

Recensement Agricole Mayotte 2010

Traitement et interprétation des données géoréférencées

Présentation des données

Une des particularités du recensement agricole 2010 de Mayotte a été le relevé GPS systématique des parcelles agricoles. Ces relevés étant réalisés sur la base d'un échantillon, le traitement des données géoréférencées implique des méthodes d'extrapolation afin de refléter les données à l'échelle du territoire et d'interpolation permettant de s'affranchir du secret statistique tout en gardant la pertinence et la fiabilité des résultats.

Ce recensement a été précédé d'un pré-recensement agricole en 2009 : étape indispensable à la réalisation du RA 2010 qui a permis d'identifier et de géoréférencer de manière exhaustive les exploitations agricoles.

Pré-Recensement 2009 : Les données sur le bâti

Historique

Pour le recensement de la population (RP) 2007, l'INSEE a utilisé et mis à jour des tables de bâtiments issues de fonds cadastraux assemblés. Sur cette base, à l'aide des photos aériennes de l'IGN 2003 et 2008 et de récepteurs GPS, les enquêteurs du SISE de la DAAF ont vérifié la position de tous les bâtiments de l'île et ont répertorié les modifications, créations ou éventuelles suppressions. La destination des bâtiments (non habité, logement non agricole, logement agricole) a aussi été renseignée. Ainsi, tous les ménages agricoles ont pu être identifiés et géo-localisés. La couche correspondante est à usage unique du SISE de la DAAF. Pour les partenaires extérieurs, une table spécifique localisant les 68 600 bâtiments de l'île, qu'ils soient habités ou non, a été mise à jour en janvier 2010. Les informations contenues dans cette base proviennent uniquement d'observations de terrain et ne sont pas des données d'enquête, donc non soumises au secret statistique. La table finale comporte deux champs : le premier, « Numéro bâtiment », est un numéro séquentiel à l'échelle communale ; le deuxième, « Type bâtiment », contient l'information de l'usage du bâtiment « Habité » ou « Non habité ».

Mises en garde

Cette opération de géo-localisation avait uniquement pour but de faciliter les opérations d'enquête ultérieures (permettre à un enquêteur de retrouver un logement donné). Les niveaux de précision dans le dessin des logements et leur localisation (2 à 5 m d'erreur de mesure, pouvant aller jusqu'à 10 m en cas de réflexion du signal GPS sur les tôles en zone de bangas) ne permettent en aucun cas d'utiliser cette couche en substitution de la couche cadastrale.

Il s'agit d'une opération ponctuelle et le bâti évolue très vite : des modifications importantes ont été observées entre le RP 2007 et le Pré-RA 2009, mais également dans certains quartiers entre le début et la fin du Pré-RA. La mise à jour régulière de cette couche n'est pas envisageable. Il faut donc prendre cette couche sur le bâti pour ce qu'elle est : une vision de l'ensemble des bâtiments, habités et non habités de l'île, présents en 2009 lors du passage des enquêteurs.

Recensement agricole 2010 : Les données parcelles

Historique

Au cours du recensement agricole 2010, parallèlement au questionnement direct de l'agriculteur, l'ensemble des parcelles cultivées de l'exploitation ont été relevées à l'aide de GPS. Toutes les données relatives à ces parcelles –surface, nombre et types d'associations de cultures, aménagements...– ont pu être intégrées en tant que données attributaires dans le système d'information géographique du RA 2010. Ces données géoréférencées permettent de disposer d'une information précieuse sur l'occupation du territoire agricole. Elles peuvent ensuite être recoupées avec d'autres systèmes d'informations comme les pentes issues du modèle numérique de terrain 2008 de l'IGN. Les résultats du recensement agricole de Mayotte constituent ainsi une contribution importante dans le cadre de l'analyse territoriale de l'agriculture locale.

Création du parcellaire agricole de l'échantillon

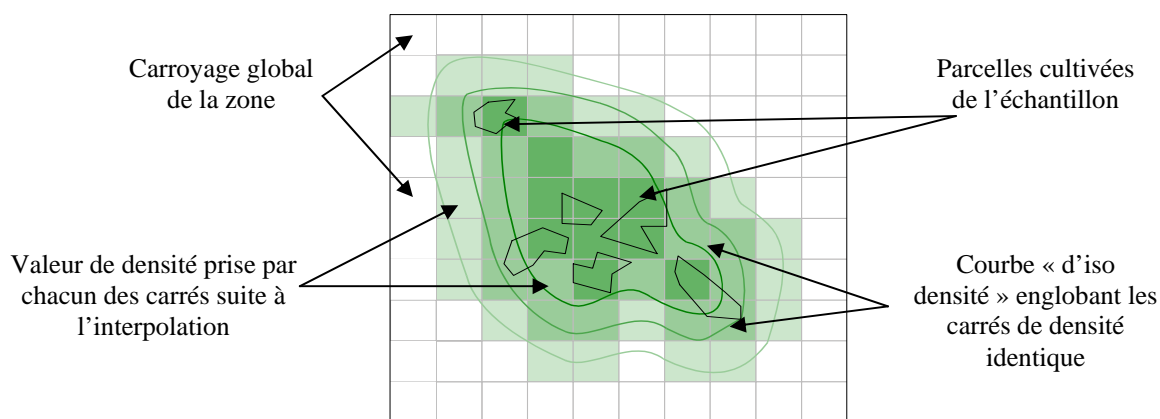
Un relevé GPS a été effectué pour chaque parcelle de chaque exploitation enquêtée. Un identifiant en lien avec le numéro du logement concerné a été attribué pour chacune de ces parcelles. La base parcellaire de l'échantillon du RA 2010 est finalement constituée de 5 690 parcelles indépendantes géographiquement, identifiées de la même manière que dans le logiciel de gestion des questionnaires (TRAPEZE).

Intégration des données du RA 2010 et d'autres sources

Les données relatives aux parcelles issues du questionnaire (TRAPEZE), ont été intégrées à la base parcellaire agricole géoréférencée. Toutes les caractéristiques du ménage agricole concerné ainsi que toutes les thématiques du questionnaire sont déclinables dans le système d'information géographique (*MapInfo*®). De même, les informations issues du MNT 2008 de l'IGN, comme la pente viennent compléter cette base. Il en résulte donc une base de données géoréférencées complète.

Découpage par carroyage de la base de données géoréférencées

Afin d'interpréter les données échantillonnées et d'aboutir à une image synthétique de l'occupation du territoire agricole mahorais, une méthode de traitement de ces données géoréférencées a été développée sur le logiciel *MapInfo*®. Elle se base sur le principe d'un carroyage du territoire dans lequel les calculs sur échantillon seront réalisés. Un maillage du territoire de Mayotte a donc été réalisé à l'aide de l'outil *GridMaker.mbx*. Se pose la question du « pas » du pixel carré : de sa taille découlera la précision des résultats de l'interpolation. Un pas trop grand (200m et plus) et l'interpolation sera non territorialisée de façon fine, un pas trop petit (20m et moins) et la précision demandée se retrouve plus restrictive que la précision des données, entraînant des artefacts dans le modèle, créant ainsi des erreurs. Le choix final d'un pas de 50m découle d'une part de plusieurs tests à des pas supérieurs non concluants et d'autre part d'avis d'experts (M. Sémécurbe notamment du pôle territorial de l'INSEE Marseille travaillant sur des sujets similaires de valorisation de données géographiques). La base parcellaire du RA 2010 a ensuite été découpée selon ce maillage créant ainsi plus de 20 000 morceaux de parcelles. La surface de chacun de ces fragments de parcelle a été recalculée. De ce fait, pour chaque fragment, sa surface ainsi que les informations sur la parcelle mère dont il est issu, permettent d'en déduire les différentes superficies allouées à chacune des cultures. L'information de la base parcellaire est ainsi morcelée dans des pixels d'aire identique, tout en gardant l'intégrité et la cohérence des relevés.



Représentation schématique du « lissage par interpolation géographique »

Extrapolation des données d'une variable surface

Pour chaque fragment de parcelle, la donnée dont on dispose est la part de la surface de la parcelle correspondante. Or, ce fragment est issu d'une parcelle ayant elle-même un certain poids dans l'échantillon en fonction de la strate dans laquelle elle se trouve. Ce poids a été matérialisé par un coefficient d'extrapolation. **L'hypothèse retenue pour la suite des travaux est que la distribution des parcelles de l'échantillon est représentative de la distribution de toutes les parcelles de l'univers considéré.** Les coefficients d'extrapolation issus des parcelles et donc des exploitations mères sont alors appliqués aux fragments de parcelles de chaque pixel. Cette étape est réalisée par un jeu de mises à jour de colonnes sur l'ensemble des données surfaciques. L'extrapolation aboutit donc à une base de données géoréférencées et extrapolées.

Interpolation et lissage géographique de la donnée extrapolée

Cette base doit maintenant être lissée par interpolation des données afin de modéliser la distribution des données considérées sur l'ensemble du territoire.

L'intérêt de ce genre de traitement est double. D'une part, il permet d'avoir une vision relativement synthétique d'une information en faisant ressortir les tendances globales, tout en s'affranchissant des limites administratives, l'unité d'analyse étant le pixel. D'autre part, il permet de s'affranchir du secret statistique qui interdit la divulgation de données individuelles d'enquête, la donnée représentée n'ayant plus de lien direct avec une exploitation.

Plusieurs méthodes permettent de développer ces nouveaux outils de modélisation. Celle utilisée dans le cadre de ce recensement a été développée tout spécialement pour celui-ci par François Sémécurbe de l'INSEE Marseille. Elle implique une méthode de lissage réalisée sous le logiciel statistique *Rgui*©.

```
1. Rayon d'interpolation
2. Fichier input = matrice d'entrée
3. Fichier output = matrice de sortie

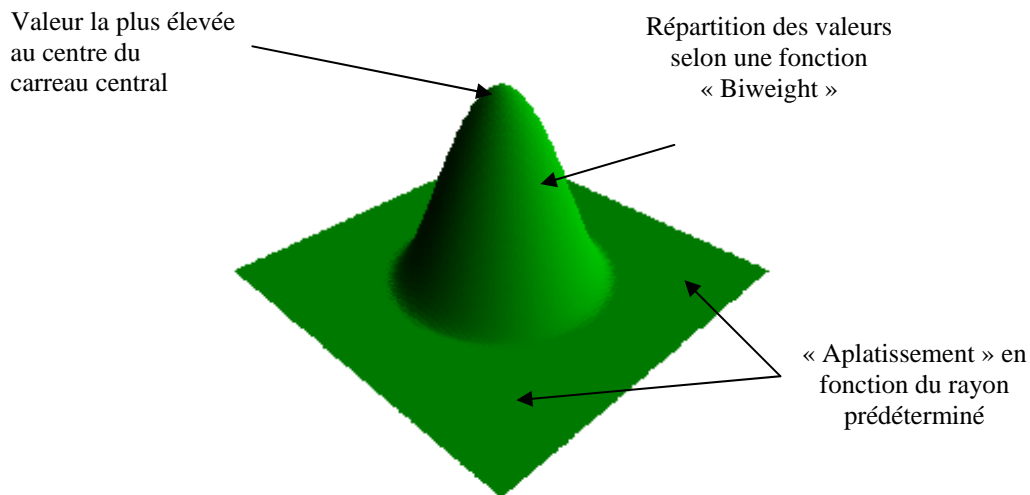
#Les paramètres
repertoire="D:/calcul R/"
rayon=1400 #attention le rayon est entre 0 et 1400

1 |
library(foreign) #pavage entrée-sortie
#chargement de la base
input=read.dbf(paste(repertoire,"bases/bases_input.dbf",sep=""),as.is=T)
(50 lignes de programme) | 2
output=data.frame(DESCRIPTION=input$DESCRIPTION,SUP_LIS=lis[cbind(input$NX,input$NY)])
write.dbf(output,paste(repertoire,"Resultats/output_c_",rayon,".dbf",sep=""))
3
```

Programmation sous le logiciel *Rgui*© (3 paramètres modifiables)

Rgui© est un outil très puissant en matière de statistique, malgré une interface relativement hermétique. Pour l'adapter au RA, seules 3 informations sont paramétrables (le rayon, le nom du fichier d'entrée et de sortie). Le format de fichier utile est le ".dbf", compatible avec le logiciel de SIG utilisé.

Le principal postulat de la méthode de lissage est que la répartition n'est pas uniforme sur le territoire mais décroît à partir de points chauds matérialisés dans le RA 2010 par les parcelles agricoles de l'échantillon. Cela se traduit par l'hypothèse que la densité de surface cultivée est décroissante selon une fonction mathématique nommée « Biweight » dès que l'on s'éloigne d'une parcelle de l'échantillon. Ainsi, il est possible de calculer avec précision cette distribution de la surface en fonction de l'éloignement, sachant qu'elle est supposée homogène en intensité quelque soit la direction prise.



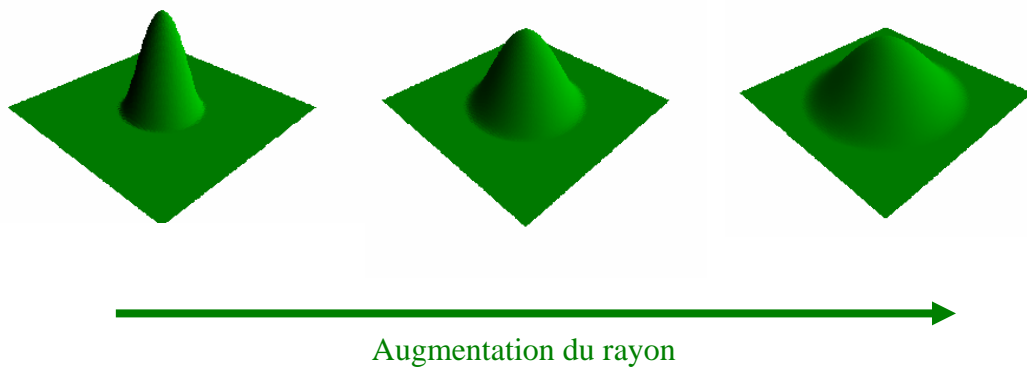
Représentation schématique du « lissage par interpolation géographique »

Pour résumer, une parcelle de l'échantillon du RA 2010 contient des données surfaciques. Une fois morcelée dans le carroyage global du territoire, un coefficient d'extrapolation y est appliqué. On en déduit donc la surface totale représentée par cette parcelle. Afin, de modéliser au mieux la distribution, on applique un lissage géographique permettant d'étendre la zone d'influence de cette donnée tout en gardant sa valeur absolue. Ainsi, la surface totale estimée est conservée, seule sa répartition géographique est influencée. On obtient ainsi une occupation du sol agricole pour la variable concernée.

Caractéristiques de la méthode de lissage géographique

Plusieurs paramètres rentrent en compte afin de calibrer la fonction mathématique.

Le plus important est le rayon de l'interpolation. En effet, la distribution est répartie de façon homogène autour d'un point chaud contenant la donnée. Néanmoins, la répartition et la dispersion de la surface est dépendante de la distance à laquelle on l'étend. Pour clarifier, comme la surface totale est conservée, plus le rayon d'impact sur le territoire est grand, plus cette surface est dispersée loin du point de mesure. A l'inverse, plus le rayon sera petit, moins il y aura de dispersion, ce qui aboutira à des valeurs relativement proches de la données mère dans un faible rayon.

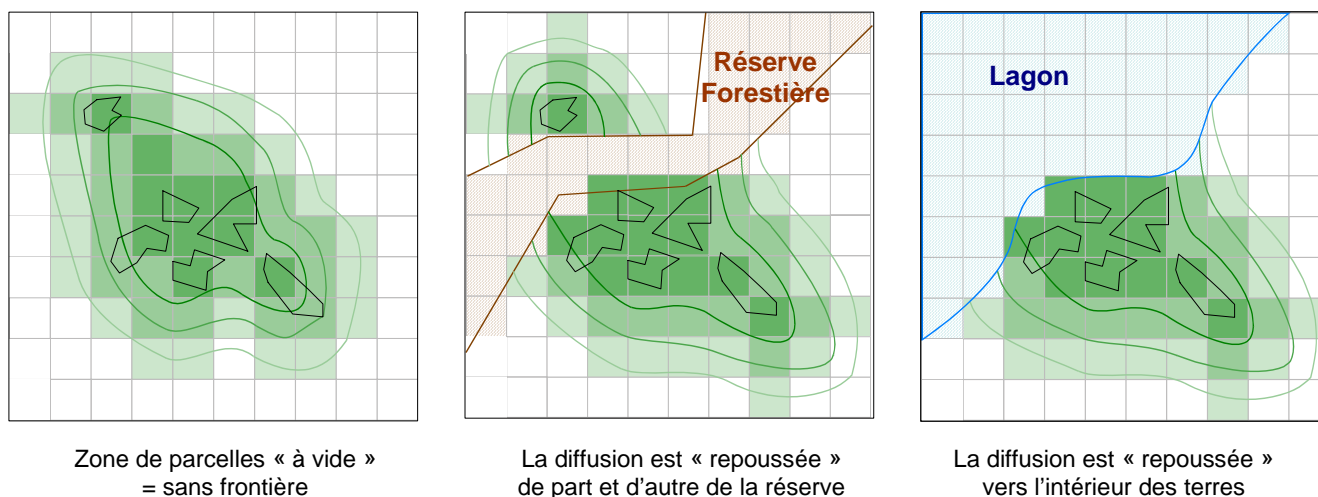


Variation de l'intensité de la dispersion en fonction de l'augmentation du rayon

Dans le cas du RA 2010, deux rayons ont été utilisés.

- Pour les cultures considérées comme à faible densité géographique, comme les cultures de vivrières qui s'étendent sur des grandes étendues, un rayon de lissage de 1 400 mètres a été utilisé.
- Pour des cultures à forte densité géographique, comme le maraichage, le rayon se doit d'être plus petit afin de reproduire au mieux les conditions de densité de cultures sur le terrain. Dans ce cas, le rayon a été fixé à 800 mètres.

Un autre paramètre est la prise en compte de l'environnement proche. En effet, d'un point de vue théorique la répartition étant uniforme quelque soit la direction, ce qui n'est pas toujours le cas dans la pratique. Ainsi, dans le cas où les parcelles sont localisées en bordure de zones urbaines ou près du littoral, la dispersion doit respecter ces limites physiques. Une solution mathématique a été trouvée et permet de renvoyer la dispersion lorsque celle-ci rencontre une barrière. Ainsi, la valeur absolue de la surface est gardée, mais sa répartition est plus forte dans les autres directions que celle de la localisation de la barrière (effet rebond). Concrètement, la diffusion du RA 2010 tient compte des « zones interdites à l'agriculture » dans lesquelles l'agriculture ne peut se mettre en place. Ce sont les plans d'eau naturels ou artificiels, les zones urbaines définies dans le pré-RA 2009, ainsi que les îlots et réserves forestières qui sont des espaces boisés remarquables et protégées pour leur caractère environnemental. Cependant, au vu du grignotage extrême de l'agriculture dans la partie nord de la réserve forestière des crêtes du nord, il a été admis pour coller à la réalité du terrain, d'autoriser la dispersion des données agricoles dans cette partie de la réserve.



Prise en compte de l'environnement lors du lissage géographique

Interprétation des résultats du lissage par un traitement SIG

Une fois les données extrapolées par pixel et interpolées entre elles avec un rayon d'influence plus ou moins grand, le logiciel délivre comme résultat un carroyage contenant pour chaque pixel une valeur représentant la part de chaque variable. L'unité de chaque carreau est donc bien une densité de présence traduisant l'occupation du sol en valorisation agricole par rapport à la surface totale du pixel mobilisable pour l'agriculture. Une fois intégré à un logiciel SIG, il est facile de regrouper différentes classes. On peut donc réaliser une analyse thématique par strates qui reflète la localisation de la variable considérée sur le territoire agricole de Mayotte. Cette stratification est réalisée sous *MapInfo*®, la détermination des seuils des différentes strates est déterminante pour l'interprétation des données. Ces seuils varient en fonction de la variable considérée, ils sont déterminés visuellement sur l'ensemble des cartes publiées dans le cadre du Recensement agricole de Mayotte 2010.

Contacts

DAAF – SISE : « Pôle Analyse, Evaluation et Prospective Territoriale »

Responsable SIG – Sébastien AUBARD,

Mél : sebastien.aubard@agriculture.gouv.fr - Tel : 02 69 61 91 11