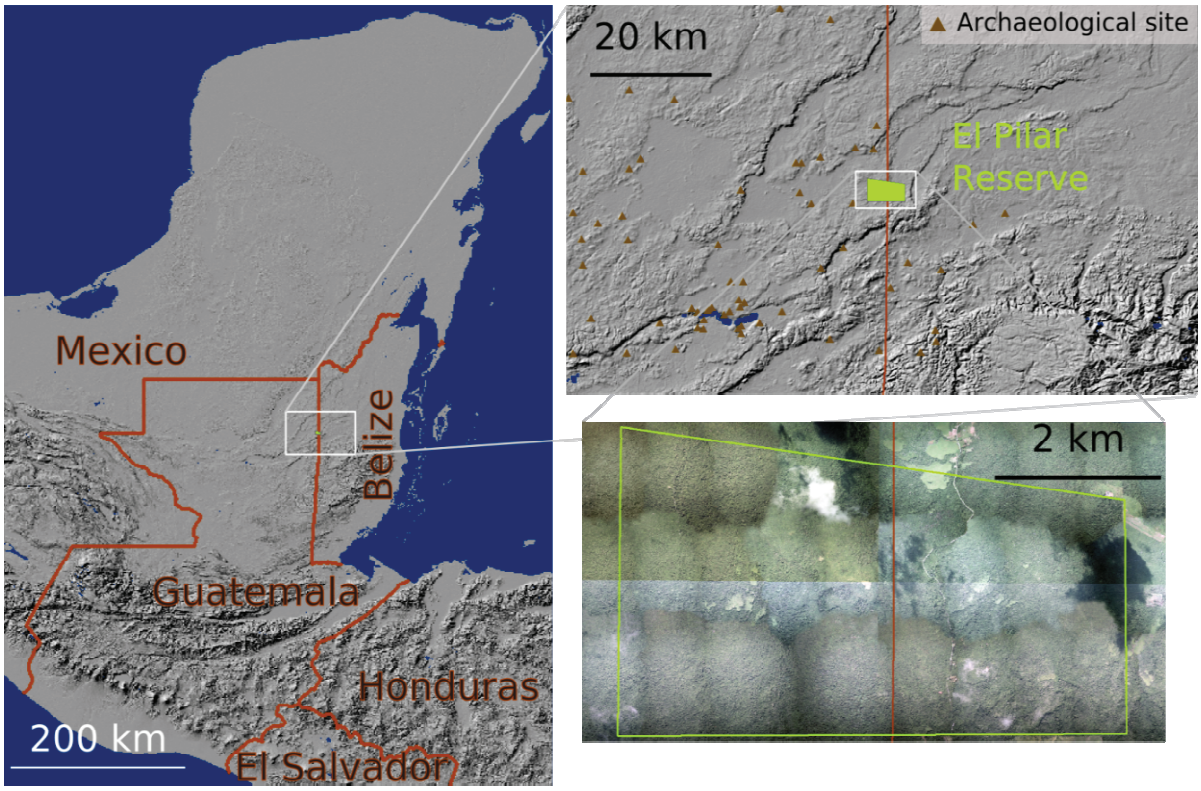
Remerciements 2
Introduction 6
I Des points à la cellule : l'importance de la taille de cellule 9
I.A Observations préliminaires et notions à propos du nuage de points de sol 9
I.A.1 Courbes de niveau et TIN face au MNT raster 9
I.A.1.a Courbes de niveau 9
I.A.1.b TIN 9
I.A.1.c MNT Raster 9
I.A.2 Pré-analyse de la densité de points de sol 10
I.A.2.a Analyse par hectare sur la réserve 10
I.A.2.b Analyse sur l'aire de levé terrain 10
I.A.3 Comment relier densité de points de sol et MNT raster ? 13
I.A.3.a 4 points en 1: concept visuel à propos des divers types de points 13
I.A.3.b Les paramètres des deux formats 14
I.A.3.c Variables entre les formats 14
I.B Des choix de tailles de cellules expliqués par trois variables : NZER, RH et CF 14
I.B.1 Contexte d'étude 15
I.B.2 NZER - "Non-Zero" Error Rate – taux de risque non nul 15
I.B.2.a Précision ou cercle de confiance 15
I.B.2.b Interprétations polynomiales 1
I.B.3 Homogénéité relative 17
I.B.3.a L'homogénéité aussi en fonction de la taille de cellule 17
I.B.3.b Proportions inverses 18
I.B.4 Fréquence de comptage 19
I.B.4.a À l'échelle de la cellule 19
I.B.4.b L'intérêt pour les cellules singulières 19
II Un modèle de surface à interpoler et à visualiser 22
II.A Sélection de l'algorithme d'interpolation selon des critères visuels et une simulation de sol 22
II.A.1 Deux boîtes à outils : Point to Raster et Raster Interpolation 22
II.A.1.a Résumé et traitement 22

II.A.1.b Présélection visuelle 25
II.A.2 Simulation d'un 'sol vrai' 25
II.A.2.a Biais d'interpolation 25
II.A.2.b L'influence du bruit 26
II.B Le besoin d'un modèle de visualisation : l'altitude ne suffit pas 28
II.B.1 Algorithmes usuels 28
II.B.1.a Color Cast (distribution des couleurs) 28
II.B.1.b Relief ombragé - Hill-shaded Relief 30
II.B.1.c Modèle de rayonnement solaire (SIM - Solar Insolation Modeling) 30
II.B.1.d Facteur de vue du ciel (SVF - Sky View Factor) 31
II.B.1.e Modélisation locale du relief (Trend Removal - Local Relief Model) 31
II.B.2 Choix a priori 32
II.C Pseudo Modèles 34
II.C.1 Traitement avec eCognition® et ArcGIS® 34
II.C.2 Les modèles mathématiques inspirés des algorithmes usuels 36
II.C.2.a Arctan Color Cast 36
II.C.2.b Pseudo SVF 37
II.C.2.c Pseudo SIM 38
II.C.2.d ACP de reliefs ombragés - PCA of Hill-shadings 38
II.C.2.e Au-delà la variation de la pente : Orientated Second Derivative extrema 39
II.C.3 Observations préliminaires 40
III Développement d'une stratégie pour la détection de sites anciens mayas 43
III.A Interprétations et validations des produits issus du LiDAR 43
III.A.1 Interprétation de sites archéologiques : l'œil est un détecteur d'objets 43
III.A.1.a Des taux de détection dans des matrices d'erreur 43
III.A.1.b Les éléments d'intérêt 44
III.A.1.c L'apport des sites découverts pour une première validation 45
III.A.1.d Le deuxième règle de validation à la recherche de sites potentiels 46
III.A.2 L'exploration du terrain pour les sites potentiels 47
III.A.2.a Équipement et notes de terrain 47
III.A.2.b Contrôle de terrain de l'interprétation du LiDAR : matrices de classification 48
III.A.2.c Troisième validation à mi-chemin entre interprétation et réalité 49
III.B Améliorations dans le protocole de détection des sites anciens mayas 50
III.B.1 Le besoin d'expérience de terrain pour détourner les pièges du LiDAR 50

III.B.2 Erreur de positionnement GPS : un frein à l'investigation? 50
III.B.3 Gestion du nuage de points et du paysage 51
III.B.3.a Post-analyse du nuage points de sol par l'étude de paramètres environnementaux 51
III.B.3.b Re-filtrage du sol 54
III.B.4 Une esquisse de protocole 56
III.B.4.a Problèmes rencontrés 56
III.B.4.b Un parcours à suivre 57
Conclusion 60
Références 62

Introduction

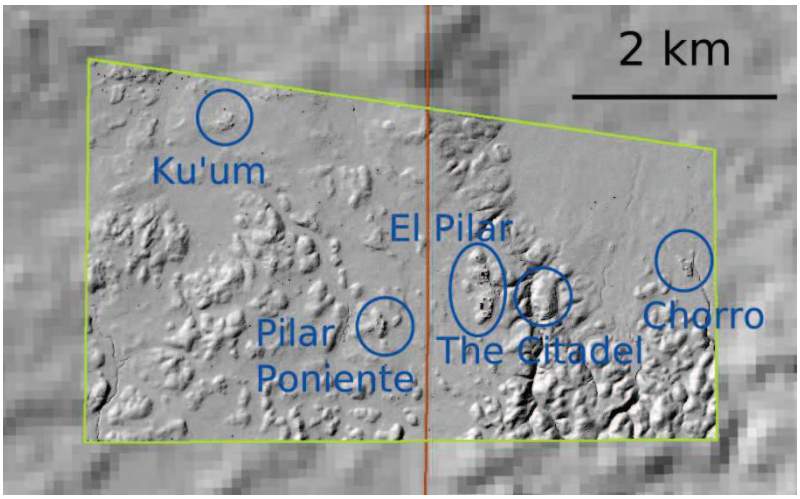
Le futur attend derrière une porte verrouillée par des clefs enfouies dans le passé. En essayant de retrouver des traces des civilisations anciennes, l'archéologie permet de trouver des explications aux événements ayant conduit à la gloire et à la chute de nos ancêtres.Dans un monde où l'intérêt pour l'environnement occupe une place de plus en plus importante au sein de la conscience collective, se tourner vers le passé nous évite bien des erreurs.Les mayas sont sans nul doute l'une des civilisations primairesqui a suscité le plus de curiosité. Depuis 4000 ans, ce peuple a appris à maîtriser l'agriculture, l'architecture, l'écriture et l'astronomie. Grâce aux *Forest Gardens*, ils ont été capables d'entretenir une relation respectueuse vis-à-vis de la nature, laissant pousser les cultures parmi la forêt tropicale, et ce selon un cycle de rotation appelé cycle du Milpa[Ford and Nigh, 2009]. Par la suite, un tragique concours de circonstances,parmi lesquelles politiques,a conduit au déclin des peuples vivant aux alentours des cités de Tikal et d'El Pilar en l'espace d'un siècle, durant la période du Classique Terminal (900-1000 ap. J.-C.), tandis que les cités du nord crûrent. Étonnamment, le système agricole a perduré, si bien qu'aujourd'hui encore des fermiers perpétuent la tradition héritée de leurs ancêtres.



Cartes de localisation de la réserve d'El Pilar

Les jardins forestiers et l'occupation du sol sont l'un des aspects essentiels du projet BRASS /El Pilar. Sous la direction de l'archéologue Dr Anabel Ford, ce projet se concentre sur la cité perdue d'El Pilar. Dans un souci de conservation archéologique, une réserve d'environ 20 km² a été délimitée à cheval sur la frontière entre le Guatemala et le Belize. Depuis sa découverte en 1983, des levés

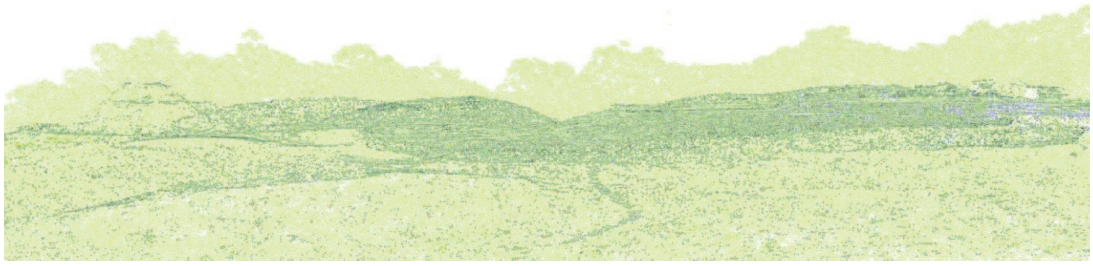
topographiques ont révélé nombre de vestiges…et beaucoup d'autres se cachent sûrement sous la



canopée. En effet, la végétation nuit aux investigations. Évidemment, les sites les plus imposants sont visibles. Mais qu'advient-il des petits sites ?
Illustration: sites majeurs au sein de la réserve d'El Pilar

Depuis une décennie, le Centre de Recherche Mésoaméricaine (MARC), a expérimenté une approche prédictive. La méthode appelée Pondération de l'information probante (Weights of Evidence) combine des variables spatiales telles la fertilité du sol, la pente topographique et le drainage[Merlet, 2009] [Ford, 2009]. Des modèles prédictifs sont ensuite créés, cartographiant la probabilité de trouver un site pour une région donnée. Les prédictions se sont montrées cohérentes mais une validation par le terrain demeure nécessaire.

En mai 2012, un hélicoptère équipé d'un système LiDAR a survolé la réservé. LiDAR signifie Light Detection and Ranging, une technique de télédétection basée sur deux principes : la distance en fonction du temps d'aller-retour d'un écho laser d'une part, et le géoréférencement d'autre part. La rayon laser est réfléchi aussitôt qu'il entre en contact avec un élément du paysage :une feuille, une roche, un papillon… Un nuage de points est ainsiacquis, chaque point correspondant à un retour d'écho vers le laser. Le couple GNSS (Global Navigation Satellite System) and IMU (Inertial Measurement Unit) géoréférence le nuage de points, c'est-à-dire déterminer les coordonnées des points. Le LiDAR est certes reconnu pour cartographier de façon automatique, tout en étant plus rapide et précis que la photogrammétrie ou le levé terrestre. Mais il est d'autant plus apprécié pour sa faculté à passer au travers du couvert forestier, pour entrevoir le sol[Crutchley, 2009].



Aperçu 3D d'un nuage de points. Végétation en vert, sol en bleu

Malheureusement, le format ponctuel ne fournit pas de support agréable pour détecter des sites archéologiques. Pour surmonter cet obstacle, le nuage de points du sol devra être traité dans un autre format. Ainsi, un modèle de surface devra être interpolé à partir des points, ce qui implique une perte de qualité.
Peut-on se fier uniquement au LiDAR pour s'atteler à la détection de sites archéologiques ? Quel processus peut-on suivre et quels sont les choix à faire pour réaliser un support propice à la visualisation ? Comment mettre en place un protocole pour valider le modèle issu du LiDAR en associant interprétation et exploration ?

Le présent mémoire aborde ces questions en trois parties.
Premièrement, nous justifierons le choix du format MNT raster (Modèle Numérique de Terrain). Ce sera l'occasion de présenter trois variables créées afin de relier les paramètres du nuage de points du sol avec ceux du MNT raster. L'analyse de ces variables nous guidera lors du choix de la taille de cellule du raster.
Deuxièmement, nous programmerons le modèle de surface en deux étapes. Une fois l'algorithme d'interpolation choisi, il s'agira de comparer les méthodes de visualisation existantes pour concevoir les rasters de visualisation.
Troisièmement, les résultats de la validation de l'interprétation des produits LiDAR aidera à la mise en place d'une stratégie de détection des sites mayas anciens.

I Des points à la cellule : l'importance de la taille de cellule

À moins d'observer des profils, la représentation ponctuelle n'aide pas à la visibilité de sites. En conséquence, un modèle de surface est nécessaire. Cette partie explique brièvementles rasions de préférer le format MNT raster aux courbes de niveau et TIN (Triangular Irregular Network). Puis, des concepts liant nuage de point du sol MNT raster vont être présentés. Enfin, des études vont accorder de l'importance à un paramètre du format MNT raster, à savoir la taille de cellule.

I.A Observations préliminaires et notions à propos du nuage de points de sol

I.A.1 Courbes de niveau et TIN face au MNT raster

Pour compenser le manque de visibilité du nuage de points, les trois formats les plus courants ont été comparés :courbes de niveau, TIN et MNT raster.

I.A.1.a Courbes de niveau

Uniquement constituées de lignes, le format de courbes de niveau dépend de l'intervalle choisi. Les sites évidents apparaissent et quelques formes caractéristiques peuvent même ressurgir (terrasses). Mais quand il s'agit de distinguer des éléments plus petits en surface et en hauteur, la détection reste difficile. Bien sûr, accroître permet de faire apparaître plus d'éléments, au détriment la visibilité globale.

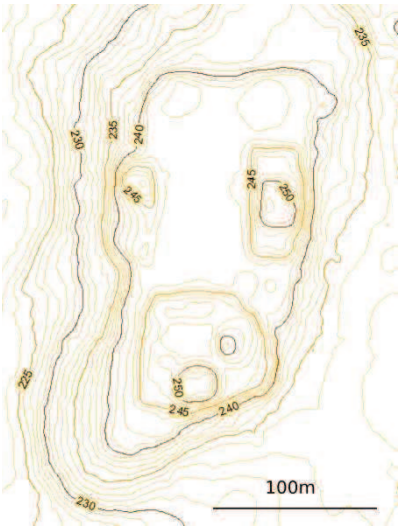


Illustration: courbes de niveau1 m - 5 m -10 msur El Pilar(Localisation UTM Zone 16Nord: E: 271890; N: 1908750)

I.A.1.b TIN

Comme indiqué ci-contre, le format TIN représente des triangles, avec un ombrage supplémentaire. Le relief créé est fidèle par rapport au nuage de points car il passe exactement par tous les points. Cependant voir des triangles nuit quelque peu à identifier des éléments, surtout pour la simple raison que les cibles à détecter sont rarement triangulaires.

I.A.1.c MNT Raster

Le MNT se caractérise par deux paramètres principaux, à savoir la méthode d'interpolation et la taille de cellule.Ainsi, avec des paramètres judicieusement choisis, cela permet de créer une image

aisément visualisable et analysable via des procédés de traitement d'image. La détection demeure difficile.

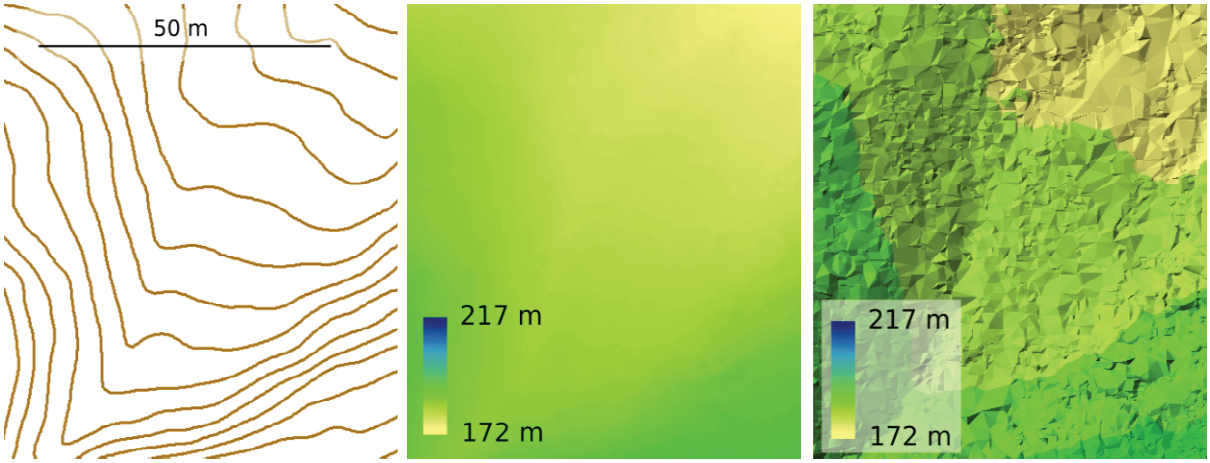


Illustration: courbes de niveau1 m/ TIN crééavec les paramètres par défaut / MNT raster avec l'algorithme Natural Neighboret une taille de cellule de 0,5 m (coordonnées: E 272320 ; N 1909350)

→Aucun des trois formats ne permet de distinguer clairement tous les types d'éléments. Contrairement aux courbes de niveau et au TIN, le MNT raster apparait comme le seul réellement utilisable. Néanmoins, il ne faudra pas négliger la perte d'information qu'entraînent l'interpolation et la discrétisation du modèle de surface. Il semble légitime d'évaluer l'incidence des choix des paramètres du MNT raster en vue de limiter la perte de fidélité. Cela commence par une analyse du nuage de points de sol, et en particulier la densité de points recouvrant le sol.

I.A.2 Pré-analyse de la densité de points de sol

L'étude de la densité de points de sol (GPD pour Ground Point Density) donne un aperçu de la répartition des points de sol. Deux analyses ont été faites : la première sur la réserve entière et la seconde focalisée sur l'aire de levé terrain.Chacune des deux zones d'étude a été découpée selon une grille régulière, dont les cellules mesurent respectivement 100x100 m² et 5x5 m². Les cartes sont illustrées ci-après.

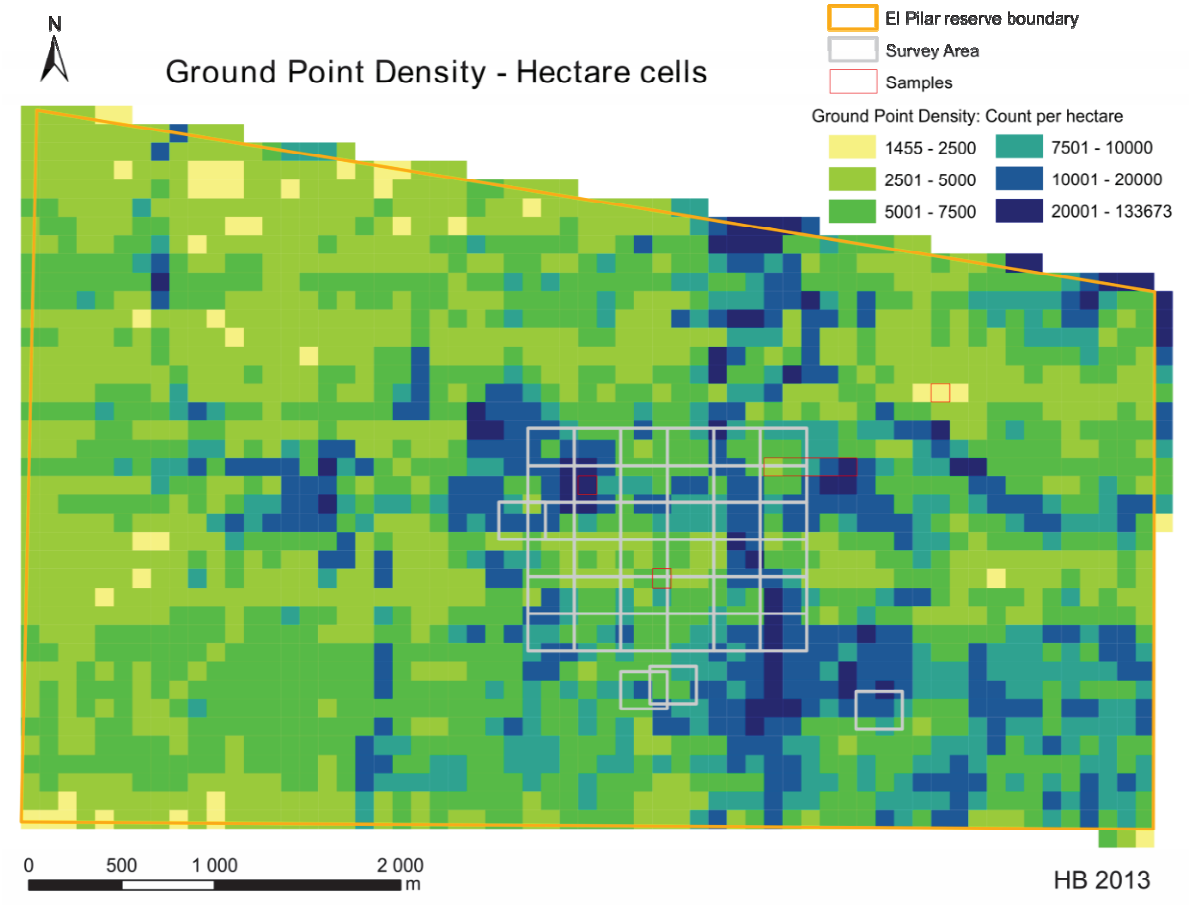
I.A.2.a Analyse par hectare sur la réserve

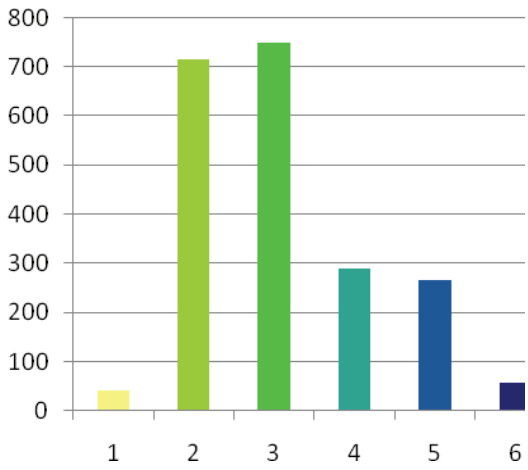
Globalement, les cellules ne répertorient que peu de points de sol, avec en majeure partie moins d'un point par m². Évidemment, plus la végétation est dense plus le nombre de points ayant touché le sol touché est faible. Les routes et les zones sans végétation ont les densités maximales.La densité varie beaucoup sur l'ensemble de la réserve.

I.A.2.b Analyse sur l'aire de levé terrain

Comme on pouvait l'imaginer, l'étude à ne plus taille de cellule permet de mieux expliquer les disparités de densité. En effet, les routes et les champs ouverts se démarquent. De plus on observe des bandes correspondant au plan de vol de l'hélicoptère.

→Les caractéristiques paysagères représentent le facteur essentiel pour expliquer une telle variation de ladensité de points de sol. Une analyse complémentaire effectuée dans la dernière partie de ce mémoire abordera l'impact de trois caractéristiques paysagères : la canopée, la relief et le sous-étage.



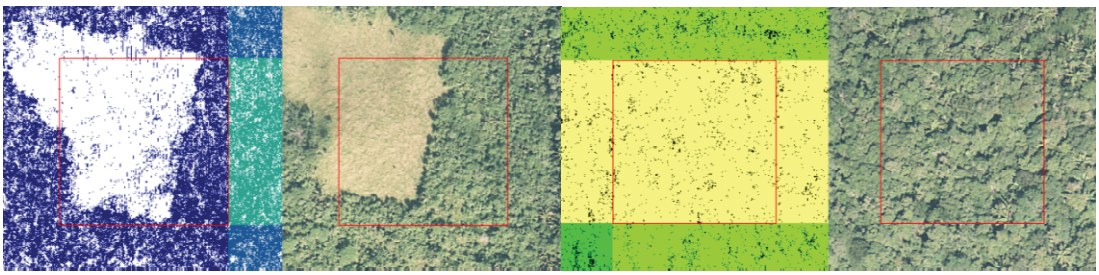


Classe	Compte de points de sol	Nombre de cellules
1	1455-2500	42
2	2501-5000	715
3	5000-7500	750
4	7501-10000	289
5	10001-20000	266
6	20001-133673	58

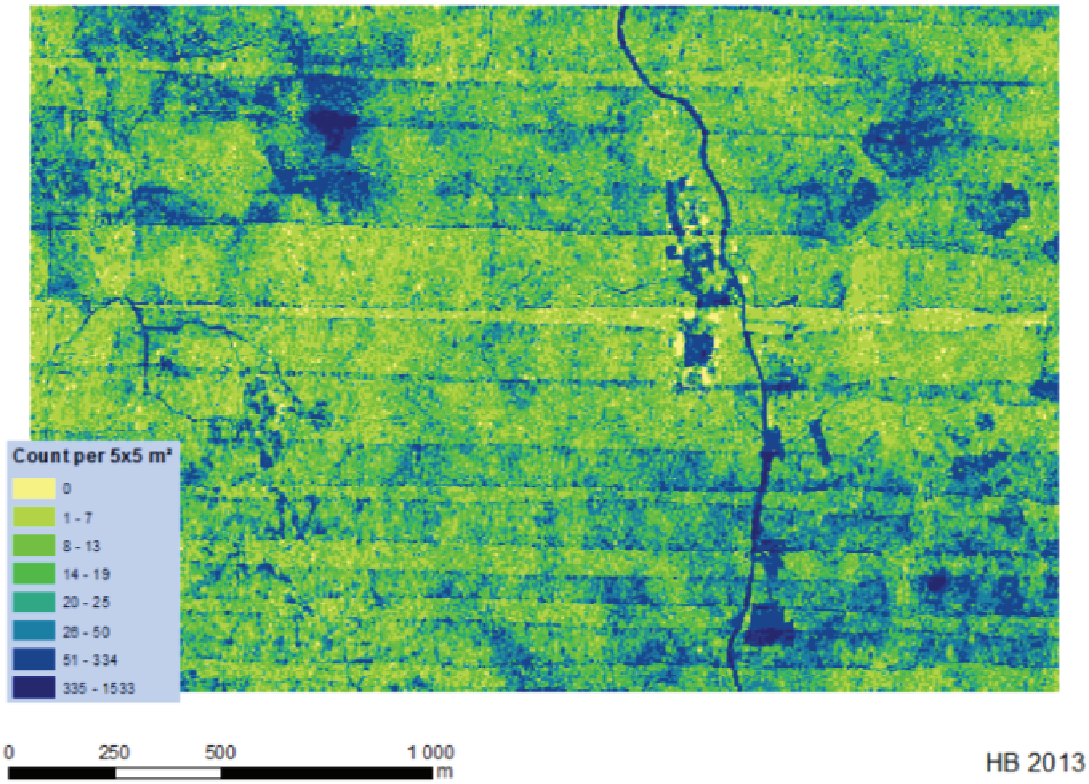
Minimum	1455
Maximum	133673
Médian	5894
Moyenne	7294
Écart-type	6228

Carte et statistiques de la densité de points de sol par hectare

NB: ensemble, les classes 2 et 3 comptent pour 69% des cellules (1465/2120).



Cellules avec les densités maximales. Les captures d'écran illustrent bien l'impact de la végétation sue la densité de points de sol. À gauche, le sol est dégagé (133673 points de sol) tandis qu'à droite la végétation semble impénétrable (1455 points de sol).



Densité de points de sol (cellules de 5x5 m²)

NB: les formes jaunes que l'on peut observer au centre-est représentent les hauts temples d'El Pilar. Pourtant il semblerait qu'aucun point n'y ait été enregistré. En réalité, le laser a touché le sommet de ces structures. Le problème est issu du filtrage des points du sol. Pour séparer les points du sol des autres points, ces filtres se basent sur des seuils associés aux pentes. Ainsi, les temples ayant des pentes importantes, les points situés dessus ont été considérés comme n'appartenant pas au sol.C'est un obstacle à la détection, quand on imagine que des terrasses puissent avoir des pentes fortes également.

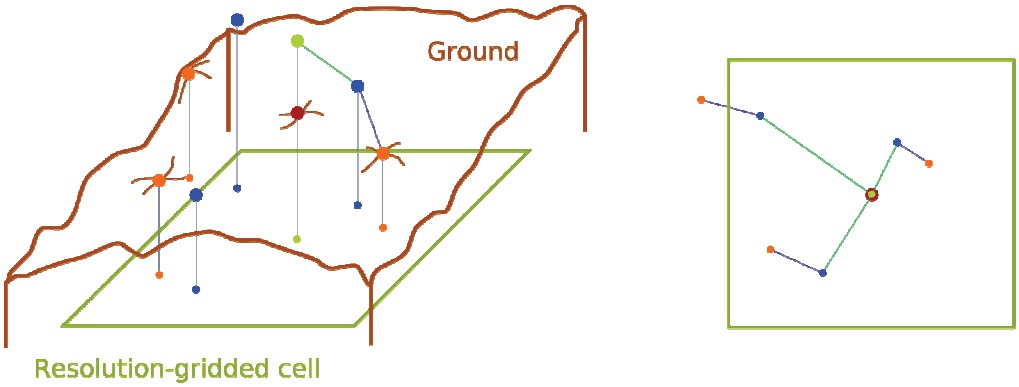
I.A.3 Comment relier densité de points de sol et MNT raster ?

Avant d'aborder les choix pour les paramètres du MNT raster, il est nécessaire d'introduire des notions qui vont lier la densité de points de sol au raster.

I.A.3.a 4 points en 1: concept visuel à propos des divers types de points

Ce paragraphe mentionne 4 types de points qualifiés Vrai (True), Mesuré (Measured), Interpolé (Interpolated) et Liant (Linking). Comprendre la relation unissant ces points constitue la clé pour analyser le lien entre nuage de points de sol et MNT raster.

Le point Vrai appartient au sol et représente la cible du laser. Le point Mesuré diffère du point Vrai de par les imprécisions du système LiDAR. Le point Interpolé caractérise le passage du format à l'autre, avec une altitude interpolée et est situé de fait au centre de chaque cellule de la grille.L'intersection de la verticale passant par ce point avec le sol constitue le dernier point, appelé point Liant.



True	Measured	Interpolated	Linking
Xt	Xm=Xt+εX	Xi=X0+rR	Xl=Xi
Yt	Ym=Yt+εY	Yi=Y0+lR	Yl=Yi
Zt	Zm=Zt+εZ	Zi=Int(Zm)	Zl=Zi+βZ

Notations :

- εX, εY, εZ: erreurs de positionnement
- (X0,Y0): originede la grille du raster
- r,l: nombre de lignes et de colonnes
- Int: fonction d'interpolation. Un ou plusieurs points mesurés sont nécessaires
- βZ: appelé biais d'interpolation. Correspond à une erreur d'altitude engendrée par l'écart de positionnement et la méthode d'interpolation.

Schéma et formules reliant les types de points

I.A.3.b Les paramètres des deux formats

Comme mentionnée précédemment, les précisions sont une partie des paramètres du nuage de points. La seconde partie se rapporte à la configuration spatiale du nuage de points, représentée par deux paramètres : la densité globale et la répartition de la densité. Par analogie avec les notions statistiques, la densité correspond à une moyenne ; la répartition à un écart-type.

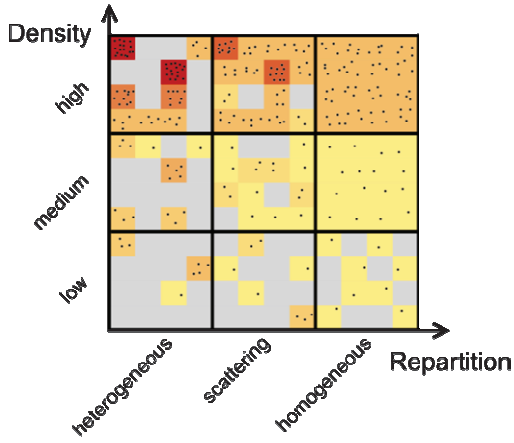
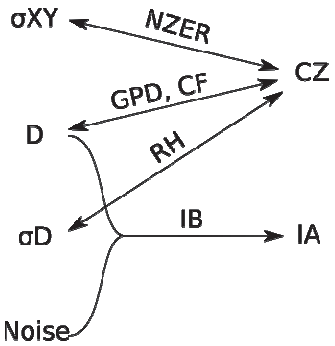


Illustration: densitéet répartition de points

À propos du MNT raster, taille de cellule et algorithme d'interpolation demeurent les seuls paramètres étudiés.

I.A.3.c Variables entre les formats

Le schéma ci-dessous illustre les variables calculées dans le but de tracer de relations entre le nuage de points de sol et le MNT raster.Chaque abréviation inscrite sur les flèches représente une variable d'analyse.



Paramètres:
- σXY: précision planimétrique (m)
- σZ: précision altimétrique (m)
- D: densité moyenne de points de sol (pt.m^{-2})
- σD: répartition (pt.m^{-2})
-CS: taille de cellule
- IA: algorithme d'interpolation

Variables:
- NZER: taux de risque non nul
- GPD: densité de points de sol par cellule
- CF: fréquence de compte
- RH: homogénéité relative
- IB: biais d'interpolation

Variables et paramètres. Le bruit (Noise) fait référence aux points filtrés par erreur comme étant des points de sol.

→Maintenant que les notions préliminaires ont été listées, l'analyse démarre avec l'étude d'un paramètre du raster toujours difficile à appréhender : la taille de cellule.

I.B Des choixde tailles de cellules expliqués par trois variables : NZER, RH et CF

Quand il s'agit de créer des rasters, le choix de la taille de cellule ne s'explique pas clairement. La décision provient soit de l'expérience soit d'une contrainte de projet. Le chapitre dévoile trois variables dans le but de calculer une ou plusieurs tailles de cellule idéales, au regard du nuage de points de sol.

I.B.1 Contexted'étude

Les variables ont été créées sous ArcGIS® 10.x. pour éviter des temps de traitement trop longs, les analyses ont été réduites à des extraits de la réserve. En somme, 8 échantillons d'hectare ont été sélectionnés : minimum, maximum, moyen et une bande de 5 hectares de long. Cette bande se caractérise par une densité de points croissante vers l'Est. Ces échantillons figurent sur la carte de densité de points de sol par hectare (en rouge).

Les tests font intervenir une liste de tailles cellules prédéfinie : 0,5 ; 0.625 ; 1 ; 1,25 ; 2 ; 2,5 ; 4 ; 5 ; 10 et 20 m.

I.B.2 NZER - "Non-Zero" Error Rate – taux de risque non nul

I.B.2.a Précision ou cercle de confiance

En une phrase, NZER peut être définie comme la proportion de points mesurés dont la probabilité que son point vrai correspondant n'appartienne pas à la même cellule soit non nulle. Visuellement, le point vrai se situe dans un cercle de confiance dont le point mesuré est le centre et la précision planimétrique le rayon. Ainsi, NZER calcule le nombre de points dont le cercle de confiance intersecte les lignes de la grille du raster.

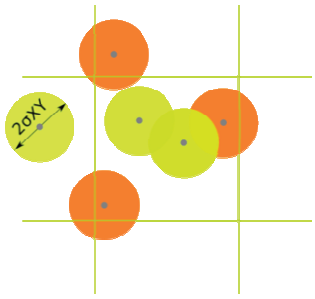
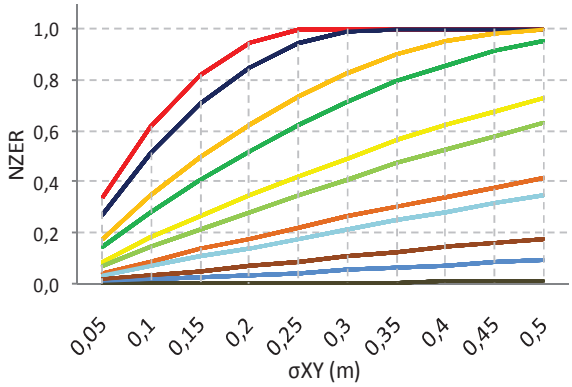


Illustration: cercles de confiance à cheval sur la grille

A priori, cette variable devrait croître avec la précision planimétrique et décroître avec la taille de cellule. Pour une taille de cellule fixée, on doit constater un seuil de saturation, ce qui signifie que la grille coupe tous les cercles.

Pour calculer ce taux, les cercles ont été dessinés sous forme de *buffers*autour des points. Plusieurs rayons ont été testés, de 0,05 m à 0,50 m par pas de 5 cm. L'outil *Fishnet*dessine les grilles pour les chacune des tailles de cellule expérimentées. Au final, une requête de sélection spatiale permet d'extraire les points « à risque », autrement dit ceux dont le buffer se trouve sur la grille. Les résultats des sélections affichent le nombre.

I.B.2.b Interprétations polynomiales



CS (m)	a	b	c	d	e	f
0.5	-234.19	304.22	-119.79	3.80	6.63	1E-02
0.625		24.82	-14.37	7.28	6.16	-2E-02
1				-3.90	3.96	-9E-03
1.25				-2.45	3.15	-7E-03
2				-0.93	1.94	-3E-03
2.5				-0.56	1.56	-4E-03
4				-0.26	0.97	1E-02
5				-0.10	0.76	-2E-03
10					0.36	-2E-04
20					0.19	-7E-04
100					0.03	-3E-04

Graphiqueet coefficients polynomiaux associés Une couleur correspond à une taille de cellule.

Les résultats confirment les attentes et vont même au-delà. En effet on constate des relations polynomiales entreNZER et σXY comme suit :

$$NZER(\sigma XY) = a\sigma XY^5 + b\sigma XY^4 + c\sigma XY^3 + d\sigma XY^2 + e\sigma XY + f$$

Pour les tailles de cellule en-deçà de 2 m, la courbe est assimilable à une droite. Au-delà, on constate un seuil de saturation, qui a tendance à incurver cette droite.

En négligeant l'effet de saturation, il est possible d'exprimer le coefficient du 1er ordre *e*, en fonction de CS. Il existe une relation inverse de la forme :

$$e(CS) \approx \frac{constant}{CS} \approx \frac{3.84}{CS}$$

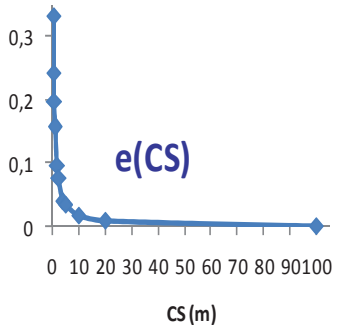


Illustration: coefficient linéaire e *en fonction deCS*

Donc, dans le cas où il n'y a pas de saturation, une formule relie CS et σXY :

$$CS \approx 3.84 \frac{\sigma XY}{NZER}$$

Cette formule empirique permet de se faire une idée du choix de la taille de cellule, et ce directement en fonction du « risque » que point vrai et point mesuré ne se trouvent pas dans la même cellule. Dans notre exemple, la précision planimétrique du LiDAR est estimée à $\sigma XY = 22.5cm$. En imaginant un taux d'erreur NZER < 50%, il vient : CS > 1,72 m.

→Bien sûr, cela dépend des échantillons présélectionnés. Aussi, le « risque » considéré est maximal et peut être atténué. À vrai dire, il serait judicieux de déterminer des aires d'erreur, c'est-à-dire la proportion du cercle de confiance qui se trouve effectivement en dehors de la cellule.

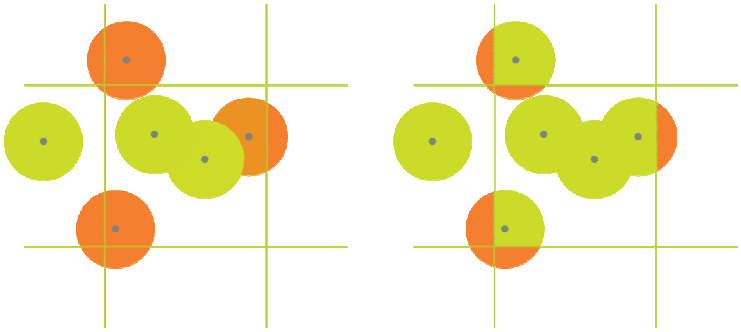


Schéma pour une autre approche du NZER. L'analyse du NZER gagnerait en cohérence si cette variable ne prenait en compte que les aires d'erreur (en orangé).

I.B.3 Homogénéité relative

I.B.3.a L'homogénéitéaussi en fonction de la taille de cellule

Comme vu précédemment, le nuage de points se caractérise par la densité de points moyenne et la répartition de la densité. Les cartes de densité de points de sol (pages 10 et 11) avaient déjà montré une variation de la répartition de la densité en fonction de la taille de cellule. L'objectif du calcul de l'homogénéité relative (notée RH) consiste à étudier cette variation, en isolant les différents échantillons. En théorie, RH devrait augmenter avec la densité moyenne.

Échantillon	Minimum	Très faible	Faible	Moyen	Medium	Élevé	Très élevé	Maximum
Compte de points de sol	1455	3636	6472	7459	8449	14524	20084	133673
D100 (pt.m^{-2})	0.15	0.36	0.65	0.75	0.84	1.45	2.01	13.37

Table: caractéristiques des échantillons

L'adjectif « relative » signifie qu'il y comparaison avec une densité de référence. Chaque échantillon a sa propre densité moyenne par hectare, notée D100 (taille de cellule =100 m)

La formule de RH ressemble de fait à celle de l'écart-type :

$$RH(CS, D100) = \sqrt{\frac{\sum_{n}^{N_{SR}} (D_{n,CS} - D100)^2}{N_{CS}}}$$

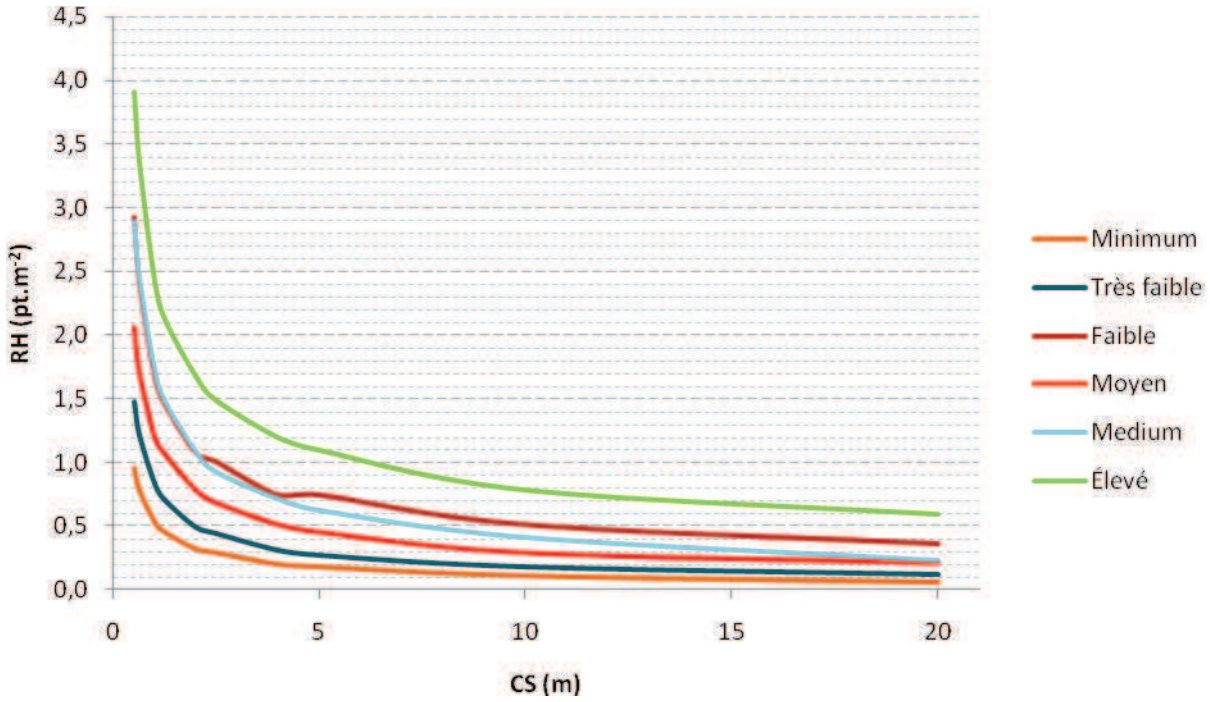
Avec:
- n:une cellule de la grille
-N_{CS}:le nombre total de cellules de la grille, qui varie avec la taille de la cellule
- $D_{n,CS} = \frac{Count(n)}{CS^2}$: la densité dans la cellule n

Sur ArcGIS®, le traitement réutilise les *Fishnets* dessinés pour le calcul de NZER. Une jointure spatiale permet de connaître le compte de points inclus dans chaque cellule. Cela fournit des tables, nommées *GPD_"CS",*qui enregistrent les comptes pour les cellules de la grille à une taille de cellule donnée.En utilisant l'outil *Summarize,* on obtient l'écart-type.

Le choix pour la meilleure taille de cellule se fera sur la base de deux critères :
1)RH < 1 pt.m^{-2}, pour la majorité des échantillons
2)pour une taille de cellule donnée, RH est indépendant de la densité par hectare. Autrement dit, l'écart-type de RH pour la série des échantillons est proche de 0.

I.B.3.b Proportions inverses

Au premier abord, RH semble inversement proportionnel à la taille de cellule, avec une relation de la forme *RH(CS) = f(α/CS)*. Le facteur doit dépendre de CS et D100. En ce qui concerne D100, les courbes s'empilent les unes sur les autres sans se recouper. Les courbes de densités maximales se situent au-dessus. L'échantillon Low fait figure d'exception.

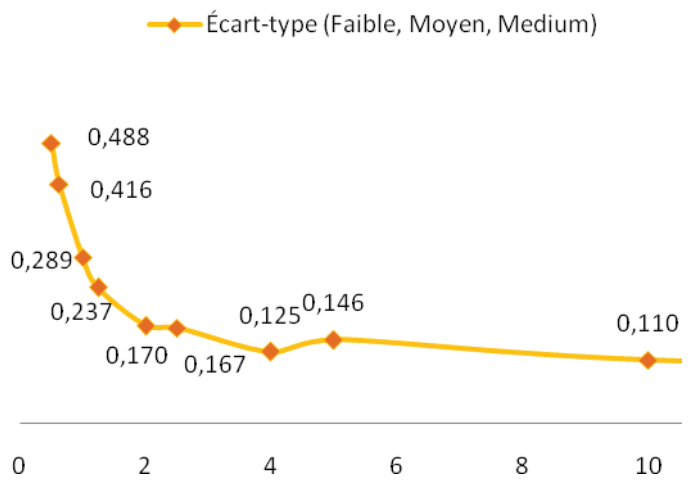


RH en fonction de CS. Les échantillons Max et Very high n'apparaissent pas pour mieux distinguer ce qui se passe pour des densités inférieures.

CS (m)	0.5	0.625	1	1.25	2	2.5	4	5	10	20
Minimum	0.95	0.78	0.53	0.45	0.32	0.29	0.20	0.18	0.11	0.06
Très faible	1.48	1.22	0.83	0.70	0.49	0.44	0.31	0.27	0.18	0.12
Faible	2.92	2.41	1.68	1.46	1.07	1.00	0.74	0.74	0.51	0.36
Moyen	2.06	1.71	1.20	1.06	0.78	0.68	0.51	0.45	0.29	0.21
Medium	2.89	2.45	1.72	1.48	1.08	0.92	0.71	0.62	0.41	0.23
Élevé	3.91	3.36	2.42	2.12	1.68	1.49	1.20	1.09	0.78	0.59
Très élevé	6.15	5.32	3.70	3.26	2.55	2.27	1.85	1.65	1.17	0.81
Maximum	21.22	19.01	16.43	15.62	14.97	14.27	13.46	13.24	12.58	12.04

Table: homogénéité relative (pt.m^{-2})par échantillon et par taille de cellule

Dans la table, les valeurs respectent le premier critère pour CS>2 mdès lors que l'on exclut les échantillons Très élevé et Maximum. De toute manière, l'extrait Maximum fournit des scores bien trop importants. Pour mettre en lumière le second critère, l'écart-type n'a été calculé qu'en considérant les échantillons appartenant aux classe majoritaires : Faible, Moyen et Medium.



Écart-type en fonction de la taille de cellule pour les échantillons Faible, Moyen et Medium.

On dénote un seuil aux alentours de CS=1,5 m. Avant cette limite, l'écart-type diminue de façon importante. Après l'écart-type semble se stabiliser vers 0,1 pt.m^{-2}

→Ni NZER ni RH ne recommandent une taille de cellule idéale. Ils donnent néanmoins des indices sur l'influence des paramètres que sont la précision planimétrique et la densité de points de sol. Une tranche de valeurs de 1,5 à 2,5 m ressort de ces analyses.

I.B.4 Fréquence de comptage

I.B.4.a À l'échelle de la cellule

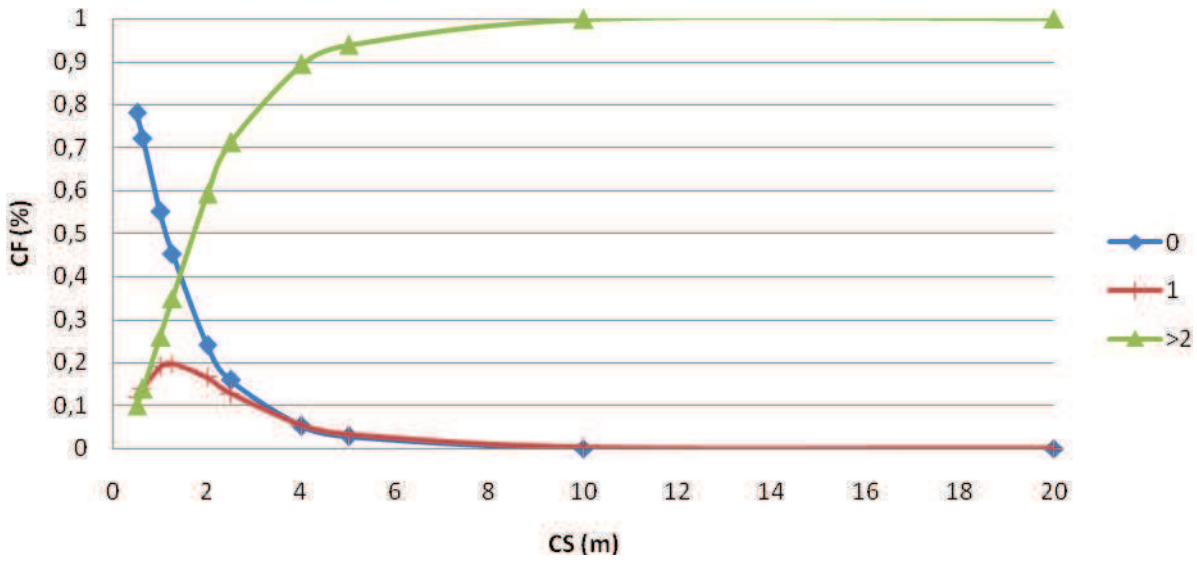
Appelée fréquence de comptage (CF pour Count Frequency), la dernière variable propose un autre point de vue pour l'analyse de la densité de points de sol. Plutôt que de s'intéresser à la répartition spatiale de la densité, CF détermine la fréquence des cellules de comptages identiques. Les champes de comptage étant déjà présents dans les tables GPD_"CS", il suffit d'itérer l'outil *Summarize,* qui va résumer les donnéespar le champ *Count*. On obtient une liste indiquant le nombre de cellules par nombre de points de sol.

La taille idéale devra gérer au mieux un compromis entre les trois critères suivants :
1)Minimumdecellules 'vides', c'est-à-dire dont le comptage vaut 0
2)Maximumde cellules 'singulières', avec exactement 1 point par cellule : noté CF1 par la suite
3)Minimumdecellules sur-denses, i.e. plus de 2 points par cellule

Théoriquement, les cellules vides se raréfient lorsque la taille de cellule augmente, contrairement aux cellules sur-denses. Le nombre de cellules singulières peut varier indépendamment de la taille de cellule.

I.B.4.b L'intérêt pour les cellules singulières

En regroupant les échantillons, on constate sans grande surprise que les minima de cellules vides et sur-denses sont atteints respectivement pour les tailles de cellule maximale et minimale. Leurs courbes se croisent vers CS=1.5 m, où seule 1 cellulesur 5 contient exactement 1 point.Les cellules singulières ont un comportement plus intéressant. En effet, on observe un pic proche de 1,25 m.

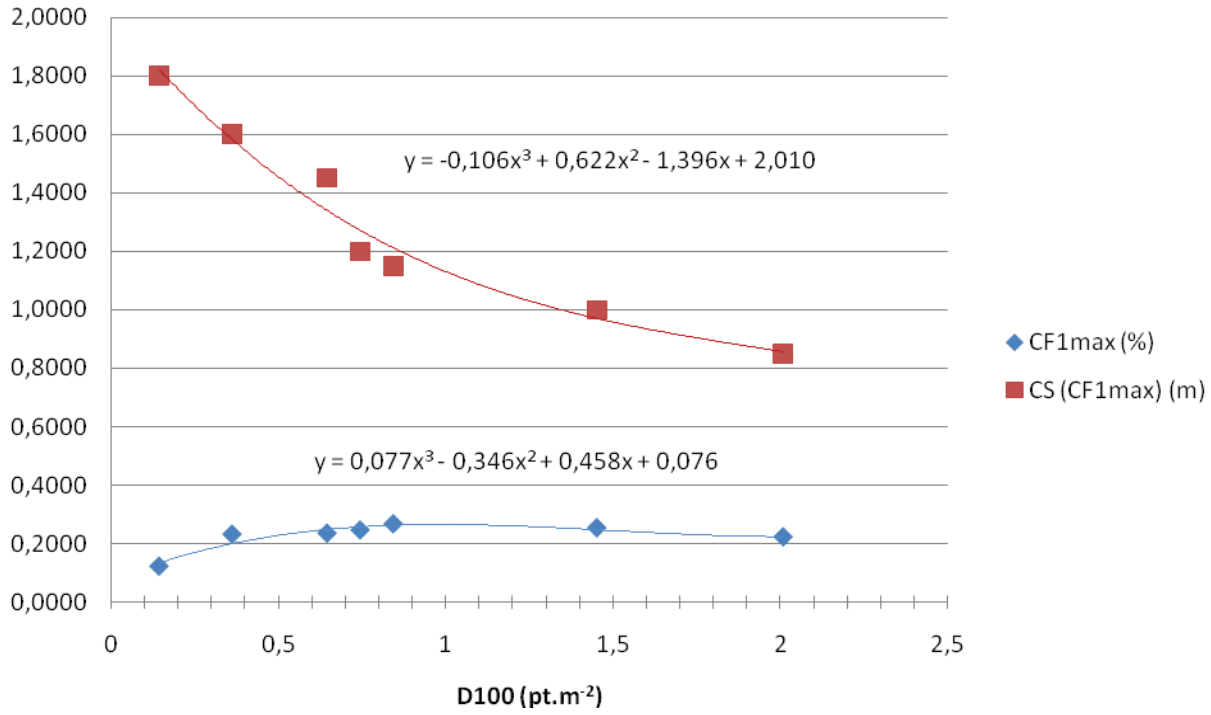


CF en fonction de la taille de cellule, tous échantillons confondus

À échantillons séparés, l'analyse de CF1 - fréquence de comptage des cellules singulières – ce maximum se déplace vers les basses tailles de cellule lorsque la densité augmente. La figure ci-dessous montre les variations du maximum de CF1, et la taille de cellule pour laquelle cette valeur maximale est atteinte. Une approche basique modélise ces courbes par des fonctions polynomiales, toutes deux de degré 3 (formules valables pour D100 < 2 pt.m^{-2}) :

$$CF1max(D100) \approx 0.0771 \times D100^3 - 0.3465 \times D100^2 + 0.4586 \times D100 + 0.0764$$

$$CS_{CF1max}(D100) \approx -0.1063 \times D100^3 + 0.6224 \times D100^2 - 1.396 \times D100 + 2.0107$$



CF1 max et la taille de cellule correspondante en fonction de la densité par hectare

En se rappelant que la densité moyenne situe par hectare aux alentours de 0,58 pt.m^{-2}, la taille de cellule idéale peut être calculée et vaudrait 1,39 m.

→Chaque variable présent un aspectintéressant pour évaluer le comportement de la taille de cellule du raster en fonction des paramètres du nuage de points de sol. Les analyses ont conduit à une estimation d'une tranche comprise allant de 1 à 2 m.Dorénavant, trois valeurs seront utilisées pour les traitements ultérieurs du MNT raster : 1 ; 1,25 et 2 m. Dès à présent, il faut s'attarder sur le second paramètre essentiel du MNT raster : l'algorithme d'interpolation.

II Un modèle de surface à interpoler et à visualiser

II.A Sélection de l'algorithme d'interpolationselon des critères visuels et une simulation de sol

Le principal objectif de cette section consiste à choisir la meilleure méthode d'interpolation du nuage de points de sol. Une présélection s'établira sur le panel d'algorithmes mis à disposition d'ArcGIS® 10.x. Ensuite, la simulation d'un « sol vrai » va mettre en lumière les faiblesses des méthodes choisies. Cette simulation permettra notamment de tester chaque algorithme sur la sensibilité à la configuration du nuage de points ainsi qu'à la présence de bruit.

II.A.1 Deux boîtes à outils : Point to Raster et Raster Interpolation

II.A.1.a Résumé et traitement

Deux approches existent dans le logiciel de SIG :*Point to Raster*et*Raster Interpolation*. L'outil *Point to Raster* calcule des rastersde variables statistiques. Par exemple, *Moyenne* déduit la moyenne des altitudes de l'ensemble des points situés dans chaque cellule. Ici, quatre statistiques importeraient : *Plus fréquent, Minimum, Maximum* et *Moyenne.* Pouvant se révéler très utile pour l'analyse statistique du nuage de points, l'outil néglige les cellules vides. En conséquence, celles-ci ne contiennent aucune information : *No Data Value.*Ceci devient un sérieux problème sachant que la proportion de cellules vides est importante lorsque la taille des cellules est inférieure à 2 m. La boîte à outils *Raster Interpolation*rassemble quant à elle des algorithmes usuels. En voici une brève description.

Algorithmed'Interpolation	Description	Paramètres additionnels
IDW (Inverse Distance Weighting)	*Pondération par la distance inverse* : Combinaison pondérée d'un ensemble de points, dont le poids est un fonction de l'inverse de la distance	Puissance: exposant de la distance Rayon de recherche des points Nombre de points voisins
Kriging *Krigeage*	Procédure géostatistique qui génère une surface estimée à partir d'un ensemble de points. Nécessite la création en amont d'un semi-variogramme qui représente la semi-variance en fonction de la distance entre couples de points.	Modèle de semi-variogramme Rayon de recherche des points Nombre de points voisins
Natural Neighbor *Voisin naturel*	Basé sur deux séries de polygones de Voronoï-Thiessen : la première série détermine le groupe de points les plus proches ; la seconde attribue des poids à chacun des points.	
Spline	Fonction mathématique qui minimise la courbure totale, via une surface passant exactement par tous les points en entrée. L'algorithme Spline avec Barrières inclue des discontinuités déjà enregistrées.	Type de Spline: lissée ou avec contraintes Poids de certains termes de la fonction Nombre de points voisins
Topo to Raster *Topo vers raster*	Représenteune surface d'écoulement naturel.	Beaucoup d'options pour cet algorithme Attention: contrairementaux autres méthodes, il requiert la création préalable d'un champ pour prendre en compte l'altitude.
Trend	Fonction polynomiale globale qui décrit les motifs à grande échelle	Degré du polynôme: jusqu'à 12 Type de régression: linéaire ou

		logistique

Table: résumé des algorithmes disponibles dans la boîte à outils Raster Interpolation. Informations extraites de l'aide d'ArcGIS® (ArcGIS® help)où les méthodes sont détaillées pour plus de renseignements.

En ce qui concerne le traitement, les MNT rasters ont été créés avec une taille de cellule de 1 m. Les paramètres par défaut ont été utilisés, excepté pour la méthode *Trend,* dont le degré du polynôme a été placé au maximum (12).

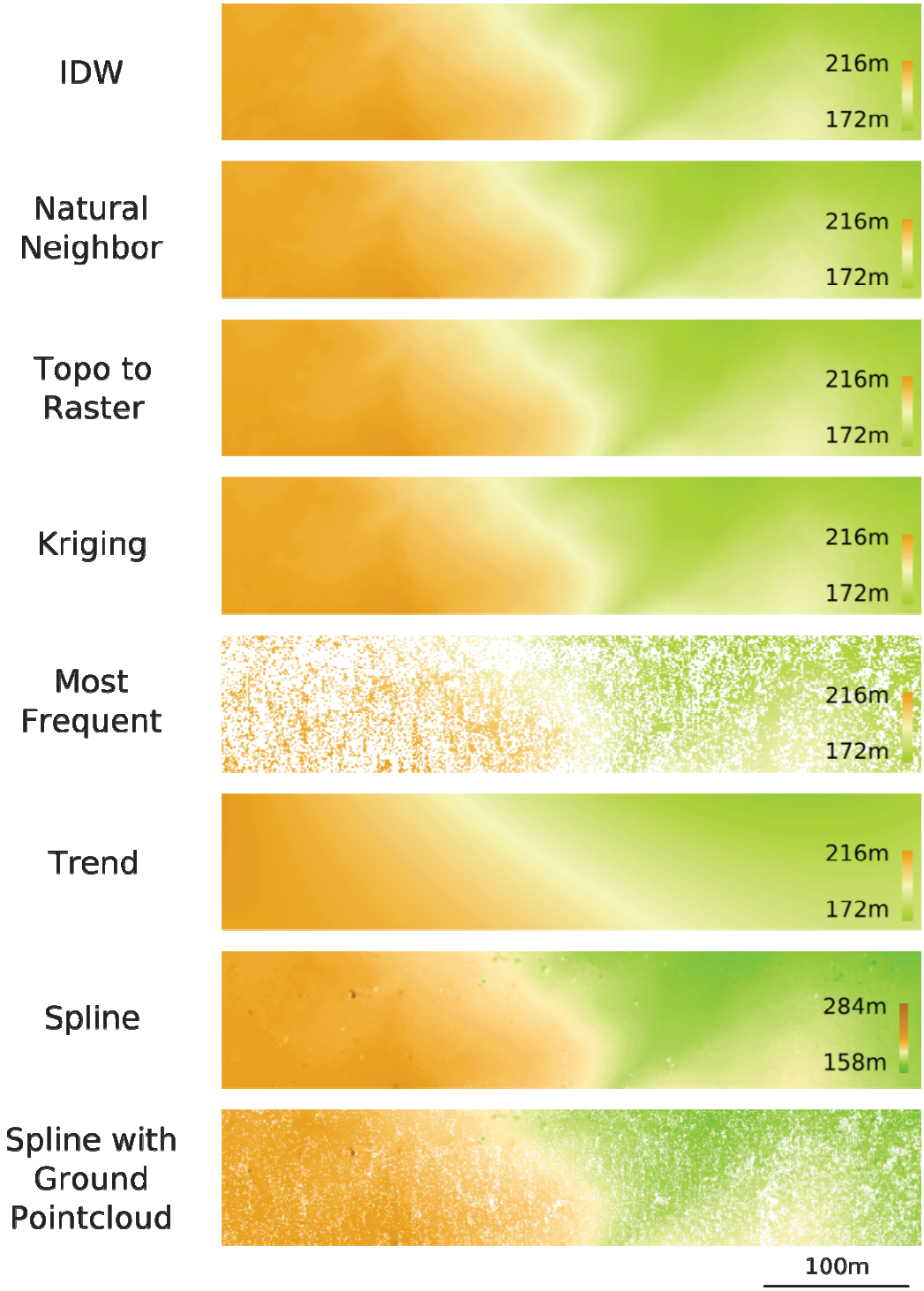


*Illustration: extraites des rasters interpolés avec les différents algorithmes.*Les cellules blanches présentes sur Most Frequent ne doivent pas être confondues avec les points de sol ajoutés sur Spline with Ground Pointcloud. Des taches apparaissent avec la méthode Spline. En les voyant également sur l'extrait Spline with Ground Pointcloud, on peut déduire une corrélation entre ces taches et les zones vides de point de sol.

II.A.1.b Présélection visuelle

Une première comparaison des rasters amène à rejeter l'algorithme Trend, puisque le modèle n'est en aucun cas réaliste et bien trop lissé. Ensuite, la méthode Spline génère des taches dans les régions vides de points de sol. Ces artéfacts élargissent l'étendue des valeurs d'altitude.Et bien entendu, l'outil Point to Raster est difficile à visualiser à cause des nombreuses cellules sans valeur.

Tous les autres algorithmes produisent des modèles similaires. Une simulation d'un 'sol vrai' va approfondir cette analyse.

II.A.2 Simulation d'un 'sol vrai'

II.A.2.a Biais d'interpolation

Le biais d'interpolation(noté IB) est une quatrième reliant nuage de points du sol et MNT raster. Plus précisément, cette variable représente un écart d'altitude entre le point Interpolé et le point Liant pour chaque cellule. À l'échelle de la grille, IB peut être estimée par analogie avec l'erreur quadratique moyenne (EQM) comme suit :

$$IB = \sqrt{\frac{\Sigma(Zl - Zi)^2}{N}}$$

Où Zl and Zi valent les altitudes des point Liant et Interpolépour chacune des N cellules de la grille.

Puisque l'altitude du point interpolé est en lien direct avec les points mesurés, les écarts de positionnement de ces derniers jouent sur l'écart d'altitude entre le point interpolé et le point liant.Devant la difficulté à modéliser des écarts de positionnement, les tests suivants se placeront dans un cas où ceux-ci sont inexistants. Autrement dit, point vrai et mesuré correspondent à la même entité.

Le but de la manœuvre est de contrôler les résultats produits par les divers algorithmes d'interpolation. Un motif en forme de monticule a été conçu et répété dans 9 blocs au total pour constituer le sol 'vrai'. Pour imiter l'acquisition LiDAR, un nuage de point de sol a été extrait sur ce sol. Chaque bloc se contient un ensemble de points dont la configuration (densité et répartition) est unique. Chacun des algorithmes a été appliqué : IDW, Kriging, Natural neighbor, Spline and Topo to raster.Une première analyse va distinguer ces algorithmes selon leur sensibilité à la configuration de points.

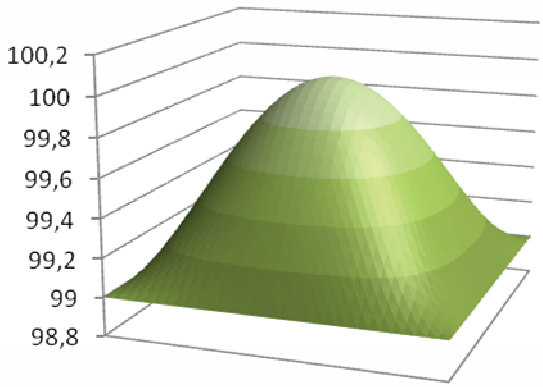
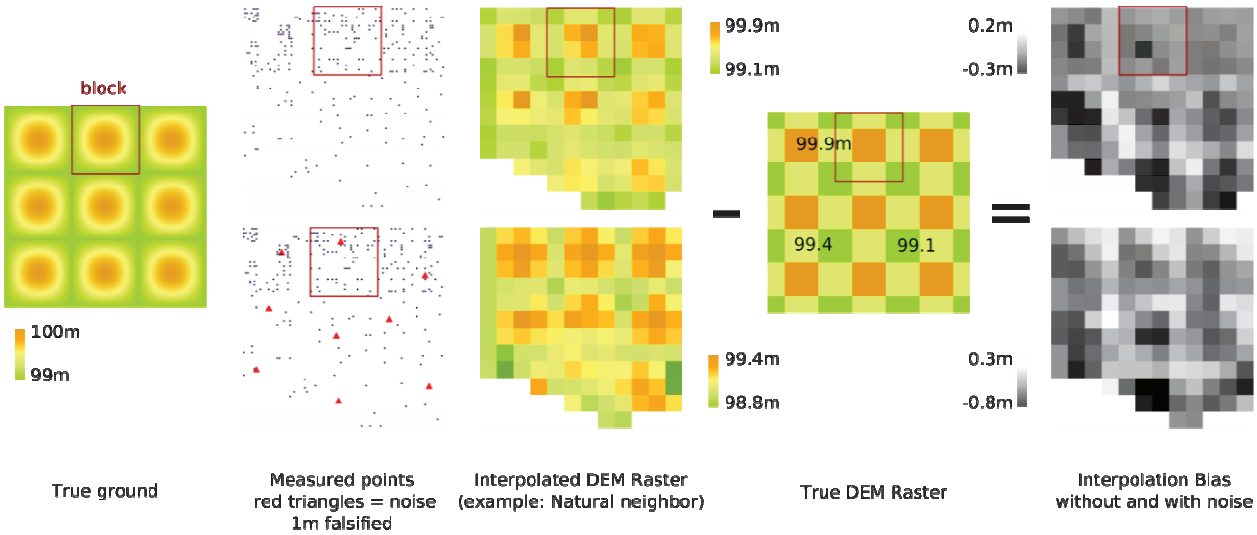


Illustration: motif de monticule

Une seconde série de tests falsifie l'altitude d'un point de chaque bloc. Ceci permettra de déceler les algorithmes sensibles au bruit.



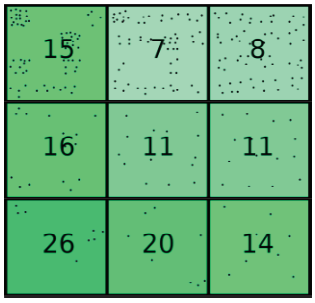
*Schema: étapes de calcul du Biais d'interpolation, dans les conditions 'normales' (au-dessus) et avec ajout de bruit (en-dessous)*NB: les cellules blaches sur le MNT raster interpolé n'ont pas de valeur (No Data value), ce qui est spécifique de l'algorithme Natrual Neighbor.

Ainsi, la décision pour le meilleur algorithme se fera selon trois critères sur IB :
1)*moyenneIB* minimaleen conditions normales
2)*écart-typeIB* minimal en conditions normales
3)*moyenne des écarts entre IB et IB bruité*minimale

NB: Les statistiques se rapportent à l'échantillon des 9 blocs.

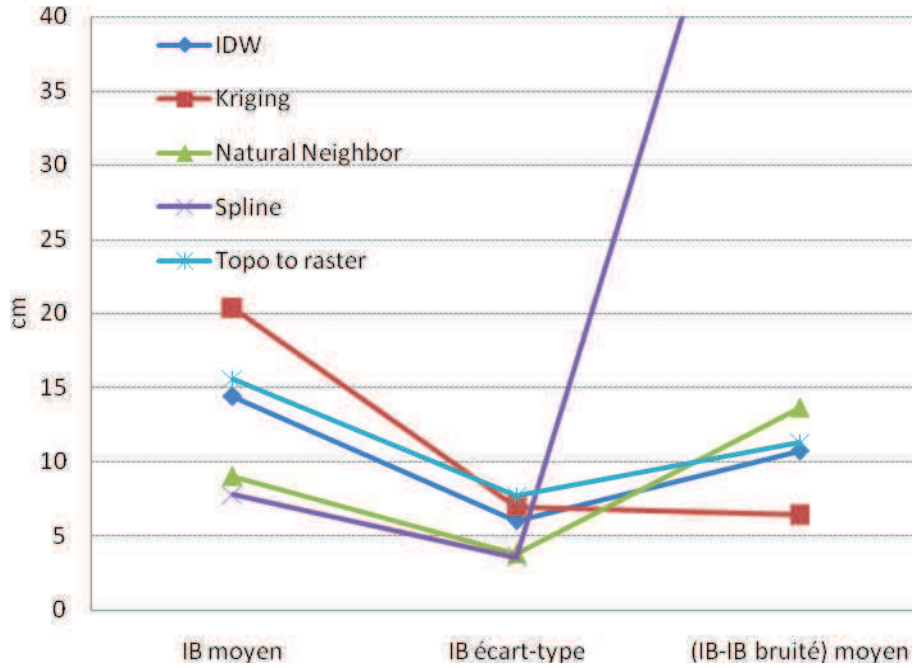
II.A.2.b L'influence du bruit

Globalement, on observe un motif unique pour décrire la variation du biais d'interpolation en fonction de la configuration de point. En effet, quel que soit l'algorithme utilisé, plus le bloc est dense et homogène plus le biais diminue. Ceci semble logique puisque l'interpolation fonctionne davantage avec un réseau régulier et contenant nombre de points. Spline et Natural Neighbor détiennent les meilleurs résultats, avec des moyennes et des écart-type au-dessous de 10 cm et 5 cm. IDW et Topo to raster suivent. Kriging a une moyenne aux alentours de 20 cm.L'écart-type de ces trois dernières demeure en-deçà de 10 cm.



IB en conditions normales (cm) - Exemple IDW

Quand le bruit s'ajoute, l'ordre s'inverse, en particulier au niveau du bloc le moins dense et homogène. Spline par exemple est très influencée par le manque de points et la présence du point falsifié entraîne une erreur autour de 9 m.Cela conduit à un écart (IB-IB bruit) moyen au-delà de 50 cm. Natural Neighbor, Topo to raster et IDW ont de bien plus faibles valeurs avoisinant 10 cm. Kriging est considéré comme l'algorithme le moins sensible au bruit, avec un écart moyen inférieur à 5 cm.



	1)	2)	3)
IDW	14	6	11
K	20	7	6
NN	9	4	14
S	8	4	66
To	16	8	11

Grapheet valeurs pour les trois critères, classés par algorithme

Natural Neighbor remplit au mieux les critères en conditions normales et démontre une certaine résistance au bruit. De plus et contrairement aux autres méthodes, il ne force pas l'interpolation d'une surface si la configuration de points est très mauvaise. Dans les blocs de faible densité, les polygones de Voronoï n'ont pu être dessinés. C'est la raison qui explique la création de cellules sans valeur. Ceci constitue un avantage intéressant, comparé à Spline qui pour le coup a créé une surface très éloigné de la réalité.

→De ces analyses préliminaires, des choix ont pu être faits en vue de réaliser le meilleur MNT raster. Si l'algorithme Natural Neighbor a été sélectionné comme le seul pouvant discrétiser l'altitude avec le minimum de perte de qualité, le choix quant à la taille de cellule a été plus large. En conséquence, les rasters seront dessinés à l'une des trois tailles de cellule suivantes : 1 m, 1.25 m and 2 m. Cependant, les rasters produits ne satisfont toujours au critère de visibilité des sites archéologiques. Certes les grandes structures se distinguent sans trop de aisément, mais les structures plus petites restent cachées dans le relief en niveaux de gris. L'étape suivante du traitement, appelée Modélisation de visualisation, a pour but d'améliorer la visibilité de tous les types de sites, en ayant recours à la pente notamment.

II.B Le besoin d'un modèle de visualisation : l'altitude ne suffit pas

Maintenant que les paramètres du MNT raster ont été ciblés, un nouvel obstacle se dresse contre la détection des sites mayas. L'altitude seule n'aide pas à distinguer d'autres structures que les sites les plus évidents. Des algorithmes dits 'de visualisation' existent pour améliorer la visibilité de la plupart de ces éléments. Les plus utilisés seront brièvement décrits et comparés. Puis, des modèles de visualisation imiteront ces algorithmes en reprenant les principes théoriques. Finalement, les rasters de visualisation obtenus amèneront à une première série d'observations, dans le but de ne garder que les modèles les plus adaptés à l'interprétation.

En plus de s'intéresser aux paramètres du MNT raster (3 tailles de cellule, 1 algorithme d'interpolation), un choix supplémentaire se fera sur la base de critères qualitatifs:

Critère	Signification
Bonne lisibilité à petite échelle	La plupart des éléments doivent être clairement identifiables à une échelle usuelle (ici 1/1000)
Lissage minimal	Les petits éléments ne doivent pas être floutés
Bruit minimal	Limiter l'influence des points faux, c'est-à-dire les points hors sol
Logique topographique	La pente et le paysage demeurent simples d'interprétation
Indépendant vis-à-vis de l'orientation	L'algorithme est valable quelle que soit l'orientation
Distinction éléments positifs/négatifs	Par exemple, les monticules et les fossés peuvent être identifiés sans ambiguïté
Temps de calcul minimal	Limiter les coûts en temps
Bonne lisibilité à grande échelle	Eventuellement, regarder à une grande échelle aide à comprendre la situation globale.

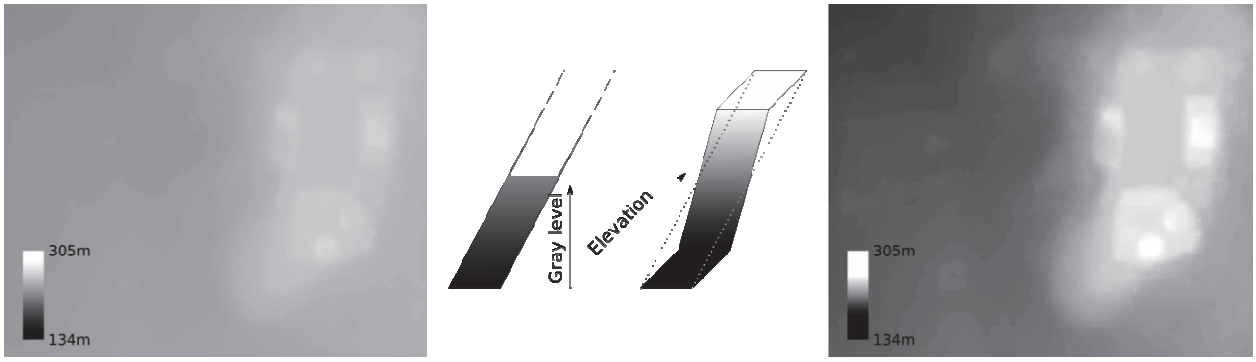
Ce qui fait un bon modèle de visualisation

II.B.1 Algorithmes usuels

Nombre d'algorithmes ont été programmés pour créer des rasters de visualisation, qu'il s'agisse des méthodes basiques et courantes ou bien de procédés plus élaborés et spécifiques. Un tour d'horizon s'appuie sur une étude récente [Kokalj,2012]pour donne un aperçu des expérimentations actuelles.

II.B.1.a Color Cast (distribution des couleurs)

Il s'agit d'une méthode basique de traitement d'image, qui modifie l'histogramme des valeurs de gris de telle sorte que le niveau de détail soit augmenté. Le cas typique consiste à employer une fonction linéaire délimitée par des seuils. La relation entre les deux échelles de gris demeure simple et quasi-proportionnelle.



Principe de Color Cast, avec l'exemple de seuils et fonction linéaire

II.B.1.b Relief ombragé - Hill-shaded Relief

La couleur en niveau de gris se fait en fonction du relief. Trois possibilités existent :

- pente (plat >blanc ; accidenté>noir): il s'agit de Slope Severity (inclinaison de la pente)
- lumière oblique : le relief est éclairé par une source artificielle dans une direction donnée (généralement : 315° horizontal et 45° vertical)
- combinaisondes deux ombrages

Le relief ombragé est une technique fréquemment utilisée pour la visualisation. La simplicité d'interprétation a un lien avec l'habitude d'observer des photographies semblables et éclairées par des lumières rasantes [Crutchley, 2009] De fait, les ombres et zones de pleine lumière sont aisément détectables, puisque les utilisateurs sont habités à en voir en imagerie.

Cependant, une seule image en relief ombragé peut ne pas fournir suffisamment d'information, parce que certains détails restent cachés dans l'ombre. La méthode statistique ACP (Analyse en composantes principales –PCA en anglais) permet de gérer un ensemble de rasters pour résumer et concentrer un maximum d'information sur 2 à 3 rasters [Devereux, 2008].

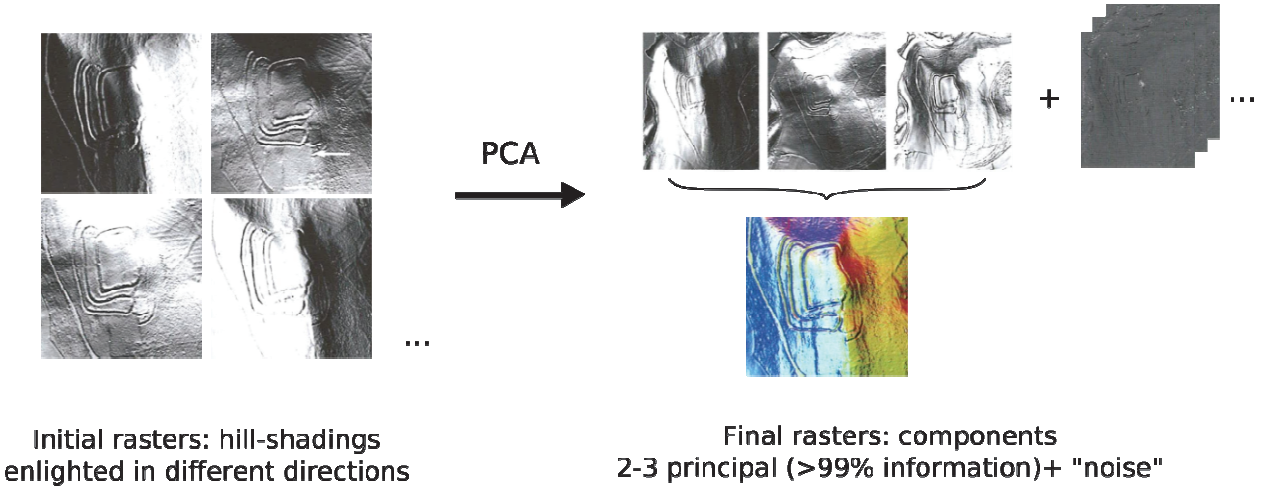


Illustration: ACP de relief ombragés(images extraites de [Devereux)

II.B.1.c Modèle de rayonnement solaire (SIM - Solar Insolation Modeling)

Le niveau de gris correspond à la quantité de lumière reçue par une surface élémentaire. Le rayonnement peut être aussi bien global que diffus.

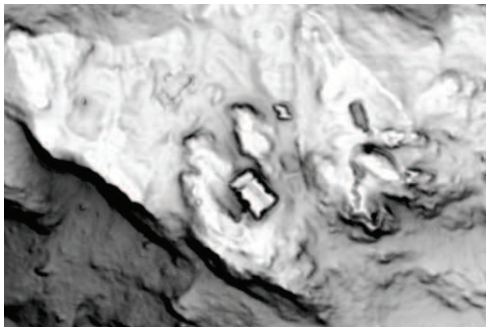


Illustration: exemple de rayonnement diffus (from Kokalj)

II.B.1.d Facteur de vue du ciel (SVF - Sky View Factor)

SVF représente la part de ciel visible depuis un point de vue. Un hémisphère centrésur ce point est divisé en quartiers. Les angles zénithaux sont calculés dans les différentes directions.SVF relie ces angles [Kokalj, 2012] :

$$SVF = 1 - \frac{\sum_{i=1}^{n} \sin \gamma_i}{n}$$

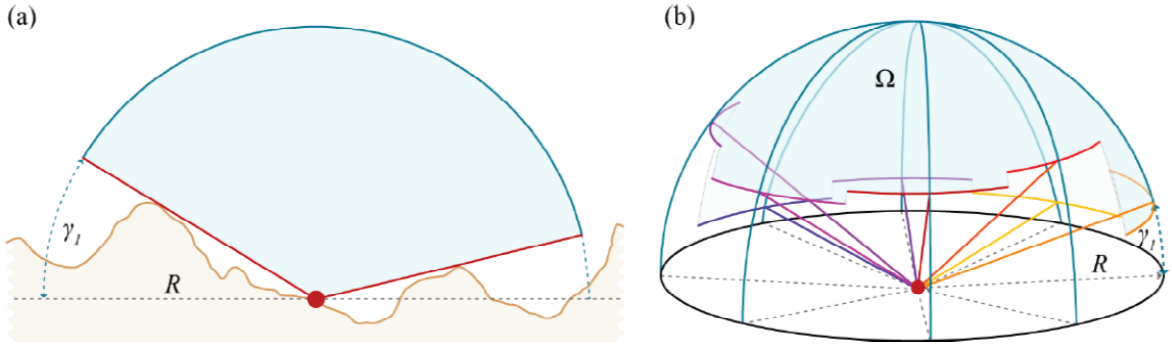


Illustration (Kokalj 2012): schémas du SVF, calculé ici dans 8 directions

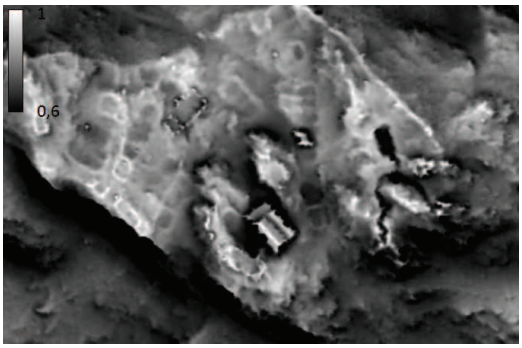


Illustration: extrait de SVF (from Kokalj)

II.B.1.e Modélisation locale du relief (Trend Removal- Local Relief Model)

Le relief est modélisé à localement et retranché à un relief à grande échelle. Ceci donne une représentation des aspérités locales du relief telles que les monticules et les fossés. Parmi les algorithmes d'interpolation les plus utilisés pour modéliser localement le relief, on dénombre : altitude moyenne, plan, bilinéaire, lissage gaussien. Hesse [2008] pense à concevoir un MNE 'purgé' (Modèle Numérique d'Élévation).

Bien que cette méthode fasse apparaître les éléments du paysage uniquement, deux inconvénients liés sont notables : l'interpolation crée une surface 'lissée' qui non seulement efface les petites structures mais cela peut créer des structures artificielles, ce qui ajoute du bruit.

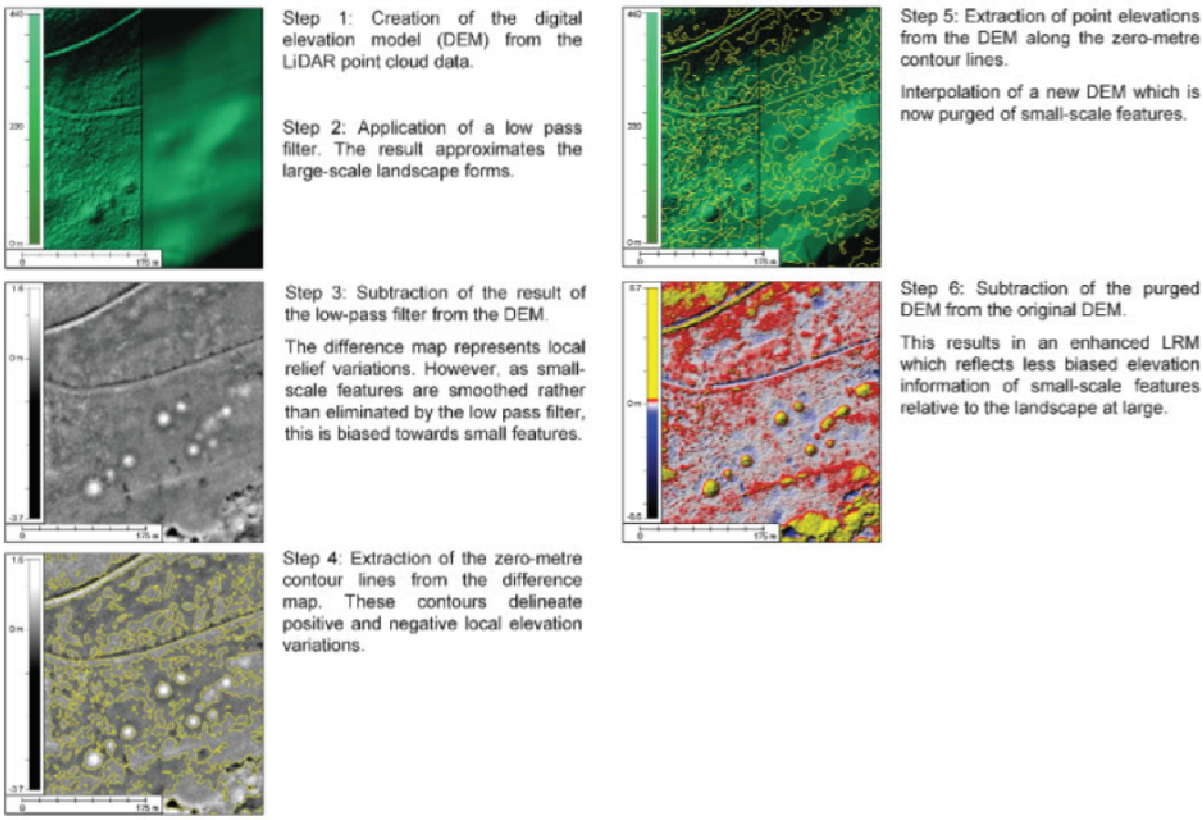


Illustration (Hesse 2008): processus du Local Relief Modeling

II.B.2 Choix a priori

Une classification fut déjà réalisée [Challis, Kokalj et al.], listant les principaux atouts et défauts de chaque algorithme. Tout d'abord, Trend Removal est à exclure pour le lissage et la création d'artifices.En ce qui concerne les autres aspects négatifs, SVF et SIM engendrent un effet parasite appelé 'Washout effect', lequel tend à effacer les structures de faible hauteur dans les reliefs plats. Color Cast résoud ce problème, en affichant plus de détails dans ces types de terrain.

Au contraire, Color Cast semble inefficace en terrain. De même, Hill-shading est certes facile à visualiser, mais les dépressions et les formes linéaires dans la direction de l'éclairement souffrent ne se démarquent pas. La méthode ACP aide à leur découverte via la superposition complexe de plusieurs rasters.

Bien que SVF and SIM soient complexes et inhabituelles pour l'utilisateur, ces algorithmes fournissent de bons résultats, que ce soit pour la gestion de l'échelle de gris ou encore la distinction d'éléments 'positifs' (monticules) et 'négatifs' (fossés). Les éléments apparaissent sans de lissage ou de bruits supplémentaires.

Méthode	Avantages	Inconvénients
Relief ombragé *Hill-shading*	▪ Facile à programmer ▪ Disponible sur la plupart des logiciels spécialisés en SIG	▪ Les zones d'ombre et de sommet demeurent sombres et claires. ▪ Difficile de détecter les éléments linéaires parallèles à la direction de l'éclairement
ACP de reliefs ombragés *PCA of Hill-shadings*	▪ Élimine la redondance d'information	▪ Difficile à interpréter lorsque le nombre de composantes (rasters) augmente
Distribution descouleurs *Color cast*	▪ Utile en terrain plat ▪ Facile à interpréter ▪ Cohérence maintenue entre le niveau de gris et l'altitude	▪ Inutile en terrain accidenté ▪ Requiert un traitement supplémentaire
Modélisation locale du relief *Trend removal*	▪ Utile pour initier un autre algorithme	▪ Lissage ▪ Bruit
Facteur de vue du ciel *Sky View Factor*	▪ Pas de saturation de gris ▪ Distinction entre dépressions and endroits surélevés ▪ Efficace pour des structures complexes ▪ Utile si les données sont très bruitées	▪ "Washout effect": sur des terrains plats, les éléments peu élevés s'estompent.
Modèle de rayonnement solaire *Solar Insolation Modeling*	▪ Topographie maintenue ▪ Adapté pour l'étude d'activités humaines	▪ Complexeet chronophage ▪ De nombreux choix à faire ▪ Washout effect

Comparaison des algorithmes de visualisation récurrents (de Kokalj 2012)

Quatre algorithmes ont été sélectionnés pour les tests ultérieurs[Challis, 2011] [Kokalj, 2012]. Deux paramètres ont été retenus, à savoir le relief et la hauteur des éléments (appelée 'intensity of earthworks'). L'algorithme Color Cast se révèle le plus adaptéen relief plat. Lorsque le terrain devient de plus en plus accidenté, SVF est à préconiser, devant Slope Severity et SIM. La méthode PCA of Hill-shaded reliefs n'est pas indiquée dans le tableau ci-dessous mais sera également utilisée.

Relief Hauteur des éléments	Plat	Accidenté
Faible	Color cast SVF SIM	SVF Slope severity SIM
Importante	Color cast Slope severity SIM	SVF Slope severity SIM

Table: choixdes algorithmes de visualisation à l'aide de 2 paramètres(from Kokalj)

II.C Pseudo Modèles

Une fois la présélection établie, les méthodes de visualisation vont être appliquées.

L'idée ici revient à comparer des modèles inspirés des vraies méthodes. Bien que les algorithmes soient disponibles dans bon nombre de logiciels spécialisés en SIG et traitement d'image, les résultats maquent souvent d'explications. Les données restent difficiles à interpréter.

Software	Operating system	Status	Colour shading	Slope analysis	Hill shading	PCA[a]	Terrain filtering	Solar insolation
ArcGIS 9/10	Windows	Commercial	X	x	x	x	x	x
ArcView 3.1	Windows	Commercial	X	x	x		x	x
Envi	Windows	Commercial	X	x	x	x	x	x
Grass	Windows/Linux/Mac	Open Source	X	x	x	x	x	x
Idrisi	Windows	Educational	X	x	x	x	x	
Imagine	Windows	Commercial	X	x	x	x	x	x
MapInfo	Windows	Commercial	X	x	x			
QGis	Windows/Linux/Mac	Open Source	X	x	x			
Saga GIS	Windows/Linux	Open Source	X	x	x		x	x

[a]Principal component analysis.

Figure: logiciels de SIG et de traitement d'image proposant des fonctions d'analyse LiDAR (d'après Challis et al., 2011, A Generic Toolkit for the Visualization of Archaeological Features on Airborne LiDAR Elevation Data, Archaeological Prospection)

Ce sont les raisons pour lesquelles des modèles vont être conçus en s'inspirant des théories. Il s'agit de créer des filtres dits de convolution, très utilisés en traitement d'image.

De plus, construire ces modèles permettrait par la suite d'analyser plus facilement les différents types d'éléments paysagers : positif/négatif, planaire/linéaire/coin/ponctuel, point faux. En observant les résultats de chacun des modèles face aux divers cas de paysage, il est possible d'imaginer une série de seuils dans l'optique d'une classification et de détection d'objets.

II.C.1 Traitement aveceCognition®etArcGIS®

La création des rasters de visualisation se base sur un principe fondamental en traitement d'image : le produit de convolution. En quelques mots, ce produit implique deux matrices : le raster initial et un filtre de convolution. Ce dernier représente une fenêtre de dimensions fixées, généralement un carré de 3x3 pixels, qui se déplace sur l'ensemble du raster. En un pixel donné, le résultat issu du produit de convolution égale la somme des produits des valeurs des pixels superposés, ainsi que le montre l'illustration.

NB: par convention, * est le symbole utilisé pour représenter le produit de convolution.

7	8	9
12	13	14
17	18	19

*

0	1	0
0	0	0
0	0	0

=

0x7 + 1x8 + 0x9
+ 0x12 + 0x13 + 0x14
+ 0x17 + 0x18 + 0x19

=

8

Raster intial

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

*

Filtre

0	1	0
0	0	0
0	0	0

=

Raster final

*	*	*	*	*
1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20

Figure: exemple de produit de convolution, avec un filtre 3x3

ArcGIS® dispose d'une fenêtre appelée *Image Analysis*, où les rasters peuvent être modifiés en améliorant notamment le contraste, la luminosité ou en appliquant les filtres de convolution usuels (Gradient, Laplacien, Sobel). Cependant, il est impossible de créer des filtres de convolution. Pour contourner ce problème, des rasters appelésMNT rasters Unités ont été calculés, et ce afin d'obtenir l'altitude dans un voisinage de 7x7 pixels. Ceux-ci ont été obtenus via eCognition®, un logiciel dédié à la classification et la détection d'objets. Par la suite, l'outil *Spatial Analyst/Map Algebra/Raster Calculator*, imitera le produit de convolution, en appliquant les modèles mathématiques et ces rasters unités.

Z(-3 -3)	Z(-3 -2)	Z(-3 -1)	Z(-3 0)	Z(-3 1)	Z(-3 2)	Z(-3 3)
Z(-2 -3)	Z(-2 -2)	Z(-2 -1)	Z(-2 0)	Z(-2 1)	Z(-2 2)	Z(-2 3)
Z(-1 -3)	Z(-1 -2)	Z(-1 -1)=Z1	Z(-1 0)=Z2	Z(-1 1)=Z3	Z(-1 2)	Z(-1 3)
Z(0 -3)	Z(0 -2)	Z(0 -1)=Z4	Z(0 0)=Z5	Z(0 1)=Z6	Z(0 2)	Z(0 3)
Z(1 -3)	Z(1 -2)	Z(1 -1)=Z7	Z(1 0)=Z8	Z(1 1)=Z9	Z(1 2)	Z(1 3)
Z(2 -3)	Z(2 -2)	Z(2 -1)	Z(2 0)	Z(2 1)	Z(2 2)	Z(2 3)
Z(3 -3)	Z(3 -2)	Z(3 -1)	Z(3 0)	Z(3 1)	Z(3 2)	Z(3 3)

Tailles de filtres: 3x3, 5x5, 7x7

Illustration:*voisinage 7x7 étudié pour produire les 49 DEM rasters Unités*Par exemple, z(-3 2) représentel'altitude au pixel situé 3 lignes au-dessus et 2 colonnes à droite du pixels centralz(0 0).

II.C.2 Les modèles mathématiques inspirés des algorithmes usuels

Détaillons dès à présent les pseudo-modèles réalisés à partir des notions des algorithmes introduits précédemment. Des extraits d'image viendront illustrer la partie suivante intitulée 'observations préliminares'.

II.C.2.a Arctan Color Cast

Pensée à la base pour mieux afficher les niveaux de gris de l'altitude, Color Cast a été appliquée ici à la pente. D'un point de vue archéologique, le raster de visualisation permettrait de faire apparaître tous les sites, et ce quelle que soit l'inclinaison du terrain. De plus, on accorde plus d'intérêt aux structures peu élevées. Le but de cette méthode se résume à distinguer trois classes de pente : plates, faibles et fortes. Comme indiqué sur l'exemple ci-dessous, des seuils séparent ces catégories, en insistant sur les pentes faibles susceptibles de recueillir le maximum de sites archéologiques. Il s'agit donc d'offrir une palette de gris maximale pour faire ressortir les sites tout en maintenant dans l'ombre et la lumière les pentes maximales, positives comme négatives. Les régions parfaitement planes sont teintées en gris intermédiaire.

Au lieu de recourir à une fonction linéaire pour rééchelonner les niveaux de gris, l'idée amène à employer une combinaison de fonctions Arctan, avec trois contraines :

- Seuil de première distinction (ε): le nombre de niveaux de gris destiné aux pentes quasi-nulles
- Point d'inflexion (δ): la pente médiane dans la classe des pentes faibles
- Pente faible minimale (α): point de départ des pentes faibles

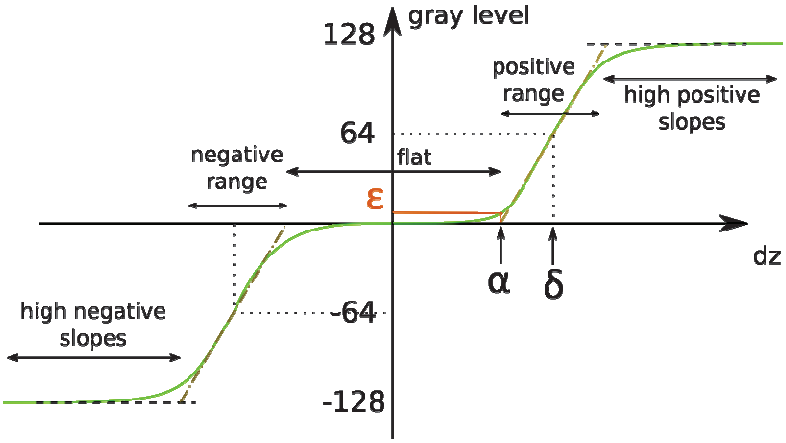


Illustration: les contraintes de Arctan Color Cast

$$CC(dz) = \frac{256}{\pi\left[p_3 \tan^{-1}\left(\frac{dz}{\delta}\right)^3 + p_5 \tan^{-1}\left(\frac{dz}{\delta}\right)^5 + p_7 \tan^{-1}\left(\frac{dz}{\delta}\right)^7\right]}$$

La fonction se compose de trois termes pondérés. Cela donne l'opportunité de déterminer la fonction idéale pour bien représenter les pentes que l'on souhait observer et selon des critères prédéfinis. Dans ce cas, deux triplets ont été testés, CC test1 and CC test2. Les rasters issus de ces tests se complètent de par le comportement de leurs fonctions.

Paramètre	ε (niveaux de gris)	δ (m)	α (m)	p3	p5	p7
Test1	10	0.75	0.5	-0.22	1.58	-0.36
Test2	3	0.2	0.4	20.46	-36.63	17.17

NB: la pente n'a été calculée que dans une direction, c'est-à-dire Ouest/Est.

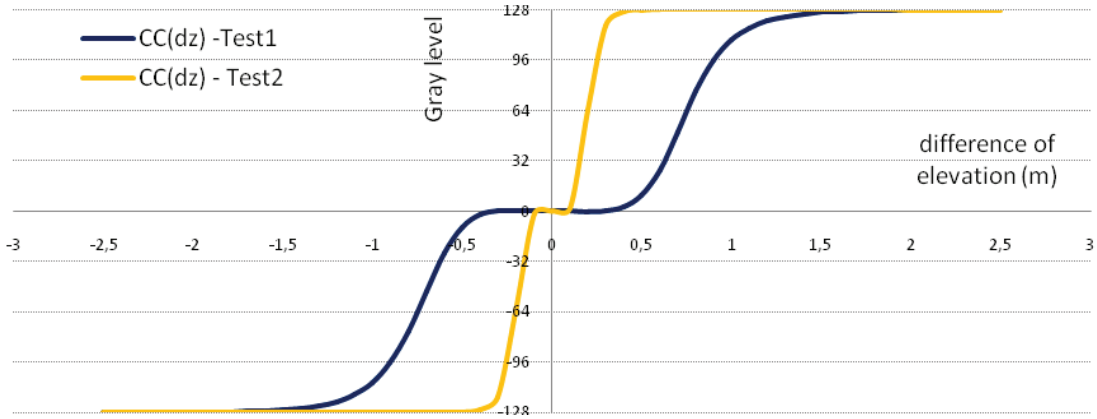
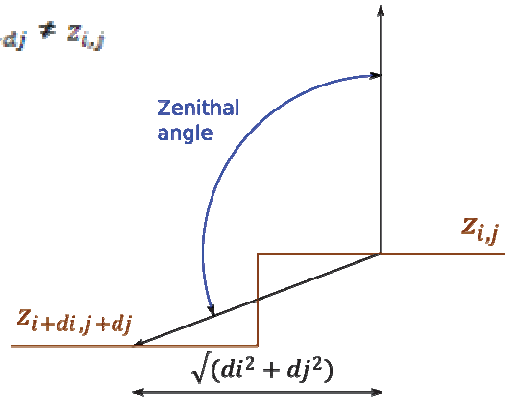


Illustration: les deux fonctions Color Cast testées et la table de leurs poids associés

II.C.2.b Pseudo SVF

Comme l'algorithme officiel, le pseudo SVF détermine le rapport entre un angle zénithal et l'hémisphère en partant d'un point de vue et dans une certaine direction:

$$SVF_{di,dj}(i,j) = \begin{cases} 1 - \frac{1}{\pi}\tan^{-1}\left(\frac{\sqrt{[(di]^2 + dj^2)}}{z_{i+di,j+dj} - z_{i,j}}\right) & if\, z_{i+di,j+dj} \neq z_{i,j} \\ \frac{1}{2} & else \end{cases}$$



Avec:

- (i,j): location du pixel central
- (di,dj): déplacement selon les lignes et les colonnes

Illustration: principe de base du SVF

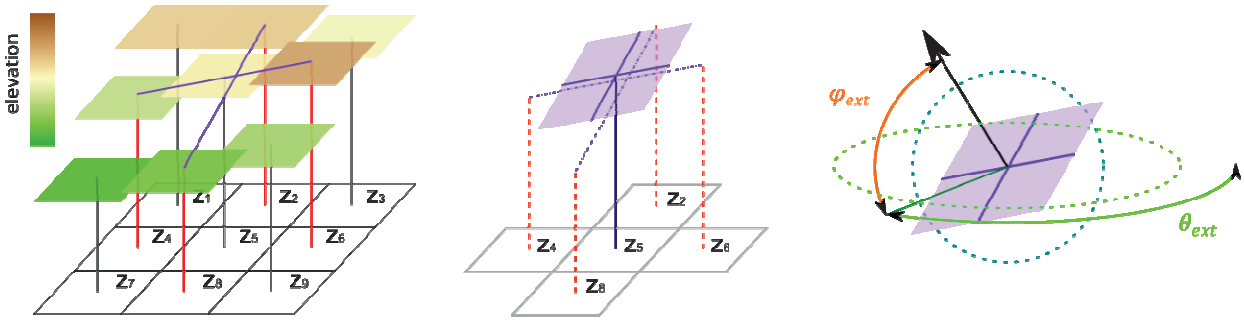
La formule relie le pixel central à l'in de ses voisins. Dans la cadre du pseudo-modèle, l'objectif est de calculer une valeur significative de l'ensemble du voisinage, pour ne dépendre d'aucune direction. L'expérience a montré que l'écart-type satisfaisait cette condition. Ainsi, cette variable a été retenue pour représenter ce pseudo-modèle. En outre, pour analyser l'influence de la taille de fenêtre, trois tailles vont y être appliquées : 3x3, 5x5 and 7x7.

NB: pour each taille de fenêtre, seuls les pixelsau bord du filtre seront considérés. Par exemple, le 5x5 prend en compte la couronne centrale du voisinage 7x7 (en orange sur l'illustration précédente)

II.C.2.c Pseudo SIM

De la même manière, la méthode SIM a été revisitée, et ne cherche pas à recalculer quelle que radiation solaire que ce soit. À un pixel donné, l'objectif consiste à déterminer les angles –horizontal et vertical respectivement notés θ et φ–pour la direction de la lumière telle que l'éclairement reçu soit maximal.

La surface considérée correspond au plan passant par le pixel central et incliné par les pentes Nord/Sud et Ouest/Est..



Illustrations: étapes dans le pseudo-modèle SIM

$$z(i,j) = \frac{(z_8 - z_2)}{2CS} i + \frac{(z_6 - z_4)}{2CS} j + z_5$$

$$\varphi_{ext} = \begin{cases} \tan^{-1}\left(-\frac{2CS}{\cos\theta_{ext}(z_8 - z_2) + \sin\theta_{ext}(z_6 - z_4)}\right)[\pi] si \tan\theta_{ext} \neq \frac{z_8 - z_2}{z_4 - z_6} \text{ ou } (z_2 \neq z_8 \text{ et } z_4 \neq z_6) \\ \frac{\pi}{2} \text{ sinon} \end{cases}$$

$$\theta_{ext} = \begin{cases} \tan^{-1}\left(\frac{z_4 - z_6}{z_8 - z_2}\right)[\pi] \text{ if } z_8 \neq z_2 \\ \frac{\pi}{2} \text{ else} \end{cases}$$

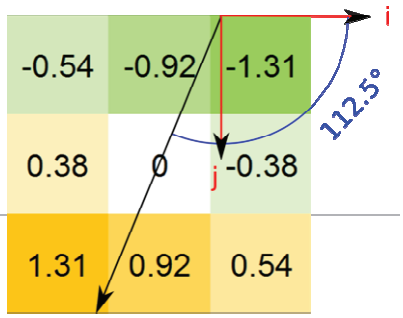
Où : etCS: taille de cellule (m)

Comme pour SVF, une variable apporte un support de visualisation intéressant. D'ailleurs, φ_{ext} semble avoir des caractéristiques similaires à celles de l'écart-type SVF.

II.C.2.d ACP de reliefs ombragés - PCA of Hill-shadings

En dépit des outils disponibles sur ArcGIS®, les reliefs ombragés ont été créés via eCognition®, en utilsant des filtres de convolution appelés Orientated Convolution Filters. Dans une direction horizontale donnée et notée θ, le filtre 3x3 applique des poids trigonométriques à chacun des pixels voisins en fonction de leur proximité de la droite passant par le pixel central et d'orientation θ. En conséquence, un ensemble de rasters, appelés Orientation First Derivative (OFD), est constitué selon le nombre d'orientations souhaité. Chaque raster correspond à un gradient : la pente du terrain peut être déduite.

$$OCF(\theta) = \left((j-2)\cos\theta + (i-2)\sin[\theta)\right]_{(i,j)\in[\![1,3]\!]^2}$$



$$= \begin{bmatrix} -\cos\theta - \sin\theta & -\sin\theta & \cos\theta - \sin\theta \\ -\cos\theta & 0 & \cos\theta \\ -\cos\theta + \sin\theta & \sin\theta & \cos\theta + \sin\theta \end{bmatrix}$$

$$OFD(\theta) = OCF(\theta) * DEM$$

$$Slope(\theta) = \frac{OFD(\theta)}{6\ SR}$$

Illustration: exemple de OCF àθ=112.5°

8 orientations ont été utilisées, de 0° to 157.5° par pas de 22.5°, à une taille de cellule de 1.25 m. Intégrés dansArcGIS®, les Orientated First Derivative rasters servent de données d'entrée pour réaliser l'ACP, grâce à l'outil *Spatial Analyst / Multivariate / Principal Components.* Le choix du nombre de composantes est libre. Les trois premières composantes comptent ; les rasters restants constituent le bruit image. Dans la cadre des modèles de visualisation, seule la première composante a été étudiée.

NB: puisque le filtre OCF est une fonction π-antipériodique, les rasters OFD produitsle sont également. Ainsi, il est inutile d'inclure les orientations rasters OFD de180° à 337.5°: 8 rasters suffisentpour l'ACP.

$$OFD(\theta + \pi) = OCF(\theta + \pi) * MNT = -OCF(\theta) * MNT = -OFD(\theta)$$

II.C.2.e Au-delà la variation de la pente : Orientated Second Derivative extrema

L'étude de la pente permet de détecter des sites : plus c'est clair, plus c'est plat. Pour aller plus loin, il serait intéressant de se consacrer à la variation de la pente. D'un point de vue archéologique, cela permettrait de distinguer les éléments naturels et culturels, selon le critère que les bords des sites culturels sont plus abrupts.

Dans ce cas, Orientated Second Derivative rasters découlent des Orientated First Derivative en appliquant une seconde fois les OCF. Les rasters OSD représentent toujours des gradients.

$$OSD(\theta) = OCF(\theta) * OFD(\theta) = OCF(\theta) * (OCF(\theta) * MNT)$$

Une analyse mathématique a démontré que cette formule pouvait se décomposer simplement en une combinaison de termes trigonométriques comme suit :

$$OSD(\theta) = (\sin[\theta])^2 A + \sin(2\theta)\, B + (\cos[\theta])^2 C$$

Avec : $A = \begin{bmatrix} 1 & 2 & 3 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ -2 & -4 & -6 & -4 & -2 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 3 & 2 & 1 \end{bmatrix} * MNT$; $B = \begin{bmatrix} 1 & 1 & 0 & -1 & -1 \\ 1 & 1 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 1 & 1 \\ -1 & -1 & 0 & 1 & 1 \end{bmatrix} * MNT$; $C = \begin{bmatrix} 1 & 0 & -2 & 0 & 1 \\ 2 & 0 & -4 & 0 & 2 \\ 3 & 0 & -6 & 0 & 3 \\ 2 & 0 & -4 & 0 & 2 \\ 1 & 0 & -2 & 0 & 1 \end{bmatrix} * MNT$

De cette expression, la dérivée seconde oscille de façon sinusoïdale entre deux extrema :

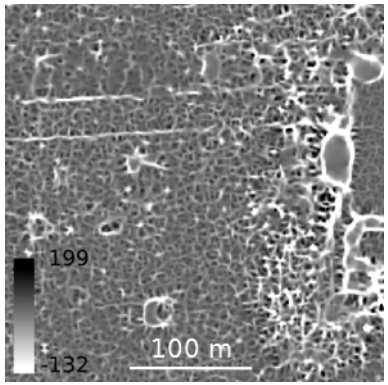
$$\theta_{ext} = \frac{1}{2}\tan^{-1}\left(\frac{2B}{C - A}\right)\left[\frac{\pi}{2}\right]$$

L'étude de la dérivée seconde amène la création de rasters complémentaires, appelés *OSDext1*et*OSD ext2*, qui renseignent sur les pentes maximale et minimale pour chaque pixel.

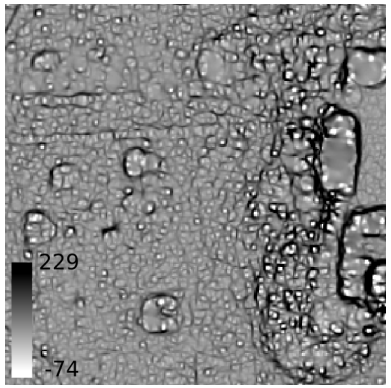
II.C.3 Observations préliminaires

Au total, 15 rasters de visualisation peuvent être comparés :

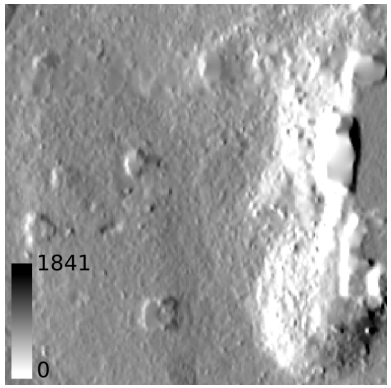
- PCA of Hill-shadings
- Color Cast, qui inclue les 2 tests : *CCtest1 et CCtest2*
- Orientated Second Derivative extrema, à la fois minimum and maximum
- SVF et SIM: 5 rasters pour chaque méthode. Deux paramètres varient, à savoir la taille du filtre (3x3, 5x5, 7x7; CS=1.25 m) et la taille de cellule (1 m, 1.25 m, 2 m; fenêtre: 3x3 px).



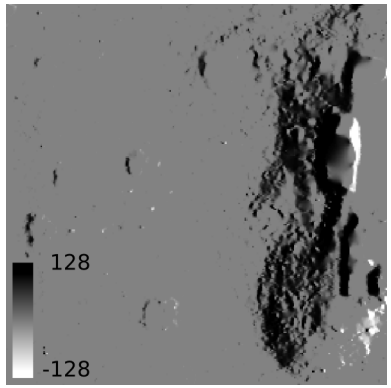
OSD ext1 1.25 m



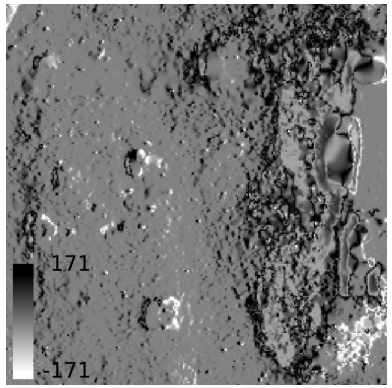
OSD ext2 1.25 m



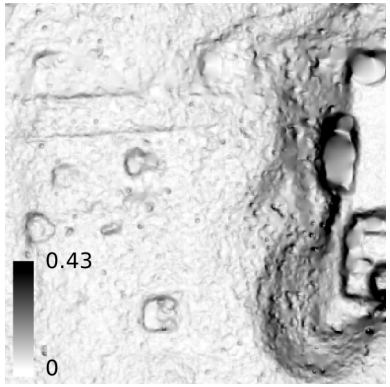
PCA of Hill-shadings 1.25 m



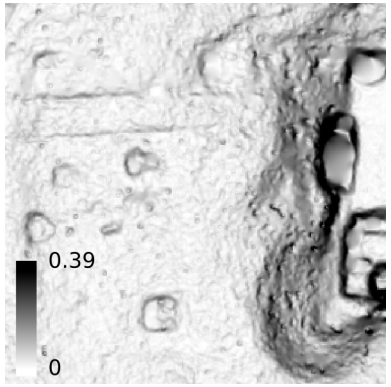
CC test1 1.25 m



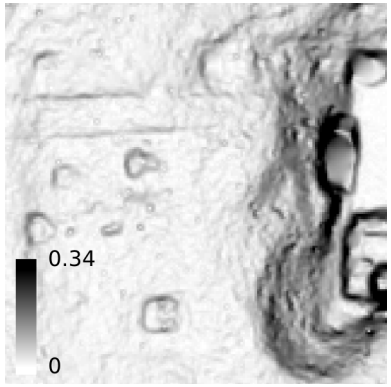
CC test2 1.25 m



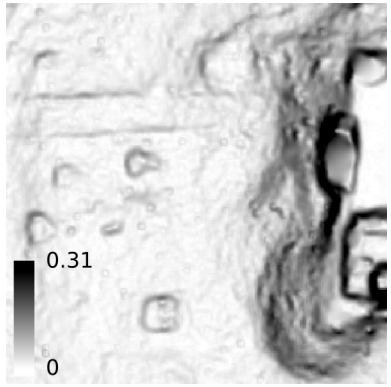
SVF écart-type 3x3 1 m



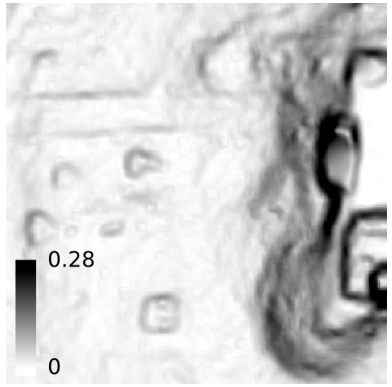
SVF écart-type 3x3 1.25 m



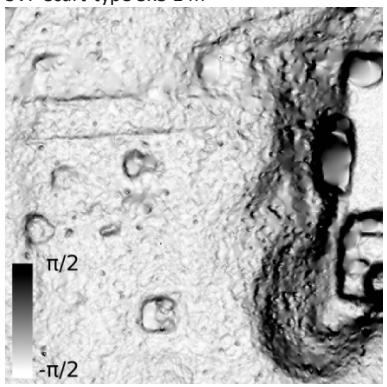
SVF écart-type 3x3 2 m



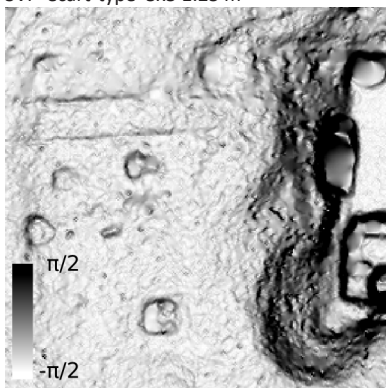
SVF écart-type 5x5 1.25 m



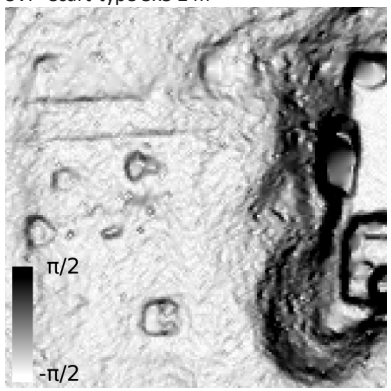
SVF écart-type 7x7 1.25 m



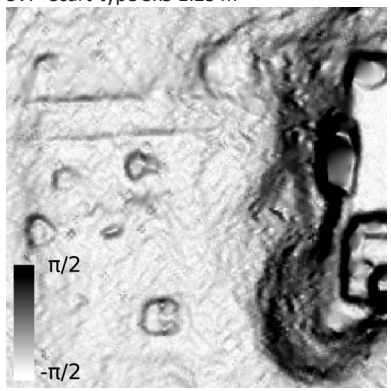
SIM phi 3x3 1 m



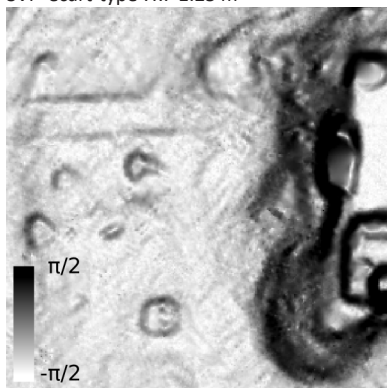
SIM phi 3x3 1.25 m



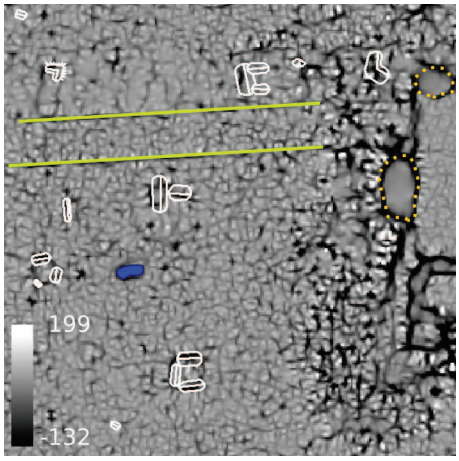
SIM phi 3x3 2 m



SIM phi 5x5 1.25 m



SIM phi 7x7 1.25 m



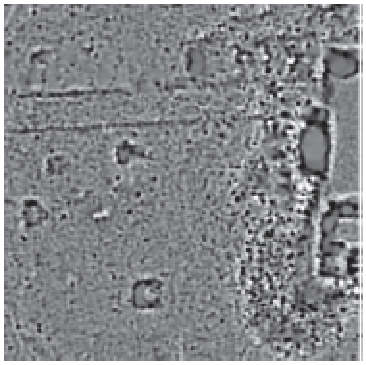
Au premier abord, l'ACP fournit le plus simple affichage. Sur les autres rasters des formes artificielles ont tendance à apparaître avec plus d'intensité, en particulier pour OSD extremaet Color Cast test2. SIM et SVF se ressemblent beaucoup.

Color Cast test1 manque clairement d'information par rapport aux autres supports. Cela vient du fait que trop de niveaux de gris ont été attribués aux terrains plats. Au contraire, le second test affiche plus de pentes faibles, et offre un aspect granuleux tout aussi nuisible à l'interprétation.

Illustration ci-contre: zoom à l'ouest d'El Pilar.

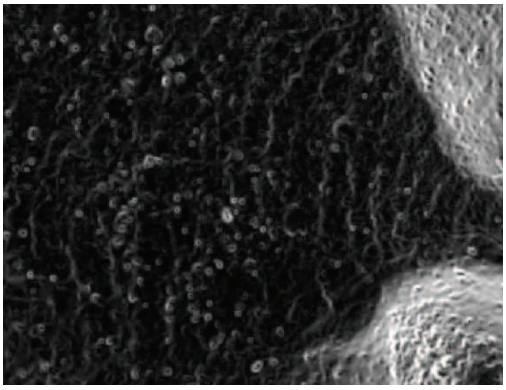
Les monticules culturels sont représentés en blanc, la voie en lignes vertes pleines, la dépression en bleu et les temples de la place Copal en ellipsoïdes jaunes pointillés.

Bien sûr, augmenter la taille de cellule ou la taille de filtre a tendance à flouter le terrain, si bien que seules les caractéristiques paysagères importantes ressortent. Cela a l'avantage d'effacer le bruit image, au détriment des sites réels, qui eux aussi ont tendance à disparaître.



OSD constitue l'unique modèle capable de distinguer éléments positifs et négatifs. Superposer les deux rasters extrema permet d'obtenir des indices intéressants sur le paysage en faisant clairement les sites majeurs. En outre, cette méthode est indépendante de l'orientation, tandis que SVF, SIM et ACP nécessitent la prédéfinition de directions. Color Cast n'a été calculée que selon un axe.

Rasters OSD extrema superposés. La dépression apparaît différemmentdes monticules voisins.



Il faut accorder une attention toute particulière aux formes semblables à des petits monticules extrêmement présents sur les modèles SVF et SIM à 1 et 1.25 m. Volontairement surnommés 'poxes' pour leur forme ronde, leur densité varie sur la réserve, allant jusqu'à 200 par hectare. À côté de cela, des lignes ressortent plus clairement en certains endroits. S'agit-il d'artifices ou bien du vrai sol ?

Illustration: 'poxes'et lignes, près de la frontière sud de la réserve

→En considération des critères préétablis, OSD est le seul modèle rejeté en raison des formes artificielles trop accentuées. Color Cast test 2est maintenu cependant.Même si ce modèle fournit trop d'information, ce raster offre une visualisation inhabituelle qui apporte un intérêt certain. Les trois autres méthodes, SVF, SIM et ACP satisfont globalement aux critères. Désormais, il faut confronter l'interprétation de ces rasters avec la validation issue du terrain. Des 15 rasters de visualisation créés, 3 ont donc été exclus : Color Cast test1 et les 2 extrema OSD.

III Développement d'une stratégie pour la détection de sites anciens mayas

LiDAR n'a eu de cesse de gagner de l'importance au sein de la communauté archéologique. Cet outil semble fournir un modèle de terrain adéquat. Cependant, si la détection de sites se fait avec d'autant plus d'aisance et de rapidité, il faut garder à l'esprit les inconvénients qui découlent de l'usage informatique, ce qui signifie ici l'ajout de bruit et la création de formes artificielles. Cette dernière partie relie l'interprétation des modèles de visualisation et la vérification sur le terrain en mettant en place une série de règles de validation. Un protocole va être proposé, comme la combinaison des expériences vécues au bureau et sur le terrain et des pistes d'amélioration et de réflexion, dans l'optique de cartographier les sites mayas anciens.

III.A Interprétations etvalidations des produits issus du LiDAR

Trois étapes composent la phase de validation. Tout d'abord, les sites archéologiques déjà découverts constituent une ressource précieuse pour comparer les rasters de visualisation, en terme de détection et d'omission. Puis, pointer des sites potentiels va amener à déterminer une deuxième règle de validation avant l'exploration sous la canopée. Enfin, les expéditions sur le terrain viendront contrôler les rasters, en indiquant la véritable nature des sites potentiels.

III.A.1 Interprétation de sites archéologiques : l'œil est un détecteur d'objets

III.A.1.a Des taux de détection dans des matrices d'erreur

La validation de l'interprétation LiDAR est essentiellement basée sur la comparaison à la réalité du terrain. Chaque site archéologique a deux caractéristiques indépendantes :

- La visibilité sur le(s) raster(s) de visualisation : oui (Yes), non (No) ou peut-être (Maybe)
- Type d'élément: culturel, naturel ou artificiel

Toute combinaison de ces variables correspond à un taux de détection. Au total, sept taux ont été définis :

		L'élément est-il visible ?		
		Oui (clairement visible)	Peut-être (visible avec incertitude)	Non (invisible)
Comment interpréter l'élément ?	Culturel (i.e. site archéologique)	Détection Positive	Détection Ambiguë	Omission Négative
	Naturel	Détection Neutre		OmissionNeutre
	Artificiel	Détection Négative		Omission Positive

Visualisation parfaite = "Tous les sites archéologiques eteux seulementsontclairement détectés"

Détection Globale= 100% Détection Positive + 0% Détection Neutre + 0% Détection Négative

Omission Globale= 100% Omission Positive + 0% Omission Neutre + 0% Omission Négative

0% DétectionAmbiguë

Taux de détection

'Peut-être' fait référence à la notion de 'logique floue' (fuzzy logic), qu'il est important d'inclure dans le processus de validation. En effet, l'interprétation ne signifie pas nécessairement que tout élément est assurément visible ou complètement invisible. En réalité, la détection visuelle repose l'association complexe de paramètres relatifs à l'affichage du raster (échelle, couleurs, rééchantillonnage éventuel) et aux caractéristiques des éléments ciblés (localisation, formes et dimensions). Bien sûr, un élément sera plus facilement vu sur un terrain plat si de plus il correspond aux attentes. Au contraire, une forme inhabituelle et floue dans un terrain accidenté a peu de chances d'être assimilé à un site culturel. En conséquence, ajouter 'peut-être' permet d'éviter de se borner à une décision binaire en prenant en compte les éléments ambigus à peine visibles. Tout comme les classifications issues de la logique floue, 'Oui' et 'Peut-être' seront réunis pour les analyses de résultats.

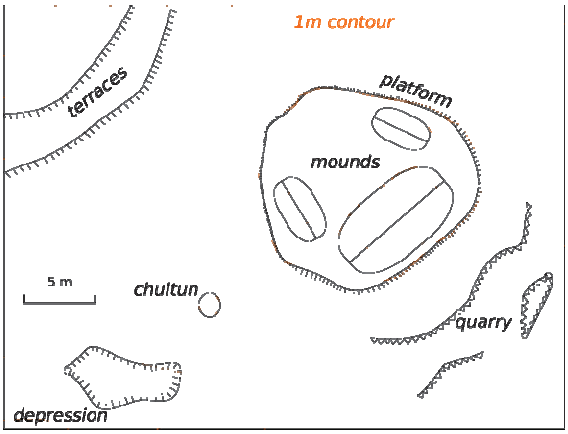
Autre rapprochement : sites culturels et naturels seront réunis pour souligner la distinction élément 'réel' et artificiel.

		L'élément est-il visible ?	
		Y+M	N
Comment interpréter l'élément ?	C+N	Détection positive	Omission négative
	A	Détection négative	Omission positive

Rapprochements pour simplifier les analyses

III.A.1.bLes éléments d'intérêt

Au sein de l'étude des sites mayas, diverses formes ont pu être observées. Celles-ci sont résumées sur l'exemple de croquis de terrain imaginaire. Les éléments d'intérêt peuvent être linéaires (terrasses, carrières), polygonaux (habitations, dépressions, chultuns). Le chultun correspond à un trou de diamètre inférieur à 1 m creusé pour l'adduction d'eau. Évidemment, les habitations se sont érodées avec le temps et ressemblent à des monticules. C'est l'un des obstacles de la détection : distinguer naturel et culturel. Les monticules représentent la plupart des sites levés jusqu'ici, avec des dimensions très variables : hauteurs entre 30 cm et 3 m – surfaces entre 6 m² et 200 m².



Exemple imaginé de croquis de terrain

En termes archéologiques, un site se compose d'une ou plusieurs structures. La logique veut que les structures situées à une distance inférieure de 20 m les unes des autres forment un site. En ce qui concerne les habitations, deux configurent ressortent le plus souvent : 2 monticules o bien 3 monticules surélevés sur une plateforme. En général, un monticule domine les autres.

III.A.1.c L'apport des sites découverts pour une première validation

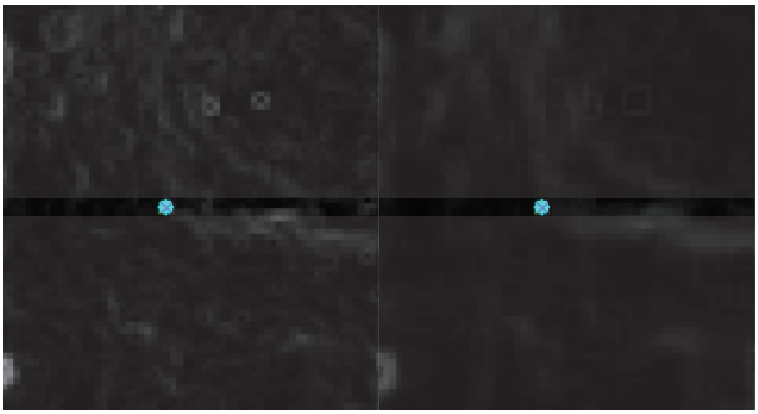
Les sites découverts permettent d'établir la première règle de validation. Par le passé, plus d'une centaine de sites ont été relevés et intégrés dans les SIG via les coordonnées GPS. Ceux-ci ont été récemment repositionnés aussi souvent que possible en se fiant à un raster de type relief ombragé. De plus, lors des expéditions sous la canopée, de nouveaux sites furent découverts. Réunis ensemble, les 165 sites anciennement et nouvellement découverts ont été observés sur les rasters de visualisation (échelle 1/1000 sur ArcGIS®).

	Y	M	N	Y+M	N
SVF 1m 3x3	75	45	45	120	45
SIM 1m 3x3	68	47	50	115	50
PCA of HS	67	54	44	121	44
SIM 1.25m 3x3	59	47	59	106	59
CC Test2	52	52	61	104	61
SVF 1.25m 3x3	51	57	57	108	57
SVF 2m 3x3	39	36	90	75	90
SIM 1.25m 5x5	37	48	80	85	80
SVF 1.25m 5x5	34	46	85	80	85
SIM 1.25m 7x7	29	41	95	70	95
SIM 2m 3x3	29	33	103	62	103
SVF 1.25m 7x7	25	34	106	59	106

Table: Les chiffres indiquent le nombre de sites clairement visibles (Y), moins visibles (M)etnon visibles (N). Les barres colorées donnent un aperçu des proportions, en considérant l'ensemble des 165 sites.

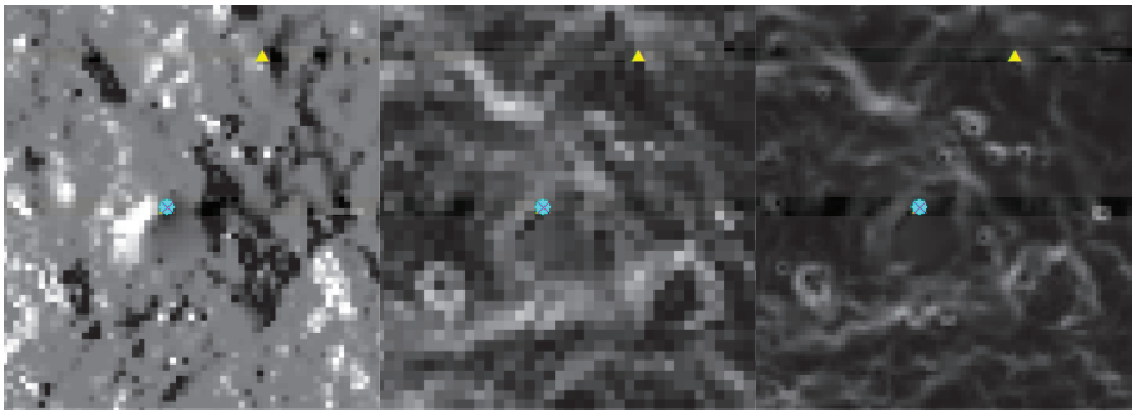
Naturellement, les tailles de fenêtre et de cellule ont une influence sur le taux de détection. Le niveau de détail augmente lorsque l'un de ces deux paramètres diminue, et ce quel que soit l'algorithme employé. En effet, SVF and SIM avec 1m etdes filtres 3x3 donne les meilleurs résultats, tout comme l'ACP. Ils détectent environ deux fois plus de sites que les rasters à 2 m de taille cellule ou des filtres 7x7. Les scores 'Peut-être' varient moins : entre33 et 57 éléments ont été classés ambigus.

Au-delà des sites visibles sur tous les rasters, la visibilité dépend du type d'élément en question. Les polygones apparaissent plus facilement sur les rasters détaillés tandis que les lignes se démarquent sur les rasters plus lisses.



SVF 1.25 3x3 / SVF 1.25 5x5

TerraceW (1909103;271105): exemple d'élément linéaire mieux visible sur les rasters lisses. Le point bleu représente l'angle du site TerraceW.



Color cast test2 / SIM 3x3 / SVF 1 3x3

Mound (1908205; 270828): élément mieux visible sur Color Cast. Le point bleu se situe à l'Est du monticule.

Parce qu'il fournit les meilleurs résultats de la première règle avec le maximum de détails, le raster SVF à 1 m et 3x3 a été préféré pour l'objet de la deuxième étape : la recherche de nouveaux sites.

III.A.1.dLe deuxième règle de validation à la recherche de sites potentiels

Grâce à l'expérience tirée de l'observation des sites découverts, la recherche de sites potentiels n'a été que plus rapide. La zone d'étude(illustréesur la carte de Ground Point Density – Hectare Cells) comprend environ 140 ha. Finalement, 125éléments ont attiré notre attention : ceux-ci ont été localisés sur le raster SVF 1 m 3x3 à la même échelle (1/1000). Le type d'élément 'a priori' en se confortant aux interprétations précédentes, reconnus ainsi soit comme culturels soit comme des 'poxes' artificiels.

La deuxième règle de validation compare les rasters avec la référence (SVF 1 m 3x3). La classification ne tient pas compte d'éléments naturels.

	Cultural (80 sites)		
	Y	M	N
SVF 1m 3x3 (Reference)	62	18	0
SIM 1m 3x3	51	20	9
SVF 1.25m 3x3	40	30	10
SIM 1.25m 3x3	38	29	13
PCA of HS	37	24	19
CC Test2	37	17	26
SVF 1.25m 5x5	27	27	26
SVF 2m 3x3	26	36	18
SIM 1.25m 5x5	24	32	24
SIM 1.25m 7x7	18	22	40
SIM 2m 3x3	18	33	29
SVF 1.25m 7x7	12	21	47

	Artificial (45 sites)		
	Y	M	N
SVF 1m 3x3 (Reference)	44	1	0
CC Test2	45	0	0
SIM 1m 3x3	43	2	0
SIM 1.25m 3x3	35	8	2
SVF 1.25m 3x3	30	15	0
PCA of HS	23	22	0
SVF 2m 3x3	9	24	12
SIM 2m 3x3	3	26	16
SVF 1.25m 5x5	1	19	25
SIM 1.25m 5x5	1	17	27
SVF 1.25m 7x7	0	2	43
SIM 1.25m 7x7	0	1	44

Tables: comptes de visibilité a priori des sites identifiés comme culturels et artificiels, avec le raster SVF à 1m et un filtre 3x3 en guise deréférence

En considérant SVF 1m 3x3 comme référence, les résultats diffèrent selon le type d'élément. D'un côté, les comptes de visibilité pour les probables sites culturels rappellent les chiffres de la première règle. Le classement des rasters n'a pas changé. D'un autre côté, les écarts pour les éléments artificiels s'accentuent entre les rasters. Par exemple, tous les 'poxes' apparaissent sur le modèle Color Cast test2 tandis que seulement un d'entre eux a été observé sur les rasters SIM à 1,25m et 7x7.

Si la table des sites culturels accorde de l'importance pour les rasters créés avec des tailles de cellule et de filtre petites, la table des sites artificiels nuance cette décision et attire l'attention sur les rasters lisses. En effet, s'il on imagine que tous les 'poxes' sont réellement artificiels, les rasters détaillés pâtiront de ce surplus d'information.Cela permet d'insister sur une constatation : un raster de visualisation doit non seulement permettre de détecter les sites existants mais devrait aussi éviter de faire apparaître les éléments artificiels. L'omission compte tout autant que la détection.

La connaissance du terrain aide nécessairement comme un moyen de contrôle pour confirmer ou informer les hypothèses établies durant l'interprétation du LiDAR.

III.A.2 L'exploration du terrain pour les sites potentiels

La troisième et dernière règle de validation est l'étape la plus importante pour préciser le rôle du LiDAR dans la découverte de sites archéologiques. Certes cette technique a l'avantage d'obtenir des informations rapidement et avec une précision acceptable. Mais, cette information contient du bruit, supposément lié aux 'poxes'. Puisque la distinction de sites artificiels et culturels dépend de l'interprétation, la réalité peut être tout autre. Ainsi, explorer les sites potentiels sur le terrain donne la solution sur la véritable nature des éléments identifiés sur les rasters.

III.A.2.aÉquipement et notes de terrain

Les missions passées ont été réalisées dans les conditions ordinaires d'exploration : GPS de randonnée, boussole, carnet de notes pour les croquis. En plus de cela, l'étude du LiDAR fut l'occasion de collecter davantage d'informations, en particulier celles relatives au paysage et la végétation. Et pour vérifier la précision des GPS, 4 GPS ont été utilisés. Des fichiers de terrain ont été

élaborés dans le but de renseigner ces informations pour tous les sites. Ce fichier est un tableau qui contient :

- Coordonnées de destination des points de sites ciblés via le SIG
- Page de l'extrait de carte imprimé où se trouve cette cible
- Colonnes dédiées aux commentaires sur la couverture de la canopée, du sous-étage ainsi quedu relief
- Espèces végétales remarquables autour de la cible
- Nombre(s) GPS : un site a pu être enregistré par un ou plusieurs GPS
- Dimensions du site (optionnel) : caractéristiques de la principale structure. Par exemple la longueur, largeur et hauteur pour un monticule ; la profondeur pour une dépression.

III.A.2.bContrôle de terrain de l'interprétation du LiDAR : matrices de classification

Les expéditions apportent un support indispensable au contrôle de ce qui a été supposé sur les rasters. Une comparaison peut être faite et résumée dans une matrice de classification. La table contiendra les comptes d'éléments selon le type a priori (de l'interprétation du LiDAR) et du type a posteriori (issu de la validation terrain). L'analyse a été faite avec le raster de référence, c'est-à-direSVF à 1m etfiltre 3x3.

		a posteriori - terrain		
		Culturel	Naturel	Artificiel
a priori LiDAR	Culturel	42	11	29
	Naturel	0	0	0
	Artificiel	1	6	36

		a posteriori - terrain	
		C+N	A
a priori LiDAR	C+N	53	29
	A	7	36

Table: classification, SVF 1 m 3x3. La seconde table fusionne éléments naturels et culturels.

En statistiques, le coefficient de Kappa détermine la corrélationentre de tels ensembles de données:

$$\kappa = \frac{N\,R - W}{N^2 - W}$$

Où:

- N: nombre d'éléments (125)
- R: somme des groupes d'éléments bien classés, i.e. types identiques(53+36=89)
- W: produit des groupesd'éléments mal classés, i.e. types différents (7x29=203)

Ici, le calcul s'effectue en fusionnant les sites culturels et naturels.

Le coefficient de Kappa vaut0,71, ce qui est plutôt important. Cependant, la quantité de mal classés est forte aussi : 29 sites potentiels sites n'avaient aucune signification sur le terrain. Au contraire, 7 poxes hypothétiques se sont révélé de véritables sites.

Cela prouve qu'il faut bien faire attention lors de l'interprétation. La distinction entre culturel, naturel et artificiel sera difficile à établir clairement. La troisième règle de validation va plus loin, en analysant la détection des sites réels sur les autres rasters de visualisation.

III.A.2.c Troisième validation à mi-chemin entre interprétation et réalité

La troisième règle de validation consiste en l'évaluation de taux de détection a posteriori, depuis le terrain vers le LiDAR. SVF à1 met 3x3 reste la référence. Trois taux globaux ont été définis : Correcte Détection (CD), Correcte Omission (CO) and Ambiguïté (A).

La Correcte Détection fait référence à la proportion de sites culturels découverts sur le terrain sur la totalité des sites vus sur les rasters LiDAR.La Correcte Omission détermine la proportion d'éléments artificiels parmi ceux qui n'ont pu être observés sur le raster.L'ambiguïté est liée à l'incertitude à l'égard de la visibilité en général, quel que soit le type d'élément en question.

Idéalement, un bon raster de visualisation doit avoir des taux élevés de Correcte Détection et Correcte Omission et un faible taux d'Ambiguïté.

$$CD = \frac{Positive\ Detection}{Global\ Detection} = \frac{Positive\ Detection}{Positive\ Detection + Neutral\ Detection + Negative\ Detection}$$

$$CO = \frac{PositiveOmission}{GlobalOmission} = \frac{PositiveOmission}{PositiveOmission + NeutralOmission + NegativeOmission}$$

$$A = \frac{Ambiguous\ Detection}{Number\ of\ sites}$$

	Correct Detection	Correct Omission	Ambiguity
SIM 1.25m 7x7	78%	64%	18%
SIM 1.25m 5x5	68%	65%	39%
SVF 1.25m 7x7	67%	59%	18%
SVF 1.25m 5x5	61%	59%	37%
SIM 2m 3x3	57%	49%	47%
SVF 2m 3x3	46%	53%	48%
SVF 1.25m 3x3	36%	30%	36%
SIM 1.25m 3x3	36%	33%	30%
PCA of HS	35%	32%	37%
SVF 1m 3x3	34%	N/A*	15%
SIM 1m 3x3	33%	11%	18%
CC Test2	28%	27%	14%

Table: taux calculés avec les informations de terrain

* Correct Omission ne peut être déterminée pour le raster de référence, étant donné que tous les éléments sont visibles en principe.

Comparé aux deux premières règles de validation, l'ordre des rasters performants s'est inversé. En effet, cette fois-ci ce sont ceux produits avec des tailles de cellule et de filtre plus grandes qui donnent les meilleurs résultats. L'ambiguïté augmente avec la taille de cellule. À propos de SVF à 2 m et filtre 3x3 par exemple, presque 1 site sur 2 a été observé avec doute. L'ACP et Color Cast ne satisfont pas, avec des taux de Correcte Détection et Correcte Omission en deçà de 40%.

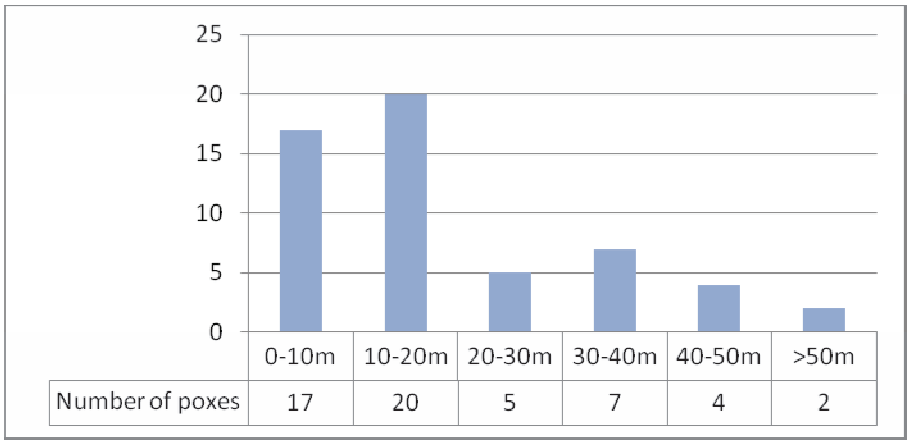
→Les trois règles de validation ne concordent pas. D'un côté, utiliser de petites tailles de cellule et de filtre crée des rasters détaillés. Ceux-ci permettent de déceler un grand nombre de sites potentiels. D'un autre côté, les rasters produits avec des paramètres plus grands engendrent moins d'erreur quant aux sites artificiels. Ils détectent peu mais efficacement. En conséquence, l'interprétation du LiDAR ne peut être employée seule : l'expérience du terrain a toujours un rôle majeur à jouer, dans le but d'assurer un contrôle tout en cernant les obstacles issus de la technique LiDAR.

III.B Améliorations dans le protocole de détection des sites anciens mayas

III.B.1 Le besoin d'expérience de terrain pour détourner les pièges du LiDAR

Les analyses précédentes ont montré l'existence de formes artificielles appelées 'poxes', lesquels apparaissent à cause de points malencontreusement classés comme appartenant au sol. L'hypothèse voudrait relier ces 'poxes' à une espèce de palmiers particulière, dont la couronne réfléchirait plus facilement les rayons laser. Pour vérifier cette proposition, une tentative de corrélation spatiale a été mise en place, entre les sites artificiels et les palmiers. Des orthophotos ont été utilisées afin de localiser ces arbres, aisément reconnaissables à leur couronne étoilée.

Sur ArcGIS®, une *jointure spatiale* relie le palmier le plus proche à chaque pox, laquelle calcule la distance entre deux. Ces distances ont été classées par tranche de 10 m.



Classification des poxes par la distance au palmier le plus proche

Le graphique montre une légère corrélation : 2/3 des poxes (37 sur 55) se situent à moins de 20 m d'un palmier. La présence de ces palmiers explique une part du bruit constaté sur les rasters mais ne peut justifier tous les sites artificiels.

III.B.2 Erreur de positionnement GPS : un frein à l'investigation?

Les GPS de randonnée peuvent nuire à la découverte à cause de l'imprécision de positionnement, indiquée aux alentours de 10 m sur les guides d'utilisation. Le LiDAR apporte une bien meilleure précision. De fait, si la précision du GPS n'est pas satisfaisante, cela peut avoir un impact sur l'investigation, en particulier lorsqu'il s'agit de rechercher des sites peu élevés. L'objectif consiste à évaluer un rayon de recherche pratique pour les expéditions à venir. Munis de 4 GPS avec des

configurations différentes, il sera possible de réaliser des comparaisons entre les coordonnées mesurées.

Deux méthodes ont été employées pour enregistrer les coordonnées: l'enregistrement instantané et la méthode dite du 'point moyenné'. La seconde méthode permet de prendre un ou plusieurs échantillon(s) de coordonnées d'un point sur une période voulue. Pendant l'expérimentation, un échantillon a été enregistré seulement, et les mesures s'arrêtaient dès lors que la 'confiance d'échantillon' atteignait 100% (entre 1 et 3 minutes).

3 des 4 GPS employés correspondent au même modèle. Deux d'entre eux ont appliqué la méthode du point moyenné. Le quatrième GPS moyennait également.

GPS	1	2	3	4
Modèle	Garmin® 62S	Garmin® 62S	Garmin® 62S	Garmin® Etrex10
Méthode d'enregistrement	Point moyenné	Instantané	Point moyenné	Point moyenné

Table: GPS et méthodes d'enregistrement adoptées

Les distances ont été calculées entre les coordonnées des points sur les rasters LiDAR et chacun des quatre GPS. La table ci-dessous fournit l'écart-type de ces distances. Étonnamment, la méthode instantanée permet de localiser des coordonnées avec autant d'erreur que la méthode du point moyenné. En effet, tous les GPS affichent un écart-type avoisinant 7 m.

GPS	1	2	3	4
Écart-type (m)	7.4	6.6	8.3	7.7

Le rayon de recherche pratique est estimé à 7 m, ce qui reste suffisamment faible quand il s'agit de trouver des sites potentiels. Aussi surprenant que cela puisse paraître, le temps d'enregistrement des coordonnées n'a pas d'influence majeure sur l'erreur de positionnement. Désormais, la méthode choisie dépendra du temps que l'on veut y consacrer.

III.B.3 Gestion du nuage de points et du paysage

Au-delà de l'expérience du terrain, une meilleure connaissance du lien unissant paysage et nuage de points de sol doit être établie. Cette section essaie d'expliquer l'incidence de caractéristiques environnementales sur la configuration du nuage de points de sol.

III.B.3.a Post-analyse du nuage points de sol par l'étude de paramètres environnementaux

Comme vu auparavant, le couvert forestier entraîne une perte de densité de points de sol. Qu'advient-il du terrain ou du sous-étage ?

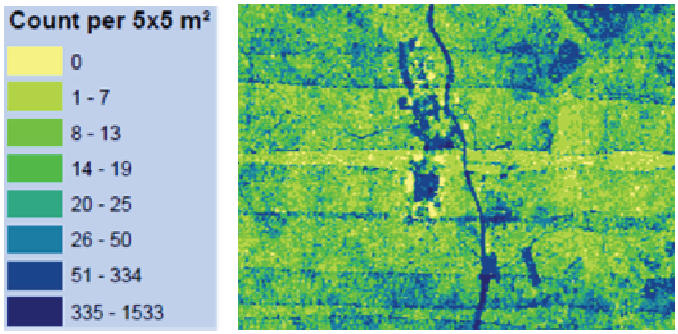
L'objectif est de réanalyser la densité de points de sol à 5x5 m² en tenant compte à la fois du terrain, de la canopée et du sous-étage. Les informations ont été collectées pour une majorité de sites visités. Des critères visuels ont permis de définir de qualifier chaque paramètre environnemental :

- 'Canopée' (Canopy)indique l'intensité du couvert forestier au-dessus du site : ouvert (Open) / medium (Medium) / dense (Dense). Un quatrième adjectif, dégagé (Clear), a été ajouté pour caractériser les endroits sans aucune canopée (herbes hautes, buissons, arbustes).

– ‘sous-étage’ (Understory)signifie quant à lui l’intensité de la végétation aux alentours du site. Les mêmes adjectifs ont été utilisés : ouvert, medium et dense.

– ‘Terrain’ qualifiela topographieautour du site :plat, pente faible, pente forte, bosselé, sommet de colline et pied de colline (base slope).

La densité de points de sol a été divisée en 8 classes (voir illustration). Avant de classer, la densité a été ‘normalisée’. En effet, l’acquisition LiDAR impose une contrainte de recouvrement des bandes de vol, ce qui fait varier la densité de points de sol. Sur le SIG, l’affichage de la densité a permis d’estimer le nombre bandes se recouvrant, allant de 1 à 4. La normalisation égale la densité divisée par ce nombre.



Extrait de la carte de densité de points de sol à 5x5 m², avec les classes correspondantes

Canopée ouverte (A2-18) ; canopée medium (A2-8N) ; canopée dense (mound?18)

Sous-étage ouvert (A2-18) ; sous-étage medium (N1W-20S) ; sous-étage dense (A2-5N)

Terrain de faible pente (close to A2-18) ;terrain bosselé (Ch A2-15S) ; terrain de pente forte (F2-1S)

Exemples de qualificatifs pour les paramètres environnementaux étudiés (nom du site entre parenthèses)

		Densité de points de sol normalisée 5x5 m²							Total
		Classe1 (0 pt/m²)	Classe2 (1-7)	Classe3 (8-13)	Classe4 (14-19)	Classe5 (20-25)	Classe6 (26-50)	Classe7 (>51)	
Couverture de canopée	Clair	0 (0%)	4 (2%)	3 (5%)	1 (4%)	1 (10%)	2 (22%)	1 (100%)	12
	Ouvert	1 (4%)	49 (26%)	24 (46%)	6 (28%)	3 (30%)	2 (22%)	0 (0%)	85
	Medium	7 (31%)	79 (42%)	16 (30%)	11 (52%)	5 (50%)	3 (33%)	0 (0%)	121
	Dense	7 (31%)	42 (22%)	5 (9%)	2 (9%)	0 (0%)	1 (11%)	0 (0%)	57
	Très dense	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0
	Non collecté	7 (31%)	10 (5%)	4 (7%)	1 (4%)	1 (10%)	1 (11%)	0 (0%)	24
	Total	22	184	52	21	10	9	1	299
Terrain	Plat / sommet de colline	8 (36%)	80 (43%)	27 (51%)	10 (47%)	1 (10%)	1 (11%)	1 (100%)	128
	Pente faible	3 (13%)	32 (17%)	11 (21%)	9 (42%)	2 (20%)	4 (44%)	0 (0%)	61
	Pente forte	1 (4%)	21 (11%)	5 (9%)	0 (0%)	1 (10%)	1 (11%)	0 (0%)	29
	Bosselé	3 (13%)	38 (20%)	5 (9%)	1 (4%)	3 (30%)	2 (22%)	0 (0%)	52
	Non collecté	1 (4%)	21 (11%)	5 (9%)	0 (0%)	1 (10%)	1 (11%)	0 (0%)	29
	Total	22	184	52	21	10	9	1	299
Densité de sous-étage	Clair	0 (0%)	5 (2%)	3 (5%)	0 (0%)	0 (0%)	1 (11%)	0 (0%)	9
	Ouvert	1 (4%)	26 (14%)	7 (13%)	6 (28%)	2 (20%)	1 (11%)	1 (100%)	44
	Medium	7 (31%)	66 (35%)	16 (30%)	4 (19%)	3 (30%)	5 (55%)	0 (0%)	101
	Dense	6 (27%)	68 (36%)	18 (34%)	10 (47%)	4 (40%)	0 (0%)	0 (0%)	106
	Très dense	0 (0%)	3 (1%)	1 (1%)	0 (0%)	0 (0%)	1 (11%)	0 (0%)	5
	Non collecté	8 (36%)	16 (8%)	7 (13%)	1 (4%)	1 (10%)	1 (11%)	0 (0%)	34
	Total	22	184	52	21	10	9	1	299

Table: comptes par classe pour chaque paramètre environnemental. Les pourcentages représentent le taux par classe. Par exemple, la canopée dense dans la classe 2 vaut 42/184 soit environ 22%.

Tout d'abord, la plupart des sites ciblés se trouvent en classe 2 (61.5%). Cela signifie que peu de points sont nécessaires pour dessiner les éléments, soit environ 10 points pour 25 m². Fait surprenant, les résultats ne confirment pas clairement l'impact de la végétation sur la perte de

densité. La répartition des pourcentages dans la table est distribuée de façon bien trop aléatoire pour y extraire des corrélations.Les constats sont rares :

- Les endroits vides de points de sol correspondent à des zones de végétation dense
- La seule cible incluse dans la plus haute classe de densité (plus de 51 pt/m²) se trouve dans un endroit ouvert et plat

En réalité, décrire la canopée ou le sous-étage a été une tâche redoutable. La distinction entre medium et dense ne coulait pas de source et souvent subjective. Sous la canopée, les lieux visités se ressemblaient beaucoup. Cette étude devrait se baser sur des critères quantitatifs en plus des critères qualitatifs définis.

III.B.3.bRe-filtrage du sol

Les poxes,etles éléments artificiels plus généralement, résultent d'une mauvaise classification du nuage de points. Les points n'appartenant pas au sol ont été inclus dans le nuage de points de sol. Cela prouve que l'algorithme de filtrage ne suffit pas. Des algorithmes de filtrage existent pour améliorer la qualité du nuage de points de sol.

Ce paragraphe se concentre sur l'un d'entre eux -SMRF (Simple MoRphological Filter) – et compare ses résultats avec les rasters de visualisation déjà créés.

L'algorithme SMRF [Pingel, 2013] appartient à la catégorie des filtres de morphologie progressive. La morphologie progressive trie les points de sol et recrée un modèle de terrain. En plus de ce traitement, SMRF finit en ne retenant que les points situés à une certaine distance du dernier modèle de terrain créé.

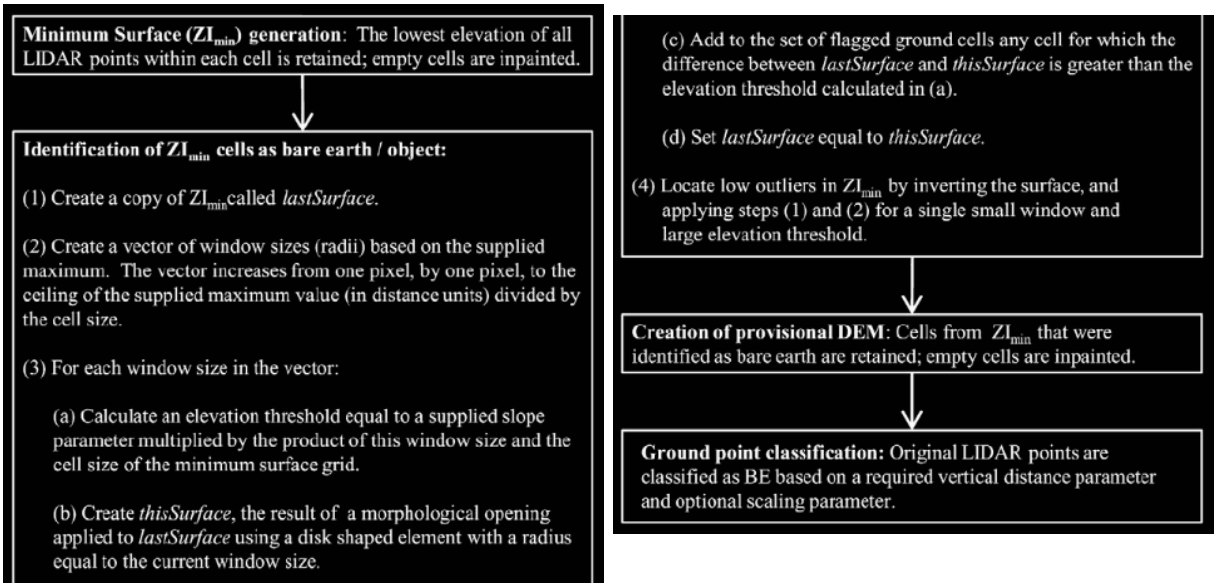
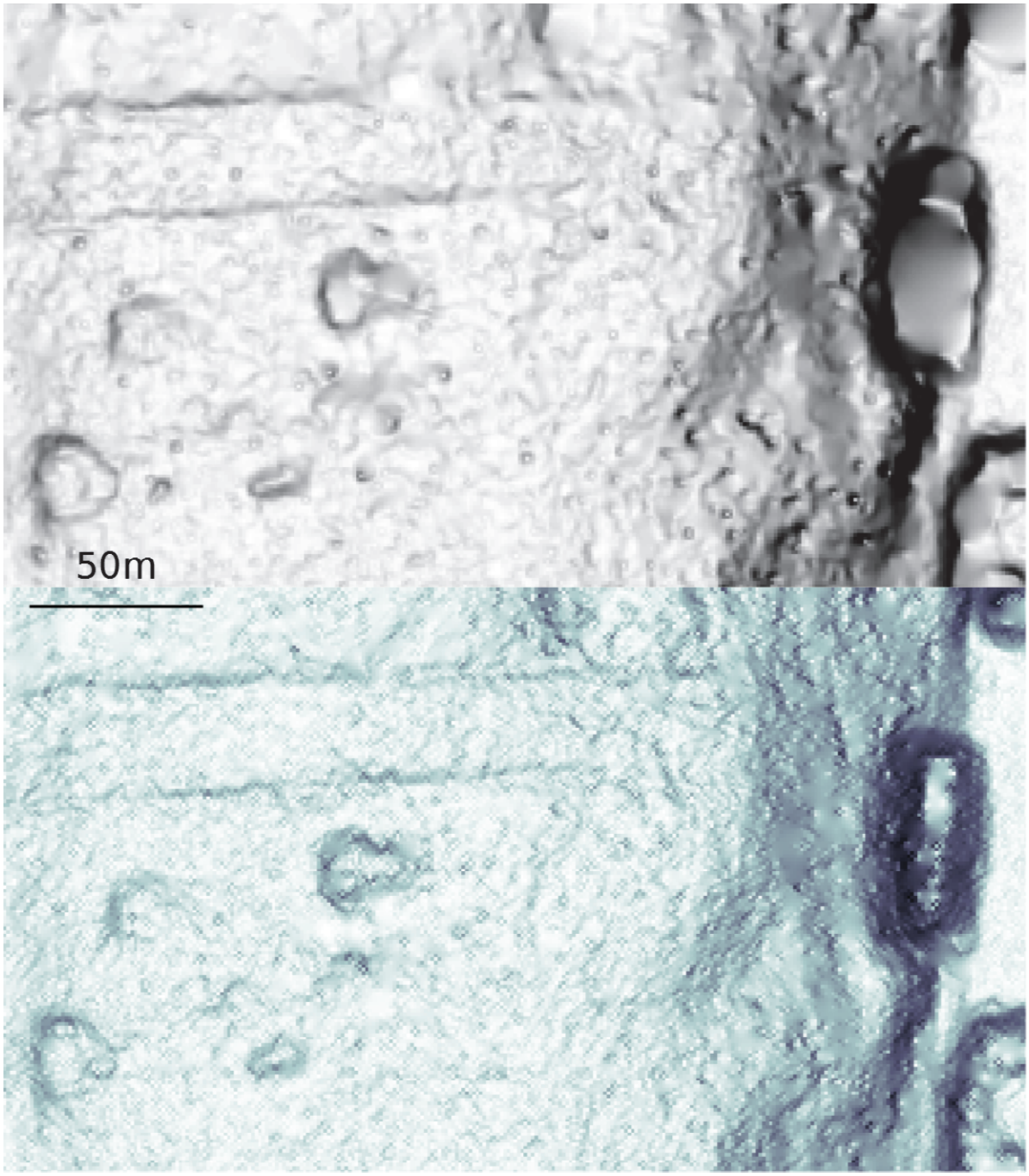


Schéma de l'agorithme SMRF. BE signifie sol (bare earth)

L'algorithme fournit un nouveau nuage de points de sol pour réaliser un raster appelé Bonemap, coloré selon la pente. L'interprétation et les règles de validation ont été réalisées à l'instar des modèles de visualisation précédents.



Comparaison visuelle du raster Bonemap avec SVF à 1m et filtre 3x3. Localisation: Ouest d'El Pilar. Échelle d'affichage sur ArcGIS® :1/1000.

Des écarts existent entre les deux cartes. La Bonemap offre un aspect plus 'granuleux' tandis que la SVF présente un terrain plus lisse. L'interprétation peut être rendue difficile par cette granularité, ce qui peut aussi éviter de détecter trop de poxes.

	Y	M	N	Y+M	N
SVF 1m 3x3	75	45	45	120	45
Bonemap	64	27	74	91	74

	Y	M	N
Cultural (80 sites)	28	19	33
Artificial (45 sites)	0	3	42

	Correct Detection	Correct Omission	Ambiguity
SIM 1.25m 7x7	78%	64%	18%
SIM 1.25m 5x5	68%	65%	39%
SVF 1.25m 7x7	67%	59%	18%
Bonemap	64%	67%	18%
SVF 1.25m 5x5	61%	59%	37%
SIM 2m 3x3	57%	49%	47%
SVF 2m 3x3	46%	53%	48%
SVF 1.25m 3x3	36%	30%	36%
SIM 1.25m 3x3	36%	33%	30%
PCA of HS	35%	32%	37%
SVF 1m 3x3	34%	N/A*	15%
SIM 1m 3x3	33%	11%	18%
CC Test2	28%	27%	14%

1ère, 2ème et 3ème règles de validation appliquées au raster Bonemap(* la Correcte Omission ne peut être calculée pour le raster de référence)

Moins de sites découverts ne sont apparus sur le raster Bonemap que sur le modèle SVF de référence (table 1). En revanche, aucun des 45 sites supposés artificiels n'a été clairement vu (table 2). Dans la dernière table, Bonemap fait partie des meilleurs modèles de visualisation : meilleure taux de Correct Omission, quatrième meilleur taux de CorrecteDétection. Le taux d'Ambiguïtéest quant à lui très faible (18%).

→L'utilisation d'algorithmes de refiltrage aide évidemment puisque cela retranche les points à l'origine de formes artificielles. Les algorithmes font nécessairement partie du processus pour la détection de sites anciens mayas.

III.B.4 Une esquisse de protocole

LiDAR a un rôle important à jouer en archéologie. Cependant le nuage de points fourni par cette technique ne suffit pas à détecter les sites archéologiques. Cette donnée brute initie un protocole destiné à visualiser, interpréter et cartographier les sites mayas. Ce paragraphe liste des difficultés rencontrées durant la mission tant du point de vue du bureau que du terrain. Il illustre ensuite un exemple de marche à suivre vers la détection de sites mayas anciens.

III.B.4.a Problèmes rencontrés

La mission de 2013 se distingue des missions précédentes de BRASS/El Pilar par le fait d'être la première à impliquer le LiDAR. Des décisions ont été faites en vue de l'analyse de la qualitéainsi que la création de modèles de surface adaptés à la visualisation. Puis, des expéditions de terrain et des règles de validationont permis d'évaluer la contribution de l'interprétation des produis LiDAR. Naturellement, cette mission innovante a été freinée par quelques difficultés, principalement dues à un problème d'organisation et de systématisme. Voici les principales difficultés.

La phase de terrain a pris le dessus sur la phase de bureau. Beaucoup de données furent collectées, en particulier des points GPS. Pour une grande d'entre eux, les renseignements restent incomplets. Toute la quantité d'information n'a pas été revue. Ce défaut peut être corrigé en reconsidérant le travail de terrain. Au-delà de la gestion de l'équipement, du temps doit être dévolu pour récupérer les données et prendre du recul sur le sens de ces données. Par exemple, jours de terrain et de bureau peuvent être séparés, avec une proportion 2/1. Les commentaires sur les données demeurent essentiels et doivent être faciles à suivre.

Une approche systématique reste à appliquer aux fichiers de notes de terrain. Pendant la mission, l'intérêt croissant pour la relation entre paysage et la densité de points de sol a conduit des feuilles de notes, lesquelles ont évolué au fur que nous progressions dans les investigations. Ces fichiers se composaient de tableaux où chaque ligne représentait un site. Le tableau contenait beaucoup de colonnes telles les observations de l'environnement, les espèces d'arbres ou encore les dimensions de la plus grande structure. Très vite, le manque de place a contraint à abréger l'information au détriment de sa clarté. Plutôt que de penser tous les sites à la fois, la solution a été de concevoir un fichier propre à chaque site. Cela donne assez d'espace pour toutes les observations, croquis et autres remarques utiles.

III.B.4.b Un parcours à suivre

Les méthodes employées pour les phases de bureau et de terrain ont fait face à défis et abouti à des résultats utiles à la détection archéologique. Cependant, la méthodologie globale reste à améliorer, depuis la gestion du nuage de points jusqu'aux sites anciens mayas.

La gestion des données LiDAR constitue l'obstacle le plus dur à surmonter. Tout d'abord, la qualité du nuage de points demeure difficile à estimer, du fait des nombreux paramètres à considérer, que ce soit l'environnement ou le système LiDAR lui-même. Ensuite, la création du nuage de points de sol ne satisfait pas encore. Les résultats ont mis en exergue la nécessité de recourir à des algorithmes de refiltrage, qui rejetteront davantage de points hors sol.

Puisque l'affichage ponctuel n'aide pas à la détection, il a fallu créer des rasters de visualisation. Cela a donc impliqué le choix de trois paramètres : la taille de cellule, l'algorithme d'interpolation et le modèle de visualisation. Des variables ont été pensées pour lier le nuage de points de sol à ces paramètres. L'analyse de ces variables a donné des directions vers une perte minimale de la qualité du raster et donc une meilleure carte à visualiser.

Le terrain apporte un moyen précieux pour contrôler les modèles issus du LiDAR. Nombre de sites artificiels sont apparus sur les rasters, sans compter que les 'poxes' ressemblaient à s'y méprendre à de vrais sites culturels. De fait, la validation par le terrain reste essentielle pour distinguer les éléments culturels.

Cette phase de validation se compose d'une succession d'étapes parmi lesquelles :

- Interprétation des rasters de visualisation: observation et pointage des sites potentiels
- Création de listes de points à explorer et fichiers de notes correspondants
- Exploration: les sites potentiels sont acceptés ou rejetés, pendant que les nouveaux sites découverts sont enregistrés au GPS
- Levé : cartographie des sites acceptés et des nouveaux sites découverts
- Intégration dans le SIG

Après l'exploration, les sites rejetés comptent tout autant que les sites acceptés. Tous deux renseignent sur la qualité de la détection fournie par le raster de visualisation. L'étude de ces éléments aidera à élaborer des motifs de reconnaissance pour une interprétation plus rapide et efficace. Certes détecter les éléments culturels est important, mais être capable de rejeter les éléments artificiels importe également.

L'interprétation du LiDAR fut basée sur la détection visuelle. Pour aller plus loin, il n'est pas impossible de penser à une détection automatique. Si l'œil peut repérer les éléments, peut-être l'ordinateur saurait-il distinguer de la même manière ? Le processus automatisé serait alors initié par une classification en seuils, une méthode courante en traitement d'image. Cette méthode trie les pixels selon des critères mathématiques. Dans le cas archéologique, le processus démarrerait avec la simulation d'un MNT comprenant divers motifs paysagères. Ces motifs décrivent le pixel et son voisinage pour de multiples configurations du paysage. Par exemple des bosses, des fossés ou des coins. Ensuite, les filtres de convolution sont appliqués pour dessiner le raster de visualisation. L'analyse des résultats devrait aider à déterminer des seuils pour dissocier les différentes configurations. Les seuils constituent la clé pour classifier les pixels. Finalement, des échantillons d'éléments peuvent être créés pour représenter les divers sites archéologiques d'intérêt tels que les terrasses, les monticules ou les dépressions. Une extraction automatique peut être lancée pour reconnaître des échantillons sur les rasters de visualisation.
Le parcours complet est illustré ci-dessous.

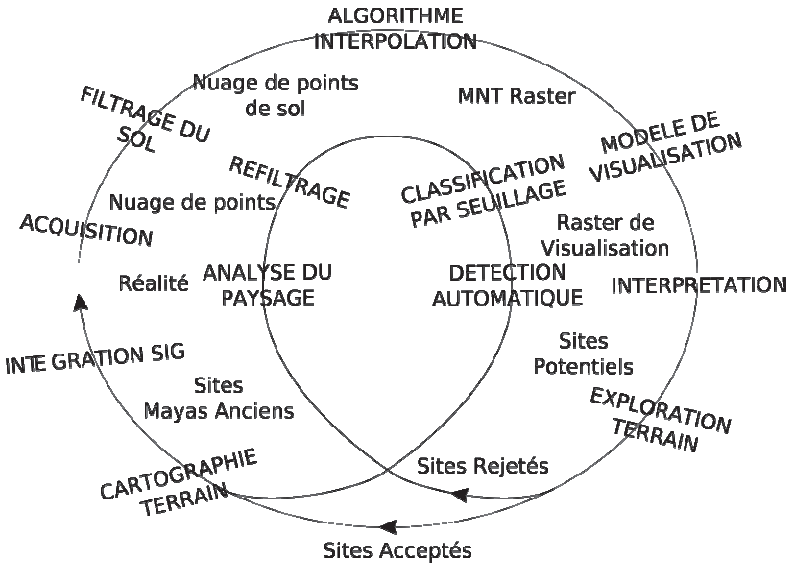


Schéma: protocole de cartographie des sites anciens mayas

Conclusion

Depuis quelques décennies, le LiDAR n'a cessé de gagner de l'importance au sein de la communauté archéologique. Il représente un intérêt évident pour la découverte dans le sens où il permet d'économiser beaucoup d'efforts et de temps dans l'investigation, surtout sous la forêt. Les progrès réguliers ont tant amélioré le développement et la compréhension de cette technique que celle-ci fournit des données de terrain relativement précises en comparaison des autres techniques de levé[Crutchley, 2009].

Ce mémoire a eu pour but d'aborder quelques défis auxquels l'utilisateur est confronté lorsqu'il doit appréhender cet « outil magique ». Quelles décisions prendre quand il s'agit d'utiliser des données LiDAR pour la prospection archéologique ?

La première décision se rapporte au format du modèle de terrain. Le format raster surpasse les autres en termes de lisibilité et de possibilités de calcul. Mais ce format implique plusieurs paramètres – taille de cellule et algorithme d'interpolation – ce qui entraîne une perte de qualité. De plus ces paramètres sont rarement choisis, c'est-à-dire que les décisions relèvent soit de l'expérience soit de contraintes de projet. De fait, la création du MNT dépend rarement de la qualité initiale du nuage de points de sol. C'est la raison pour laquelle des variables ont été pensées pour tracer des liens entre le nuage de points de sol et le MNT au format raster. Pourtant, l'altitude ne fournit pas suffisamment d'indices pour détecter les sites archéologiques.
Ceci amène au deuxième défi : le besoin de visualisation. En dépit des nombreuses possibilités offertes par les logiciels de SIG et de traitement d'image, elles adressent partiellement l'aspect de la détection. Le relief ombragéfait figured'exception : ces modèles fournissent un support adapté à la visualisation. Mais il a aussi des défauts, la dépendance vis-à-vis de la direction de l'éclairement par exemple. Plusieurs recherches se sont récemment focalisées sur la création d'un algorithme de visualisation idéale. Finalement, l'algorithme parfait n'existe pas et dépend aussi bien du terrain que des dimensions des sites à découvrir[Challis, 2012]. Ici, aucun de ces algorithmes n'a été utilisé : les notions théoriques ont été traduites en filtres de convolution, un principe pas si évident à mettre en place pour réaliser les MNT rasters. En regardant ces rasters, quelques sites ont été facilement reconnus et d'autres formes particulières ont attiré notre attention. Appelés 'poxes' pour leur présence inattendue, ces formes rondes ont rapidement été assimilées à des éléments artificiels. Comment confirmer une telle supposition qui viendrait affecter l'interprétation du LiDAR ?
Trois règles de validation ont été mises en place afin de répondre à cette interrogation. En vue de la comparaison entre l'interprétation LiDAR et l'investigation terrain, les règles ont défini tous types de détection et d'omission les résultats ont clairement prouvé la nécessité de contrôler « l'outil magique » qu'est le LiDAR. Des expéditions sur le terrain demeurent indispensables pour confirmer ou rejeter les suppositions issues du LiDAR. En effet, si les rasters dérivés du LiDAR aident à la détection de sites culturels, une grande partie des formes observées relevait de l'artificiel. Cela provient d'une mauvaise classification des points de sol lors de la création du nuage de points de sol. Les algorithmes de refiltrage constituent un domaine de recherche majeur dans la quête de l'amélioration de la qualité LiDAR.

Un protocole propose une démarche pour les archéologues désireux d'utiliser la technique LiDAR vers la prospection de sites mayas. La méthode contient des idées générales applicables à tout projet localisé en

région forestière. Néanmoins, certaines pistes de réflexion pourraient intéresser tout spécialement le projet BRASS/El Pilar.

La connaissance du paysage et de la géomorphologie a une influence sur l'interprétation. Ceci devrait aider à analyser les formes illogiques engendrées par le bruit pour mieux se concentrer sur les formes culturelles. De plus, nombre de sites potentiels étaient en réalité des sites naturels. Détecter des formes telles que des dépressions ou des monticules atteste l'efficacité du modèle de visualisation. Mais ces éléments ne représentent pas un grand intérêt pour la découverte archéologique. Pour éviter une accumulation d'informations superflues, l'idée consiste à identifier des différences notables entre les sites culturels et naturels découverts, dans le but de générer des motifs caractéristiques.

Les précédentes missions BRASS se sont spécialisées sur la méthode dite « Poids des preuves » (Weights of Evidence) afin de prédire la localisation de sites grâce à la combinaison de variables spatiales. La réserve a fait l'objet d'une classification en classes de preuve, dans l'optique de qualifier la probabilité de trouver des sites archéologiques. Cette méthode, bien que probabiliste, a donné des résultats cohérents. Ainsi, ces modèles prédictifs peuvent être inclus en donnée d'entrée, ne serait-ce que pour initialiser les recherches dans les régions à plus haute probabilité.

BRASS/El Pilar n'est pas le seul projet archéologique mésoaméricain. Une autre cité perdue maya a attiré l'attention d'archéologues. Diane et Arlen Chase ont exploré la cité de Caracol depuis 1985. Ils ont introduit récemment LiDAR parmi leurs méthodes de levé. Les données collectées par des capteurs hyperspatiaux et LiDAR ont été comparées avec celles provenant des levés précédents. Trois objectifs comptent : 1) détecter des éléments archéologiques autour de la cité et au-delà; 2) déterminer les caractéristiques de la forêt tropicale dite 'forêt de pluie' et 3) distinguer habitation et végétation sous la canopée[Chase 2012). L'usage d'imagerie satellitale, IKONOS et Landsat TM, guide l'analyse de la signature spectrale. Cela permetde créer des motifs pour reconnaître les formes artificielles vers une extraction automatique. Les capteurs LiDAR et hyperspatial renseignent des informations complémentaires, respectivement au-dessous et au-dessus de la canopée.

Dans la lignée du projet Caracol, la mission BRASS/ El Pilar 2013 a rencontré des difficultés. Du temps aurait dû être économisé avec une meilleure gestion des données. Ce qu'il faut retenir, c'est l'innovation et les pistes de recherche tirées d'untel projet. En effet, l'entrée du LiDAR en 2012 a modifié la façon de penser en termes d'interprétation et de découverte. Un tout nouveau protocole a dû être développé en vue la prospection archéologique. Comme nous faisions confiance à cet « outil magique », la première méthode n'abordait que la détection d'éléments. Toutes les formes aperçues étaient associés à des sites potentiels, sans supposer l'existence d'artifices. Mais finalement, l'existence des 'poxes' a soulevé le problème des pièges dans l'interprétation du LiDAR. Ô combien rapide et précise cette technique puisse paraître, l'expédition sur le terrain doit persister en guise de contrôle de l'interprétation du LiDAR.

Le présent mémoire a représenté une opportunité exceptionnelle pour bon nombre de raisons. Travailler dans le cadre d'un projet archéologique m'a offert la possibilité de comprendre le monde de la recherche dans un cas concret. En outre, j'ai pu y découvrir une curiosité à l'égard des produits issus du LiDAR et la gestion de leur qualité. Le LiDAR a un rôle majeur à jouer pour les décennies à venir mais il reste encore tant de zones d'ombre…

Les expériences du bureau et du terrain sont une richesse que l'école ne peut apporter. Et, par- delà les problèmes vécus lors la collecte et l'organisation des données, j'ai retenu bien des leçons, tant sur le plan professionnel que personnel.Beaucoup de pistes à améliorer sur le chemin de l'ingénieur.

Références

BRASS / El Pilar

Mémoires de recherche

MERLET Sébastien. *Étude de l'utilisation des SIG dans le cadre d'un projet archéologique multi-échelles*. Mémoire de travail de fin d'études présenté en vue de l'obtention du Diplôme d'Ingénieur de l'ESGT, Le Mans : ESGT, 2009, 98 p.

Articles de périodiques électroniques

FORD Anabel, CLARKE Keith C. & RAINES Gary. *Modeling Settlement Patterns of the Late Classic Maya Civilization with Bayesian Methods and Geographic Information Systems*. Annals of the Association of American Geographers [en ligne], 2009, vol. 99, no 3, pp. 496-520.
Disponible sur : http://www.tandfonline.com/doi/abs/10.1080/00045600902931785. (consulté le 07.02.2013)

FORD Anabel & NIGH Ronald. *Origins of the Maya Forest Garden: MayaResource Management*. Journal of Ethnobiology, [en ligne], 2009, vol. 29, no 3, pp. 213-236. (consulté le06.02.2013)

Sites web

ESP. Exploring Solutions Past – The Maya Forest Alliance, [en ligne].
Disponible sur : www.espmaya.orghttp://www.espmaya.org/. (consulté le 16.11.2013)

ESP. The El Pilar Forest Garden Network, [en ligne].
Disponible sur : www.mayaforestgardeners.org. (consulté le 16.11.2013)

FORD Anabel. El Pilar Conserving the Past for the Future, [en ligne].
Disponible sur : elpilarprograms.blogspot.com. (consulté le 24.11.2013)

ISBER/MesoAmerican Research Center. BRASS/El Pilar program, [en ligne].
Disponible sur : http://www.marc.ucsb.edu/elpilar/Join.php. (consulté le 16.11.2013)

ISBER/MesoAmerican Research Center. Meso American Research Center, [en ligne].
Disponible sur : www.marc.ucsb.edu. (consulté le 16.11.2013)

Caracol

Articles de périodiques électroniques

CHASE Arlen F., CHASE Diane Z., FISHER Christopher T., LEISZ Stephen J. & WEISHAMPEL John F.. *Geospatial Revolution and Remote Sensing LiDAR in Mesoamerican Archaeology*. PNAS, [en ligne], 2012, vol. 109, no. 32, pp. 12916-12921.
Disponible sur : www.pnas.org/cgi/doi/10.1073/iti3212109. (consulté le19.02.2013)

CHASE Arlen F., CHASE Diane Z., WEISHAMPEL John F., DRAKE Jason B., SHRESTHA Ramesh L., SLATTON K. Clint, AWE Jaime J., CARTER William E.. *Airborne LiDAR, Archaeology, and the Ancient Maya Landscape at Caracol, Belize*. Journal of Archaeological Science, [en ligne], 2011, vol. 38, pp. 387-398.
Disponible sur : http://www.caracol.org/include/files/chase/ChaseEtAl2011.pdf. (consulté le 25.02.2013)

Sites web

CHASE Arlen, CHASE Diane. Caracol, [en ligne].
Disponible sur : http://www.caracol.org/index.phphttp://www.espmaya.org/. (consulté le 19.02.2013)

Communication dans un congrès
WEISHAMPEL John F., CHASE Arlen F., CHASE Diane Z., DRAKE Jason B., SHRESTHA Ramesh L., SLATTON K. Clint, AWE Jaime J., HIGHTOWER Jessica, ANGELO James. Remote Sensing of Ancient Maya Land Use Features at Caracol, Belize Related to Tropical Rainforest Structure [en ligne]. In : FORTE Maurizio, CAMPANA Stefano, LIUZZA Claudia (éd.), *Space, Time, Place* ,Third International Conference on Remote Sensing in Archaeology, Tiruchirappalli (India), 17-21 Aout 2009. Archaeopress, 2010, pp. 45-52. Disponible sur : http://www.caracol.org/include/files/chase/SpaceTimePlace.pdf . (consulté le 19.02.2013).

LiDAR
Articles de périodiques électroniques
PINGEL Thomas J., CLARKE Keith C. & McBRIDE William A.. *An Improved Simple Morphological Flter for the Terrain Classification of Airborne LiDAR Data*. ISPRS Journal of Photogrammetry and Remote Sensing, [en ligne], 2013, vol. 77, pp. 21–30.
Disponible sur : http://www.sciencedirect.com/science/article/pii/S0924271613000026 (consulté le04.03.2013)

SITHOLE George & VOSSELMAN George. *Experimental comparison of filter algorithms for bare-Earth extraction from airborne laser scanning point clouds*. ISPRS Journal of Photogrammetry and Remote Sensing, [en ligne], 2004, vol. 59, pp. 85-101.
Disponible sur :
http://champs.cecs.ucf.edu/Library/Journal_Articles/pdfs/Experimental%20comparison%20of%20filter%20algorithms%20for%20bare-Earth%20extraction%20from%20airborne%20laser%20scanning%20point%20clouds.pdf (consulté le01.03.2013)

SULAIMAN Nurul Shahida, MAJID Zulkepli & SETANHalim. *DTM Generation from LiDAR Data by Using Different Filters in Open-source Software*. Geoinformation Science Journal, [en ligne], 2010, vol. 10, no 2, pp. 89-109.
Disponible sur :
http://eprints.utm.my/27792/1/NurulShahidaSulaiman2010_DtmGenerationfromLidarData.pdf. (consulté le01.03.2013)

Sites web
The Light Fantastic: Using Airborne Laser Scanning in Archaeological Survey , 2009. In : CRUTCHLEY Simon & CROW Peter. . English Heritage, [en ligne].
Disponible sur : http://www.english-heritage.org.uk/publications/light-fantastic/light-fantastic.pdf . (consulté le 06.02.2013)

LASer (LAS) File Format Exchange Activities, 2012. In : ASPRS (American Society for Photogrammetry and Remote-Sensing). ASPRS Online, [en ligne].
Disponible sur : http://asprs.org/Committee-General/LASer-LAS-File-Format-Exchange-Activities.html . (consulté le 15.02.2013)

Visualisation
Articles de périodiques électroniques
CHALLIS, K., FORLIN, P., KINCEY, M. *A Generic Toolkit for the Visualisation of Archaeological Features on Airborne LiDAR Elevation Data*. Archaeological Prospection, [en ligne], 2011, vol.18, no4, pp.279-289.
Disponible sur : http://onlinelibrary.wiley.com/doi/10.1002/arp.421/abstract . (consulté le 21.02.2013)

Articles de périodiques imprimés
DEVEREUX B. J., AMABLE G. S., CROW P. Visualisation of LiDAR terrain models for archaeological feature detection, Antiquity. 2008, vol.82, no316, pp 470-479.

Sites web
Analyse en composantes principales. In : Wikipédia [en ligne]. Dernière modification de cette page le 22.12.2012 à 21h34.
Disponible sur : http://fr.wikipedia.org/wiki/Analyse_en_composantes_principales#Poids . (consulté le 11.02.2013)

Case Studies: DerwentWISE - Sky-View Factor - NERC Barlings Lidar. In : MALONE S. J. Archaelogoical Projects Services, [en ligne].
Disponible sur : http://www.apsarchaeology.co.uk/services/lidar/index.php?page=Services_LIDAR_CS1 . (consulté le 15.02.2013)

HESSE Ralf. *LiDAR-derived Local Relief Models – A new tool for archaeological prospection,* Archaeological Prospection, [en ligne], 2010, vol.17, no 2, pp.67-72.
Disponible sur :
http://onlinelibrary.wiley.com/doi/10.1002/arp.374/abstract . (consulté le 21.02.2013)

KOKALJ Žiga, ZAKŠEK Klemen & OŠTIR Krištof. Application of Sky-view factor for the Visualisation of Historic Landscape Features in LiDAR-derived Relief Models [en ligne], 2011.
Disponible sur : http://www.thefreelibrary.com/Application+of+sky-view+factor+for+the+visualisation+of+historic...-a0252190858 . (consulté le 25.02.2013)

LiDAR DEM Visualizations. In : KOKALJ Žiga, OŠTIR Krištof & ZAKŠEK Klemen. Archaeological Landscapes, [en ligne], 2012.
Disponible sur : http://www.archaeolandscapes.eu/index.php/en/interpret/topography/207.html . (consulté le 25.02.2013)

Présentations
KOKALJ Žiga, OŠTIR Krištof & ZAKŠEK Klemen. DEM Visualization Techniques for Archaeological Interpretation [en ligne], 2012.
Disponible sur :http://iaps.zrc-sazu.si/en/svf#v . (consulté le 21.02.2013)

Divers
Sites web
Accord entre 2 juges. In : BONNARDEL P. Coefficient de Kappa, [en ligne].
Disponible sur : http://kappa.chez-alice.fr/Kappa_2juges_Def.htm . (consulté le 22.02.2013)

ArcGIS® Help, http://resources.ArcGIS®.com/en/help/main/10.1/index.html#//00qn0000001p000000 . (consulté le 10.03.2013)

GPS Garmin® 62s, User guide (p.36)
http://www.tramsoft.ch/downloads/Garmin®/manuals/Garmin®_gpsmap62-series/gpsmap62-series_owners-manual_fr.pdf . (consulté le 01.05.2013)

La Division des périodes historiques de la civilisation Maya. In : Michael D. Coe. Noname, [en ligne].
Disponible sur : http://www.noname.fr/cisco/HistMaya.html. (consulté le 19.02.2013)

Supports de cours

BACCINI Alain. Statistique Descriptive Multidimensionnelle, 2010 . Université Paul Sabatier (Toulouse). Disponible sur : http://www.math.univ-toulouse.fr/~baccini/zpedago/asdm.pdf . (consulté le 11.02.2013)

Interprétation du LiDAR : les défis dans l'identification des anciens sites mayas

Hugo BIHR- ESGT

Depuis quelques décennies, le LiDAR n'a eu de cesse de gagner en importance au sein de la communauté archéologique. Plus précise que la photogrammétrie et plus rapide que les méthodes de levé terrestres, cette technique laser plaît surtout pour sa capacité à dévoiler le sol caché par la végétation épaisse, sous la forme d'un nuage de points[figure 1]. Certes, un affichage judicieux de ce nuage de points permet de distinguer les édifices les plus imposants. Mais quand il s'agit d'observer les sites de dimensions moindres et aux formes moins prononcées, la détection s'annonce plus délicate. L'étude aborde diverses questions, afin de faire des choix successifs dans la réalisation d'un modèle de surface à visualiser. Viennent ensuite des règles de validation, basées à la fois sur l'interprétation des modèles et les expéditions de terrain en guise de contrôle.

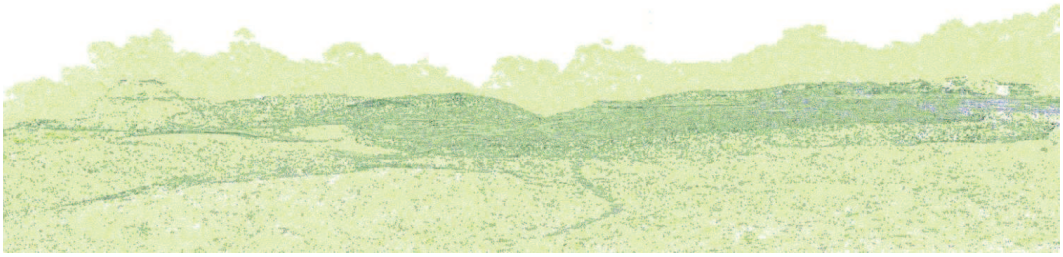


Figure 1 : Aperçu 3D d'un nuage de points. Végétation en vert, sol en bleu

BRASS/El Pilar : l'archéologie sous la canopée

Le projet *BRASS/El Pilar* (Belizean River ArchaeologicalSettlement Survey) se situe en Amérique Centrale, à cheval sur la frontière Belize-Guatemala [figure2]. Il s'agit d'une réserve de 3,5x6 km² recouverte exclusivement par la forêt tropicale. Sous la direction du Dr Anabel FORD, archéologue de l'UCSB (Santa Barbara, ÉU), le projet a deux objectifs principaux : 1) conserver les sites anciens mayas de la cité d'El Pilar et des environs ; 2) étudier l'occupation du sol et en particulier la technique agricole des *Jardins de la forêt*. Depuis 1983, de nombreuses expéditions se sont concentrées au centre, révélant ainsi une partie des vestiges. En mai 2012, la réserve a été survolée par hélicoptère muni d'un système LiDAR : le nuage de points 3D acquis contient un demi-milliard de points dont seulement 3% appartiennent au sol.

→Le LiDAR trouve sa place dans un projet de découverte archéologique freiné par la végétation dense, mais ne peut permettre de détecter directement les vestiges mayas. La création d'un modèle de surface s'impose.

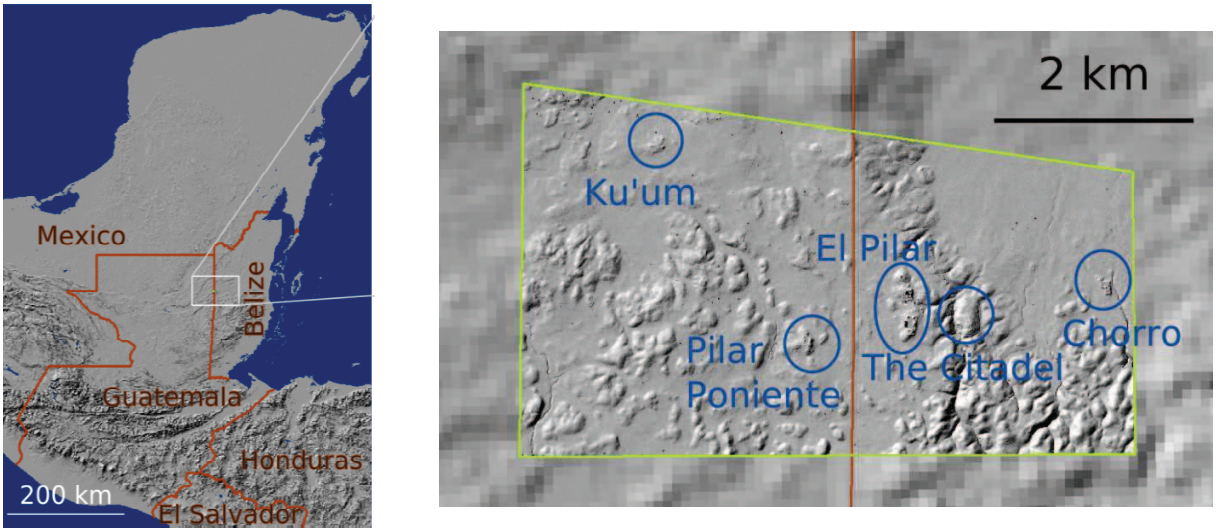


Figure 2 : Carte de localisation de la réserve d'El Pilar et de ses sites majeurs

Un modèle de surface en fonction du nuage de points de sol

La question du format se pose en premier lieu : raster, courbes de niveau ou TIN (TriangularIrregular Network) [figure 3]. Les courbes de niveau font ressortir les aspérités du terrain à condition d'utiliser un pas inférieur au mètre, ce qui entraîne une abondance d'information dans les zones de fortes pentes. La représentationen triangles du TIN est certes fidèle au semis de points initial, mais les sites archéologiques ressortent difficilement. Leformat raster n'affiche pas plus de détails. C'est son aspect plus « lisse » et les nombreuses possibilités de traitement qui en font le format adéquat.

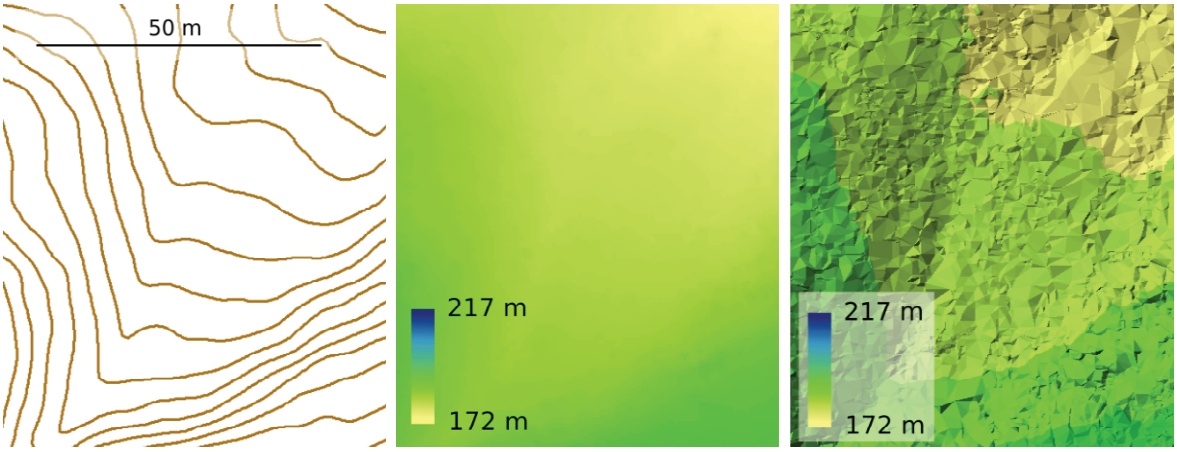


Figure 3 : courbes de niveau 1 m / TIN créé avec les paramètres par défaut / MNT raster avec l'algorithme Natural Neighbor et une taille de cellule de 0,5 m

La création du raster implique deux paramètres : algorithme d'interpolation et pas d'échantillonnage (appelé aussi taille de cellule). En réalité, le choix de ces paramètres est souvent placé en arrière-plan, l'utilisateur se fiant à l'expérience ou bien à des contraintes de projet. Dans un contexte archéologique, il est légitime d'accorder plus d'importance à un paramètre pouvant avoir une influence sur la détection de sites.

De fait, l'accent a été mis sur la relation entre nuage de points de sol et le raster, en créant des variables en fonction de deux paramètres issus de chaque format. Pour le nuage de points, les paramètres concernent la répartition des points, en termes de densité et d'homogénéité, ainsi que les précisions de positionnement (horizontale et verticale).

Une variable a retenu notre attention : *Count Frequency*(Fréquence de comptage). Une grille de taille de cellule fixe divise le nuage de points. L'objectif consiste à déterminer le nombre de cellules contenant exactement 1 point de sol. L'analyse a été réalisée sur une partie de la réserve pour réduire les temps de calcul. Plusieurs tailles de cellule ont été utilisées, allant de 50 cm à 5 m. Les résultats montrent un maximum de cellules à point de sol unique, estimé aux alentours de 1,2 m [figure 4].

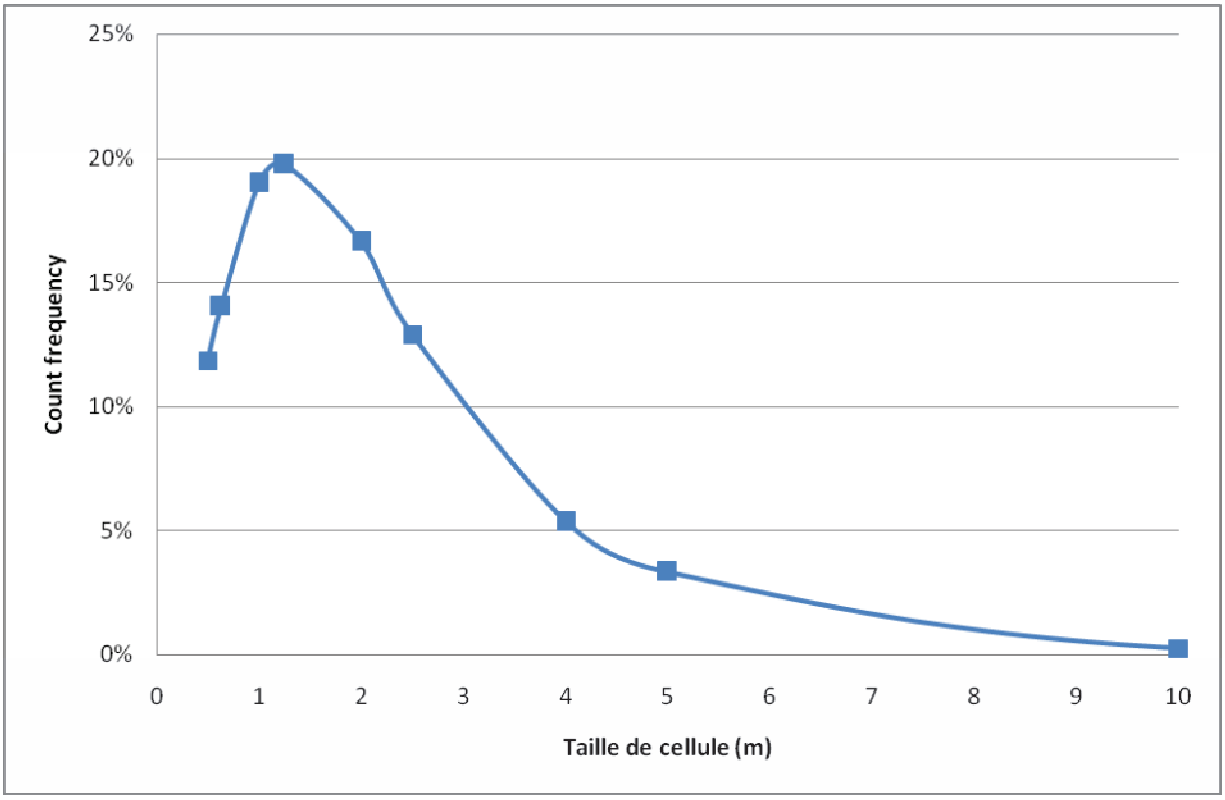


Figure 4 : Fréquence de comptage en fonction de la taille de cellule. Les pourcentages représentent la proportion de cellules à point de sol unique sur les échantillons testés à différentes tailles de cellule (carrés).

Quant à l'algorithme d'interpolation, la sélection s'est établie selon trois critères :

- Critère préliminaire visuel : rejet des algorithmes présentant des défauts majeurs
- Sensibilité à la configuration du nuage de points, c'est-à-dire la densité et l'homogénéité de la répartition des points
- Sensibilité au bruit, c'est-à-dire la présence de points hors-sol mais non filtrés en amont

Pour tester ces sensibilités, la simulation d'un sol (motif de colline) a permis de concevoir un nuage de points avec des densités et homogénéités variables. Dans un second temps, certains points ont été falsifiés, avec une altitude diminuée de 1 m. Les algorithmes usuels ont été comparés, dont *Voisin*

Naturel(*Natural Neighbor*), *Spline* et *Krigeage*. Il est donc possible de déterminer l'écart entre l'altitude vraie et l'altitude interpolée.

Au premier abord, tous les algorithmes affichent les mêmes résultats. Globalement, la configuration de points a la même incidence quel que soit l'algorithme. Comme on peut s'y attendre, plus le réseau est dense et homogène plus les erreurs diminuent (écart moyen variant entre 5 cm et 30 cm selon l'algorithme).

Le cas le plus intéressant se rapporte à l'influence des points de bruit sur la réalisation du modèle de surface. En effet, le comportement diffère selon la méthode d'interpolation. *Spline* est bien plus sensible à la présence de points faux que les autres algorithmes avec des erreurs atteignant 9 m. *Voisin Naturel* se dégrade très légèrement avec des erreurs moyennes de 20 cm. Le *Krigeage* se révèle être la plus résistante au bruit, en affichant des valeurs d'écarts semblables au test sans point faux.

→Trois tailles de cellule ont été retenues : 1 m, 1,25 m et 2 m. Quant à l'algorithme, la rapidité du traitement a fait pencher la balance en faveur de la méthode *Voisin Naturel*. Cependant, la détection de sites n'est pas améliorée pour autant. Une phase de traitement supplémentaire est nécessaire : l'algorithme de visualisation.

Le recours aux rasters de visualisation

L'altitude ne représente pas une variable adaptée à la découverte de sites potentiels. Si les structures les plus grandes apparaissent naturellement, les petites se fondent dans le paysage. Cela implique la création de modèles dits *rasters de visualisation*, dans le but de faire ressortir toutes les structures quelles que soient leurs dimensions. Pour ce faire, plusieurs algorithmes ont été développés, faisant intervenir directement ou non à la pente.

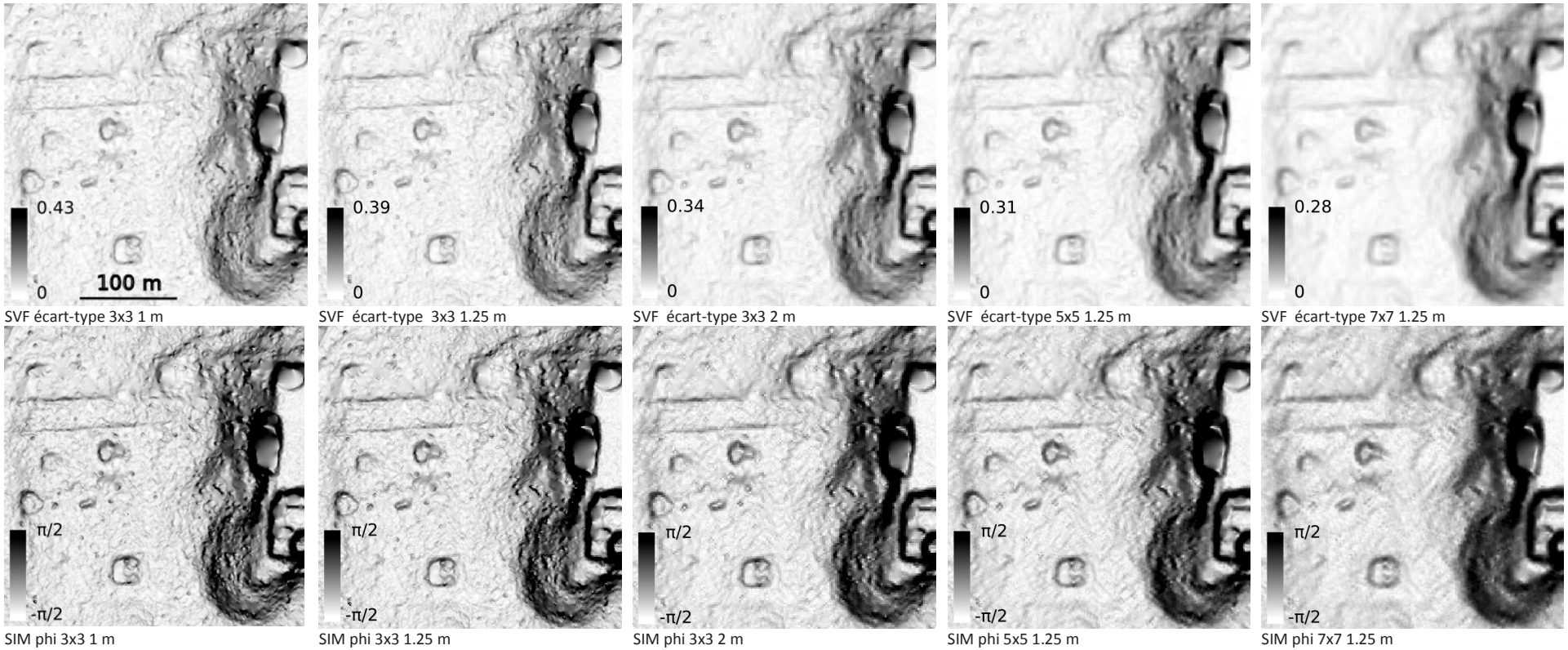
Dans le cadre de l'étude, cinq modèles ont été façonnés sous la forme de filtres de convolution à appliquer au raster d'altitude en entrée. Le filtre de convolution est carré et adopte 3 tailles prédéterminées : 3x3 px, 5x5 px et 7x7 px. Les modèles s'inspirent aussi bien des algorithmes usuels que d'expérimentations récentes. Parmi eux, deux modèles ont offert des résultats probants en terme de détection de sites, nommés *pseudoSVF* et *pseudoSIM*.

Le premier tient ses origines de l'algorithme *SkyView Factor* (SVF), qui détermine la part de visibilité moyenne du ciel depuis un point de vue. Le calcul de cette moyenne se fait pour une série de directions données. Un sommet de colline aura un *SVF* proche de 1 tandis qu'un point bas de dépression vérifiera un *SVF* quasi nul. Dans le cadre de l'étude, il a été découvert que l'écart-type faisait apparaître plus de détails que la moyenne.

Le second s'inspire de l'algorithme *Solar Insolation Modeling*(SIM), dont le but est de déterminer différentes valeurs de radiations solaires. Ici, il s'agit de déterminer la direction lumineuse tel que l'éclairement soit maximal, et ce pour chaque surface élémentaire (représentée pas un plan centré sur chaque pixel). Deux angles sont ainsi calculés : l'angle horizontal et l'élévation. Noté φ_{max}, ce dernier fait également ressortir nombre de détails.

Ces modèles donnent lieu à des rasters de visualisation créés pour les 3 tailles de cellule et les 3 tailles de filtre de convolution choisies [figure 5].

Figure 5 : Extraits de rasters de visualisation.



SVF écart-type: pseudo modèle *SkyView Factor*
SIM phi: pseudo modèle *Solar Insolation Modeling*

Tailles de filtre de convolution : 3x3 px, 5x5 px et 7x7 px
Tailles de cellule : 1 m, 1,25 m et 2 m.

Comme on peut s'y attendre, les tailles de cellule et de filtre ont un impact sur la visibilité des sites. Lorsque la taille de cellule augmente, la lisibilité du raster diminue, du fait de l'effet pixellisé. La taille du filtre fait intervenir les pixels du voisinage : une taille plus grande aura donc tendance à aplanir le terrain, les pentes s'estompant.

→Ce qui compte surtout, c'est l'interprétation des rasters. Il faut accorder plus d'importance aux rasters de tailles petites, qui offrent nombre de détails qui laissent à penser que de nombreux sites archéologiques se trouvent dans la réserve. Cependant, une connaissance préalable du terrain permet de déceler des formes artificielles apparaissant sur les rasters. En raison d'une mauvaise classification des points de sol, le nuage de points inclue des points hors-sol et engendre des formes pouvant être assimilées à tort à des sites archéologiques.
Une validation sur le terrain permet de vérifier que l'interprétation des sites correspond bel et bien dans la réalité.

Trois règles pour valider l'interprétation des produits du LiDAR

La détection et l'interprétation de sites archéologiques fut visuelle. A priori, la visibilité de ces sites dépend de nombreux facteurs (dimension et forme des sites, paramètres de création du raster, conditions d'affichage : couleurs, échelle). Dans le cadre de notre étude, la détection sur les rasters issus du LiDAR a été combinée avec la validation sur le terrain, de manière à apporter un moyen de contrôle.

Trois règles de validation ont ainsi été mises en place, faisant intervenir deux paramètres :

- La visibilité sur les rasters de visualisation : oui, non, peut-être
- Le type de site : culturel (site archéologique à proprement parler), naturel (formes paysagères) et artificiel (forme visible sur les produits LiDAR mais n'existant en réalité)

La première règle part du terrain vers les rasters de visualisation. Elle concerne les sites archéologiques découverts à l'issue des précédentes expéditions sur le terrain. Déjà intégrés sous SIG, ils peuvent ainsi être observés sur les rasters. Cela permet de tester la visibilité des 165 sites.

Les deuxième et troisième règles suivent le chemin inverse. En observant un raster de référence, une liste de sites potentiels (aussi bien culturels qu'artificiels) est dressée. Les coordonnées de ces sites sont intégrées dans des GPS de navigation pour l'exploration. Le raster de référence choisi est celui offrant le maximum de détails, autrement dit le raster dont les tailles de cellule et de filtre sont minimales (pseudo-SVF 1 m et 3x3 px). 80 sites sont supposés culturels ; 45 sites ont interprétés comme étant artificiels. La deuxième règle détermine des taux de détection *a priori*, en prenant en compte l'interprétation des formes. La troisième règle résume les taux de détection *a posteriori*, en intégrant les informations issues des expéditions sous la canopée.

Les trois règles de validation se sont contredites. D'un côté, les deux premières donnent l'avantage aux rasters de petites tailles de cellule et de filtre. Puisqu'ils sont détaillés, ils permettent de déceler un grand nombre de sites potentiels. D'un autre côté, la troisième règle met en lumière les rasters produits avec des paramètres plus grands engendrent moins d'erreur quant aux sites artificiels. Ils détectent peu mais efficacement.

→En conséquence, l'interprétation du LiDAR ne peut être employée seule : l'expérience du terrain a toujours un rôle majeur à jouer, dans le but d'assurer un contrôle tout en cernant les obstacles issus de la technique LiDAR.

Une démarche à adopter en vue de la détection des sites anciens mayas

Le LiDAR joue un rôle majeur dans la détection de sites sous la canopée. Cependant, il est important de rappeler que si cette technique est rapide et plutôt précise, elle recèle encore nombre de zones d'ombre qui peuvent représenter des freins à la prospection archéologique. En premier lieu, l'affichage en format de nuage de points n'est pas adapté. Il a fallu procéder à des traitements, sous format raster car plus facilement utilisable. De fait, cela implique des choix qui vont entraîner une perte de fidélité vis-à-vis de la réalité du terrain. Le choix de la taille de cellule n'a été fait que pour un échantillon de la réserve : l'analyse gagnerait en pertinence en considérant l'intégralité.

La qualité du nuage de points fait déjà l'objet de nombreuses recherches. Des algorithmes sont développés dans le but de mieux classifier les points de sol et ainsi filtrer les points de bruit à l'origine des formes artificielles. Lors de cette étude, des modèles se sont inspirés de ces algorithmes. Le modèle parfait n'existe pas : chacun des rasters dérivés du LiDAR ont mis en évidence les sites archéologiques avec plus ou moins de clarté. Tous confortent l'idée que la détection des sites repose en grande partie sur l'étude de la variation de la pente. Pour aller dans ce sens, il paraît opportun d'initialiser la détection par une classification par seuillage. Les pixels des rasters seraient ainsi regroupés en fonction de la variation de la pente. Ceci permettrait alors de configurer les classes : d'un côté le sol et de l'autre les éléments d'intérêt (avec si possible une distinction entre culturel, naturel et artificiel). Encore faut-il trouver les seuils différenciant ces classes.

À l'issue de ce travail de recherche, un frein demeure : l'interprétation des produits dérivés du LiDAR doit toujours être accompagnée d'expéditions sur le terrain. Sans l'appui d'un contrôle en explorant la forêt tropicale,rien ne permet d'assurer l'existence des formes aperçues sur l'écran d'ordinateur. Analyser les différences entre les sites existants et les formes artificielles engendrées par le nuage de points offre un outil d'interprétation complémentaire de la classification par seuillage. Cette étude visuelle initie le processus de développement d'une méthode de détection automatique, laquelle s'intéresse aussi bien aux sites existants que les éléments artificiels engendrés par le nuage de points de sol. Faire ressurgir les sites archéologiques est une notion essentielle en détection. Plus importante est la faculté de savoir rejeter les sites qui ne le sont pas.

Un protocole illustre sommairement le cycle des étapes pour aboutir à la cartographie des sites archéologiques [figure 6]. Ce processus propose de considérer aussi bien les sites dits 'acceptés' (sites culturels et naturels identifiés sur le terrain) que les sites 'rejetés' (non identifiés sur le terrain donc artificiels) dans le but decaractériser l'influence de paramètres environnementaux – végétation de la canopée, densité du sous-bois, aspérités du relief - sur la qualité du nuage de points de sol. Ces sites rejetés interviendront ensuite dans le processus de détection et rejet des formes artificielles.

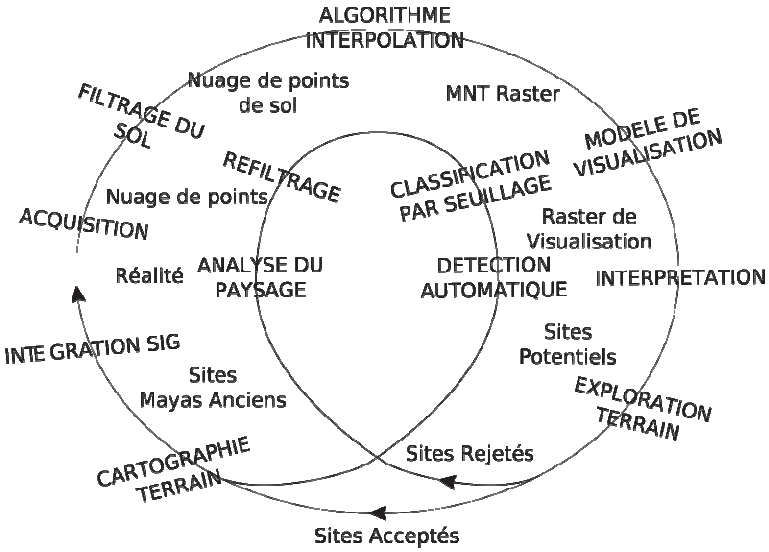


Figure 6 : Protocole de cartographie des sites anciens mayas