



HAL
open science

Construction d'un modèle d'habitat du Pinson des arbres (*Fringilla coelebs* L.) en zone forestière et applications

Jacques Lauga, Jean Joachim, Gerard Balent

► To cite this version:

Jacques Lauga, Jean Joachim, Gerard Balent. Construction d'un modèle d'habitat du Pinson des arbres (*Fringilla coelebs* L.) en zone forestière et applications. La forêt paysanne dans l'espace rural : Biodiversité, paysages, produits, 29, INRA, 268 p., 1996, Etudes et Recherches sur les Systèmes Agraires et le Développement, 2-7380-0684-1. hal-02840984

HAL Id: hal-02840984

<https://hal.inrae.fr/hal-02840984v1>

Submitted on 7 Jun 2020

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.

Construction d'un mod le d'habitat du Pinson des arbres (*Fringilla coelebs* L.) en zone foresti re et applications

Jacques LAUGA*
Jean JOACHIM
G rard BALENT

*UMR 9964, Ecologie Terrestre, Universit  Toulouse III, 31062 Toulouse Cedex.

R sum 

Le but de cette  tude  tait de construire un mod le d'habitat du Pinson (*Fringilla coelebs*) en zone foresti re dans le sud-ouest de la France. La faible couverture foresti re de la r gion et son morcellement laissaient supposer une structuration possible de la population r gionale de Pinsons en *m tapolopulations* au sens de Levins. La construction d'un mod le d'habitat devait aider   identifier les noyaux isol s de populations.

La couverture foresti re de la zone  tudi e (2327 km²) a  t  extraite d'une sc ne Landsat MSS et ramen e   une grille finale de 5865 quadrats de 650 m tres de c t  chacun. Pour chaque quadrat une mesure locale combinant la couverture foresti re et sa fragmentation a  t  calcul e : l'Indice Forestier, I.F. Un plan d' chantillonnage sur 556 quadrats a  t  ensuite d fini. Il a permis de recenser le statut (pr sence/absence) de l'esp ce durant la saison de reproduction.

Une carte d'incidence a  t  ensuite  tablie sur l'ensemble de la zone, reliant l'IF de chaque quadrat   la probabilit  de rencontre de l'esp ce.

La recherche du seuil optimal de d cision sur la s rie continue de probabilit  a  t  r alis e par l'analyse de la courbe R.O.C. Une fois ce seuil optimal d termin , la zone a  t  partag e en deux ensembles : l'un de pr sence probable de l'esp ce, l'autre d'absence probable. Cinq foyers de pr sence, convexes et isol s, ont  t  d termin s. Ces r sultats ont permis de guider la suite de l' chantillonnage en vue d' prouver l'hypoth se d'une structuration en m tapolopulations.

Mots cl s : mod le d'habitat, oiseaux, fragmentation foresti re, incidence.

Abstract

Building a model of Chaffinch habitat in forested areas ; some applications. *The aim of this study was to develop a model of Chaffinch (*Fringilla coelebs*) habitat in forested areas of south western France. Due to poor forest cover and forest fragmentation it was assumed that Chaffinch populations were structured in "metapopulations" according to Levins's concept. It was hypothesized that building a species-habitat model would help identify the different patches where local/isolated populations could be found.*

The forest cover of the study area (2327 sq. km) was extracted from a Landsat MSS scene and then superimposed on a grid of 5865 square quadrats, each 650m wide. For each quadrat a local measure combining forest cover and forest fragmentation was computed (I.F., standing for Indice Forestier). A sampling scheme of 556 quadrats was next defined on the area, and the status of the Chaffinch (Present/Not present) was assessed on each during the reproductive season.

An incidence map was then computed on the area, linking the IF of each quadrat to the probability of finding the species.

The optimal decision-threshold on the continuous range of probability was found using R.O.C. methodology. Once it had been determined, the area was divided into zones where the Chaffinch was likely to be present from zones where it was likely to be absent. Five convex and isolated zones of presence were found. These results were in turn used for conducting further studies aimed at proving the metapopulation hypothesis.

Keywords: species-habitat model, birds, forest fragmentation, incidence.

1. Introduction

Le maintien durable de la biodiversité et des ressources renouvelables suppose de prévoir, qualitativement et quantitativement, les effets à long terme de la modification des habitats sur les espèces vivantes.

Idéalement la résolution de ces problèmes imposerait de connaître pour chaque espèce :

- les besoins en ressources,
- les réponses démographiques aux modifications de ces ressources,
- les effets induits par les successions naturelles dans l'habitat de l'espèce,
- les effets d'interaction entre ces ressources et d'autres facteurs externes (prédation, parasites, climat, etc.).

Ce cadre n'a de valeur qu'au plan théorique. Il est trop vaste pour être jamais complètement élucidé ; son seul mérite est de délimiter ce qui, dans l'idéal, devrait être connu des relations entre les espèces et leur habitat, pour permettre de comprendre, prédire et gérer la dynamique des espèces vivantes.

Opérationnellement le premier problème à résoudre est celui de l'identification et de la modélisation des besoins et des réponses des espèces à certains paramètres-clés de leur habitat.

Dans ce domaine les données sont très nombreuses : toutes les connaissances accumulées depuis des décennies sur la biologie des espèces sont dignes d'intérêt, mais elles ont été trop souvent recueillies de manière non coordonnée, sous forme descriptive ou seulement qualitative. La plupart d'entre elles sont faiblement utilisables dans une perspective quantitative et modélisatrice.

C'est vers le milieu des années 70 que des efforts coordonnés, visant à générer et à intégrer des informations chiffrées sur les espèces et leurs habitats apparaissent aux Etats-Unis. Même si les résultats obtenus n'ont pas toujours été à la hauteur des espérances (Thomas, 1986), le mérite de ces études a été de poser les problèmes, de montrer leur urgence et de déclencher, dans la communauté des biologistes, une

série de tentatives et de recherches nouvelles.

Verner, Morrison et Ralph (1986), Nelson et Salwasser (1982) dressent un bilan de ces tentatives, qu'il est commode d'englober sous le terme général de *modèles d'habitat*.

Les modèles d'habitat visent à résoudre un même problème, mais par des voies souvent différentes. Ainsi rencontre-t-on "l'Habitat Suitability Index" ou SI (Fish and Wildlife Service 1981), "l'Habitat Capability model" ou HC model, (Hurley *et al.*, 1982) ; le "Pattern Recognition Model" ou PATREC (Williams *et al.*, 1978) ; le "Wildlife and Fish Habitat Relationships program" ou WFHR (Nelson et Salwasser, 1982) ; les "Habitat Evaluation Procedures" ou HEP (Schamberger et Farmer, 1978 ; Fish and Wildlife Service, 1980a, 1980b) pour ne citer que les principaux.

Quelle que soit la démarche adoptée, l'espèce étudiée ou le milieu considéré, le modèle d'habitat établit une relation quantitative entre une ou plusieurs variables indépendantes : les *variables d'habitat* et une variable dépendante : *la réponse de l'espèce*. Cette réponse n'est pas unique. Selon le but recherché elle pourra s'exprimer par une biomasse, une densité, un effectif, une probabilité de présence etc. pour ne citer que quelques exemples (Schamberger et O'Neil, 1986).

Construire et valider un modèle d'habitat est en définitive une tâche complexe qui n'échappe pas à certains arbitrages. Mais la méthode, malgré ses défauts, possède au moins quatre avantages :

- elle impose d'explorer aussi complètement que possible la biologie des espèces en vue d'identifier les principales variables d'habitat qui agissent sur la réponse considérée ;
- elle impose de planifier soigneusement l'échantillonnage ;
- elle permet d'examiner plusieurs types de liaisons fonctionnelles entre les variables d'habitat et la réponse ;
- enfin le modèle, une fois établi, peut être validé, *donc* réfuté.

Le but de ce chapitre est de montrer, sur un exemple précis, les étapes de construction d'un modèle d'habitat en milieu forestier.

L'espèce choisie est un passereau forestier commun, le Pinson des arbres (*Fringilla coelebs* L.); la réponse étudiée est la présence/absence; la variable d'habitat considérée est la couverture forestière et son morcellement. On examinera une à une les différentes étapes de construction de ce modèle "simple": les pièges et certains arbitraires apparaîtront plus nettement.

La construction d'un modèle d'habitat sur cette espèce de passereau forestier n'était pas gratuite. Le modèle, une fois établi, devait servir d'outil au sein d'une étude plus large concernant les rapports entre le morcellement forestier et l'apparition de métapopulations (Levins, 1969). On sait que la fragmentation, des habitats d'une espèce réduit la superficie des zones favorables, isole des populations distribuées en taches et augmente les probabilités d'extinction locale (Hanski et Gilpin, 1991; Gilpin et Hanski, 1991).

La mise en évidence de noyaux isolés de populations, au sein d'une zone géographique donnée, pouvait être obtenue en lui appliquant un modèle d'habitat de type présence/absence. Les noyaux, une fois révélés, guidaient le recueil de données en vue d'éprouver l'hypothèse de métapopulation. Ce chapitre ne développera pas cet aspect de l'étude, il se concentrera uniquement sur la méthodologie de construction du modèle.

2. Zone étudiée, recueil des données relatives au morcellement forestier

La zone étudiée se situe au nord-ouest de l'agglomération toulousaine et dessine un rectangle de 53,5 sur 43,5 kilomètres, soit une superficie totale de 2327 km². La couverture forestière de cette région n'est que de 4% et se répartit en trois types de formations :

- Des massifs forestiers (plus de 1000 ha) localisés au nord (Buzet) et au sud (Bouconne) de la zone. Les essences représentées sont principalement le chêne pédonculé, le charme et le frêne. Les conifères y sont peu nombreux et localisés.

- La ripisylve de la Garonne. Formée d'aulne à l'état spontané et de cultures de peupliers, elle coupe obliquement la zone selon une direction nord-ouest sud-est.

- Une mosaïque de bosquets aux superficies variées.

Pour être utilisables dans la construction d'un modèle d'habitat les données devaient être disponibles sous forme numérique. On a donc utilisé une image de la région prise par le satellite Landsat (scène # 214-029 du 30 juillet 1981) sur laquelle on a appliqué les méthodes classiques d'identification de la couverture forestière (Barrett et Curtis, 1976). A l'issue de ce traitement l'image est devenue binaire. Chaque élément (ou pixel) est classé "forêt" ou "non forêt".

Le pixel du satellite Landsat a une superficie au sol trop petite pour pouvoir y situer un plan d'échantillonnage. On a donc regroupé les pixels en quadrats de 650 mètres de côté sur l'ensemble de la zone étudiée. Le résultat obtenu se présente sous forme d'une grille de 85 x 65 = 5525 quadrats de 42 ha chacun. A cette échelle un plan d'échantillonnage peut être correctement situé sur la grille avec une erreur négligeable.

Les données disponibles sur chaque élément de la grille s'expriment en *pourcentage de couverture forestière*. Ces valeurs sont calculées d'après le nombre de pixels classés "forestiers" dans chacun des quadrats (figure 1).

Si la fragmentation forestière apparaît nettement sur cette carte, elle n'est pas mesurée pour autant. Chaque quadrat de la grille ne montre que ses caractéristiques propres. Mesurer la fragmentation exige davantage : il faut caractériser chaque quadrat par une mesure qui prenne en compte la couverture forestière des éléments voisins. Cette opération peut être réalisée par application d'un filtre de lissage à l'ensemble de la zone. Le filtre utilisé ici est de forme exponentielle.

Numérotons par $i=0,1,2,\dots,N$ les N voisins d'un quadrat donné. L'indice $i=0$ représente le quadrat lui-même. Chacun des N voisins est situé à une distance moyenne d_i et présente une couverture forestière f_i .

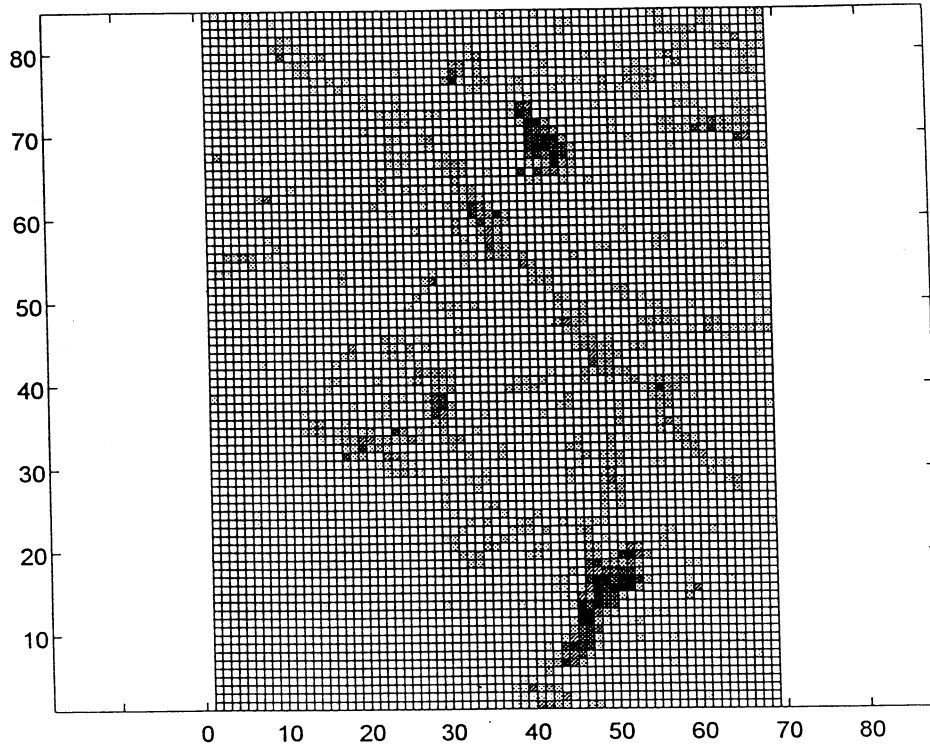


Figure 1 : Pourcentage de couverture forestière dans chacun des quadrats de la zone étudiée. Le codage est en niveaux de gris. Chaque quadrat mesure 650 mètres de côté et la superficie totale est de 2327 km². Deux grands massifs forestiers sont bien visibles : Bouconne au sud, Buzet au nord.

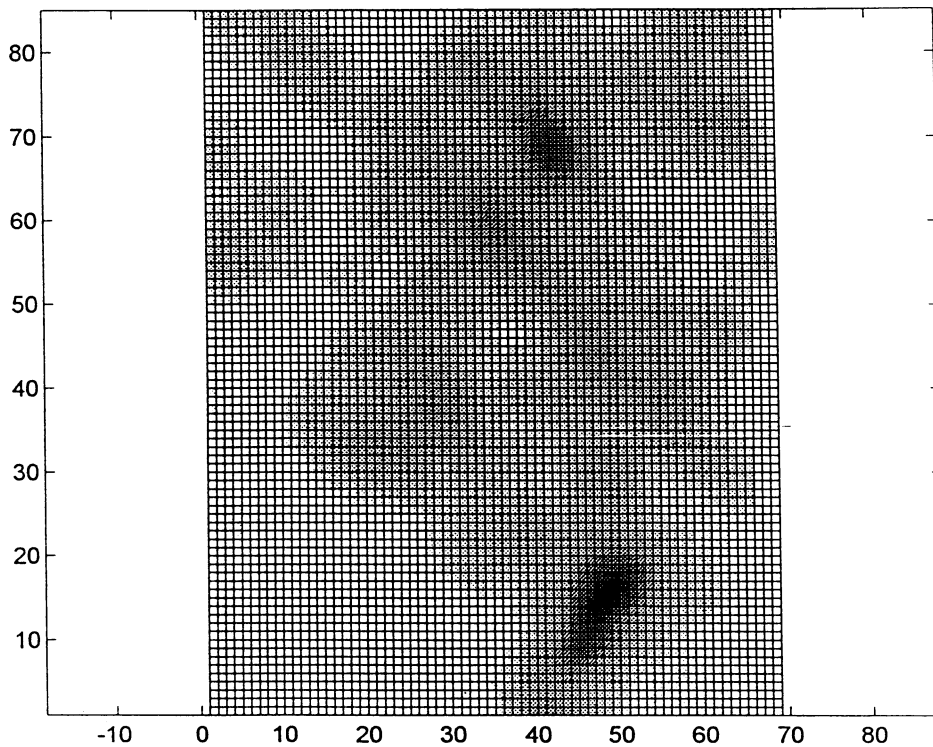


Figure 2 : Valeurs de l'Indice Forestier (IF) dans chacun des quadrats de la zone. Le codage est réalisé en niveaux de gris.

On caractérise alors chaque quadrat par la quantité suivante, nommée Indice Forestier (IF) :

$$IF = \sum_{i=0}^N f_i \cdot e^{-\alpha \times d_i}$$

L'indice forestier intègre à la fois la couverture forestière et son morcellement. Mathématiquement l'opération est une "convolution discrète" ressemblant à un lissage des données (Lauga et Joachim, 1992). Le terme α , pris ici égal à 0,563 pondère la décroissance de la contribution d'un quadrat avec son éloignement. Le résultat obtenu (figure 2) est plus régularisé que celui de la seule mesure locale de couverture forestière. Les valeurs de l'IF ont été normées entre 0 et 100. Malgré leur ressemblance les figures 1 et 2 ne partagent qu'une assez faible proportion d'information commune. La corrélation entre les valeurs du pourcentage de couverture forestière (figure 1) et de l'IF (figure 2) pour les 5865 quadrats de la zone n'est que de $r=0,759$, soit un r^2 de 0,576 : la variance commune aux deux images est inférieure à 60%. C'est la preuve que la prise en compte du voisinage forestier d'un quadrat apporte une information propre.

Les données de la figure 2 constituent finalement la variable indépendante prise en compte dans la construction du modèle d'habitat. On voit *qu'il ne s'agit pas d'une donnée simple, mais au contraire d'une variable construite* qui s'efforce d'intégrer, au mieux, le facteur clé auquel semble réagir l'espèce.

3. L'échantillonnage

Echantillonner "correctement" un ensemble de 5865 quadrats posait une triple question : celle de l'effort à consentir, celle de la répartition spatiale des points-échantillons au sein de la zone prospectée, celle enfin de la stratification à respecter vis-à-vis de la variable d'habitat considérée, l'IF.

L'effort d'échantillonnage, c'est-à-dire son coût, ne peut se déterminer automatiquement en appliquant des formules sim-

ples (Snedecor et Cochran, 1980 ; Krebs, 1989). Il doit être proportionné à la variabilité du phénomène étudié et croît avec elle. Mais des facteurs externes comme le temps disponible le limitent vers le haut. On a considéré dans cette étude que l'effort consisterait à échantillonner 10 % de la population parente. Cette limite, en partie arbitraire, correspondait aux moyens mis à la disposition de l'étude. Au total 556 points d'écoute ont été réalisés sur l'ensemble de la zone, soit 9,48 % des 5865 quadrats.

La répartition spatiale des relevés-échantillons au sein de la zone (leur densité) devait être la plus homogène possible. Mais une stricte homogénéité, bien que concevable, aurait été dépourvue de sens. Placer des relevés au milieu de zones cultivées, privées de toute formation boisée, aurait conduit à gaspiller du temps et des ressources, plus rentables ailleurs. La répartition finalement adoptée a tendu à l'homogénéité sans l'atteindre complètement. Concrètement on n'a négligé aucun secteur boisé et la zone a été visitée sur l'ensemble de son étendue.

La stratification des relevés selon les valeurs de l'indice forestier posait un dernier problème. La figure 3 montre la répartition comparée des valeurs de l'IF dans les 5865 quadrats de la zone (A) et dans les 556 points-échantillons (B). Il apparaît que les relevés :

- ont privilégié les zones à faible IF (valeurs inférieures à 10) car elles sont majoritaires dans l'ensemble de la zone (N=188, 33,8%) ;

- ont largement exploré les quadrats d'IF faible, moyen ou peu élevé (classes 2 à 8 de l'histogramme, figure 3 B) car c'est précisément dans ce type de quadrats que l'incertitude sur la présence de l'espèce est la plus forte. Dans le problème analysé ici, une stratification rigoureuse de l'échantillonnage selon les valeurs de l'indice forestier aurait été dépourvue de sens. Elle aurait consisté à disposer la majorité des relevés (78%) dans les quadrats d'IF inférieur à 10, c'est-à-dire dans des zones où l'on est à peu près certain de ne jamais rencontrer l'espèce et à négliger les situations d'IF intermédiaire où l'incertitude sur la présence du Pinson est la plus grande. On reviendra sur ce point en conclusion.

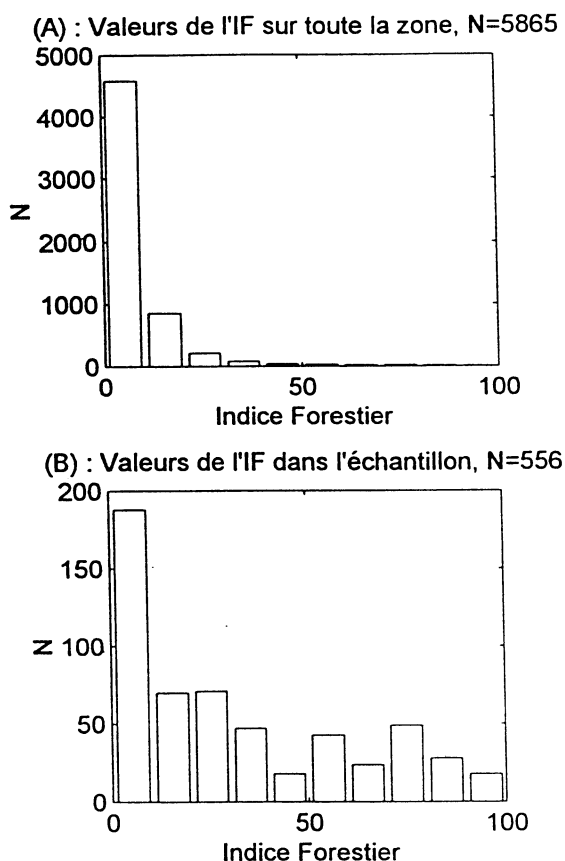


Figure 3 : Répartition des valeurs de l'Indice Forestier (IF) sur :
 (A) l'ensemble de la zone étudiée (N=5865 quadrats) ;
 (B) l'ensemble des points-échantillons (N=556). L'IF est ramené à l'intervalle {0..100}.

A l'issue de la campagne de terrain, menée au printemps lorsque les couples sont cantonnés, on disposait donc de 556 relevés ornithologiques. Ils ont permis de contacter un total de 72 espèces. Le Pinson est reconnu dans 343 relevés (61,7%) et non reconnu dans 213. Bien que l'absence de contact ne prouve pas l'absence de l'espèce, on considérera dans la suite de ce travail que le Pinson est effectivement absent dans les 213 points où il n'a pas été reconnu.

Les données disponibles consistaient finalement en un tableau de 556 lignes et 2 colonnes, indiquant pour chacun des relevés : l'IF (valeurs de 1 à 100) et la présence ou l'absence du Pinson (valeurs 0 ou 1). C'est à partir de ce tableau qu'on se proposait de construire un modèle probabiliste de rencontre de l'espèce, interpolé à l'ensemble de la zone.

4. Le modèle d'incidence du Pinson

La notion d'incidence est empruntée à l'épidémiologie, et elle semble avoir été introduite en écologie par Diamond (Diamond, 1975 ; Gilpin et Diamond, 1981).

L'incidence est une probabilité de présence, elle-même fonction d'une ou plusieurs variables indépendantes à valeur explicative. Dans le cas le plus simple qui est aussi celui du Pinson, l'incidence est supposée être fonction d'une seule variable : l'indice forestier. En notant p l'incidence et x l'indice forestier on a donc :

$$p = f(x) \text{ avec } 0 \leq p \leq 1 \text{ et } 0 \leq x \leq 100(1)$$

f est la fonction d'incidence.

Les formes possibles de f sont en principe très nombreuses, mais en pratique on a restreint f à :

a- être monotone croissante ou décroissante,

$$\left(\frac{df}{dx} \geq 0 \text{ ou } \frac{df}{dx} \leq 0\right),$$

b- appartenir à des familles "connues". La famille des fonctions de type logistique est le plus souvent utilisée (Taylor, 1991) et c'est celle que l'on emploie ici. Mais d'autres formes sont également utilisables (Jongman, Ter Braak et Van Tongeren, 1987). On reviendra sur ce point en conclusion.

Ayant choisi le modèle logistique et supposant f croissante (ce qui traduit un accroissement de la probabilité de rencontre p avec l'augmentation de l'indice forestier x) l'équation (1) s'écrit :

$$p = \frac{e^x}{1 + e^x} \quad (2)$$

Telle quelle cette équation est inutilisable en raison de sa symétrie qui impose $p=0,5$ pour $x=0$. Deux paramètres permettront de l'adapter au problème posé et aux données disponibles. Ils reviendront à faire subir à

(2) un changement d'échelle (paramètre α) et de localisation (paramètre b) :

$$p = \frac{e^{\alpha x + b}}{1 + e^{\alpha x + b}} \quad (3)$$

Dans (3), p mesure une probabilité de présence du Pinson pour la valeur x de l'indice forestier. On montre aisément que son complément à 1, $(1-p)$, ou probabilité d'absence s'écrit :

$$(1-p) = \frac{1}{1 + e^{\alpha x + b}} \quad (4)$$

L'expression (4) est monotone décroissante.

Le problème revient donc à estimer les paramètres a et b dans (3) à partir du jeu de 556 relevés dont on dispose. C'est affaire de calcul numérique, mais avant tout de statistique car les méthodes ordinaires d'estimation par les moindres carrés sont ici en défaut. En effet, dans (3), p varie avec x . Au voisinage de x , p se comporte comme une variable de Bernoulli de variance $p(1-p)$. La variance de la fonction étudiée n'est pas constante sur l'étendue des x . Elle passe par un maximum pour $p = 0,5$ comme le montre l'examen de la dérivée. Il faut donc estimer les paramètres a et b par d'autres méthodes. On utilise en général le maximum de vraisemblance. Les logiciels de statistique répandus comme SAS, BMDP (SAS 1989, Dixon 1990) et bien d'autres réalisent facilement ce type de calcul. En sortie chaque paramètre, outre sa valeur estimée, est assorti de sa variance asymptotique et fait l'objet d'un test de signification contre zéro qu'éprouve le c^2 de Wald.

Le modèle d'Incidence du Pinson sur notre échantillon s'écrit finalement :

$$p(x) = \frac{e^{0.004x - 0.6763}}{1 + e^{0.004x - 0.6763}} \quad (5)$$

Chacun des deux paramètres a et b est hautement significatif (Tableau 1).

Paramètre	Valeur	Ec. Type	Khi-2 de Wald	p
a	0.044	0.00483	83	<0.0001
b	-0.6763	0.1403	23	<0.0001

Tableau 1 : Résultat de l'ajustement d'une fonction logistique aux données de présence/absence du Pinson (N=556 points échantillons). Les deux paramètres a et b sont assortis d'un test de signification contre zéro. La courbe obtenue est représentée à la figure 4.

La fonction d'incidence, forme régularisée de probabilité de présence du Pinson selon les valeurs de l'Indice forestier, peut être calculée pour chacune des N=556 valeurs de l'échantillon (figure 4). L'augmentation de probabilité de rencontre de l'espèce avec les valeurs croissantes de l'IF est clairement mise en évidence sur ce graphique.

5. La carte d'Incidence du Pinson

Puisque (5) associe une probabilité de rencontre à toute valeur de l'Indice forestier x , de la même manière qu'on a calculé ces probabilités sur les 556 valeurs de l'échantillon, on peut les calculer sur les 5865 valeurs de la grille puis cartographier le résultat. Cette carte est représentée à la figure 5. On y a tracé les courbes de même probabilité de rencontre (d'iso incidence) entre les valeurs $p=0,4$ et $p=0,9$ par pas de 0,1. Les noyaux à fort Indice forestier se retrouvent évidemment sur cette carte. Mais on se rend compte que la ligne d'iso incidence $p=0,4$ les relie tous au sein d'un même ensemble. Autrement dit, au seuil de 0,4, on n'a affaire ici qu'à un bloc en complète continuité. Il n'y a pas, à ce seuil là, de morcellement en sous ensembles disjoints. Des noyaux disjoints n'apparaissent qu'au-delà de l'iso incidence 0,4, vers 0,5 et au-dessus.

La carte d'incidence est riche d'enseignements mais elle ne clôt pas le problème. Elle précise surtout des probabilités de présence et représente leurs variations dans l'espace. Ces probabilités évoluent continûment entre environ 0,3 et 0,95.

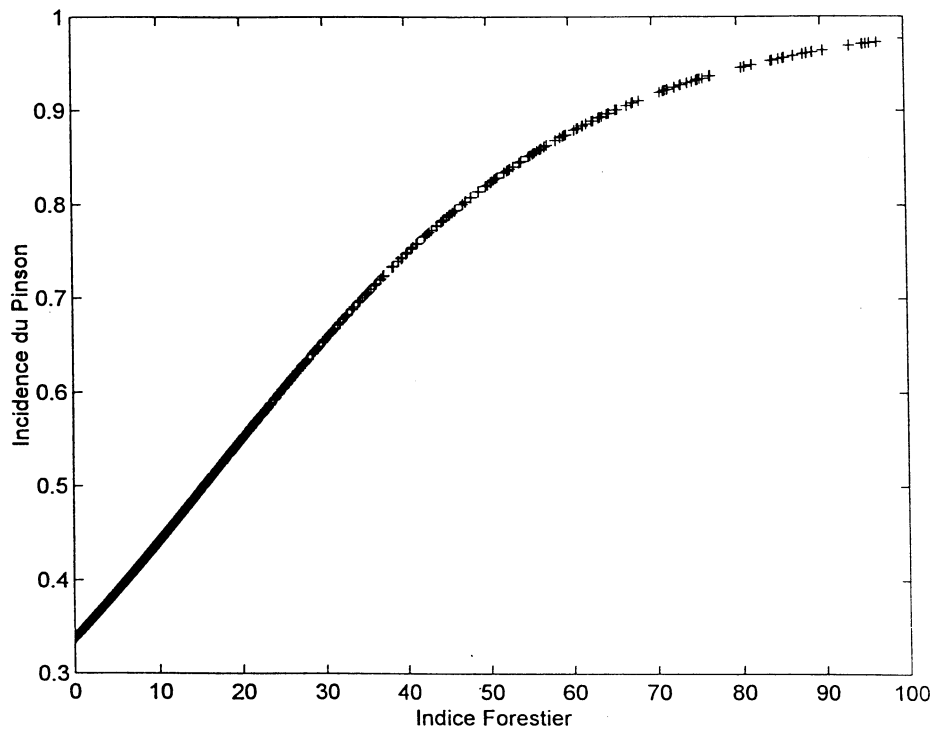


Figure 4 : Fonction d'Incidence du Pinson calculée à partir des 556 points échantillons. En ordonnées l'Incidence (probabilité de rencontre de l'espèce dans un quadrat), en abscisses l'Indice Forestier (IF) du quadrat. La courbe ajustée est de type logistique.

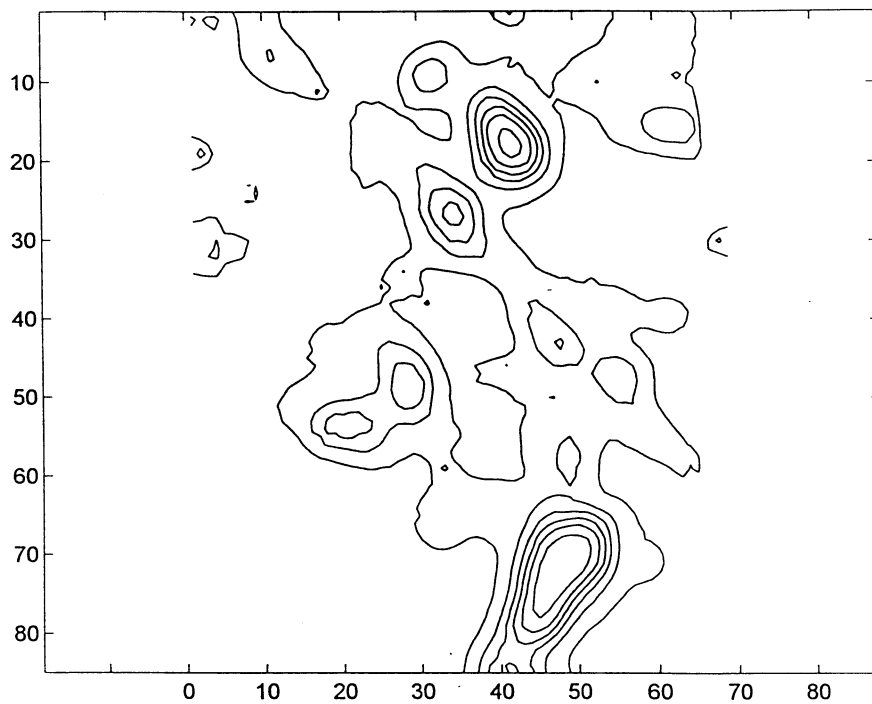


Figure 5 : Carte d'iso-incidence du Pinson sur l'ensemble de la zone étudiée. Les lignes d'iso-incidence sont tracées de 0,4 à 0,9 par pas de 0,1.

La question posée en introduction ne peut se satisfaire d'une réponse aussi graduée. Il faut à présent déterminer sur le continuum de probabilités 0,3...0,95 le "meilleur" seuil de décision, au sens statistique du terme, qui découpera la carte en deux types de zones : "Pinson déclaré présent", "Pinson déclaré absent".

6. Seuil de décision sur la fonction d'Incidence : la courbe R.O.C.

Le problème à résoudre consiste en ceci : étant donné une série continue des valeurs d'une variable (ici p), à quelle valeur particulière faut-il fixer le seuil de décision séparant les données en deux ensembles disjoints ? Une solution intuitive consisterait à considérer la valeur de $p=0,5$ comme seuil de décision. Ce n'est pas la meilleure, et le problème est plus compliqué qu'il n'y paraît.

Avant d'entrer dans des détails, examinons sur un exemple comment se présentent les données. Fixons un seuil de décision : par exemple $p=0,7$. Pour un point d'Incidence supérieure à 0,7 nous déciderons que le Pinson est présent sinon qu'il est absent. Confrontons les résultats obtenus en appliquant cette règle de décision aux 556 points échantillons (Tableau 2).

Appelant a , b , c , d les quatre cases de la table 2×2 obtenue, on voit qu'il est possible de calculer plusieurs grandeurs caractéristiques. La proportion de réponses correctes $(a+d) / (a+b+c+d)$, ici égale à 0,66 est loin d'être la seule à prendre en compte, et ce n'est pas forcément la plus pertinente. Deux quantités supplémentaires, au moins, sont à prendre en considération pour déterminer la "valeur intrinsèque" de la décision :

- la sensibilité S . $S = a / (a+c)$

C'est le rapport du nombre de Pinsons correctement détectés (a) au nombre total de Pinsons présents dans l'échantillon ($a+c$).

- la spécificité s . $s = d / (d+b)$

C'est le rapport du nombre de Pinsons correctement déclarés absents (d) au nombre total de Pinsons "absents" ($d+b$).

Les deux valeurs de la sensibilité S et de la spécificité s sont directement dépendantes du seuil de décision. En effet, lorsqu'on élève la valeur de ce seuil ($p=0,7...0,8...0,9$) le nombre de Pinsons détectés (i.e. la sensibilité) diminue alors que le nombre de Pinsons déclarés absents (i.e. la spécificité) augmente. Il apparaît donc que toute modification du seuil améliore un des deux paramètres mais dégrade l'autre. Le seuil de décision "idéal" que l'on cherche consiste à trouver un bon compromis entre sensibilité et spécificité.

Une grandeur, dans notre échantillon, ne dépend toutefois pas du seuil : la proportion de Pinsons présents : $343/556 = 0,617$. Cette valeur, pour une raison qu'on évoquera plus loin, est appelée prévalence. Il s'agit en fait, pour un point donné, de la probabilité *a priori* de présence du Pinson.

Ce type de problème est étudié depuis longtemps en statistique médicale, car il permet de fixer les seuils de décision pour le diagnostic d'une maladie à partir des valeurs d'une variable continue, comme les résultats d'un dosage. L'analogie avec le problème du Pinson est évidente. Ici il s'agit de préciser la valeur d'un dosage q qui décidera si le sujet est malade ou non, là il s'agit de fixer la valeur de l'Incidence p qui décidera si le Pinson est présent ou non. La notion de prévalence, vue plus haut, vient précisément de l'épidémiologie. La théorie statistique de cette question est celle des courbes R.O.C. (Receiver Operating Characteristics). Le lecteur trouvera dans Egan (1975) les détails théoriques -souvent complexes- de la méthode. On montrera simplement ici, et sans les justifier, les trois étapes de détermination du seuil optimal de décision.

1) On a vu que, selon les valeurs de ce seuil, sensibilité et spécificité évoluaient en sens inverse. Pour des raisons de commodité on souhaite disposer de grandeurs qui évoluent dans le même sens lorsqu'on déplace le seuil. On considérera donc la sensibilité S et le complément à 1 de la spécificité : $(1-s)$. S et $(1-s)$ croissent avec l'augmentation du seuil de décision.

Seuil de décision : 0,7

		Observe		Total
		Présent	Absent	
Calcule	Présent	170 (a)	18 (b)	188
	Absent	173 (c)	195 (d)	368
Total		343	213	556

Correct 65,6 %
 Sensibilité 0,49
 Spécificité 0,91

Tableau 2 : Résultats obtenus après segmentation de l'échantillon (N=556) au seuil d'incidence $p=0,7$. Si l'incidence d'un quadrat est supérieure à 0,7, le Pinson est déclaré présent, sinon il est déclaré absent. La table 2x2 obtenue permet de calculer plusieurs grandeurs, dont la spécificité et la sensibilité, utilisées dans l'établissement de la courbe R.O.C.

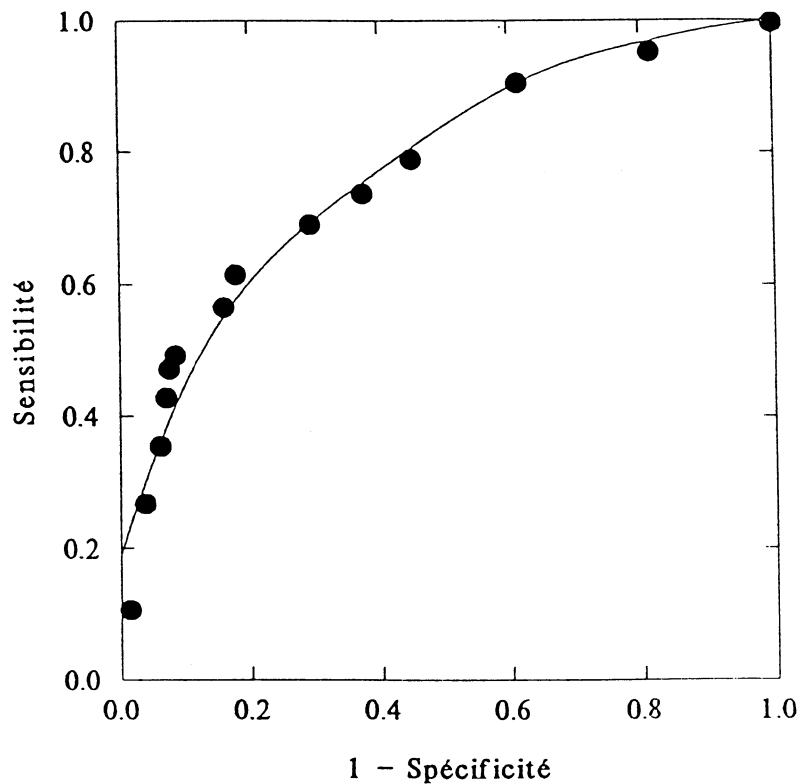


Figure 6 : Courbe R.O.C. de rencontre du Pinson sur l'ensemble des N=556 points échantillons. Quatorze seuils de décisions différents ont été étudiés. Chacun d'eux permet de calculer une spécificité et une sensibilité comme il est expliqué au tableau 2. La courbe d'interpolation est appelée courbe R.O.C. (Explications dans le texte).

2) Sur l'échantillon dont on dispose on fait régulièrement varier le seuil et l'on calcule les différentes valeurs correspondantes de S et $(1-s)$. On représente la double série de points obtenus : c'est la courbe ROC (figure 6). On voit qu'elle est monotone croissante.

3) La théorie montre que le seuil "idéal" est celui pour lequel la pente de la courbe ROC est égale à la prévalence. On adoptera donc comme seuil de décision celui pour lequel la pente de la figure 6 vaudra 0,617.

Comme on ne dispose, ici du moins, d'aucun modèle paramétrique plausible pour la courbe ROC, la détermination du seuil se réalise par interpolation graphique. La précision obtenue est suffisante.

Le seuil finalement obtenu vaut $p = 0,46$ (tableau 3).

Sans doute observera-t-on que la valeur obtenue n'est pas très éloignée de 0,5, la valeur "naïve" de décision évoquée plus haut. C'est un heureux hasard, elle aurait pu être nettement différente.

Il est maintenant possible de retracer la carte d'Incidence de la figure 5 en ne fixant qu'une seule ligne de contour : l'iso Incidence de 0,46.

Le résultat obtenu est représenté à la figure 7. Deux types de régions sont maintenant nettement distingués : "Pinson déclaré présent", "Pinson déclaré absent".

La détermination optimale du seuil de présence/absence du Pinson sur l'ensemble de la zone permet donc d'y reconnaître au moins cinq foyers probables de rencontre régulière de l'espèce, isolés les uns des autres (figure 7). La mise en évidence de ces foyers soutient l'hypothèse de départ, posée en introduction, celle d'une structuration régionale en métapopulations. Rappelons que ce résultat est obtenu avec un effort d'échantillonnage de 10% seulement de ce qu'aurait représenté le recensement exhaustif de la zone étudiée.

L'étape suivante est naturellement celle de la validation du modèle obtenu. Elle ne sera pas développée ici puisqu'elle fait l'objet d'un chapitre séparé de l'ouvrage (Joachim, *et al.*). Il suffira de préciser seulement que l'examen des populations de Pinsons dans chacun des foyers reconnus sur la figure 7 permet d'y reconnaître des différences sen-

sibles. Le marqueur utilisé est celui des variantes locales du chant (les "dialectes").

Le modèle d'habitat finalement obtenu répond à l'usage qu'on en attendait : fournir à partir de l'étude d'une variable-clé le système de répartition le plus probable de l'espèce sur toute la zone, par suite guider l'identification des noyaux de populations, isolés les uns des autres, où des divergences locales peuvent apparaître et se maintenir.

7. Conclusion

L'intérêt des modèles d'habitat n'est plus à démontrer (Verner, Morrison et Ralph, 1986). Leur usage est aujourd'hui intégré dans les démarches des scientifiques comme des questionnaires. Pour autant les problèmes que soulèvent leur établissement, leur perfectionnement et leur usage ne sont pas complètement résolus. On évoquera, dans cette conclusion, les critiques adressées à la méthode.

Parmi les étapes de construction d'un modèle d'habitat, certaines jouent un rôle décisif en terme de crédibilité et de validité des résultats. On aura perçu sur l'exemple examiné ici que trois étapes au moins sont cruciales : le choix de la variable d'habitat, l'échantillonnage et le choix du mode de réponse de l'espèce.

Ramener la présence d'une espèce au résultat du jeu d'une seule variable d'habitat est évidemment trop réducteur. Une manière possible de contourner la difficulté et d'améliorer la capacité de prédiction des modèles consiste à prendre en compte plusieurs variables d'habitat et non une seule. Mais il existe au moins deux manières de procéder et aucun accord n'existe sur la supériorité de l'une par rapport à l'autre (Van Horne, 1986). On peut d'un côté, augmenter la capacité de prédiction des modèles d'habitat en étudiant davantage de sites, sur de plus longues périodes, c'est-à-dire en utilisant une stratégie extensive. Mais on peut aussi augmenter la capacité de prédiction du modèle à construire en examinant plus finement, sur un site donné, la biologie de l'espèce : la stratégie est ici intensive.

Seuil de décision : 0,46

		Observe		Total
		Présent	Absent	
Calcule	Présent	247 (a)	95 (b)	362
	Absent	76 (c)	118 (d)	194
Total		343	213	556

Correct 69,2 %
 Sensibilité 0,78
 Spécificité 0,56

Tableau 3 : Résultat final de la segmentation des données au seuil d'incidence $p=0,46$. Ce seuil est déterminé par l'exploitation de la courbe R.O.C.. Il est optimal pour l'échantillon dont on dispose. (Explications dans le texte).

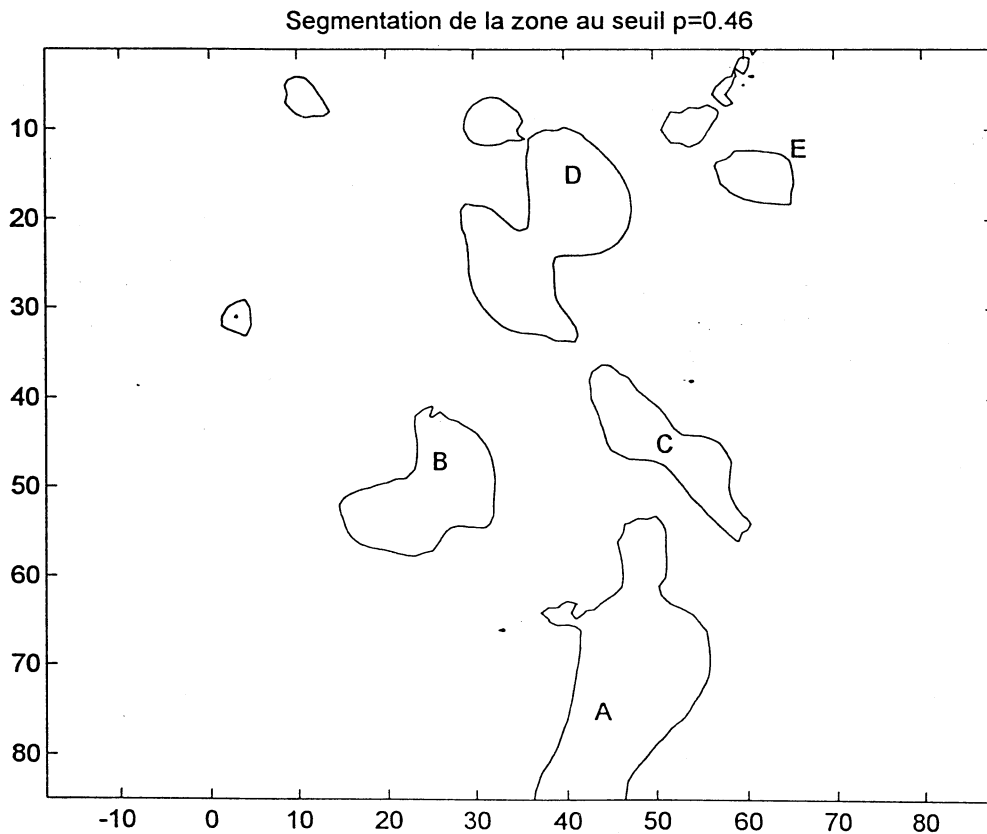


Figure 7 : Segmentation finale de la zone pour l'incidence $p=0,46$. Cette valeur de segmentation résulte de l'analyse de la courbe R.O.C. représentée à la figure 6. Les foyers potentiels de populations sont notés A, B, C, D, E.

Aucun consensus n'existe aujourd'hui sur la meilleure méthode à adopter et il est probable que la recherche d'une solution générale est dépourvue de sens. A notre point de vue, la construction d'un modèle d'habitat restera toujours la construction d'un modèle particulier, dans un cadre donné, à une échelle donnée, pour une espèce précise.

L'échantillonnage pose d'autres problèmes. La question du coût a été évoquée, on ne dira quelques mots que de la répartition des unités-échantillon. On a vu, sur l'exemple du Pinson, qu'une répartition de l'échantillonnage stratifiée selon la variable d'habitat aurait conduit au gaspillage des efforts. Il est plus habile d'adapter l'échantillonnage au problème posé, c'est-à-dire de sous-échantillonner les zones où l'on possède une quasi-certitude sur la présence ou l'absence de l'espèce et de sur-échantillonner là où l'incertitude est maximale.

Mais la détermination des zones à sur-échantillonner ... suppose de déjà connaître la répartition étudiée. Le problème est circulaire, il ne peut être résolu que par les méthodes statistiques classiques de l'échantillonnage en deux étapes. Un premier échantillon est constitué : son étude fournit une approximation. Cette approximation, à son tour, guide la suite de l'échantillonnage.

Quelles unités faut-il suréchantillonner ? Celles où l'incertitude sur la présence de l'espèce est maximale, donc autour de la probabilité $p=0,5$.

En effet la variance d'une épreuve de Bernoulli de paramètre p est proportionnelle à $p.(1-p)$: elle passe par un maximum pour $p=0,5$. C'est autour de cette valeur que l'information apportée par une unité d'échantillonnage sera la plus grande. En effet, l'information qui vaut :

$$p.\log(p).(1-p).\log(1-p),$$

passé également par un maximum pour $p=0,5$.

Le choix de l'équation de réponse de l'espèce à la (ou les) variable d'habitat est également un élément clé de la validité du modèle. On a utilisé ici l'équation logistique parce qu'elle montre une grande souplesse et, empiriquement, se comporte de manière

satisfaisante. Mais d'autres modèles sont également utilisables et, en toute rigueur, le choix définitif de la relation utilisée devrait s'appuyer sur une étude comparée. Jongman, Ter Braak et Van Tongeren (1987) examinent différents modèles possibles souvent rencontrés en statistique. Mais en dehors des distributions classiquement rencontrées, on doit aussi envisager d'éprouver des équations *ad-hoc*, spécifiques au problème analysé : par exemple des modèles à seuil.

En définitive la construction d'un modèle d'habitat n'échappe pas à des arbitraires : choix de la variable considérée, échantillonnage, choix de l'équation de réponse.

Il n'est donc pas surprenant que dans certaines situations le résultat obtenu soit un échec. Hutchinson (1978), Carey (1984) ont examiné les raisons générales par lesquelles un modèle d'habitat ne répond pas aux attentes :

- la variable d'habitat est seulement corrélée au paramètre étudié sur l'espèce, elle n'a pas, directement, de valeur causale. Le modèle s'appuie alors uniquement sur des coïncidences et non sur des relations de cause à effet. Il est dépourvu de valeur générale et ne donne que de mauvaises prédictions.

- Les espèces migratrices sont sous l'influence de deux milieux. L'étude de l'un d'entre eux seulement, par exemple l'aire de nidification printanière pour un passe-reau, ignore les effets possibles de l'autre sur la démographie de l'espèce.

- L'influence des variations climatiques, qui est en général ignorée dans les modèles d'habitat.

- La répartition spatiale des espèces en taches ou en mosaïque peut entraîner de grandes variations locales d'abondance. Ces situations sont difficiles à échantillonner correctement.

- Les facteurs contrôlant la démographie d'une espèce peuvent varier sur la zone étudiée, surtout lorsqu'elle est étendue. En outre les interactions entre facteurs limitants peuvent avoir une grande influence sur la réponse que l'on modélise.

- Les phénomènes stochastiques (génétiques, démographiques, climatiques ou environnementaux au sens large) ne peuvent

être pris en compte, par définition, dans les modèles d'habitat.

Construire un modèle d'habitat *consiste en définitive à tenter l'impossible* : apporter une réponse simple à un problème complexe. Pour autant s'agit-il d'une approche simpliste des problèmes ? Nous pensons que non. Quels que soient leurs imperfections, leurs défauts et leur part d'arbitraire, les modèles d'habitat tentent de formaliser la réponse d'une espèce à des facteurs clés.

Leur part de réussite, même minime, est précieuse : un modèle approximatif vaut mieux que pas de modèle du tout . Leur part d'échec l'est tout autant : elle conduit à s'interroger sur la représentativité des données dont on dispose, sur la forme de la réponse et finalement invite à davantage étudier et à mieux connaître la biologie des espèces.

Bibliographie

Barrett E.C., Curtis L.F., 1976. Introduction to Environmental Remote sensing. Chapman and Hall, Londres.

Carey A.B., 1984. A critical look at the issue of species-habitat dependency. *In* : Proceedings of the 1983 Convention of the Society of American Foresters, Portland, Oregon. Society of American Foresters, Bethesda, MD, 346-351.

Diamond J.M., 1975. *In* : Cody M.L. et Diamond J.M. (eds), *Ecology and Evolution of Communities*, 342-444. Harvard University Press, Cambridge, Massachusetts.

Dixon W.J., Brown M.B., Engelman L., Jennrich R.I., 1990. BMDP Statistical Software Manual, Vol 2, University of California Press, Berkeley, Californie. 1385 pages.

Egan J.P., 1975. Signal Detection Theory and ROC analysis. Academic Press, New-York. 277 pages.

Fish and Wildlife Service, 1980 a. Habitat as a Basis for Environmental Assessment. Ecological Services Manual #101. US Department of Interior, Fish and Wildlife Service, Division of Ecological Services. Gov. Printing Office, Washington. 28 pages.

Fish and Wildlife Service, 1980 b. Habitat Evaluation Procedures (HEP). Ecological Services Manual #102. US Department of Interior, Fish and Wildlife Service, Division of Ecological Services. Gov. Printing Office, Washington. 84 pages.

Fish and Wildlife Service, 1981. Standards for the development of Suitability Index Models. Ecological Services Manual #102. US Department of Interior, Fish and Wildlife Service, Division of Ecological Services. Gov. Printing Office, Washington. 68 pages.

Gilpin M., Diamond J.M., 1981. Immigration and extinction probabilities for individual species : relation to incidence function and species colonization curves. *Proceedings of the National Academy of Sciences, USA*, 78 : 392-396.

Gilpin M., Hanski I. (eds.), 1991. Metapopulation dynamics : Empirical and Theoretical investigations. Academic Press, New-York. 336 pages.

Hanski I., Gilpin M., 1991. Metapopulation dynamics: brief history and conceptual domain. *Biological Journal of the Linnean Society*, 42,3-16.

Hurley J.F., Salwasser H., Shimamoto K., 1982. Fish and Wildlife habitat capability models and special habitat criteria. *Cal-Neva Wildlife Transactions*, 82 : 40-48.

Hutchinson G.E., 1978. An Introduction to Population Ecology. Yale University Press, New Haven CT. 260 pages.

Jongman R.H.G., Ter Braak J.F., Van Tongeren O.F.R., (eds) 1987. Data analysis in community and landscape ecology. Centre for Agricultural Publishing and Documentation, Wageningen. 299 pages.

Krebs C.J., 1989. Ecological Methodology. Harper Collins, New-York. 654 pages.

Lauga J., Joachim J., 1992. Modelling the effects of forest fragmentation on certain species of forest-breeding birds. *Landscape Ecology*, 6 : 183-193.

Levins R., 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America*, 15 : 237-240.

Nelson R.D., Salwasser H., 1982. The Forest Service Wildlife and Fish Habitat Relationship program. *Transactions of The North American Wildlife and Natural Resources Conference*, 47: 174-183.

SAS Institute, 1989. SAS/STAT User's Guide, Version 6, 4^e édition, Volume 2. Cary, NC. 846 p.

Schamberger M., Farmer A., 1978. The habitat evaluation procedures: their application in project planning and impact evaluation. Transactions of the North American Wildlife and Natural Resources Conference, 43: 274-283.

Schamberger M.L., O'Neil L.J., 1986. Concepts and constraints of Habitat-Model Testing. In : Verner, Morrison et Ralph (eds), 1986, Wildlife 2000 : *Modeling habitat Relationships of Terrestrial Vertebrates*, 5-11. University of Wisconsin Press.

Snedecor G.W., Cochran W.G., 1980. Statistical methods. 7^e édition. Iowa State University Press, Ames. 507 pages.

Taylor B., 1991. Investigating species incidence over habitat fragments of different areas - a look at error estimation. *Biological Journal of the Linnean Society*, 42: 177-191

Thomas J.W., 1982. Needs for and approaches to wildlife habitat assessment. Transactions of the North American Wildlife and Natural Resources Conference, 47,35-46.

Thomas J.W., 1986. Wildlife-Habitat Modeling : Cheers, Fears and Introspection. In : Verner, Morrison et Ralph (eds), 1986, Wildlife 2000 : *Modeling habitat Relationships of Terrestrial Vertebrates*, xix-xxii. University of Wisconsin Press.

Van Horne B., 1986. When Habitats failed as Predictors : The Researcher's Viewpoint. In : Verner J., Morrison M.L., Ralph C.J. (eds), 1986. Wildlife 2000 : *Modeling habitat Relationships of Terrestrial Vertebrates*. University of Wisconsin Press, 257-258.

Verner J., Morrison M.L., Ralph C.J. (eds), 1986. Wildlife 2000 : Modeling habitat Relationships of Terrestrial Vertebrates. University of Wisconsin Press, 470 pages.

Williams G.L., Russel K.R., Seitz W.K., 1978. Pattern recognition as a tool in the ecological analysis of habitat. In : Marmelstein A. (ed.) *Classification, Inventory and Analysis of Fish and Wildlife Habitat*, FWS/OBS-78/76, US Department of Interior, Fish and Wildlife Service, Office of Biological Services, Gov. Printing Office, Washington, 521-531.

