



HAL
open science

Sensibilité des indices de diversité à l'agrégation

Ilène Mahfoud, Didier Josselin, Bruno Fady

► **To cite this version:**

Ilène Mahfoud, Didier Josselin, Bruno Fady. Sensibilité des indices de diversité à l'agrégation. *Revue Internationale de Géomatique*, 2007, 17 (3-4), pp.293-308. 10.3166/geo.17.293-308 . hal-02659883

HAL Id: hal-02659883

<https://hal.inrae.fr/hal-02659883>

Submitted on 5 Jan 2022

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.

Sensibilité des indices de diversité à l'agrégation

Ilène Mahfoud * Didier Josselin* Bruno Fady**

**UMR 6012 ESPACE – Université d'Avignon*

74, rue Louis Pasteur 84029 Avignon cedex 1

mahfoud_ilene@yahoo.fr, didier.josselin@univ-avignon.fr

***INRA d'Avignon, Unité de recherche forestière méditerranéenne*

Site Agroparc Domaine Saint Paul 84914 AVIGNON cedex 9

RÉSUMÉ. Cette recherche vise à évaluer la pertinence de l'utilisation des mesures de diversité sur des images satellitaires en abordant le problème du MAUP (Modifiable Areal Unit Problem). Nous calculons tout d'abord un indice de végétation sur une image SPOT 5 de 5 mètres de résolution (image 1). Puis, nous créons une image répartissant aléatoirement ces mêmes pixels observés sur une zone de même taille (image 2). Deux images extrêmes sont également utilisées pour comparaison : une image de la diversité nulle (tous les pixels ont la même valeur, image 0) et une image de diversité « maximale » (tous les pixels sont différents, image 3). Nous découpons alors les images 1 et 2 en grilles de différentes résolutions, croissantes et imbriquées. Considérant les pixels de l'image initiale contenus dans chaque maille, nous calculons 3 indices de diversité, Shannon, Simpson et de Dominance. Les résultats présentés sont les moyennes, médianes et distributions statistiques des valeurs de diversité des mailles de chaque image. Ils permettent de mesurer l'ampleur et l'évolution de la sensibilité de ces indices à travers les échelles et selon les types d'images. Nous comparons aussi les écarts et les rapports entre les valeurs des diversités des trois séries. Les résultats permettent d'introduire et de discuter le concept d' « échelle pertinente ».

ABSTRACT. This research aims to evaluate the relevance of using the diversity measurements on satellite images taking into account the problem of the MAUP (Modifiable Areal Unit Problem). We calculate a vegetation indice for an image SPOT 5 of 5 meters resolution. Then, we cut out the image into grids of different aggregated resolution. For each image, we calculate 3 diversity index : Shannon, Simpson and Dominance. We process a resampling of the observed data and design another image corresponding to the "maximal" diversity. Same calculations of diversity index are applied on these two images. We then measure the extent and the evolution of the phenomenon through the scales according to the types of images and the statistical distributions of the pixels. The results show how sensitive is the diversity assesment to the MAUP and how we may identify what we call a "pertinent scale".

MOTS-CLÉS : MAUP, agrégation, indices de diversité, échelle, entropie, ré-échantillonnage, robustesse.

KEYWORDS : MAUP, aggregation, diversity index, scale, entropy, resampling, robustness.

1. Introduction

En écologie de paysage, l'on a fréquemment besoin d'utiliser des données agrégées pour analyser les phénomènes spatiaux. Ainsi, la recherche en écologie est conduite à travailler sur de petites échelles (l'échelle des paysages et de la région), pour traiter notamment des dynamiques de paysages, de la biodiversité et du changement global (Jelinski et Wu 1996).

Le problème du MAUP (Modifiable Areal Unit Problem) a été observé depuis les années trente, avec davantage d'acuité par l'écologie du paysage, plus récemment. Il identifie la difficulté de produire un indice statistique synthétique à une échelle donnée, indice dont la valeur sera plus ou moins sensible à l'agrégation de l'information source. Ce problème, dans la pratique, est accentué par le développement des systèmes d'information géographique et des logiciels d'analyse spatiale et d'images, qui permettent de manipuler avec une certaine aisance le processus d'agrégation sans se poser la question de la fiabilité des mesures produites. Ces différents éléments ont favorisé l'engouement des chercheurs pour la résolution de ce problème (Openshaw, 1984, Rastetter *et al.*, 1992 ; Reynolds, 1998).

Le problème du MAUP se décompose en deux aspects imbriqués. Premièrement, l'effet de l'échelle ou du niveau d'agrégation considéré, pour lequel l'information fournie par entité géographique représente une situation moyenne des unités agrégées dans celles-ci. Cela conduit à perdre une partie de l'information selon le niveau de l'agrégation (Reynolds, 1998). Le deuxième aspect est la méthode d'agrégation elle-même. Le choix de cette méthode a un impact à nouveau sur les indices statistiques fournis à un niveau agrégé. Ainsi, dans le cas d'un nombre constant d'unités, le résultat de l'agrégation est différent (Openshaw, 1984 ; Dusek, 2005). Nous pouvons ajouter à ces deux points bien connus l'effet de l'efficacité statistique. En effet, le nombre et les distributions spatiales et statistiques des données ont une influence sur la robustesse de la méthode employée.

Plusieurs études géographiques et statistiques ont essayé de tester la sensibilité des analyses ayant recours à des données agrégées. La première observation du MAUP remonte en 1934 quand Gehkle et Biehl ont remarqué un rapport entre les coefficients de corrélation estimés et le niveau des unités de recensement utilisées dans le calcul. Des résultats similaires ont été observés par Yule et Kendall en 1950, par l'étude de la corrélation entre les rendements de blé et de pommes de terre pour 48 régions anglaises et la mise en évidence d'un effet de taille sur les mesures. Après ces observations, Robinson (1950) a montré l'effet de l'échelle sur les valeurs absolues de la corrélation quand le nombre d'observations diminue et la taille de l'unité observée augmente. En 1976, Clark et Karen ont affirmé l'importance de l'agrégation spatiale sur les analyses de la corrélation et de la régression d'un phénomène spatiale.

D'autres chercheurs ont montré les effets du MAUP sur une variété de contextes d'analyse spatiale. En 1994, Marceau *et al* ont vérifié l'impact de la résolution spatiale et du niveau d'agrégation sur la précision de la classification appliquée sur des données satellitaires. Ce travail a montré que la précision par classe a été considérablement affectée par le changement de la résolution spatiale et le niveau d'agrégation. Les données de télédétection se présentent donc comme un cas particulier du problème des unités spatiales modifiables, et y sont également particulièrement sensibles (Jelinski et Wu, 1996 ; Wu, Gao & Tueller, 1997). D'autres travaux ont analysé le lien entre le MAUP et d'autres méthodes d'analyse spatiales comme la Geographical Weighted Regression (Charleux, 2004).

En fait, beaucoup de recherches ont réussi à mettre en évidence le MAUP, mais peu offrent de solutions génériques et pratiques pour traiter les données agrégées en évitant l'effet de l'agrégation. Une première approche consiste à abandonner les méthodes statistiques traditionnelles considérées comme sensibles au MAUP, et élaborer de nouvelles méthodes d'analyse thématique (Jelinski et Wu, 1996). Openshaw (1984), quant à lui, a proposé de considérer des entités non modifiables, qui seraient insensibles à l'effet écologique d'agrégation. Cette solution nécessite beaucoup de connaissances sur les objets géographiques manipulés. Une autre solution proposée est d'appliquer un module optimal d'agrégation qui maximise la variation inter zones et minimise la variation intra zones. Le problème de cette approche provient de la méthode d'agrégation, qui, bien qu'optimale pour une variable ne l'est peut-être pas pour une autre (problème de non reproductibilité). La dernière approche porte sur l'analyse de sensibilité des calculs appliqués sur des données agrégées, afin de déterminer le poids du MAUP dans les mesures produites. D'autres pistes prometteuses sont à explorer dans le domaine des méthodes mathématiques et des simulations (Amrhein, 1995, Reynolds, 1998).

Notre recherche propose de prendre une orientation similaire aux travaux relatés dans le paragraphe précédent, en recherchant, par simulation, à éliminer le MAUP et identifier les entités pertinentes (non modifiables) correspondant à une échelle adéquate.

2. Méthodologie

2.1. Calcul des indices de biomasse et de diversité

Nous abordons le problème du MAUP en testant la robustesse de 3 indices de diversité (Shannon, Simpson et Dominance) à l'effet d'agrégation. Ces indices sont calculés à partir d'une image satellitaire, dont les pixels sont agrégés à différents niveaux d'échelle imbriqués. L'objectif visé est d'une part, de tenter d'évacuer ou

réduire l'impact du niveau d'agrégation sur les indices en question, d'autre part, de rechercher l'échelle qui permet de mesurer le mieux possible la diversité.

L'image source est une extraction d'une fenêtre de l'image SPOT 5 (résolution de 5 mètres, dimensions de 388 x 276 pixels) dans les canaux du vert, du rouge et du proche infra-rouge sur le versant Sud du Mont Ventoux, montagne du sud-est de la France (Vaucluse). A partir de cette image, nous avons calculé à l'aide du logiciel GRASS 5 l'indice de végétation normalisé (NDVI), nous permettant de réaliser une analyse de la diversité de la biomasse.

Le NDVI (indice de végétation normalisé) représente la différence de la réflectance entre les deux bandes spectrales du rouge (B3) et de l'infra rouge (B2) divisé par la somme des réflectances :

$$NDVI = (B3 - B2) / (B3 + B2) \quad [1]$$

Les indices de diversité choisis dans cette étude décrivent un aspect particulier de la diversité dans l'écologie du paysage et sont complémentaires. Ainsi, chaque indice a un intérêt écologique différent. L'indice le plus couramment utilisé dans la littérature est celui de Shannon. Son importance vient du fait qu'il combine la richesse taxonomique et l'équitabilité. En effet, l'indice de Shannon (H') est minimal si par exemple, dans un peuplement, une espèce est dominante et les autres espèces sont représentée par un seul individu, et maximal quand toutes les espèces sont variées et réparties de façon équitable (Frontier, 1983). De plus, cet indice est sensible à l'existence des espèces rares dans le peuplement (Farina, 2000). Nous avons utilisé la forme suivante, la plus classique, de cet indice :

$$H' = - \sum_{i=1}^m p_i * \ln(p_i) \quad [2]$$

Où : p_i est la fréquence de l'attribut i dans le peuplement

m est le nombre des attributs

Le deuxième indice de diversité auquel nous nous intéressons est celui de Simpson. Il mesure la probabilité avec laquelle deux individus sélectionnés au hasard appartiennent à la même espèce (Baker, 1997). Cet indice est sensible à l'existence d'une espèce dominante (Farina, 2000). Il est basé sur la formule:

$$1/S = 1 / \sum_{i=1}^m p_i^2 \quad [3]$$

Où : p_i est la fréquence de l'attribut i dans le peuplement,
 m est le nombre des attributs.

Nous utiliserons cette forme de l'indice de Simpson (le rapport $1/S$). D'autres formes complémentaires existent ($1-S$) mais ne sont pas testées dans cet article.

Le dernier indice calculé dans ce travail est celui de la dominance. Il est lié à l'indice de Shannon, mais correspond à une évaluation d'un aspect complémentaire de l'entropie, en mettant en évidence l'importance d'espèces qui dominent dans le groupe. Il se calcule selon la formule suivante :

$$D = \ln(n) - H' \quad [4]$$

Où : n est le nombre des attributs possibles dans le peuplement,
 H' = indice de Shannon.

2.2. Une approche systématique de la diversité à travers les échelles

Afin d'étudier l'effet de l'agrégation sur les indices de diversité, nous avons découpé l'image en grilles de résolutions croissantes imbriquées, avec des agrégats de pixels de longueur de côté de 10, 20, 40, 80, 160 et 320 mètres. Puis, nous avons procédé au calcul des trois indices de diversité (Shannon, Simpson et de dominance) aux 6 niveaux d'agrégation. Cette méthode nous permet de calculer, pour chaque échelle, les indices de diversité pour tous les pixels inclus dans toutes les mailles de l'image considérée. Les résultats du calcul permettent de construire les images résultantes, pour chaque indice. Pour pouvoir comparer nos résultats, celles-ci sont re-classifiées à l'aide d'une discrétisation commune à l'ensemble des images.

A ces images, nous avons ajouté le calcul de la moyenne et de la médiane des diversités calculées pour les agrégats de pixels à une échelle donnée, ainsi que la distribution statistique de ces valeurs. Ainsi, il est possible d'évaluer les sensibilités respectives des indices de diversité aux différents niveaux d'agrégation testés systématiquement (figure 1). Toutefois, l'objectif est de minimiser l'impact de

l'agrégation sur les indices de diversité ne peut être obtenu avec cette première approche.

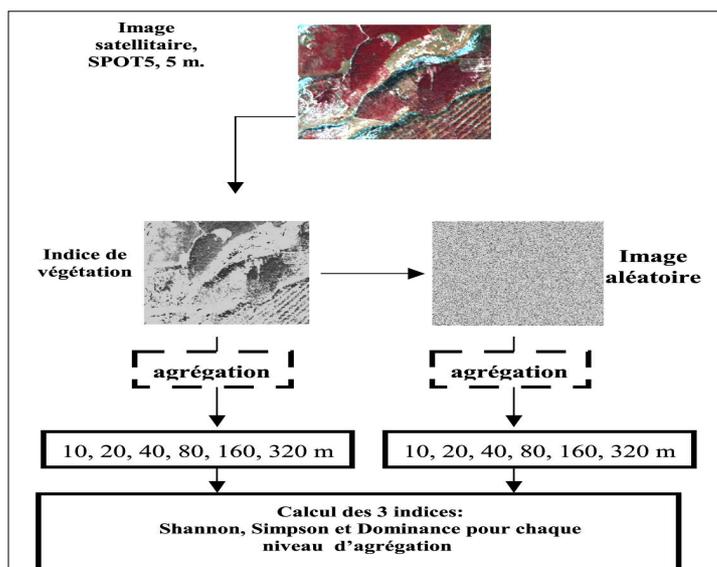


Figure 1. Méthode suivie pour l'estimation de la sensibilité des indices de diversité à l'agrégation.

2.3. Série d'images construites à l'aide de méthodes différenciées de distribution de pixels

Pour tester de l'évaluation de la sensibilité des indices de diversité à l'agrégation des pixels, nous avons développé une méthodologie de ré-affectation des pixels (notamment par ré-échantillonnage), afin de chercher à identifier, au sein de la valeur mesurée de la diversité, la part effective de celle-ci portée par la structure spatiale, indépendamment du niveau d'agrégation. Il nous fallait en effet des images de référence correspondant à différents cas de répartition de diversité.

Ainsi, nous avons conservé, pour l'ensemble des images, la taille de la première image de l'indice de biomasse (même nombre de cellules, même forme et grille identique). Puis, nous avons construits 3 autres images.

La première image est constituée d'un ensemble de pixels identiques. Dans ce cas, quelle que soit l'échelle considérée, la diversité est nulle.

A l'opposé, nous avons construit, pour chaque niveau d'agrégation, une image qualifiée de « diversité maximale », ce terme restant discutable pour les indices de Simpson et de dominance. Sur l'ensemble de la grille, nous avons disposé des pixels systématiquement de valeurs différentes. Ce cas correspond à une variété maximale, mais également à une forme d'homogénéité statistique, puisqu'il n'y a notamment aucune espèce dominante (unicité des occurrences de valeurs).

Enfin, nous avons construit, pour chaque image, son « équivalent aléatoire », en re-distribuant les pixels observés de l'image de biomasse (NDVI) au hasard dans l'image. Ce processus a été reproduit 100 fois, et des tests ont été réalisés sur plusieurs images, convergents en termes de résultats. L'image résultante représente le cas où les valeurs calculées à partir des données se retrouvent dépourvues de leur structure et de leur autocorrélation spatiale. Elle intègre ainsi, d'un côté, la part de la diversité due à la taille de l'image et au niveau d'agrégation considéré, mais également la part de la diversité due aux valeurs des pixels observés. La part de la diversité effectivement portée par la structure spatiale a théoriquement disparu suite à la distribution aléatoire. Il s'agit donc d'une situation intermédiaire avec une diversité généralement assez importante, plus élevée que celle de l'image initiale (pixels observés) mais plus faible que celle l'image de diversité « maximale ».

La comparaison des trois cas particuliers d'images (diversité nulle, maximale et aléatoire sous contrainte des valeurs de pixels observés) avec l'image des observations (diversité de l'indice de biomasse selon les 6 niveaux d'agrégation) nous permet de dégager des pistes prometteuses pour la prise en compte de l'effet du MAUP dans les mesures de diversité à travers les échelles.

3. Résultats

3.1. Forte sensibilité globale des indices

Les résultats obtenus sont présentés dans les figures qui suivent. Nous disposons ainsi de 4 types d'images étudiés pour les 3 indices de diversité et pour 6 niveaux d'agrégation :

- Le cas n° 0 représente l'homogénéité et la diversité nulle (valeurs remplacées ici par la moyenne des pixels de l'image d'origine) ;
- Le cas n° 1 est celui des valeurs de la diversité observée ;
- Dans le cas n° 2, les valeurs des pixels observés sont réparties au hasard dans l'image ;
- Enfin, le cas n° 3 est le cas de l'hétérogénéité « maximale » (cas où les pixels de l'image sont tous différents).

L'analyse de la sensibilité du calcul de diversité au MAUP est abordée de deux façons. D'une part, nous mesurons l'ampleur et l'évolution du phénomène à travers les échelles selon les types d'images et les distributions statistiques des pixels. D'autre part, nous comparons les trois types d'image, en essayant d'évaluer la part de la diversité portée par la structure spatiale et celle incluse dans le processus d'agrégation lui-même. Pour ce faire, nous croisons dans un graphique les moyennes des indices de diversité selon les types d'image avec les différentes résolutions spatiales. Puis nous calculons les écarts et les rapports entre les valeurs des courbes ainsi obtenues. Cette méthode vise à identifier la part de l'échelle dans les estimations de diversité, en vue de limiter, voire éliminer, l'effet d'agrégation. Il s'agit également d'identifier d'éventuels points d'inflexion pouvant correspondre à des échelles pertinentes pour la mesure de la diversité sous ses différentes formes décrites par les indices.

Les résultats, dans leur ensemble, montrent que la structure spatiale des données n'est pas le seul facteur d'explication des valeurs de diversité. Le niveau d'agrégation est parfois même très fortement porteur d'explication (jusqu'à plus de 90% de la diversité). Cela n'est pas étonnant compte tenu du problème clairement identifié par les différentes recherches préalablement citées. On peut toutefois constater l'importance de ce facteur dans les différents résultats obtenus. Les indices d'évaluation de la (bio)diversité sont donc à manipuler avec une grande précaution, si l'on considère le poids du support ou du découpage de l'information spatiale dans les valeurs mesurées de ces indices.

3.1. Indice de Shannon et le concept d' « échelle pertinente »

Ainsi, l'indice de Shannon montre une forte sensibilité à l'agrégation, puisque la part de la diversité liée à la structure spatiale elle-même est réduite à environ 5-15% de la diversité totale.

On observe par ailleurs que la moyenne des indices de Shannon augmente sensiblement avec la résolution dans les cas 1, 2 et 3 (Figure 2). A titre d'exemple, à la résolution de 10 mètres, la moyenne des indices de Shannon est de respectivement 1.265, 1.354 et 1.386. En revanche, à la résolution de 40 mètres, la moyenne s'élève à 3.154, 3.633 et 4.158. A la résolution de 320 mètres, la moyenne enregistre des valeurs allant de 4.155, 4.425 à 8.361 pour ces mêmes images.

Par ailleurs, dans le cas n° 1, on constate que le nombre de classes des valeurs possibles de l'indice de Shannon augmente dès que la résolution augmente, ce phénomène paraissant plus marqué que dans le cas aléatoire (Figure 2). En fait, la répartition aléatoire assure une distribution équilibrée des pixels sur la totalité de la zone. Cela renforce la probabilité d'engendrer des nouveaux agrégats à forte diversité. On voit déjà l'influence de la structure spatiale sur le nombre de classes à travers les échelles générées par l'agrégation.

Cette impact est également visible lorsqu'on compare les deux courbes correspondant aux cas aléatoire et observé (figure 3). La structure spatiale (cas n° 1) via son autocorrélation, a tendance à réduire les valeurs de diversités pour l'indice de Shannon : l'écart entre les cas 1 et 2 identifie donc la baisse de la diversité due à la structure. Il permet d'éliminer l'effet du niveau d'agrégation, puisque le cas 2 (aléatoire sous contrainte du cas n° 2) n'inclut que la diversité portée par le niveau d'agrégation donné, la taille de l'image et les valeurs de pixels observées, alors que le cas n° 2 inclut l'impact de l'organisation spatiale.

Cela nous amène à proposer l'identification de ce que nous appellerons une « échelle pertinente ». Dans les recherches sur la « Spatial Data Usability » (Josselin, 2003), la pertinence (qui pourrait se décliner entre autres comme la « relevance », la « fitness to use » en termes anglais) peut être assimilée à une valeur associée à la relation entre une méthode, un ensemble de données ou un objet et un objectif fixé généralement par un utilisateur. Dans notre acception, dire qu'une échelle est « pertinente » signifie qu'elle permet de maximiser la capacité d'un indice à atteindre son objectif de mesure, et sous-entend qu'elle n'est valide que pour l'objet étudié qui possède ses caractéristiques propres. Il ne s'agit donc pas d'une échelle pertinente dans l'absolu, mais seulement relative à l'objet étudié, et, qui plus est, dans les conditions de son analyse. Par ailleurs, cette approche, même si elle reste dans cet article, à un niveau d'analyse statistique, peut s'appliquer à terme à l'évaluation de la biodiversité, en lien avec les pratiques de gestion forestière et de conservation écologique. Ces aspects font d'ailleurs partie des travaux entrepris, par la prise en compte de relevés forestiers.

Plus précisément, pour ce qui est de l'estimation de la diversité, l'échelle pertinente correspond au niveau d'agrégation pour lequel l'écart entre les cas 1 et 2 est le plus élevé, c'est à dire qu'il identifie la part maximale de la diversité liée à la structure spatiale de l'image considérée, image caractérisée par sa taille, sa résolution, et l'organisation spatiale de ses observations (valeurs de l'indice de végétation). Pour le cas de l'indice de Shannon, cette échelle correspond à une résolution située entre 40 et 80 mètres de côté de maille (figure 3). De plus, l'écart relatif, rapporté à l'image aléatoire (2-1/2) confirme le résultat précédent en présentant un pic saillant vers ces deux échelles (figure 4).

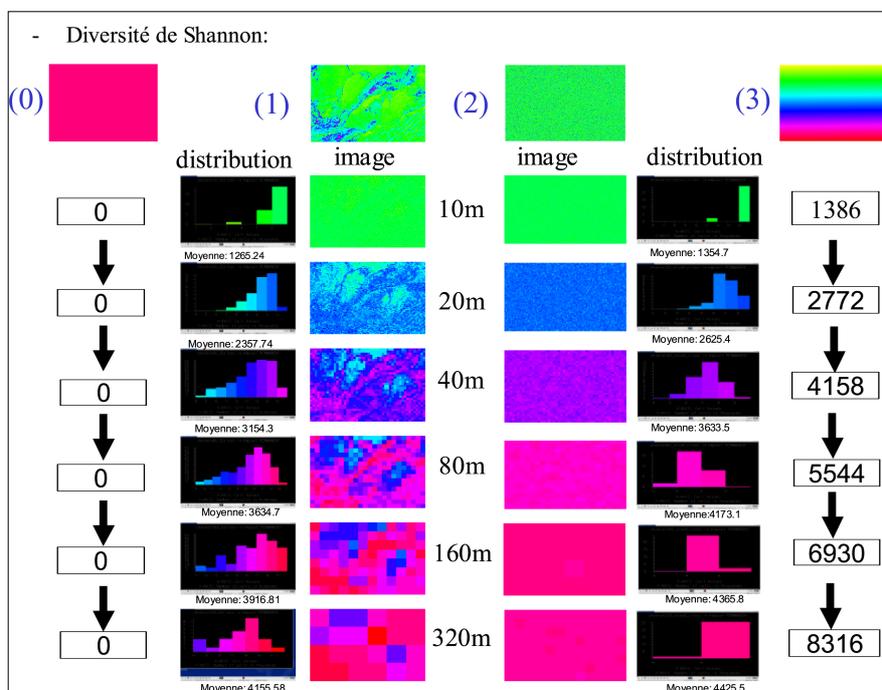


Figure 2. Diversité de Shannon à travers les échelles imbriquées (x 1000).

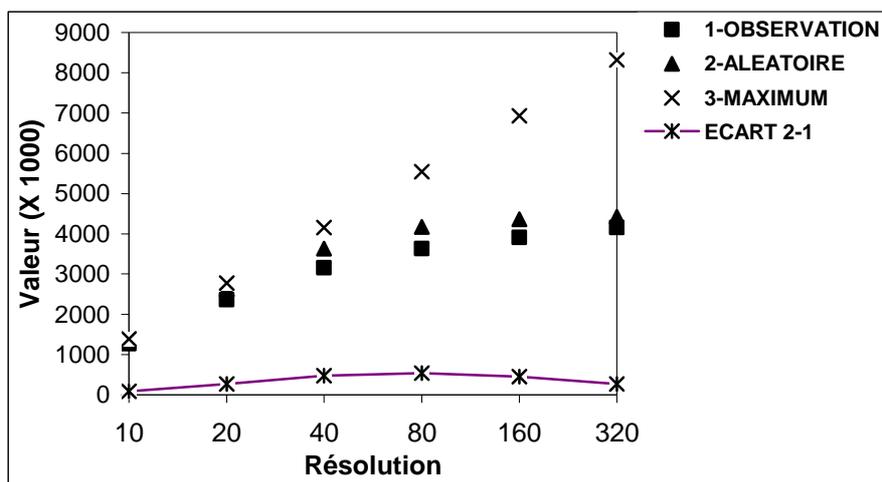


Figure 3. Indice de diversité de Shannon selon les niveaux d'agrégation et les images de référence.

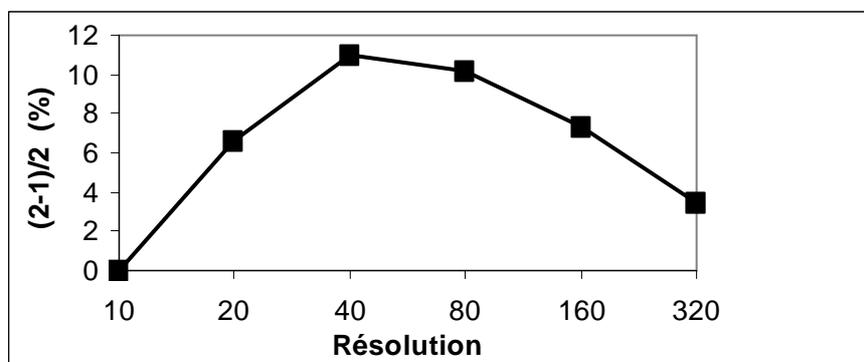


Figure 4. *Ecart relatif entre les indices de diversité moyenne de Shannon des valeurs observées et distribuées aléatoirement.*

3.2. Indice de Simpson

Le comportement de l'indice de Simpson n'est pas tellement différent de celui de Shannon (figures 5, 6 et 7). Sa sensibilité à l'effet d'agrégation est marquée, puisque la part de la diversité liée à l'espace reste inférieure à 35 % pour toutes les résolutions.

Le résultat le plus intéressant vient de la comparaison des cas n° 1 et 2, où l'indice de Simpson résiste mal à la procédure d'agrégation. Tout comme pour l'indice de Shannon, la moyenne croît à mesure que la résolution augmente. La diversité de Simpson « aléatoire » (cas 2) est plus importante que la diversité observée (cas 1). Parfois même, elle dépasse la diversité dite « maximale » (figure 6). Cela vient de la construction de la formule et du fait que le cas n°3 peut être considéré par cet indice comme relevant d'une certaine homogénéité, alors que tous les pixels sont différents. Pour estimer la diversité portée par la structure spatiale dans ce cas, nous proposons d'utiliser la même approche que celle prônée pour l'indice de Shannon, en utilisant les écarts absolus et relatifs de diversité. Ceux-ci montrent que la diversité structurelle arrive à son maximum à nouveau vers les résolutions 40 et 80 mètres (Figure 6). Les résultats sont donc peu différents de ceux de l'indice de Shannon et une échelle « pertinente » peut être à nouveau identifiée entre 40 et 80 mètres.

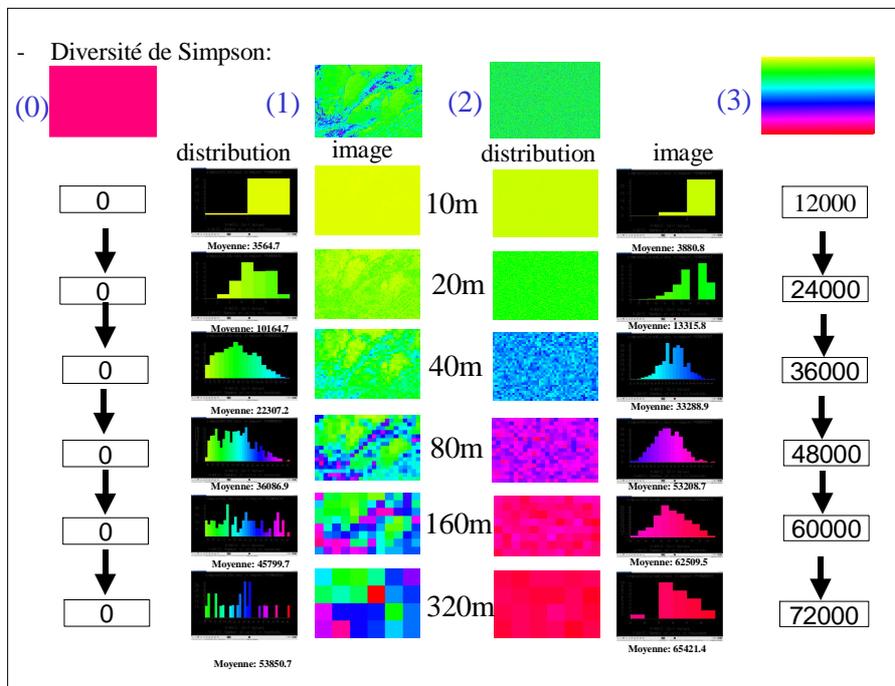


Figure 5. Diversité de Simpson à travers les échelles imbriquées (x 1000).

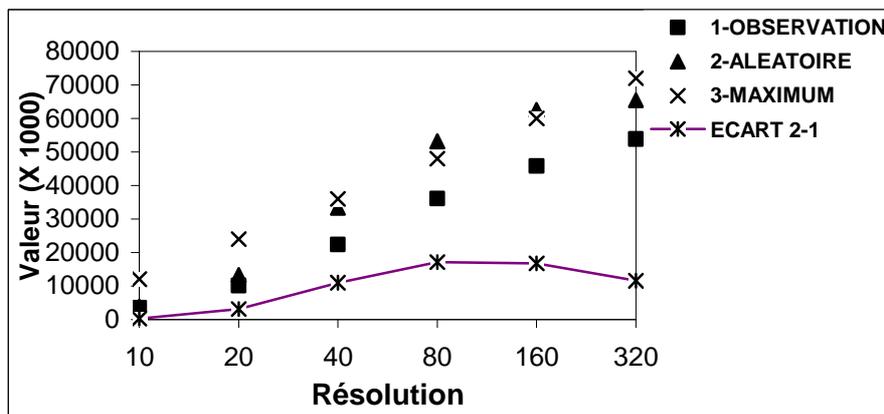


Figure 6. Indice de diversité de Simpson selon les niveaux d'agrégation et les images de référence.

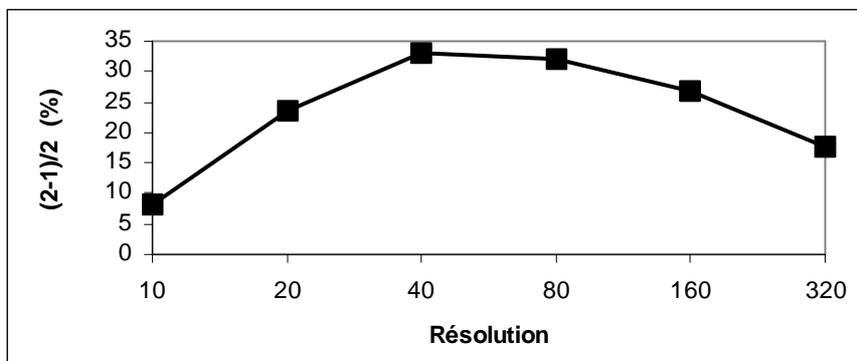


Figure 7. *Ecart relatif entre les indices de diversité moyenne de Simpson des valeurs observées et distribuées aléatoirement.*

3.3. Indice de Dominance et relativité de l' « échelle pertinente »

Il convient d'interpréter les différents cas de figures de l'indice de dominance avec un autre angle de vue, une dominance croissante correspondant à une forme particulière de baisse de diversité. L'image aléatoire n°2 constitue par conséquent un cas où la dominance va être relativement faible, puisque cette fois-ci, l'aléa répartit de façon équitable les valeurs observées dans l'image, réduisant ainsi le niveau d'autocorrélation. Les images 0 et 3 constituent quant à elles deux cas extrêmes de dominance, nulle dans le premier cas et uniquement lié à la variété des valeurs possibles dans le second cas. Les écarts observés entre les images 1 et 2 changent de signe par rapport aux deux indices de diversité précédents.

Globalement, l'indice de dominance se comporte différemment de l'indice de Shannon, tout en restant lui aussi extrêmement sensible à l'effet d'agrégation. Cette fois-ci, la dominance portée par la structure spatiale est, par construction, très importante dans le calcul. En effet, elle varie de 1 fois (aux grandes résolutions) à 2,5 fois (aux petites résolutions) la dominance aléatoire.

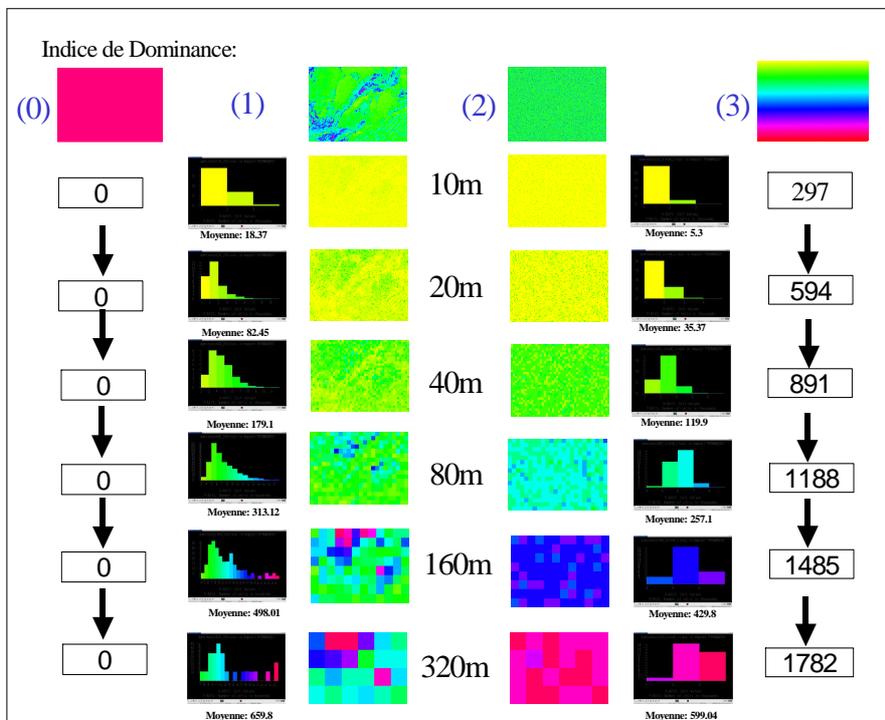


Figure 8. La dominance à travers les échelles imbriquées (x 1000).

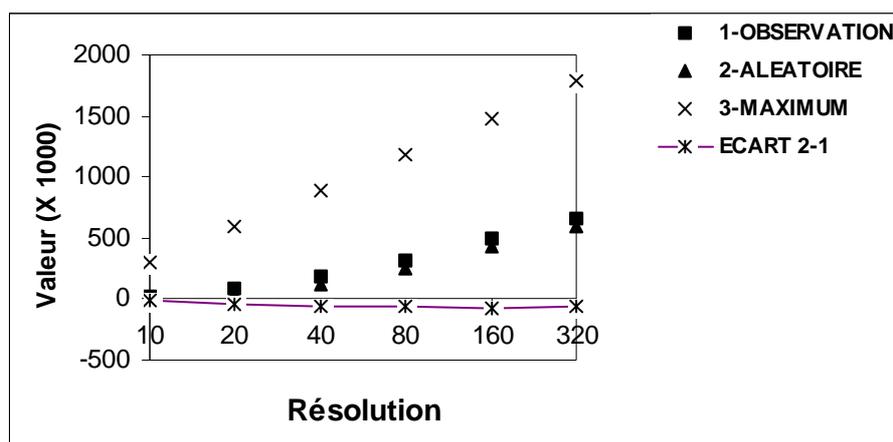


Figure 9. L'indice de dominance selon les niveaux d'agrégation et les images de référence.

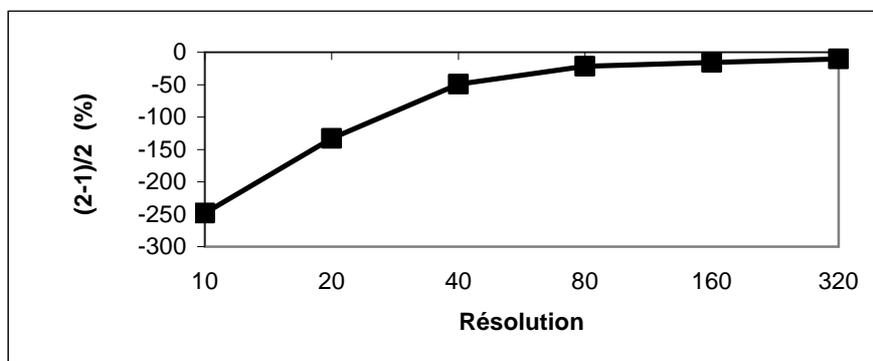


Figure 10. *Ecart relatif entre les indices de diversité moyenne de Dominance des valeurs observées et distribuées aléatoirement.*

Dans le détail, la moyenne de dominance à 10 m de résolution est respectivement de 18.4, 5.3 et 297 dans les cas n° 1, 2 et 3. A 80 et 320 mètres, la moyenne des dominances devient plus élevée (313.1, 257.1 et 1188, puis 659.8, 599 et 1782). On constate que la dominance est cette fois-ci plus faible pour le cas 2 (répartition aléatoire des pixels observés).

Si les écarts de dominance entre les cas 1 et 2 mettent en évidence la stabilité de la dominance à partir des résolutions autour de 80 mètres (Figure 9), les écarts relatifs $(2-1)/2$ relativisent ce résultat. En effet, on constate qu'ils augmentent au fur et mesure de l'agrégation. La détermination de l'échelle pertinente pour l'indice de dominance paraît de fait plus délicate, voire impossible, puisque la courbe reste monotone. Dans ce cas, tout l'enjeu consiste à trouver une échelle pour laquelle on puisse disposer à la fois d'un rapport et d'un écart de diversité favorables, tout en conservant le plus d'information possible, c'est à dire un nombre élevé de mailles ou d'agrégats. Il faut donc considérer le point d'inflexion de la courbe des écarts relatifs de dominance. Dans notre cas, nous conserverons donc l'échelle de 80 mètres, puisqu'au-delà de cette valeur, les rapports de diversité augmentent peu, mais la perte d'information due à l'agrégation est très importante. On constate toutefois que dans ce cas particulier, comme probablement beaucoup d'autres, la recherche de l'échelle pertinente telle que nous la concevons s'avère délicate à mener. Soit elle devient caduque, soit elle nécessite une redéfinition, telle que celle que nous venons de proposer, faisant entrer en jeu d'autres considérations que les valeurs des écarts absolus et relatifs des diversités.

4. Conclusion et perspectives

Par cette recherche, nous avons pu montrer dans quelle mesure les trois indices de diversité étudiés (Shannon, Simpson, et dominance) sont sensibles au MAUP. Puis, à l'aide d'une méthode de ré-échantillonnage, nous avons tenté d'éliminer l'impact du niveau d'agrégation sur l'estimation de la diversité. Enfin, par l'étude des écarts et des rapports entre les valeurs de diversité, nous avons pu élaborer une méthode permettant d'identifier une « échelle pertinente » à laquelle l'estimation de la diversité pourrait être envisagée, dans des conditions de détermination maximale de la diversité portée par la structure spatiale. Cette échelle, dans notre étude, se situerait entre 40 et 80 mètres de résolution pour les indices de Shannon et Simpson (côté de l'agrégat à considérer) pour le cas particulier d'un indice de végétation calculé à partir d'une image SPOT 5 à 5 mètres de résolution, sur le secteur géographique considéré du Mont Ventoux et compte tenu des caractéristiques de l'image source utilisée (taille, forme et résolution).

La difficulté à identifier l'échelle pertinente pour le cas de l'indice de dominance nous permet de proposer une alternative à l'utilisation de notre méthode multiscalaire. En effet, si l'on cherche à mesurer la diversité globale absolue portée par la structure spatiale, on peut considérer qu'elle est quantifiée par la surface déterminée par la partie inférieure de la courbe des écarts absolus (figures 3, 6, 9), qui correspondrait à l'intégrale de la diversité, s'il s'agissait d'une fonction. Cette valeur de surface a l'intérêt de sommer, par paliers, les diversités à toutes les échelles. Plus elle est importante, et plus la diversité multiscalaire l'est également. On pourrait également appliquer ce raisonnement à l'estimation de la diversité relative, en utilisant cette fois-ci l'écart des diversités observées et aléatoires rapporté à la diversité aléatoire (figures 4, 7, 10).

Dans les perspectives de ce travail, nous proposons d'évaluer l'influence de la résolution initiale des données sources, la taille des images, la répartition spatiale des valeurs des pixels et le choix de discrétisation des distributions sur l'échelle pertinente déterminée par notre méthode. Il nous semble aussi intéressant de chercher à construire un indice de diversité le plus adéquat et le moins sensible possible à l'effet d'agrégation. Enfin, nous appliquerons ces différentes approches méthodologiques à l'évaluation de la biodiversité en lien avec des traitements d'images (classification supervisée et photo-interprétation) en ajoutant d'autres indices de diversité, comme celui de la richesse, par exemple, qui définit la biodiversité selon le nombre d'espèces présentes dans chaque unité spatiale.

5. Remerciement

Ilène Mahfoud est financée, dans le cadre de sa thèse pluridisciplinaire en géographie et en écologie, par le gouvernement syrien.

6. Bibliographie

- Amrhein C., « Searching for the elusive aggregation effect: evidence from statistical simulations », *Environment and Planning A*, 27, 1995, p. 105-119.
- Baker W., « The r.le Programs, A set of GRASS programs for the quantitative analysis of landscape structure ». Version 2.2, 1997, University of Wyoming, USA. http://grass.itc.it/gdp/terrain/r_le_22.html
- Charleux L., « GWR, MAUP et lissage par potentiels », *Acte de colloque CASSINI'04, Géomatique et analyse spatiale*, 2-4 juin 2004, Grenoble, pp. 119-124.
- Clark W.A., Karen L., « The effects of Data Aggregation in Statistical Analysis ». *Geographical Analysis*, vol. VIII, 1976, p.429-438.
- Dusek T., « The Modifiable Areal Unit Problem in Regional Economics ». *The 45th Congress of the European Regional Science Association*, 2005, Amsterdam.
- Farina A., *Landscape Ecology in Action*, Kluwer Academic Publishers, 2000, London.
- Frontier S., *L'échantillonnage de la diversité spécifique. In Stratégie d'échantillonnage en écologie*, Frontier et Masson édit, 1983, Paris (Coll. D'Ecologie), XVIII + 494 p.
- Gehlke C.E., Biehl, K., « Certain effects of grouping upon the size of the correlation coefficient in census tract material ». *Journal of the American Statistical Association*, 1934, p. 169-170.
- Jelinski D.E., Wu J., «The modifiable areal unit problem and implications for landscape ecology ». *Landscape Ecology*, vol. 11 no. 3, 1996, p. 129-140.
- Josselin D., « Spatial Data Exploratory Analysis and Usability ». *Codata, Data Science journal*, vol. 2, 2003.
- Marceau D.J., Howarth P.J., Gratton D.J., Remote sensing and the measurement of geographical entities in a forested environment; part 1 The scale and spatial aggregation problem, *Remote Sensing of environment*, Vol. 49, N° 2, 1994, p. 93-104.
- Openshaw S., *The modifiable areal unit problem*. 1984, Number 38, Geo Books, Norwich.
- Rastetter E.B., King A.W., Cosby B.J., Hornberger G.M., O'Neill R.V., Hobbie J.E., « Aggregating fine-scale ecological knowledge to model coarser-scale attributes of ecosystems ». *Ecological Applications* 2, 1992, p. 55-70.
- Reynolds, H. D., « The modifiable areal unit problem: empirical analysis by statistical simulation ». 1998, Thèse de doctorat, University of Toronto.
- Robinson A.H., « Ecological correlation and the behaviour of individuals ». *American Sociological Review*, N° 15, 1950, p. 1-357.
- Wu J., Gao W., Tueller P.T., « Effects of changing spatial scale on the results of statistical analysis with landscape data: A case study », *Geographic Information Sciences*, no 3, 1997, p. 30-41.
- Wu J., Levin S.A., « A spatial patch dynamic modelling approach to pattern and process in annual grassland ». *Ecological Monographs*, no 64, 1994, p. 447-467.

18 Géomatique. Volume X – n° x/2007

Yule, G.U. and Kendall, M.G., *An introduction to the theory of statistics*, 1950, Griffin, London.