

Zeitschrift: Saussurea : journal de la Société botanique de Genève
Herausgeber: Société botanique de Genève
Band: 33 (2003)

Artikel: Simuler la répartition géographique des espèces et de la végétation : ou "Si De Candolle avait eu un ordinateur..."
Autor: Guisan, Antoine
DOI: <https://doi.org/10.5169/seals-1098869>

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

Download PDF: 17.04.2026

ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>

Simuler la répartition géographique des espèces et de la végétation

(ou “Si De Candolle avait eu un ordinateur...”)

L'écologie végétale repose sur la botanique et cette dernière nécessite une bonne mémoire, c'est un fait avéré : la Suisse compte à elle seule plus de trois mille espèces végétales. En ce début de troisième millénaire, les écologues et les botanistes sont également amenés à maîtriser un autre type de mémoire, d'un type plus virtuel, et relevant autant de la virtuosité de leurs dix doigts que d'une bonne organisation de leurs neurones. L'informatique, pour ne pas la nommer, a en effet pris ces dernières années une grande place en écologie, comme dans le reste des sciences, donnant naissance à de nombreuses disciplines hybrides. L'une d'entre elles, la modélisation biogéographique, est à l'interface entre l'écologie, la géographie, l'informatique et les statistiques.

Antoine Guisan

Des débuts de la biogéographie à nos jours : deux siècles nous séparent

Au début du XIX^e siècle, Alexander von Humboldt et son ami Aimé de Bonpland accomplissent un voyage de découverte qui les mène notamment en Amérique. Durant ce périple, ponctué par quelques ascensions prestigieuses, ils sont frappés par la tendance qu'ont les végétaux à se distribuer en fonction du climat. A leur retour, ils publient leur *Essai sur la géographie des plantes* (1807). Ce constat sera repris et développé près d'un demi-siècle plus tard par Augustin Pyramus De Candolle - un genevois devenu célèbre - dans son traité de *Géographie botanique raisonnée* (1855). Tous ces éminents botanistes remarquèrent que la composition de la flore pouvait rester constante sur de grandes zones homogènes, puis changer drastiquement lorsqu'une discontinuité géographique (ex. topographique ou climatique) imposait une transition. Ces observations, originales pour l'époque, ont poussé les scientifiques à penser qu'il devait être possible de faire ressortir ces différences floristiques sur des cartes, donnant ainsi naissance à la biogéographie végétale, et notamment aux premières cartes de végétation. Par la suite, les travaux des différents géobotanistes ont abouti, au cours du XX^e siècle, à un enrichissement considérable des connaissances sur la répartition géographique (latitudinale, longitudinale et altitudinale) des différents types de végétation se répartissant à la surface du globe, sur les principaux paramètres qui régissent cette répartition (facteurs climatiques, géologiques, pédologiques) et sur les adaptations qui permettent aux plantes de survivre sous ces différents climats (formes biologiques, caractéristiques morphologiques, structure verticale de la végétation ; voir p. ex. Walter, 1985).

A l'aube du XIX^e siècle, les outils au service de cette discipline ont grandement progressé. Grâce au même type de technologie qui nous permet désormais, dans un véhicule, de nous faire dicter la route à suivre par un ordinateur de bord, soit le couplage d'un système de positionnement satellitaire (GPS) avec un système d'information géographique (SIG), et de techniques de statistiques performantes, il est dorénavant devenu possible de prédire la distribution géographique des espèces.



Des premiers pas aux nombreux modèles actuels : deux décennies ont suffi

La première simulation informatique de la répartition d'organismes vivants semble devoir être attribuée au modèle de distribution de la végétation du "Glacier National Park", dans le Montana (USA), développé par à Stephen R. Kessel en 1976. Peu après, en 1981, Elgene Box fut l'un des premiers à proposer un modèle informatique satisfaisant de la distribution des principaux biomes¹ à la surface terrestre. Ces travaux initièrent une nouvelle école biogéographique, plus prédictive. Une décade plus tard, stimulées par les progrès effectués dans l'étude des relations espèce-environnement (voir Austin, 1987), les études se portèrent également sur la prédiction d'espèces individuelles (ex. Hill, 1991) et devinrent de plus en plus fréquemment intégrées à l'approche biogéographique. C'est cependant durant la dernière décennie, avec l'avènement de l'informatique, et notamment des logiciels de traitement et d'analyse de l'information géographique (les systèmes d'information géographiques : SIG), que l'on assista véritablement à une explosion du nombre d'études ayant recours à cette approche, comme en témoigne le nombre croissant d'articles, de livres ou d'atlas publiés sur le sujet durant la même période (voir Franklin, 1995, Guisan & Zimmermann, 2000, Scott et al., 2002). Ainsi, en Suisse, des modèles ont été développés pour prédire la distribution de communautés végétales alpines et subalpines (Binz & Wildi, 1988, Fischer, 1990, Zimmermann & Kienast, 1999), de communautés forestières (Brzeziecki et al., 1993) ou d'espèces individuelles, qu'elles soient végétales (ex. Guisan et al., 1998) ou animales (ex. Jaberg & Guisan, 2001). Les atlas des mammifères (Hausser, 1995) et des reptiles (Hofer et al., 2001) de Suisse, qui contiennent, outre les cartes de distribution classiques, une carte du domaine potentiel de chaque espèce, sont des exemples bien diffusés dans le grand public de telles applications.

Ce récent engouement contribua fortement à une plus grande diversification des méthodes prédictives. Outre les nombreuses approches statistiques classiques existantes (voir ci-après), de nouvelles techniques alternatives – inspirées notamment des techniques d'intelligence artificielle² ou d'algorithmes d'analyse génétique – sont apparues et commencent aujourd'hui à être employées.

S'il existe ainsi une grande variété de modèles, une série d'étapes communes caractérise néanmoins toujours l'approche générale, nécessaire pour mener à bien ce type de modélisation. On y retrouve, dans l'ordre de succession : (1) la définition de la (ou les) hypothèse(s) à tester, ainsi que du modèle conceptuel (théorique), (2) l'acquisition et la préparation des données (comprenant les observations géographiques de l'espèce et les descripteurs environnementaux associés), (3) le choix d'une technique de modélisation et l'ajustement d'un modèle aux données (i.e. l'estimation de ses paramètres), (4) l'évaluation de la qualité des prédictions, et donc du modèle, et (5) la définition du domaine d'application du modèle. Dans ce qui suit, seul l'exemple de la modélisation des distributions d'espèces végétales individuelles sera considéré.

Le point de départ : la définition du modèle conceptuel

Les buts recherchés dans la modélisation de la distribution d'une ou plusieurs espèces végétales sont variés. Il peut s'agir, par exemple, de suggérer de nouvelles stations pour des espèces rares ou menacées, de préparer des scénarios d'impact des changements climatiques sur la distribution des espèces (ex. Guisan & Theurillat, 2001), d'identifier des zones à haut potentiel de biodiversité (les *hotspots* en anglais) (ex. Wohlgemuth, 1998), ou enfin de tester

¹ Les biomes sont les regroupements, discernables à l'échelle planétaire, de formations végétales dont la distribution est fortement corrélée avec les caractéristiques régionales du climat. On parle ainsi communément de *tundra* pour la végétation herbacée des régions arctiques ou alpines, de *taïga* pour les forêts boréales de conifères, et ainsi de suite, pour caractériser l'ensemble des principales formations végétales terrestres (voir Walter 1985).

² Et particulièrement les réseaux neuronaux artificiels.

certaines hypothèses biogéographiques (ex. Birks, 1993, Leathwick, 1998). Des exemples concrets de telles applications seront discutés plus loin.

Dans toutes ces situations, on part d'un même *postulat de travail*, dont l'acceptation est un préalable nécessaire à l'application de la méthode. On considérera en effet que, dans les limites de la période et de la région d'étude considérées, la distribution de l'espèce modélisée est supposée être en équilibre avec son environnement. Par équilibre, on entend que l'environnement soit favorable à l'espèce dans la région d'étude et que celle-ci ne s'y maintienne pas comme un vestige de conditions environnementales passées (p. ex. un climat différent). L'éventualité de cette dernière situation peut être minimisée si, en chaque point d'observation, il est possible de considérer l'abondance de l'espèce (plutôt que sa simple présence ou absence) et/ou sa vitalité (p. ex. sa capacité reproductive), deux critères supplémentaires qui permettent de mieux traduire la potentialité réelle des habitats échantillonnés. Il est important de noter ici que seules les espèces de fin de succession³ (*late successional species*; à l'opposé des espèces pionnières), et donc composant l'état final d'un écosystème (climax), peuvent être considérées en équilibre.

Il existe cependant plusieurs cas de figure pouvant faillir à ce postulat. Une espèce peut par exemple être en voie de colonisation (à l'exemple des néophytes⁴) et donc ne pas encore occuper tous les sites potentiellement favorables dans la région. Ces derniers sites ne devraient donc pas constituer de véritables absences au même titre que celles correspondant à des sites réellement défavorables à l'espèce. Certaines méthodes prédictives sont plus appropriées que d'autres pour traiter de tels cas. Le site peut également avoir subi une perturbation importante ayant ramené le système à un état antérieur de végétation. Cette situation est particulièrement fréquente dans un paysage anthropisé – p. ex. les pâturages – où le type d'utilisation du sol (*landuse*) conditionne fortement le type de végétation observé et doit être pris en compte dans l'ajustement des modèles. Les perturbations peuvent aussi être d'origine naturelle (avalanche exceptionnelle, glissement de terrain, etc).

La répartition géographique des espèces à la surface du globe étant, cependant, le fruit de processus écologiques dynamiques, de tels équilibres n'ont probablement jamais existé si l'on considère l'échelle des temps géologiques et les changements climatiques associés. On parle donc plutôt de *pseudo-équilibre* lorsque les changements de répartition peuvent être considérés comme négligeables sur l'échelle de temps considérée (la période d'échantillonnage des observations), généralement très courte en comparaison des échelles durant lesquelles la vie a évolué.

Ce postulat de pseudo-équilibre est implicite au principe même de cette approche de modélisation, largement basée sur la *théorie de la niche* développée par Hutchinson en 1957. De manière simplifiée, Hutchinson a défini la niche d'une espèce comme le "nuage" circonscrivant l'ensemble des conditions qui lui sont favorables au sein d'un espace à n dimensions (un hyper volume) formé théoriquement des n variables environnementales ayant un impact significatif sur la survie et la vitalité de l'espèce (fig. 1). Dans le cas des plantes terrestres vivant sous nos latitudes, trois variables environnementales importantes définissant la niche sont: (i) la lumière, (ii) la température et (iii) le substrat (dont principalement sa capacité à retenir l'eau, comme les sols à forte teneur en argiles).

La quantité de lumière agit principalement sur la photosynthèse, alors que la température peut agir soit par le biais de températures minimales (et notamment, chez nous, par la fréquence des gels tardifs au printemps), soit par

³ En un site dépourvu de végétation (p. ex. terre labourée), une *succession* d'espèces – et donc également de types de végétation – prend place dans le temps jusqu'à atteindre un état stable. On parle ainsi d'espèces pionnières pour les premières à coloniser le site et d'espèces climaciques pour celles définissant l'état final. La durée totale de la succession dépend généralement du type de climat et de substrat.

⁴ Les néophytes sont des espèces nouvelles pour une région, introduites par l'homme ou (plus rarement) venues naturellement; un exemple pré-occupant, en raison de son caractère fortement allergène, est l'Ambrosie (*Ambrosia artemisiifolia* L.) d'Amérique du Nord, introduite accidentellement en Europe (voir Jeanmonod & Lambelet, 2002).

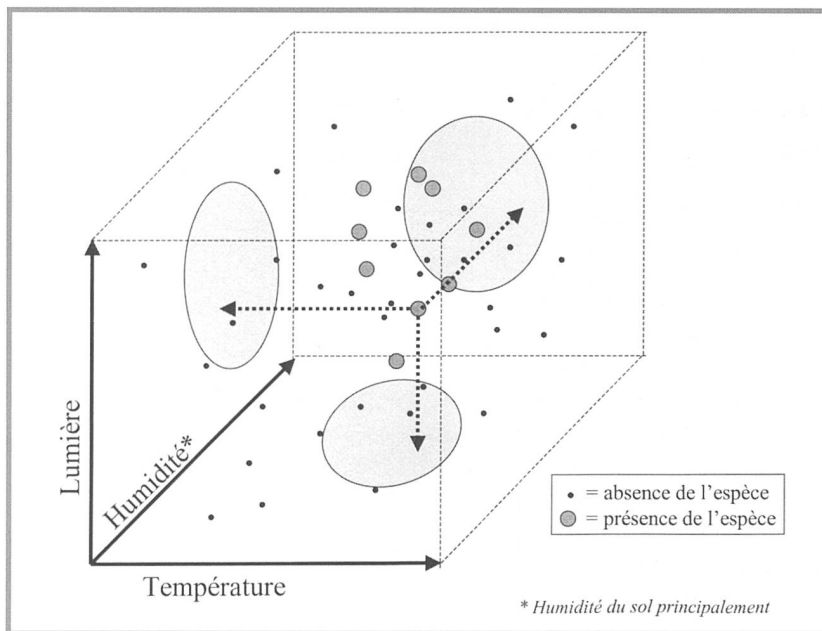


Figure 1 : Niche d'une espèce hypothétique en fonction de trois prédicteurs (variables environnementales). Le nuage des points de présence de l'espèce peut être projeté (flèches en traitillé) dans chacun des trois plans définis par les prédicteurs pris deux à deux (ellipses grises).

la somme de chaleur accumulée qui définit alors la durée de la saison de croissance (p. ex. dans les régions alpines). Enfin, plusieurs éléments doivent être pris en compte en ce qui concerne la nature du substrat, tels que son humidité relative, sa nature chimique (acide ou basique) et sa teneur en nutriments, cette dernière agissant principalement en modifiant les rapports de compétition entre espèces (par exemple, pour la lumière). Chaque espèce occupe un nuage différent dans l'espace environnemental (fig. 2). Bien que cela ne soit pas le cas dans la figure 2 (par souci de simplification), les nuages de différentes espèces se superposent souvent partiellement.

L'identification de l'intervalle des valeurs environnementales tolérées par une espèce peut se faire en laboratoire, en soumettant par exemple l'espèce à différents traitements de lumière, de température ou de la teneur du sol en nutriments (phosphore et azote principalement) ou en eau. Il en résulte des courbes de réponse exprimant la vitalité de l'espèce (mesurée par exemple par le nombre de graines ou la biomasse produite) en fonction de ces différents facteurs. En milieu naturel, on observe cependant souvent que de nombreuses espèces n'occupent pas l'intégralité de leur niche théorique, probablement en raison du fait qu'elles en sont exclues par d'autres espèces plus compétitives, ce qui est représenté en une seule dimension environnementale dans la figure 3. On parle ainsi de *niche réalisée* (ou niche écologique) pour la niche d'une espèce mesurée en milieu naturel, et de *niche fondamentale* (ou niche théorique) pour celle mesurée en laboratoire. La niche réalisée (observée sur le terrain) est donc généralement considérée comme une fraction de la niche fondamentale, et les principales causes de cette restriction de niche sont attribuées aux interactions négatives avec d'autres organismes vivants (p. ex. compétition avec d'autres espèces végétales).

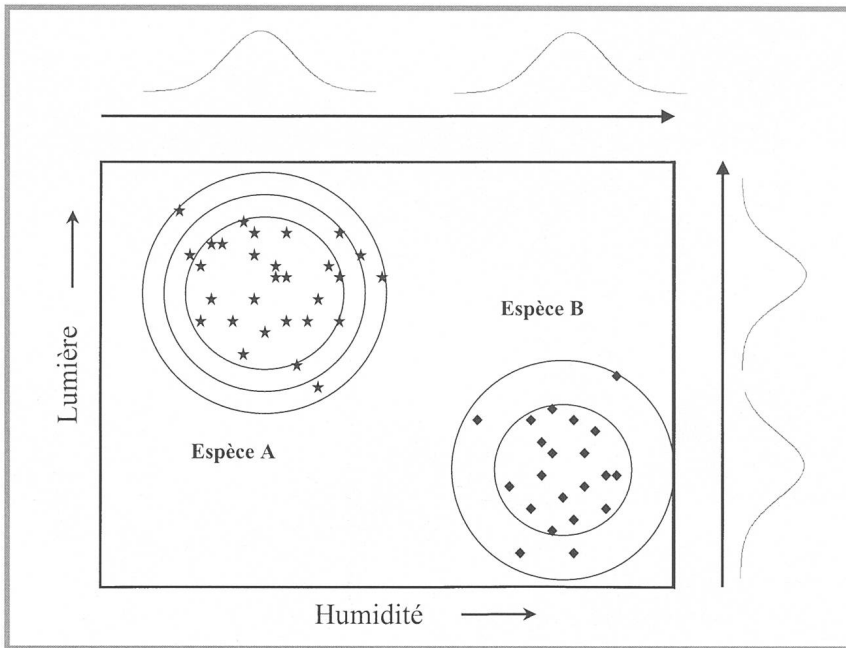


Figure 2: Niches respectives de deux espèces ayant des exigences écologiques très distinctes dans l'espace de deux prédicteurs: l'humidité du sol et la lumière. Les différents cercles concentriques illustrent différents niveaux de probabilités d'observer l'espèce et les courbes en position verticale et horizontale à droite et au-dessus représentent les courbes de réponse de chaque espèce en fonction du prédicteur correspondant (fonctions de probabilité).

Le principe de l'approche par modélisation dont il est question ici est de mettre en relation des observations géographiques d'une espèce avec certains des attributs environnementaux associés aux sites d'observation (fig. 4).

On modélise donc nécessairement ici la niche réalisée de l'espèce, en partant du constat précédent que sa répartition observée est déjà le fruit des éventuelles contraintes biotiques restreignant sa niche fondamentale.

D'un point de vue théorique, il n'est cependant pas exclu que la niche observée sur le terrain puisse, dans certains cas, se révéler plus large que la niche mesurée en laboratoire, si des processus de facilitation par d'autres espèces peuvent intervenir et permettre à l'espèce de survivre dans des conditions environnementales considérées comme peu favorables du point de vue expérimental. De tels processus de facilitation de niche pourraient notamment résulter du fait de symbioses, telles que certains types de mycorhizes – l'alliance de champignons et de plantes, qui semblent concerner près de 60 % des espèces végétales (Sanders, 2002) – ne pouvant pas être reproduites en laboratoire.

La première étape de modélisation est donc de définir un modèle conceptuel de la niche de l'espèce étudiée. Cela revient souvent à définir un diagramme des relations plante-environnement tel qu'illustré dans la figure 5. Ce dernier permet notamment de suggérer quelles variables environnementales devraient être idéalement à disposition, et cartographiées sur l'ensemble de la région d'étude⁵, pour permettre l'ajustement du meilleur modèle prédictif possible.

Un autre aspect théorique de première importance à considérer dans l'élaboration du modèle conceptuel est l'identification des principales hypothèses biogéographiques ou écologiques à tester ou à considérer lors de l'interprétation finale des modèles. Une question majeure est notamment de

⁵ De manière à ce que le modèle soit extrapolable à l'ensemble de la région.

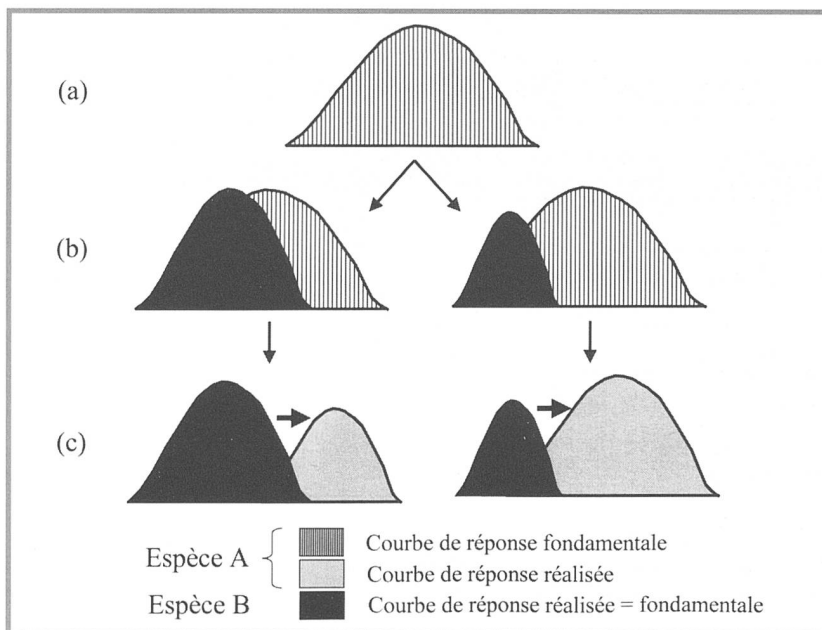


Figure 3: (a) Courbe de réponse fondamentale (surface hachurée) d'une espèces hypothétiques A en fonction d'un prédicteur hypothétique. (b) L'espèce A est exclue d'une partie de sa niche par une espèce B, plus compétitive pour le prédicteur concerné; deux scénarios différents de tolérance écologique pour l'espèce B sont considérés ici; (c) Courbes de réponse réalisées de l'espèce A obtenues par exclusion compétitive en considérant les deux tolérances écologiques définies précédemment pour l'espèce B.

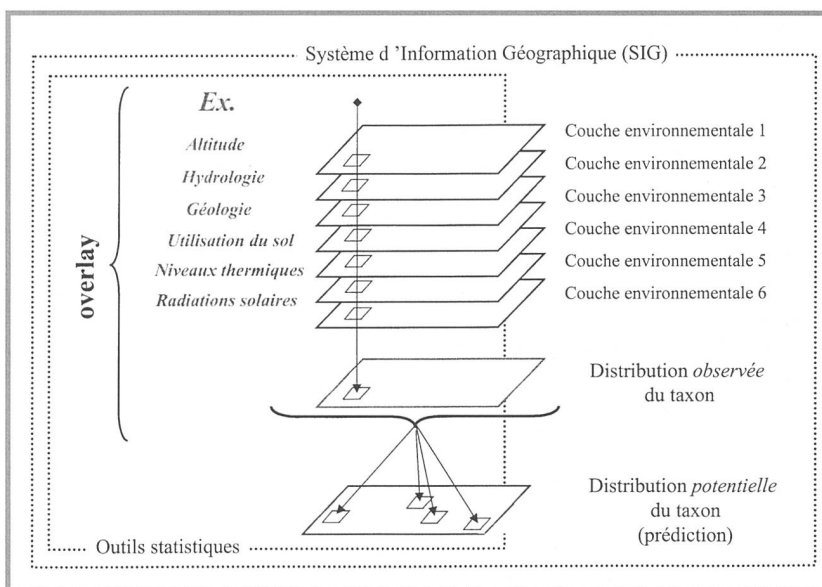


Figure 4: Mise en relation des couches environnementales - les prédicteurs - avec la distribution observée de l'espèce - la variable réponse - permettant de prédire la distribution potentielle du taxon modélisé dans toutes les unités cartographiques (ou pixels) de la zone.

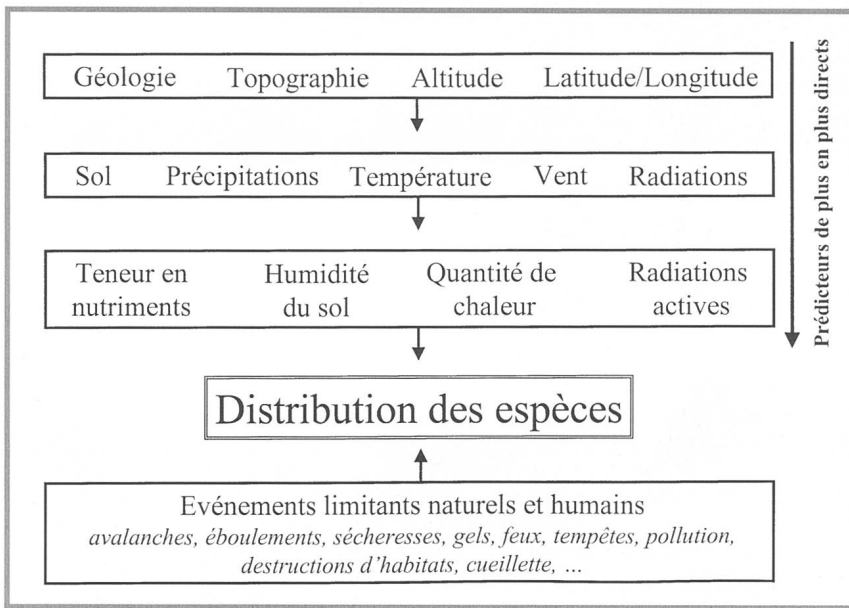


Figure 5: Modèle conceptuel simplifié des variables environnementales déterminant la répartition des espèces végétales (modifié de Guisan & Zimmermann, 2000). Les trois groupes de prédicteurs dans la partie supérieure ont une influence de plus en plus directe sur la distribution des espèces (de haut en bas). Cette dernière est également influencée par des événements occasionnels, et parfois extrêmes, d'origine naturelle ou anthropique. Ce schéma n'est pas exhaustif, tous les types de prédicteurs n'ayant pas été considérés.

déterminer l'importance relative des processus biotiques par rapport aux facteurs abiotiques (les prédicteurs environnementaux) dans la détermination des limites d'aires de distribution des espèces. Brown et al. (1996) nous rappellent sur ce point que la distribution de la majorité des espèces le long des principaux gradients environnementaux pourrait bien être limitée d'un côté par un stress physique et d'un autre côté par un stress biologique (principalement la compétition avec d'autres espèces), comme l'avait déjà suggéré Ronald H. MacArthur pour les gradients latitudinaux dans son traité de 1972. Cette hypothèse a également été proposée pour expliquer la limite inférieure de répartition de certaines espèces dans le cas des gradients altitudinaux (Guisan et al., 1998). Les modèles de distribution actuels se basant sur une quantification de la niche écologique des espèces, ils ne permettent pas de simuler cette compétition, sinon de l'inclure *de facto* dans leurs prédictions. Cette problématique constitue un axe de recherche en développement actuellement.

Les ingrédients de base: les données géoréférencées

Les données utilisées dans ce type d'approche appartiennent à deux grandes catégories. D'une part, la variable que l'on cherche à prédire, appelée la *variable réponse*, et d'autre part l'ensemble des variables utilisées pour prédire cette variable réponse, que l'on nomme les *prédicteurs*. Les termes variable expliquée ou variable dépendante sont des synonymes couramment utilisés du terme variable réponse et les termes variables explicatives ou variables indépendantes sont similairement des synonymes du terme prédicteur. Par soucis de clarté, ils ne seront cependant plus utilisés par la suite.

Pour pouvoir construire des modèles de distribution, ces variables doivent impérativement être géoréférencées, et donc associées à des coordonnées géo-

graphiques aussi précises que possible. Des progrès considérables ont été rendus possibles avec l'apparition, au début des années nonante, des systèmes de positionnement satellitaire (*Global Positioning System*, GPS). Un nouveau pas a été franchi en 1999, lorsque les États-Unis ont cessé de brouiller la qualité des signaux reçus, ce qui limitait considérablement la précision du positionnement.

Une même unité cartographique de base doit aussi être définie pour toutes les couches de données, de manière à les rendre superposables dans un système d'information géographique (figure 4). Ceci implique (i) que l'ensemble des données possède un même système de projection et (ii) que certaines variables soient agrégées pour n'offrir qu'une valeur (un attribut) par unité cartographique. L'unité cartographique régulière de base est communément appelée *pixel* (pour *pictural element*) et le choix de sa taille (définissant la résolution de l'étude) devrait dépendre en premier lieu de l'espèce modélisée (la résolution pertinente n'est pas la même pour une mousse et pour un arbre). Dans la pratique, cependant, le choix de la taille du pixel est souvent déterminé par la résolution du prédicteur le plus grossier, et donc par les données à disposition. Il est également possible de considérer une unité de modélisation irrégulière, tels que des polygones définis dans le paysage (p. ex. des unités de végétation), ou des pixels d'une forme autre que le carré (p. ex. des hexagones), mais de tels exemples sont encore rares dans la littérature.

La variable réponse est généralement constituée d'observations géographiques de l'espèce étudiée. Lorsqu'un inventaire, prenant en compte l'ensemble des espèces végétales en chaque site d'observation, en est la source (relevés phytosociologiques⁶, réseau de carrés permanents), il est généralement possible et relativement facile⁷ de bénéficier également de points de "non observation" de l'espèce. Ceux-ci sont dès lors utilisés comme indication d'absence de l'espèce dans le site correspondant. Ces absences permettent généralement une meilleure distinction par le modèle entre sites favorables et défavorables à l'espèce. Mais il arrive également fréquemment que seules des observations individuelles de l'espèce soient disponibles. C'est notamment le cas de bon nombre de données accumulées dans les grandes bases de données nationales, tel que le Centre du réseau suisse de floristique (CRSF, Genève) en Suisse. C'est aussi tout particulièrement le cas des espèces rares. Dans ce cas de figure, aucun site ne pouvant être utilisé pour quantifier les critères d'absence de l'espèce, le recours à des méthodes de modélisations particulières est rendu nécessaire (voir section suivante).

En plus de la simple présence de l'espèce en un site, son abondance peut également être mesurée précisément ou évaluée. Cette information écologique supplémentaire peut se révéler d'un grand intérêt pour le modélisateur, en lui permettant de répartir les sites d'observation sur un gradient plus nuancé de "favorabilité", selon qu'un petit ou un grand nombre d'individus ait été répertorié. La vitalité de l'espèce est parfois également mesurée et peut également constituer une source d'information complémentaire de grand intérêt du point de vue écologique. Ces différents types de variables réponse sont cependant mesurés sur des échelles de valeurs numériques différentes (0 ou 1 pour la présence-absence, l'intervalle de 0 à $+\infty$ pour les comptages d'individus et, par exemple, des classes successives codées entre 0 et 8 pour la vitalité), de telle sorte que le choix de l'une ou l'autre de ces variables implique parallèlement le choix d'un modèle approprié (Guisan & Zimmermann, 2000). Le choix d'un modèle inadéquat peut en effet avoir de fâcheuses répercussions sur la qualité des prédictions (Guisan, 2002).

Parmi les prédicteurs les plus souvent utilisés, l'altitude figure de loin en première place. Lorsque celle-ci est représentée en continu sur l'ensemble de

⁶ La phytosociologie est l'étude des communautés végétales : elle nécessite donc de procéder à des relevés exhaustifs des espèces présentes et à la quantification de leur importance respective par des mesures d'abondance-dominance.

⁷ Ce n'est par contre pas le cas de nombreuses espèces animales, souvent bien plus difficiles à observer que les plantes, même lors d'inventaires considérés comme exhaustifs.

la région d'étude, on nomme la couche géographique correspondante un *modèle numérique de terrain* (MNT; en anglais: *Digital Elevation Model, DEM*). En Suisse, la résolution du MNT national est rapidement passée de 250 m à 25 m, et un modèle au mètre devrait être prochainement disponible. A partir d'un tel MNT, il est extrêmement aisé de dériver, dans un SIG, d'autres variables topographiques, comme la pente, l'exposition, la courbure du terrain ou la position topographique le long d'un gradient caractérisant les fonds de vallée, les milieux de pente et les sommets. Ces variables n'auront, bien sûr, pas tout à fait la même signification selon la résolution du MNT utilisée.

D'autres variables (radiations solaires, température, précipitations) peuvent être obtenues de manière similaire, mais généralement au prix de procédures plus complexes (e.g. Thornton et al., 1997), nécessitant des observations complémentaires (par exemple un réseau de mesures météorologiques; voir Zimmermann & Kienast, 1995). C'est au niveau de ces procédures complexes que s'exprime notamment toute la puissance des SIG. On peut mentionner ici les analyses de voisinage permettant d'attribuer à un pixel donné une valeur basée sur les pixels voisins, selon une fenêtre de voisinage de forme définie par l'utilisateur, par exemple un carré ou un cercle, ou même une forme asymétrique.

Enfin, d'autres prédictors peuvent encore provenir de sources cartographiques existantes, au préalable numérisées dans un SIG. C'est notamment le cas des cartes géologiques, géomorphologiques ou pédologiques⁸, ainsi que des bases de données fédérales sur l'utilisation du sol (GEOSTAT pour la Suisse) ou des objets particuliers d'importance écologique (p. ex. les inventaires fédéraux des bas-marais ou des zones alluviales). Une dernière catégorie de prédictors est générée par des techniques d'imagerie satellitaire (ou télédétection), en utilisant directement les bandes spectrales d'images satellites ou de photographies aériennes, ou en les recombinaison pour obtenir de nouvelles variables, comme c'est le cas des indices de végétation (p. ex. le *Normalized-Difference Vegetation Index* abrégé NDVI ou le *New Vegetation Index* abrégé NVI).

Il existe donc une grande quantité de sources de données possibles et le choix des bonnes variables n'est pas toujours aisé. Cette complexité nous renvoie au modèle conceptuel vu précédemment, et à la nécessité de se baser, dès le début de l'étude, sur des hypothèses écologiques crédibles permettant d'orienter le choix des prédictors. On préférera notamment les prédictors ayant un effet physiologique ou écologique *direct* sur l'espèce modélisée, plutôt que des prédictors ayant un effet indirect, car agissant par le biais d'autres facteurs (Austin, 2002). C'est, par exemple, le cas de l'altitude qui n'a aucun effet direct sur les plantes, comme nous le démontre le cas des espèces arctico-alpines comme la renoncule des glaciers (*Ranunculus glacialis*; fig. 6).



Figure 6: La renoncule des glaciers (*Ranunculus glacialis* L.), une espèce arctico-alpine indifférente à l'altitude (on la trouve dans l'arctique) mais sensible aux températures.

⁸ La pédologie est la science des sols.

En effet, celles-ci sont aussi bien distribuées à très haute altitude dans les Alpes qu'à basse altitude dans l'Arctique. Aucun modèle basé sur l'altitude ne pourra donc expliquer leur distribution conjointe sous ces différentes latitudes, alors qu'un modèle basé sur des prédicteurs écologiquement importants pour les espèces (comme, par exemple, la somme de température disponible durant la saison de croissance) aura beaucoup plus de chance d'aboutir à des résultats significatifs, puisque les valeurs de ces prédicteurs devraient s'avérer équivalentes entre les milieux alpins et arctiques, *a priori* si différents. Le fait que ces espèces se trouvent distribuées dans les deux types de milieux mentionnés suggère, en tous les cas, la nécessaire convergence de certains paramètres environnementaux de portée globale (aussi dit "proximaux"). L'approche par modélisation cherche justement à identifier de tels facteurs proximaux, afin de permettre d'expliquer et de prédire la distribution des espèces dans des types de milieux aussi variés que possible. En revanche, dans une situation locale (p. ex. une vallée alpine), les prédicteurs indirects (distaux) s'avèrent encore très souvent meilleurs que les prédicteurs directs pour simuler la distribution des espèces, car ils sont habituellement générés avec moins d'erreur. Tout l'art est donc de trouver les meilleurs prédicteurs en fonction de l'échelle d'étude et de l'interprétation que l'on veut ensuite faire du modèle et de ses prédictions.

La phase technique : le choix et l'ajustement du modèle

Il y a vingt ans, l'analyse des relations entre la distribution des espèces et leur environnement était en grande partie basée sur des techniques classiques, telles que la régression ordinaire des moindres carrés (voir les explications ci-dessous) ou l'emploi de simples coefficients de corrélation (mesures de covariation). La question du choix d'un modèle statistique ne se posait donc pour ainsi dire pas et l'essentiel de l'effort pouvait être mis sur le choix des prédicteurs et sur la qualité des données biologiques. La régression ordinaire, encore largement utilisée aujourd'hui dans certaines situations, avait cependant le défaut de nécessiter que la variable réponse suive une loi de probabilité mathématique particulière, la *fonction de Gauss* ou loi Normale⁹. Depuis, il s'est avéré que de nombreuses variables en écologie ne suivent pas une telle loi (dont les variables de présence-absence qui sont typiquement binaires) et donc que le postulat même d'application de cette méthode est violé lorsqu'elle est appliquée à des variables non-Gaussiennes (Austin, 2002, Guisan et al., 2002). De nouvelles approches statistiques ont donc été développées. Il en existe aujourd'hui une telle multitude que le choix d'un modèle statistique adéquat s'avère parfois difficile.

Les principales approches permettant de modéliser la distribution des espèces peuvent être regroupées en sept catégories: (i) les systèmes experts, basés sur les connaissances théoriques de spécialistes, (ii) les enveloppes géographiques, (iii) les enveloppes climatiques, (iv) les méthodes de régression multiple, (v) les arbres de classification, (vi) les méthodes ayant recours à des mesures d'association multivariées, et enfin (vii) les méthodes d'intelligence artificielle. Il est cependant difficile, et hors du contexte de cet article, d'expliquer l'ensemble de ces approches (voir pour cela Guisan & Zimmermann, 2000 ou Elith, 2002). Les méthodes (i) à (iii) seront brièvement présentées et seule l'approche (iv) sera discutée et illustrée.

Les modèles les plus simples, qui sont aussi les premiers à avoir été largement utilisés, sont les systèmes experts, les enveloppes géographiques et les enveloppes climatiques. Les *systèmes experts* sont de simples ensembles de

⁹ Ou plus exactement, les erreurs (ou résidus) d'un modèle de régression ordinaire doivent suivre une distribution normale.

règles, souvent booléennes (oui/non), utilisées pour combiner des couches géographiques choisies au préalable, de manière à aboutir à une carte potentielle assez primaire de la distribution de l'espèce. Les *enveloppes géographiques* au sens large comprennent l'ensemble des méthodes cherchant à dessiner automatiquement le *contour* de la distribution d'une espèce sur une carte (*home range*). Il s'agit donc, assez simplement, de relier tous les points d'observation, ou une partie d'entre eux, situés les plus en bordure d'aire, de manière à définir un polygone fermé (ou éventuellement plusieurs) circonscrivant l'aire de distribution observée de l'espèce. C'est l'approche qui a succédé au dessin manuel de ces limites sur une carte géographique. Différentes règles peuvent être utilisées pour relier les points entre eux, aboutissant par exemple à une surface minimale si les points sont reliés entre-eux de manière à minimiser l'aire du polygone. Les *enveloppes climatiques* représentent le pendant de l'approche précédente, mais en considérant cette fois l'aire de l'espèce dans un espace climatique plutôt que géographique. Les méthodes de calcul de polygones sont donc les mêmes, mais appliquées à l'espace climatique. En ce sens, l'espace climatique de l'espèce ainsi défini constitue une première circonscription de sa niche réalisée, qui nécessiterait encore d'être restreinte par d'autres variables environnementales importantes (ex. types de sols). Le problème principal des méthodes d'enveloppes est qu'un même poids est généralement attribué à tous les prédicteurs, ce qui s'avère irréaliste d'un point de vue écologique (Elith, 2002). Par contre, un de leur grand avantage réside dans le fait de ne nécessiter que des données de présence, ce que les méthodes qui suivent – de type discriminantes – ne permettent pas.

Un type de modèle plus élaboré permettant d'attribuer des poids différents aux prédicteurs est la régression multiple. Le principe de la régression est d'ajuster un coefficient pour chaque prédicteur et d'additionner leur produit dans une formule du type :

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

où y est la variable réponse que l'on cherche à modéliser, α est une constante, $\{x_1, x_2, \dots, x_k\}$ sont les prédicteurs, $\{\beta_1, \beta_2, \dots, \beta_k\}$ sont les coefficients multipliant les prédicteurs et ε est le vecteur des erreurs que l'on cherche à minimiser. En remplaçant les x_i dans la formule par les valeurs mesurées en chaque site on obtient une valeur prédite de Y . Les coefficients sont estimés mathématiquement de manière à minimiser la différence globale (ε) entre les valeurs observées et les valeurs prédites sur l'ensemble des sites d'observation. On parle de régression simple lorsqu'un seul prédicteur (x_1, x_2, \dots) est considéré et de régression multiple lorsque plusieurs prédicteurs sont considérés.

Les approches par régression multiple ont été, de loin, les plus utilisées ces dernières années pour prédire la distribution des espèces. Ce succès est dû en grande partie à la généralisation de l'approche proposée au début des années quatre-vingt, à travers le développement des modèles linéaires généralisés (*generalized linear models*, GLM; voir Guisan et al., 2002). Les GLMs représentent une famille de modèles basés sur le principe de la régression multiple, mais utilisant un autre algorithme mathématique d'estimation des paramètres (technique du maximum de vraisemblance). Au contraire de la régression traditionnelle qui nécessite exclusivement des variables réponses normales (gaussiennes), les GLMs permettent de considérer un grand nombre de variables réponses différentes rencontrées en écologie, telles que : (i) la présence-absence d'une espèce (données binaires 0/1), (ii) les comptages d'individus ou

d'espèces (qui suivent souvent une loi de probabilité de Poisson¹⁰) ou, de prime importance pour la valorisation de certaines données phytosociologiques, (iii) les classes successives (ordonnées) d'abondance (une variable semi-quantitative; voir Guisan, 2002).

L'un des intérêts majeurs des méthodes de régression multiple est également de permettre de considérer un grand nombre de transformations possibles au niveau des prédicteurs. Ainsi, outre la prise en compte évidente d'une relation linéaire entre la variable réponse et un prédicteur donné (fig. 7a), correspondant à l'inclusion d'un terme simple (x) du prédicteur dans le modèle, une relation unimodale (qui signifie à une seule "bosse" ou mode) et symétrique sera tout aussi facilement simulée par l'inclusion d'un terme supplémentaire quadratique (x^2) du même prédicteur dans le modèle (fig. 7b). Des réponses plus complexes, telles qu'asymétrique ou multimodales, pourront similairement être prises en compte en rajoutant des termes cubiques (x^3) ou de degré polynomial supérieur dans le modèle (x^4, x^5, \dots). Des transformations plus complexes ainsi que des interactions entre prédicteurs, permettant de conditionner l'importance d'un prédicteur par celle (conjointe ou inverse) d'un autre prédicteur, peuvent aussi être considérées.

Un exemple de régression multiple est illustré de manière simplifiée par le modèle (sans interactions) du trèfle alpin *Trifolium alpinum* L. dans la région de Belalp (Valais, Suisse; voir photo et carte potentielle en figure 8):

$$\text{Prob}(T. alpinum) = \beta_1 (\text{température}) + \beta_2 (\text{température})^2 + \beta_3 (\text{radiations}) + \beta_4 (\text{radiations})^2 + \beta_5 (\text{position topographique}) + \beta_6 (\text{position topographique})^2 + \beta_7 (\text{bande spectrale no 2 de la photo aérienne infrarouge}) + \beta_8 (\text{bande spectrale no 2 de la photo aérienne infrarouge})^2$$

où les β_i sont, comme dans l'équation précédente, les coefficients estimés lors de l'ajustement du modèle et $\text{Prob}(T. alpinum)$ est la probabilité d'observer l'espèce. Dans un GLM, c'est une transformation de cette probabilité (assurée par une fonction dite "de lien") qui est mise en relation linéaire avec la combinaison des prédicteurs (partie de droite de l'équation) expliquant le plus de variation dans les données. Dans le modèle de *T. alpinum*, la forme carrée de chaque prédicteur a également été retenue. La deuxième bande spectrale d'une photographie aérienne infrarouge illustre bien l'utilisation possible de données de télédétection dans ce type d'approche.

Une extension des GLMs – les modèles additifs généralisés (*generalized additive models*, GAM) – permet justement de remplacer ces coefficients fixes par des *fonctions de lissage*, autrement dit des fonctions plus libres se calquant mieux aux données modélisées. En règle générale, les GAMs sont donc des modèles mieux ajustés aux données à disposition que les GLMs, mais moins généralisables à d'autres situations et parfois d'interprétation plus difficiles (fig. 7c).

La sélection des prédicteurs dans un GLM ou dans un GAM peut être effectuée de différentes manières et selon différents critères. Elle se base généralement sur la proportion de la variance totale (i.e. la variation au sein des données) expliquée par chaque prédicteur et sur la significativité statistique de celle-ci. Cette sélection peut être automatisée ou manuelle. On notera ici que, comme dans le cas de la régression multiple, l'ordre d'entrée de prédicteurs partiellement corrélés¹¹ dans un GLM ou un GAM a une influence sur leur sélection. Par exemple, l'altitude et la température moyenne annuelle sont deux variables très fortement corrélées, la première permettant de calculer en partie la deuxième. Ce problème disparaît dès lors que l'ensemble des pré-

¹⁰ Simeon-Denis Poisson (1781-1840) fut professeur de mathématique à l'Ecole polytechnique de Paris et à la Sorbonne. Il participa au développement de la théorie des probabilités, si bien qu'une loi de probabilité discrète porte aujourd'hui son nom. En écologie, cette dernière s'applique particulièrement bien aux *comptages* d'espèces ou d'individus.

¹¹ La corrélation statistique est une mesure de co-variation - et donc de dépendance - entre deux variables; une valeur de 0 n'indique aucune corrélation alors qu'une valeur de 1 indique une corrélation parfaite.

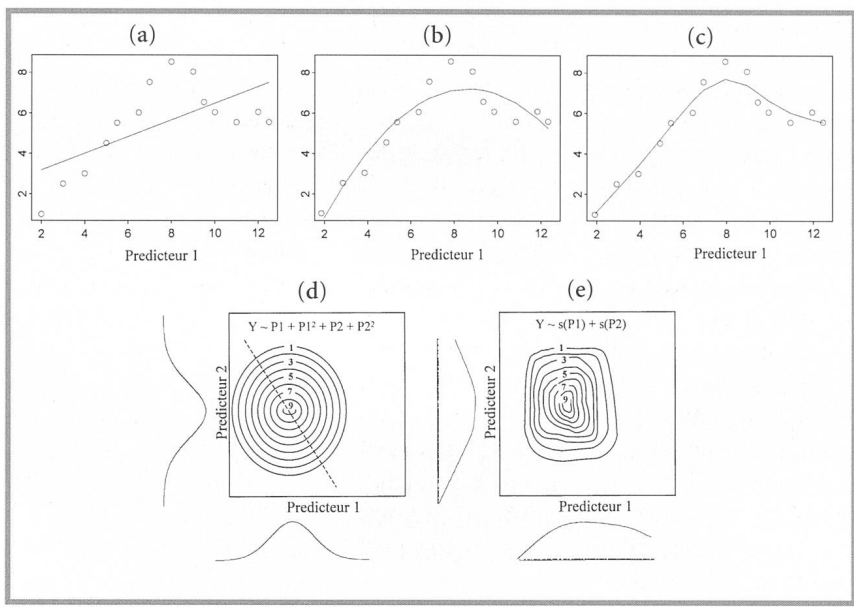


Figure 7: (a) Courbes de réponse de l'abondance d'une espèce hypothétique à un prédicteur environnemental (P1) dans les cas d'un GLM Poisson avec un terme linéaire simple x ; (b) idem mais avec un terme quadratique x^2 additionnel; (c) cas d'un GAM (voir texte) avec une fonction de lissage de type "spline" (s). Les GAM permettent d'ajuster des courbes de réponse "plus proche des données" que les GLM, mais qui sont souvent moins généralisables à d'autres situations. (d) Représentation de la réponse d'une espèce à deux prédicteurs environnementaux P1 et P2 obtenue dans le cas d'un GLM Poisson. Les isohypses d'abondance décroissante à partir du centre définissent une surface de réponse dans l'espace à deux dimensions. (e) Idem, mais dans le cas d'un GAM (voir texte).

dicteurs sont totalement non-corrélés entre eux, ce qui peut être obtenu par une procédure particulière de réduction et d'orthogonalisation des données (p. ex. par une analyse en composante principale) générant de nouvelles variables totalement non-corrélées. Dans le cas de deux prédicteurs fortement corrélés entre eux et avec la variable réponse, le premier à être introduit dans le modèle peut être conservé au détriment de l'autre. Il peut cependant arriver que le prédicteur non retenu puisse avoir, en réalité, une influence plus directe (car plus "causale" du point de vue biologique) sur la distribution de l'espèce modélisée, ce que la méthode mathématique ne peut, en soi, détecter. On voit donc à nouveau l'importance de se baser sur une hypothèse définie *a priori* des relations que l'on veut modéliser et qui permette un choix préalable des prédicteurs ainsi que l'utilisation d'éventuelles règles de préséance les concernant (hiérarchie de leurs effets).

La phase critique: l'évaluation des modèles et de leurs prédictions

L'évaluation d'un modèle et de ses prédictions est la phase la plus sensible de la procédure de construction d'un modèle, dans la mesure où cette étape constitue le lien indispensable avec le domaine d'application du modèle. Un modèle ne devrait, par exemple, pas être appliqué pour la protection d'une espèce si sa validité n'a pas été au préalable assurée. La notion de validité est

cependant subjective et sous-entend que l'on connaisse précisément la réalité, ce qui est rarement le cas, à l'exception peut-être de certaines populations reliques dont l'emplacement et le nombre total d'individus sont précisément connus. Dans la grande majorité des cas, on parlera donc plutôt d'évaluation et on cherchera à obtenir un modèle dont le risque d'erreur est raisonnablement bas. Le modèle peut être optimisé pour minimiser un type d'erreur particulier, tel que le risque de prédire une espèce rare comme absente d'un site alors qu'elle s'y trouve, ce qui peut se traduire par un *coût écologique* particulièrement élevé si le site est ensuite dégradé par faute de n'y avoir pas appliqué les mesures de protection adéquates.

Le modèle mathématique doit tout d'abord être lui-même évalué. Des mesures statistiques permettent de quantifier la part de variation expliquée par le modèle. On parlera dans ce cas de *qualité de l'ajustement*¹². Il est également possible de tester dans quelle mesure les paramètres du modèle pourraient avoir été obtenus par chance. On procède pour cela de manière itérative en reconstituant un grand nombre de fois les données au hasard (par *permutation*), de manière à obtenir autant de jeux de données reconstruits aléatoirement et de modèles associés, et en observant si certains de ces modèles, et combien, s'avèrent meilleurs que le modèle basé sur les données originales.

Deux types d'évaluation – visuelle et quantitative – peuvent également être conduits au niveau des prédictions du modèle.

L'évaluation visuelle la plus simple consiste à représenter les prédictions sous une forme cartographique, correspondant à la carte de distribution potentielle de l'espèce, puis de les comparer à la distribution observée de l'espèce et finalement de les confronter au jugement d'un expert. Un raffinement possible de l'évaluation visuelle, bien qu'encore peu appliqué, est de représenter également les erreurs associées à chaque prédiction sous forme cartographique, constituant ainsi une ou plusieurs cartes complémentaires à la carte des prédictions. Ces erreurs peuvent être calculées directement à partir du modèle mathématique (comme dans le cas des GLM et des GAM) ou estimées par des procédures, plus complexes de ré-échantillonnage des données (le *bootstrap*) et plus coûteuses en temps de calcul.

L'évaluation quantitative est faite en confrontant les prédictions aux observations réelles et en calculant un ou plusieurs indices permettant de mesurer la force de cette corrélation, si possible sur un jeu de données indépendant de celui ayant servi à construire le modèle. Différents indices peuvent être utilisés selon le type de la variable réponse. L'exemple le plus fréquent est celui des modèles de présence-absence. Dans ce dernier cas, les prédictions probabilistes (soit des valeurs allant de 0 à 1) doivent être ramenées à des données de présence-absence (des valeurs de 0 ou 1 exactement) en utilisant un seuil de reclassification. Les valeurs inférieures au seuil (p. ex. 0.5) sont recodées en 0 et les valeurs supérieures au seuil sont recodées en 1. Le choix de ce seuil peut être arbitraire ou optimisé en utilisant des seuils successifs (entre 0 et 1; p. ex. tous les 0.01). Pour chaque seuil, les prédictions recodées en 0/1 sont ensuite comparées aux observations réelles, prenant elles aussi des valeurs 0/1, dans un tableau croisé à deux entrées (tab. 1). Quatre cas de figures peuvent alors être observés: a) une présence est bien prédite, b) une absence est faussement prédite comme présence, c) une présence est faussement prédite comme absence, d) une absence est bien prédite. Le tableau 1 énumère le nombre de cas correspondant à chaque situation. Les deux seuls types d'erreurs possibles sont donc les situations b et c.

¹² Une mesure classique de la qualité de l'ajustement d'un modèle dans le cas des régressions multiples est le coefficient de détermination (R^2) qui donne une mesure de la variation expliquée par le modèle.

Tableau 1 : Table de contingence (ou matrice de confusion) à deux entrées, permettant la comparaison entre valeurs observées et valeurs prédites recodées en présence/absence à un seuil de probabilité donné. Pour la signification de a, b, c et d, se référer au texte.

| | | | |
|-------------------------|---|--------------------------|---|
| | | <i>Valeurs observées</i> | |
| | | 1 | 0 |
| <i>Valeurs prédites</i> | 1 | a | b |
| | 0 | c | d |

$N = a+b+c+d$

Différentes mesures peuvent être calculées à partir d'un tel tableau, comme le pourcentage de valeurs bien prédites, calculé par le rapport $(a + d)/N$, la sensibilité, calculée par le rapport $a/(a + c)$, la spécificité, calculée par $d/(b + d)$ ainsi que de nombreuses autres mesures similaires ou plus complexes (p. ex. Kappa), dont certaines permettent d'attribuer un poids différents à chaque type d'erreur (p. ex. Kappa pondéré), ce qui peut être souhaitable selon le type d'organisme considéré si le risque de prédire faussement une absence n'est pas jugé équivalent à celui de prédire faussement une présence. D'autres mesures, plus objectives car indépendantes du choix d'un seuil, peuvent être obtenues en intégrant les valeurs de certaines mesures sur l'ensemble des seuils considérés pour obtenir, finalement, une seule valeur synthétique d'évaluation, à l'exemple de l'approche graphique ROC qui combine sensibilité et spécificité (mesure prenant des valeurs entre 0.5 = prédiction nulle et 1 = prédiction parfaite; voir Fielding & Bell, 1997).

Ces mesures d'évaluation quantitatives s'avèrent particulièrement utiles lorsqu'on souhaite comparer entre eux les modèles de différentes espèces, ou pour comparer plusieurs modèles de la même espèce incluant différentes combinaisons de prédicteurs environnementaux.

Quelles applications pour ces modèles?

Il existe de nombreuses applications à ces modèles, permettant de fournir des réponses à différentes questions en écologie spatiale et en biogéographie. Trois cas d'études sont présentés ci-dessous. Ils concernent: (i) le test d'une hypothèse biogéographique, (ii) l'impact des changements climatiques sur la distribution des espèces et (iii) la gestion d'espèces rares et menacées.

Une première application de cette approche de modélisation consiste à tester des hypothèses en biogéographie. L'intérêt ne se porte donc pas, ici, sur la production finale de cartes de distribution potentielle, mais sur l'interprétation écologique des modèles statistiques eux-mêmes et le pouvoir déductif que l'on peut en tirer. Une hypothèse encore très discutée en biogéographie européenne est celle de l'existence de refuges de haute altitude (les nunataks) pour certaines espèces durant les périodes glaciaires. Elle s'oppose à l'hypothèse alternative de *tabula rasa* stipulant l'émigration aux plus basses latitudes et/ou l'extinction de certaines espèces au moment des glaciations puis une re-colonisation post-glaciaire des régions concernées. Il peut en effet s'avérer difficile d'expliquer certains types de distribution ou la rapidité de re-colonisation d'un certain nombre d'espèces sans avoir recours à l'existence de tels refuges glaciaires. Dans les Alpes en particulier, une chaîne de montagne orientée d'est en ouest et ayant constitué une barrière sévère aux migrations de nombreuses espèces végétales et animales durant les glaciations, la question de l'existence – encore fortement contestée par certains auteurs – de ces

nunataks semble aujourd'hui confortée par des analyses génétiques (p. ex. Stehlik et al., 2001). Au contraire des Alpes, les montagnes norvégiennes sont orientées selon un axe nord-sud et n'ont ainsi pas constitué de véritable barrière à la migration. En 1993, H.J.B. Birks montra, par une étude statistique faisant recours à des techniques similaires aux procédures de modélisation présentées précédemment, que la distribution actuelle d'un grand nombre d'espèces dans ces montagnes scandinaves pouvait être largement expliquée au seul moyen des variables climatiques, géologiques et topographiques. Par une approche purement statistique et déductive, il montra ainsi que la distribution actuelle des taxons végétaux dans ces montagnes ne nécessite aucun recours à l'existence de refuges glaciaires. Dans son étude, ces derniers étaient simulés par une carte numérique des zones théoriquement non-recouvertes de glace durant les périodes froides et testée par rapport aux autres variables dans ses modèles de distribution d'espèces. Des études de génétique moléculaire appliquées à la phylogéographie¹³ de ces régions permettront probablement de valider ou d'invalider la valeur déductive de ces résultats.

Une autre question, plus brûlante d'actualité, est de savoir comment la distribution des espèces pourrait être affectée, dans le futur, par les changements climatiques en cours¹⁴ (voir Rebetz, 2002). Une bonne illustration de l'importance économique d'une telle question est de savoir comment la distribution des vignobles pourrait être affectée ou favorisée dans le futur (Kienast & Brzeziecki, 1994). Le fait de changer les paramètres climatiques (par ex. une augmentation de température) dans un modèle prédictif de distribution permet justement de suggérer de tels scénarios d'impact et, ainsi faisant, de fournir un premier élément de réponse. Les résultats montrent non seulement les positions futures des zones d'habitats favorables à l'espèce modélisée, telles qu'elles pourraient être modifiées par un réchauffement climatique, mais également les zones d'habitat actuellement favorables qui pourraient devenir défavorables (fig. 8). En montagne, une telle situation pourrait être causée par l'exclusion d'une espèce de certaines de ses positions actuelles (notamment à sa limite altitudinale inférieure) par des espèces plus compétitives provenant d'altitudes inférieures au bénéfice d'un adoucissement progressif du climat. Des études de sensibilité peuvent être menées de la même manière en faisant varier l'intensité du changement climatique dans le modèle (p. ex. différents niveaux de réchauffement) et en visualisant les effets correspondants au niveau de la distribution de l'espèce. Une telle démarche a par exemple été appliquée dans les Alpes valaisannes sur 62 espèces végétales alpines et subalpines (Guisan & Theurillat, 2001), en considérant trois scénarios de réchauffement: +1,5 °C, +3 °C et +4,5 °C (fig. 8).

Les résultats ont montré qu'un impact majeur sur la distribution de nombreuses espèces végétales pouvait être attendu, même dans le cas du réchauffement le plus faible. Selon ces scénarios, les espèces alpines et nivales devraient être plus menacées d'extinction locale que les espèces subalpines ou que celles distribuées jusqu'à plus basse altitude. En effet, une analyse numérique des surfaces couvertes par les différents étages de végétation a montré que la surface totale actuellement occupée par l'étage subalpin serait seulement légèrement réduite après un réchauffement modéré alors que les surfaces occupées par les étages alpin et nival pourraient être très fortement réduites (Theurillat & Guisan, 2001). Des prédictions similaires pour les espèces végétales ont été obtenues à l'échelle de l'Europe entière (voir par exemple Bakkenes et al., 2002).

Le dernier exemple d'application touche à la problématique de la conservation de la nature. Un des problèmes majeur de la gestion des espèces rares et

¹³ La phylogéographie est l'étude des relations entre l'histoire évolutive des taxons (éclairé notamment par les récents résultats de la génétique moléculaire) et leur distribution géographique.

¹⁴ La réalité de ces changements est aujourd'hui officiellement reconnue par la commission intergouvernementale sur les changements climatiques (IPCC), une émanation de l'ONU; voir <http://www.ipcc.ch>.

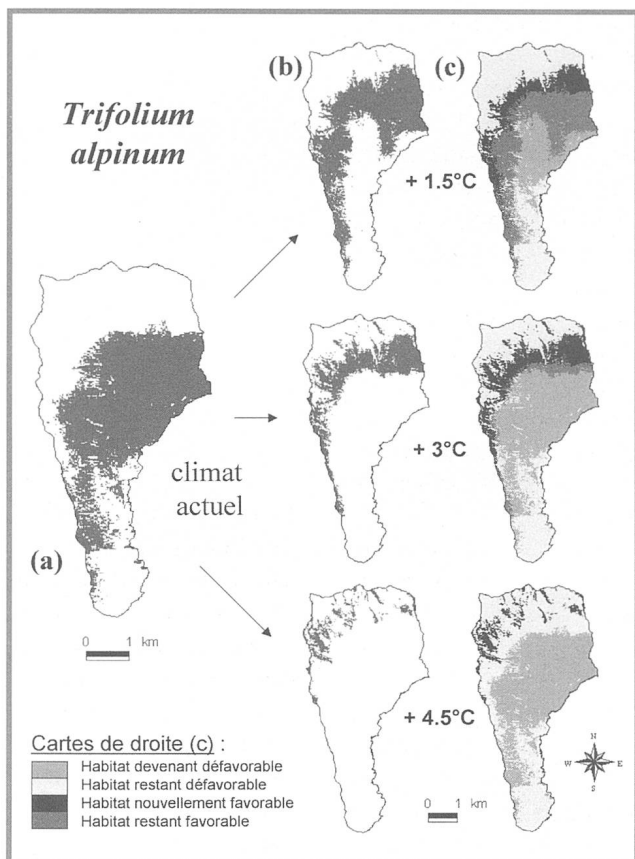


Figure 8: Cartes de la distribution potentielle de *Trifolium alpinum* L. dans la région de Belalp (VS, Suisse), (a) sous le climat actuel et (b) dans l'hypothèse de trois scénarios de changement climatique différents; les zones grisées indiquent les sites ayant une probabilité supérieure à 50 % d'abriter l'espèce; (c) détails des scénarios montrant les quatre situations possibles (voir légende intégrée dans la figure) pour chaque unité cartographique modélisée après le réchauffement correspondant; (d) *Trifolium alpinum* L., une Fabacée de couleur rose.

menacées est de connaître la position géographique d'un nombre aussi grand que possible de populations (idéalement toutes), de manière à pouvoir évaluer objectivement le statut de menace de l'espèce et gérer ses populations de manière appropriée. Cette situation idéale est loin d'être celle qui prévaut actuellement. Jusqu'à présent, une partie non négligeable des observations stockées dans les banques de données floristiques nationales, comme le Centre du réseau suisse de floristique (CRSF) en Suisse, a été fournie par des naturalistes bénévoles. Les espèces rares n'ayant généralement pas fait l'objet d'une recherche systématique, il n'est dès lors pas impossible que d'autres stations existantes pour une espèce soient absentes de ces inventaires. Ces stations manquantes pourraient cependant s'avérer difficiles à localiser sur le terrain sans le soutien d'une analyse préalable du paysage permettant d'identifier les sites favorables pouvant éventuellement abriter une population non répertoriée de l'espèce concernée. Des étudiants de mon groupe à l'Université de Lausanne ont ainsi récemment évalué, en collaboration avec le CRSF (Beat Bäumler) et l'Office fédéral de l'Environnement, des Forêts et du Paysage (OFEFP), la possibilité d'utiliser de tels modèles prédictifs pour suggérer de nouveaux sites d'échantillonnage pour les espèces rares. Deux espèces menacées et relativement rares ont été choisies pour des essais préliminaires: L'ancolie des Alpes (*Aquilegia alpina*; fig. 9) à l'échelle des cantons de Vaud et du Valais (Febbraro & Broennimann, 2002) et le panicaud des Alpes ou chardon bleu (*Eryngium alpinum*; fig. 10) à l'échelle suisse (Engler & Rechsteiner, 2002). Cette dernière espèce fait également l'objet d'études génétiques aux Conservatoire et Jardin botaniques de la Ville de Genève (CJB, Dr Y. Naciri-Graven). Sur ces bases, un nouveau travail a été effectué pour généraliser la méthode et permettre ainsi de



Figure 9: *Aquilegia alpina* L.
(photo F. Clot).

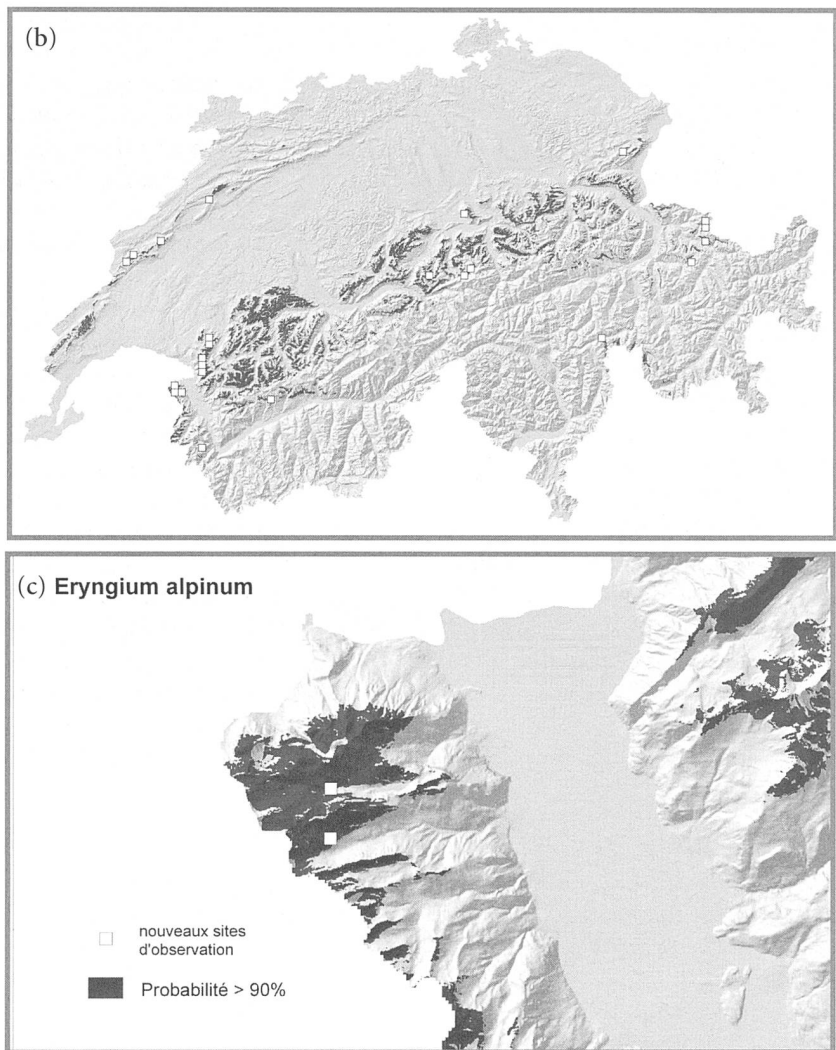
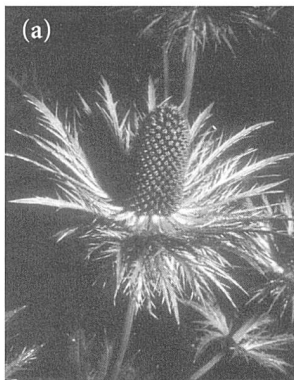


Figure 10: (a) *Eryngium alpinum* (photo F. Clot). (b) Carte de distribution potentielle d'*Eryngium alpinum* en Suisse modélisée par régression multiple (GLM) à une résolution de 25 m (Broennimann, 2003). Seule les probabilités d'observer l'espèce supérieure à 90 % sont représentées ici en noir. Les points de présence de l'espèce, agrégées par maille de 3x3 km, sont superposés à la carte. Fond de carte: relief ombré, calculé à partir du modèle numérique de terrain (MNT25, © 1996, Office fédéral de topographie). Les six plus importants prédicteurs inclus dans le modèle sont ici (par ordre d'importance): (i) la température moyenne de juillet, (ii) la somme des températures moyennes supérieures à 3 °C (degree-days) (iii) un indice de précipitation traduisant la sécheresse estivale, (iv) le bilan hydrique du sol, (v) le nombre de jours de gel par année, (vi) le nombre de jours de précipitation par année. Les prédictions du modèle ont ensuite été filtrées par les classes de variables environnementales qualitatives (p. ex. la géologie) sur lesquelles l'espèce n'est jamais observée. (c) Agrandissement d'une partie de la carte précédente, illustrant l'emplacement de deux nouvelles stations pour *Eryngium alpinum*.

simuler la distribution d'un plus grand nombre d'espèces rares et menacées en Suisse (Broennimann, 2003). Les cartes de distribution potentielle obtenues se sont révélées être de bonnes bases sur lesquelles poursuivre l'étude de ces espèces, comme attesté notamment sur le terrain par la découverte de deux nouvelles stations d'*E. alpinum*, préalablement suggérées par un premier modèle puis utilisées pour construire un second modèle, plus précis (fig. 10c).

Quelles perspectives pour l'avenir?

De grands progrès restent à faire pour améliorer la fiabilité de ces modèles, actuellement très variables d'une espèce à l'autre (de moins de 10 % à plus de 90 % de variation spatiale expliquée), notamment par l'obtention de meilleurs prédicteurs environnementaux et à travers la possibilité de conduire, dans un très proche avenir, ce type d'études à une résolution plus fine que l'actuelle (par exemple 1 m au lieu de 25 m). Il serait également vain de vouloir améliorer la qualité des prédictions sans recourir à de nouvelles études de terrain assurant la collecte de nouvelles données selon des protocoles d'échantillonnage plus adaptés à l'objectif final de modélisation (Hirzel et Guisan, 2002).

Parmi les nombreuses autres perspectives intéressantes, il me semble important de citer le couplage de ces modèles spatiaux avec des modèles dynamiques, intégrant la dispersion des espèces dans le paysage ainsi que la succession des espèces et la dynamique de leur population en chaque site.

Cette double approche, spatio-temporelle, devrait permettre de générer des modèles de distribution des espèces plus réalistes, en particulier dans le contexte des changements climatiques.

Un autre défi important sera celui d'appliquer plus activement ces modèles à la conservation de la nature, notamment en intégrant les résultats de la recherche en génétique des populations au sein de cette approche spatiale. Deux directions, parmi de nombreuses autres, actuellement explorées par le laboratoire SIG de l'Institut d'Ecologie (<http://www.unil.ch/iesig>) et le Laboratoire de Biologie de la Conservation (LBC; <http://www.unil.ch/lbc>) de l'Université de Lausanne.

Remerciements

A Niklaus E. Zimmermann, Pascal Vittoz et André Schlüssel pour leurs remarques critiques et constructives sur une première version du manuscrit. A Olivier Broennimann pour la préparation des figures 9c et 9d. Et à ma femme, Saskia Weber Guisan, pour sa relecture attentive du manuscrit.

Références

- AUSTIN, M.P. (1987). Models for the analysis of species' response to environmental gradients. *Vegetatio*, 69, p. 35 - 45.
- AUSTIN, M.P. (2002). Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, 157(2-3), p. 101 - 118.
- BAKKENES, M., J. R. M. ALKEMADE, F. IHLE, R. LEEMANS & J. B. LATOUR (2002). Assessing the effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. *Global Change Biology*, 8, p. 390 - 407.
- BINZ, H.R. & O. WILDI (1988). *Das Simulationsmodell MaB-Davos*. Schlussbericht. Schweiz. MaB-Programm Nr. 33, Bundesamt für Umweltschutz, Bern.
- BIRKS, H.J.B. (1993). Is the hypothesis of survival on glacial nunataks necessary to explain the present-day distributions of Norwegian mountain plants? *Phytocoenologia*, 23, p. 399 - 426.
- BOX, E.O. (1981). *Macroclimate and Plant Forms: An Introduction to Predictive Modeling in Phytogeography*, Junk, The Hague.
- BROENNIMANN, O. (2003). *Modéliser la distribution potentielle des espèces végétales rares et menacées*. Travail de diplôme, Université de Lausanne.

- BROWN, J.H., G. C. STEVENS & D. M. KAUFMAN (1996). The geographic range: size, shape, boundaries, and internal structure. *Annual Review of Ecology and Systematics*, 27, p. 597 - 623.
- BRZEZIECKI, B., F. KIENAST & O. WILDI (1993). A simulated map of the potential natural forest vegetation of Switzerland. *Journal of Vegetation Science*, 4, p. 499 - 508.
- DE CANDOLLE, A.P. (1855). *Géographie botanique raisonnée*. Masson, Paris.
- ELITH, J. (2002). *Predicting the distribution of plants*. PhD thesis, University of Melbourne, Australia.
- ENGLER, R. & L. RECHSTEINER (2002). *Distribution potentielle de Eryngium alpinum en Suisse*. Travail de certificat en botanique, Institut d'Ecologie, Université de Lausanne.
- FEBBRARO, I. & O. BROENNIMANN (2002). *Suggestion de nouveau sites d'échantillonnages pour Aquilegia alpina, une espèce menacée: une approche par modélisation*. Travail de certificat en botanique, Institut d'Ecologie, Université de Lausanne.
- FIELDING, A.H. & J. F. BELL (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24(1), p. 38 - 49.
- FISCHER, H.S. (1990). Simulating the distribution of plant communities in an alpine landscape. *Coenoses*, 5(1), p. 37 - 43.
- FRANKLIN, J. (1995). Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, 19(4), p. 474 - 499.
- GUISAN, A. (2002). Semi-quantitative response models for predicting the spatial distribution of plant species, p. 315 - 326. In: J. M. SCOTT, P. J. HEGLUND, J. B. HAUFLE, M. MORRISON, M. G. RAPHAEL, W. B. WALL & F. SAMSON (eds.), *Predicting species occurrences: issues of accuracy and scale*, Chapter 25, Island Press, Covelo, California.
- GUISAN, A., T. C. EDWARDS & T. HASTIE (2002). Generalized linear and generalized additive models in studies of species distribution: setting the scene. *Ecological Modelling*, 157(2-3), p. 89 - 100.
- GUISAN, A., J.-P. THEURILLAT & F. KIENAST (1998). Predicting the potential distribution of plant species in an alpine environment. *Journal of Vegetation Science*, 9, p. 65 - 74.
- GUISAN, A. & J.-P. THEURILLAT (2001). Assessing alpine plant vulnerability to climate change: A modeling perspective. *Integrated Assessment*, 1, p. 307 - 320.
- GUISAN, A. & N. E. ZIMMERMANN (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135(2-3), p. 147 - 186.
- HAUSSER, J. (ed.). (1995). *Mammifères de la Suisse*. Birkhäuser Verlag, Bâle.
- HILL, M.O. (1991). Patterns of species distribution in Britain elucidated by canonical correspondence analysis. *Journal of Biogeography*, 18, p. 247 - 255.
- HIRZEL, A. & A. GUISAN (2002). Which is the optimal sampling strategy for habitat suitability modelling? *Ecological Modelling*, 157(2-3), p. 331 - 341.
- HOFER, U., J.-C. MONNEY & G. DUSEJ (eds). (2001). *Reptiles de la Suisse*. Birkhäuser Verlag, Bâle.
- HUMBOLDT, A. VON & A. BONPLAND (1807). *Essai sur la géographie des plantes*. Masson, Paris.
- HUTCHINSON, G. E. 1957. Concluding remarks. Cold Spring Harbor Symp. *Quant. Biol.*, 22, p. 415 - 427.
- JABERG, C. & A. GUISAN (2001). Modelling the distribution of bats in relation to landscape structure in a temperate mountain environment. *Journal of Applied Ecology*, 38(6), p. 1169 - 1181.

- JEANMONOD, D. & C. LAMBELET (2002). Alerte à l'ambrosie! *Saussurea*, 32, p. 132 - 133.
- KESSELL, S.R. (1979). *Gradient modeling: resource and fire management*. Springer Verlag, New-York.
- KIENAST, F. & B. BRZEZIECKI (1994). D'où proviendra le vin de notre pays si les températures augmentent de 2 °C? *Informations du secteur de recherche écologie du paysage*, 22.
- LEATHWICK, J.-R. (1998). Are New-Zealand's *Nothofagus* species in equilibrium with their environment? *Journal of Vegetation Science*, 9, p. 719 - 732.
- MCARTHUR, R.H. (1972). *Geographical Ecology: Patterns in the Distribution of Species*. Harper and Row, New York, NY.
- REBETEZ, M. (2002). *La Suisse se réchauffe*. Presses Polytechniques et Universitaires Romandes, Lausanne.
- SANDERS, I. (2002). Ecology and evolution of multigenomic arbuscular mycorrhizal fungi. *American Naturalist*, 160, p. 128 - 141.
- SCOTT, J. M., P. J. HEGLUND, J. B. HAUFLER, M. MORRISON, M., M. G. RAPHAEL, W. B. WALL & F. SAMSON (eds.) (2002). *Predicting species occurrences: issues of accuracy and scale*. Island Press, Covelo, California.
- STEHLIK, I., J. I. SCHNELLER & K. BACKMANN (2001). Resistance or emigration: response of high alpine plant *Eritrichium nanum* (L.) Gaudin to ice age within the Central Alps. *Molecular Ecology*, 10, p. 357 - 370.
- THEURILLAT, J.-P. & A. GUISAN (2001). Potential impact of climate change on vegetation in the European Alps: A review. *Climatic Change*, 50, p. 77 - 109.
- THORNTON, P.E., S. W. RUNNING & M. A. WHITE (1997). Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of Hydrology*, 190(3-4), p. 214 - 251.
- WALTER, H. (1985). *Vegetation of the Earth and ecological systems of the geobiosphere*. 3 Ed. Springer, Heidelberg.
- WOHLGEMUTH, T. (1998). Modelling floristic species richness on a regional scale: a case study in Switzerland. *Biodiversity and Conservation*, 7, p. 159 - 177.
- ZIMMERMANN, K. & F. KIENAST (1995). Das Klima lässt sich nicht kartieren - Klimakarten werden gerechnet. *Informationsblatt des Forschungsbereiches Landschaftsökologie, Eidg. Forschungsanstalt für Wald, Schnee und Landschaft, Birmensdorf*, 27, p. 1 - 4.
- ZIMMERMANN, N. E. & F. KIENAST (1999). Predictive mapping of alpine grasslands in Switzerland: species versus community approach. *Journal of Vegetation Science*, 10, p. 469 - 482.

Texte et illustrations:
Antoine Guisan

Institut d'Ecologie, Université
de Lausanne
Bâtiment de biologie,
CH-1015 Lausanne, Suisse

Antoine.guisan@ie-bsg.unil.ch