



Méthodes probabilistes pour l'étude de la biodiversité

Article rédigé par Charline Carlier, Ingénieur de Recherche SCM

Juillet 2007

Résumé

La rédaction de cet article fait suite à une étude que nous a confiée l'Agence Européenne pour l'Environnement sous le thème « *Developing the principles of probabilistic assessment of terrestrial habitats in view of preparing ecosystem and habitats accounting* ».

La question posée est : « Comment localiser un habitat et suivre son évolution dans le temps et dans l'espace ? ». Un habitat est une zone définie par des attributs biotiques (végétation, couverture du sol...) et abiotique (type de sol, élévation,...).

Jusqu'à présent, une seule méthode existe pour répondre à cette question : c'est la méthode déterministe.

Dans cette approche, on considère que l'habitat est présent dans un pixel (ou une zone donnée) si et seulement si tous les attributs qui le composent sont réunis dans ce pixel. Cette approche est simple et facile à comprendre, mais ne donne pas les meilleurs résultats et surtout ne représente pas ce qui se passe réellement dans la nature.

La méthode probabiliste, que nous avons mise en place dans le cadre du contrat avec l'Agence Européenne de l'Environnement, peut sembler plus complexe mais elle consiste en fait en une mise en forme des données d'observation des habitats. Dans cette approche, on utilise toutes les données fournies par les observations de terrain pour déterminer la probabilité de trouver un habitat dans une zone définie par un attribut, deux attributs... En utilisant des données « réelles », elle fournit les résultats les plus proches possible de la réalité.

Introduction

L'étude de la biodiversité n'est pas simple : les espèces (faune et flore) se déplacent, disparaissent et apparaissent et ceci en grande partie à cause de la détérioration et de la disparition de leur habitat.

Des politiques environnementales, au niveau européen, ont été mises en place afin de responsabiliser les pays face à ces problèmes. La question posée est la suivante : « Quelle est la méthode la plus robuste pour localiser les habitats en Europe et suivre leur évolution dans le temps et l'espace avec les données actuellement disponibles ? ».

Pour répondre à cette question deux méthodes sont envisagées : la méthode déterministe et la méthode probabiliste. La première, simple mais critiquable, est celle utilisée actuellement. La seconde, plus robuste, est celle que la SCM a mise en place dans le cadre d'un contrat avec l'Agence Européenne pour l'Environnement.

Dans cet article, nous présentons la méthodologie pour l'étude des habitats mais les mêmes techniques s'appliquent aussi bien à l'étude de la faune et de la flore.

I La méthode déterministe

1. Données en entrée

Les données en entrée dépendent, bien entendu, de ce que l'on recherche. Dans notre cas il s'agit d'un habitat. La définition d'un habitat est la suivante :

« Zones terrestres ou aquatiques qui se distinguent par des caractéristiques géographiques, biotiques et abiotiques, qu'elles soient naturelles ou semi naturelles. »

Il faut donc connaître les caractéristiques biotiques (ce qui est vivant) et abiotiques (ce qui n'est pas vivant) des habitats.

De nombreuses classifications existent mais celle de l'Annexe I de la Directive Habitat est la plus employée. Elle est composée de trois niveaux : 9 classes au premier niveau, 31 classes au second niveau et 198 classes au dernier niveau ; et donne une description plus ou moins détaillée des habitats en terme de type de sol, végétation caractéristique, distribution biogéographique, altitude, hydrographie...

Nous citons quelques unes des bases de données nécessaires à l'application de la méthode déterministe :

- Type de sol (attribut abiotique) : Base de données « FAO-UNESCO Soil map » et « European soil map » ;
- Élévation (attribut abiotique) : Base de données « Digital Elevation Model » ;
- Couverture du sol (attribut biotique) : Base de données « Corine and Cover » ;
- Répartition d'espèces végétales (attribut biotique) : Base de données « Atlas Flora Europeaea » ;

- Localisation potentielle des habitats : Base de donnée « Potential Natural Vegetation » ;
- Etc...

2. Description de la méthode

La méthode déterministe est souvent choisie pour sa simplicité d'utilisation et de compréhension, mais, comme nous allons le voir, les résultats ne reflètent pas toujours la réalité.

On dispose d'une carte de l'Europe découpée en carré de 1 km de côté. L'approche déterministe est simple : on souhaite connaître la localisation d'un habitat donné en Europe.

Les étapes de la méthode sont les suivantes :

❖ Etape 1 : Choix de l'habitat

❖ Etape 2 : Sélection des attributs

On sélectionne les attributs de l'habitat choisi. Il faut donc savoir de quels attributs il est composé.

Par exemple :

- Type de sol : Calcaire ;
- Couverture de sol : Forêt ;
- Etc...

❖ Etape 3 : Recoupement des bases de données

On recoupe les bases de données de chaque attribut. Le résultat final de la méthode déterministe est le lieu où toutes les bases de données se recoupent.

Prenons un exemple simple : un habitat défini par seulement deux attributs (les attributs vert et bleu). La figure ci-dessous présente leur localisation. Les attributs bleu et vert occupent chacun 20 pixels.

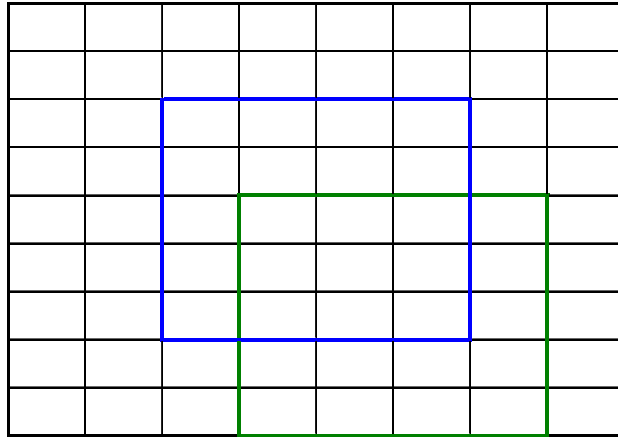


Figure 1 : Localisation des attributs bleu et vert

Le recoupement de ces deux attributs définit la localisation de l'habitat (en rouge sur la figure ci-dessous).

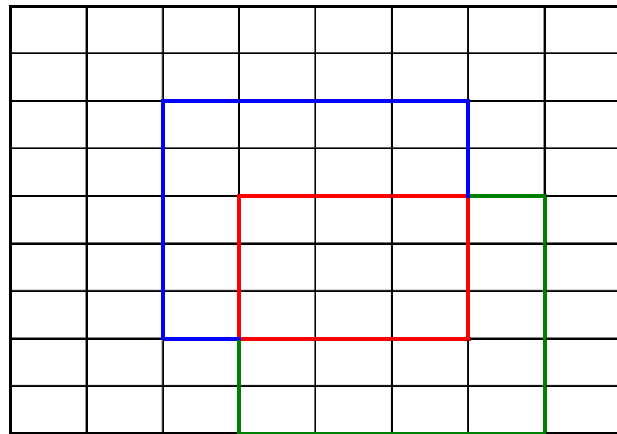


Figure 2 : Localisation de l'habitat

On considère que l'habitat est présent uniquement dans les 9 pixels de la zone rouge. La méthode déterministe considère donc que les autres pixels ne contiennent pas l'habitat même si ceux-ci sont très proches de la zone rouge.

3. avantages et inconvénients de la méthode

Les avantages de cette méthode sont :

- Simplicité d'utilisation et de compréhension ;
- Résultat clair.

Les inconvénients de cette méthode sont :

- Résultat binaire : l'habitat est absent ou non ;
- Ne reflète pas toute la réalité.

La nature évolue, change, s'adapte. Il n'est pas correct de considérer qu'un habitat est présent uniquement dans les sites où tous ses attributs sont présents.

L'utilisation d'une méthode déterministe dans ce cas-là n'est donc pas pertinente ; il est nécessaire de mettre en place une méthode plus robuste qui utilise les caractéristiques des habitats définies par la Directive Habitat mais également des données d'observation.

II La méthode probabiliste

1. Données en entrée

La méthode probabiliste, en plus d'utiliser les bases de données des attributs, utilise des données provenant des observations des habitats sur le terrain (base de données Natura2000 et Corine Biotope).

L'utilisation des données de terrain permet de se rapprocher le plus possible de la réalité. On ne fabrique pas une loi, mais on tire une loi des données existantes : tel habitat a telle probabilité d'exister sur un pixel contenant deux attributs, etc.

2. Description de la méthode

Comme pour la méthode déterministe, on dispose au départ d'une carte de l'Europe découpée en carré de 1 km de côté.

On souhaite connaître la probabilité de trouver un habitat sachant qu'un pixel est défini par un ou plusieurs attributs (par exemple sol calcaire ou élévation de plus de 200 m).

La méthode se déroule en plusieurs étapes très simples :

❖ Etape 1 : Choix de l'habitat

❖ Etape 2 : Sélection d'un attribut

On sélectionne l'attribut pour lequel on veut calculer la probabilité d'occurrence de l'habitat. Comme pour la méthode déterministe, il faut connaître les attributs qui le caractérisent.

On veut, par exemple, calculer la probabilité de trouver l'habitat « *pelouses naturelles* » (niveau 2 de l'Annexe I de la Directive Habitat) sur un sol de type calcaire.

On peut également décider de travailler avec plusieurs attributs en même temps. On sélectionne uniquement les pixels contenant tous les attributs dont on souhaite faire l'étude. On regarde alors la probabilité de trouver un habitat sachant qu'un pixel est composé de deux attributs ou plus.

❖ *Etape 3 : Sélection des pixels avec un habitat recensé dans les observations*

Parmi tous les pixels sélectionnés dans l'étape 2, nous retenons uniquement ceux qui appartiennent à un site recensé dans l'une de ces bases de données. Le nombre de pixels ainsi retenu est noté N_{At} .

Dans la figure ci-dessous, les pixels représentent un attribut donné. Les pixels en gris représentent des zones où l'on connaît l'habitat grâce aux observations de terrain. Sur les 225 pixels de l'attribut, les habitats sont connus pour 40. Dans ce cas : $N_{At} = 40$.

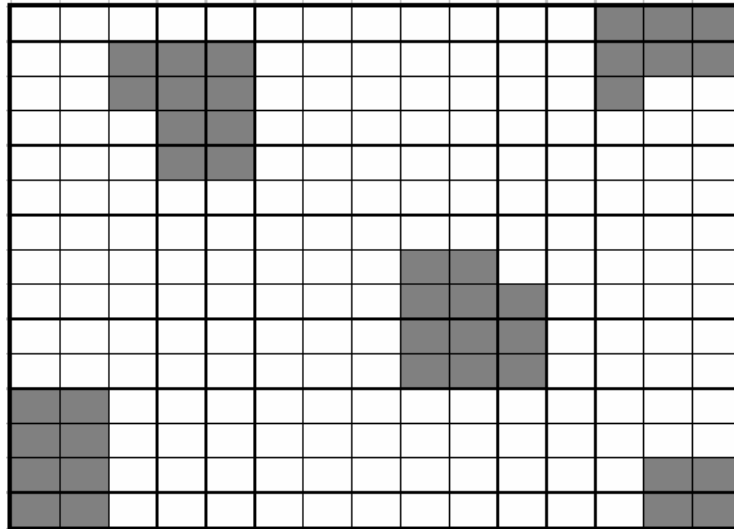


Figure 3 : Localisation de l'habitat grâce aux observations de terrain

❖ *Etape 4 : Sélection des pixels avec l'habitat recherché*

Parmi les N_{At} pixels dénombrés dans l'étape précédente, nous retenons uniquement ceux qui appartiennent à l'habitat recherché. Le nombre de pixels ainsi retenu est noté N_{Hab} .

❖ *Etape 5 : Calcul des probabilités*

A partir des observations d'un échantillon de données, nous souhaitons déduire les résultats pour toute une zone.

La formule généralement utilisée pour calculer cette probabilité est la suivante :

$$p = \frac{\text{Nombre de cas favorables}}{\text{Nombre de cas possibles}} = \frac{\text{Nombre de pixels de l'Habitat X}}{\text{Nombre total de pixels}} \quad (1)$$

Dans la formule ci-dessus, lorsque aucun habitat X n'a été observé dans l'échantillon, alors la probabilité de le trouver toute la zone d'étude est nulle.

Bien évidemment l'échantillon renseigne sur la composition de l'ensemble mais un échantillon n'est qu'une observation partielle. Il n'y a donc aucune raison de considérer la probabilité comme nulle. Prenons un exemple simple : on souhaite étudier la taille de

la population française. On décide de prendre un échantillon de 1000 personnes car bien évidemment on ne peut pas mesurer toute la population. Dans cet échantillon aucune personne ne mesure 1,65 m ; mais on ne peut pas en déduire qu'aucun Français ne mesure 1,65 m.

Nous recommandons donc d'utiliser une méthode basée sur l'évaluation du taux de risque ; méthode décrite en détail dans le livre de Bernard Beauzamy : « *Méthodes probabilistes pour l'étude des phénomènes réels* ».

Cette méthode, adaptée à notre cas dit la chose suivante :

Sachant qu'il existe N_{Hab} pixels d'un habitat X parmi les N_{At} pixels d'un échantillon de données, la probabilité d'avoir N'_{Hab} pixels de l'habitat X en Europe parmi N'_{At} pixels en Europe est donnée par la formule :

$$p(N'_{Hab}, N'_{At}; N_{Hab}, N_{At}) = \frac{(N_{At} + 1)}{(N_{At} + N'_{At} + 1)} \frac{\binom{N'_{At}}{N'_{Hab}} \binom{N_{At}}{N_{Hab}}}{\binom{N_{At} + N'_{At}}{N_{Hab} + N'_{Hab}}}$$

Rappelons que le résultat recherché est la probabilité d'avoir 1 habitat X dans 1 pixel ; soit N'_{Hab} et N'_{At} égaux à 1. Après simplification on obtient la formule suivante :

$$P_{Hab} = \frac{N_{Hab} + 1}{N_{At} + 2} \quad (2)$$

La formule qui, au début, pouvait sembler complexe, devient au final très simple.

Dans le cas de grands nombres cette formule est très proche de la formule (1) présentée précédemment. Par contre, lorsque aucune observation de l'habitat X n'a été faite sur l'échantillon de données, la probabilité de trouver cet habitat n'est pas nulle mais égale à :

$$P_{Hab} = \frac{1}{N_{At} + 2}$$

Au final, on représente sur une carte la probabilité de trouver un habitat donné en fonction de la composition de chacun des pixels. La figure ci-dessous présente un exemple de résultat : l'attribut sélectionné est la région Biogéographique et l'habitat recherché l'habitat « Forêt de résineux ».

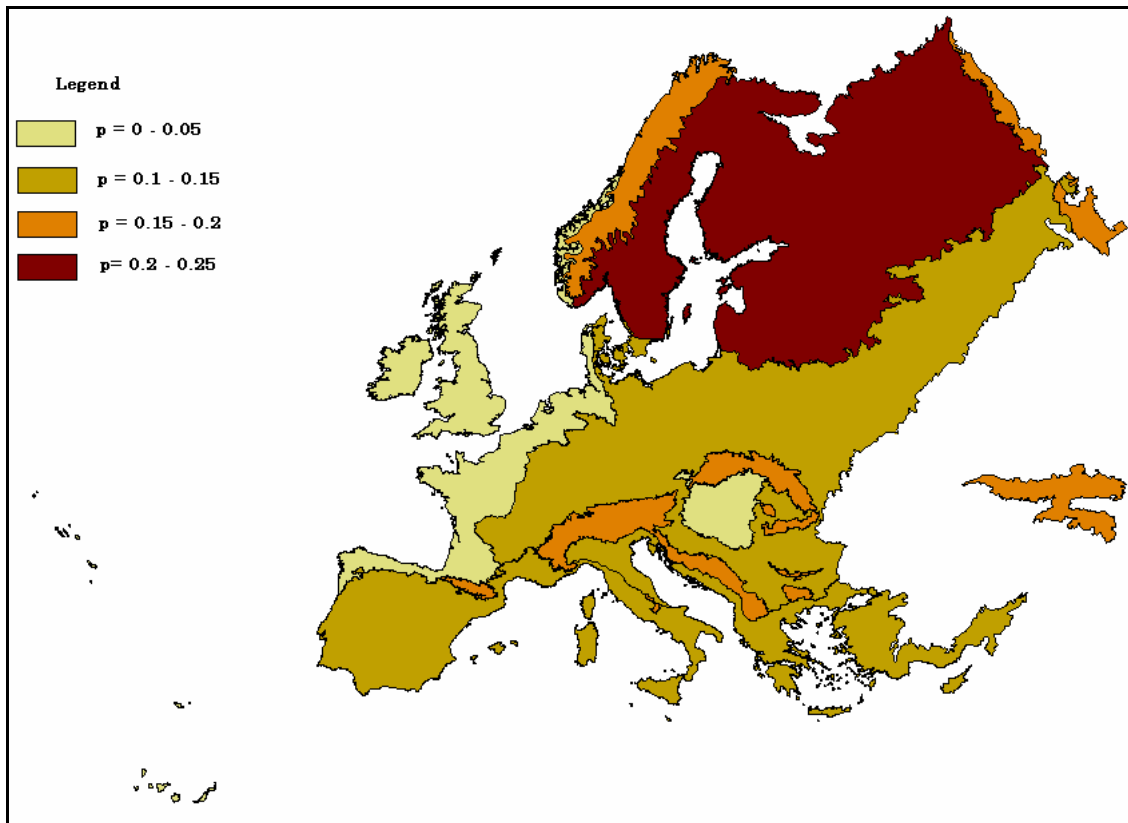


Figure 4 : Probabilité de trouver l'habitat « Forêt de résineux »

On remarque sur cette carte que les probabilités de trouver l'habitat sont assez faibles. Le nombre d'habitats qui peuvent être trouvés sur une zone caractérisée par un seul attribut est très grand, la probabilité de trouver un habitat spécifique sera alors très petite. Au contraire, une zone est caractérisée par plusieurs attributs (type de sol, végétation, élévation...), le nombre d'habitats pouvant être trouvés sera restreint et donc la probabilité de trouver l'habitat caractérisé par ces attributs sera élevé.

3. *Avantages et inconvénients de la méthode*

Les avantages de la méthode sont les suivants :

- Méthode plus robuste ;
- Résultat non binaire ;
- Méthode plus complexe mais formule de calcul très simple.

Le principal inconvénient de la méthode est un temps de calcul plus long que la méthode déterministe. Le nombre de résultats pour chaque habitat est plus grand ; il est donc logique que la méthode soit un peu plus longue à mettre en œuvre.