

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

FONCTIONS DE DIVERSITÉ APPLIQUÉES AUX ÉCOSYSTÈMES :
SYNTHÈSE ET RÉCONCILIATION DE LA LITTÉRATURE ÉCONOMIQUE

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN ÉCONOMIQUE

PAR
MAXIME LEFEBVRE

MARS 2007

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Je tiens tout d'abord à remercier mes parents, Nicole et Guy et mon frère Marc-Alexandre qui ont toujours cru en moi et qui m'ont encouragé tout au long de mes études. J'aimerais ensuite remercier Monsieur Pierre Lasserre, mon directeur, pour son soutien et la qualité de son encadrement durant la rédaction de ce mémoire. J'aimerais aussi le remercier pour m'avoir incité à présenter ce travail lors du 46^e Congrès annuel de la Société canadienne de science économique. Je tiens finalement à saluer tous mes amis que je côtoie depuis plusieurs années.

TABLE DES MATIÈRES

LISTE DES FIGURES ET TABLEAUX	v
RÉSUMÉ	vi
INTRODUCTION	1
CHAPITRE I LE CADRE GÉNÉRAL DE LA THÉORIE DE LA DIVERSITÉ : L'APPROCHE HIÉRARCHIQUE	4
1.1 Mesure de diversité basée sur la notion de distances entre écosystèmes	6
1.1.1 Exemple 1 : non-unicité de la mesure de diversité	9
1.2 Fonction de diversité	10
1.3 Fonction de diversité alternative	12
1.3.1 Algorithme sous jacent à la fonction de diversité	14
1.3.2 Exemple 2 : calcul de la fonction de diversité $V(S)$	14
1.4 Transformation des distances en distances ultramétriques	16
1.5 Représentation graphique d'un ensemble d'écosystèmes	18
CHAPITRE II L'APPROCHE LEXICOGRAPHIQUE	20
2.1 Vecteur de distance $D(S)$	20
2.2 Fonction de diversité issue du vecteur de distance	23
2.2.1 Exemple 3 : calcul de la fonction de diversité $U(S)$	24
CHAPITRE III MODÉLISATION DE LA DIVERSITÉ	26
3.1 Mesure de différence de base	28
3.1.1 Interprétation de la Figure 1.1	29

3.2 Mesures de différences normalisées	30
3.3 Mesure basée sur le concept d'endémisme	31
CHAPITRE IV	
DISCUSSION	33
4.1 Caractéristiques des fonctions de diversité	33
4.2 Politiques de préservation	34
CHAPITRE V	
QUELQUES EXTENSIONS ET LIMITES	37
5.1 La diversité des écosystèmes en fonction des espèces ou des gènes ?	37
5.2 Les mesures de diversité des objets multidimensionnels	38
5.3 Les mesures de distance pour objets multidimensionnels	39
5.3.1 Exemple 4 : Modification de la structure de la diversité suite à un changement d'unité	42
5.3.2 La distance euclidienne pondérée	45
5.4 La diversité comme problème d'agrégation : l'exemple de la super-espèce	46
CONCLUSION	52
BIBLIOGRAPHIE	54

LISTE DES FIGURES ET TABLEAUX

	Page
Figure	
1.1 Représentation graphique de l'ensemble S	19
5.1 a) Graphique initial	44
5.1 b) Graphique modifié	44
 Tableau	
1.1 Matrice des distances de l'ensemble S	15
1.2 Matrice des distances ultramétriques de l'ensemble S	17
3.1 Matrice des données relatives à ε_i^x	27
4.1 Valeur des fonctions de diversité	35
5.1 Les principales mesures issues de la distance de Minkowski	40
5.2 Valeurs de deux caractéristiques appliquées à quatre objets	42
5.3 a) Matrice de distance initiale	43
5.3 b) Matrice de distance modifiées	43
5.4 a) Matrice de distance ultramétriques	43
5.4 b) Matrice de distance ultramétriques modifiées	43
5.5 a) Matrice des descripteurs x_j^i de l'écosystème x	48
5.5 b) Matrice des descripteurs y_j^i de l'écosystème y	49
5.5 c) Matrice des descripteurs z_j^i de l'écosystème z	49
5.6 Valeurs des distances issues de la distance de Minkowski selon les formes additive et indicielle de h	50

RÉSUMÉ

En science économique, le problème de la préservation de la biodiversité peut être analysé sous la forme d'un problème d'optimisation sous contrainte budgétaire où la principale difficulté repose sur le choix de la fonction à optimiser. Ce mémoire s'inscrit à la suite des travaux produits par Weitzman (1992), Bossert, Pattanaik et Xu (2001) et Weikard (2002) qui proposent différentes fonctions qui permettent de mesurer la diversité d'un ensemble d'objets. Généralement, nous pouvons considérer la diversité comme étant le résultat d'une certaine forme d'agrégation des différences parmi les objets d'un ensemble. Puisque la biodiversité est un concept plutôt vague, nous envisageons le cas où les écosystèmes représentent les objets à préserver. Le concept de différence, pour sa part, est présenté dans sa forme la plus simple par une mesure de distance entre deux écosystèmes ou entre un écosystème et un ensemble et nous construisons ces distances sur la base de l'information obtenue par la présence ou l'absence d'espèces parmi les différents écosystèmes d'un ensemble. Nous présentons le cadre général permettant de construire de telles fonctions de diversité. En plus de mesurer la valeur d'un ensemble d'écosystème, ces fonctions permettent de déterminer l'apport, en terme de diversité marginale, de chaque écosystème par rapport à l'ensemble considéré. Lorsqu'elles sont appliquées au cas particulier des écosystèmes, celles-ci décrivent certains concepts de diversité généralement utilisés en écologie. Nous présentons finalement quelques politiques de préservation sous-jacentes à ces fonctions de diversité.

Mots clés : diversité marginale, écosystème, fonction de diversité, mesure de distance, préservation de la biodiversité.

INTRODUCTION

La préservation de la biodiversité est un domaine de recherche qui prend de l'ampleur en économie de l'environnement. Cette préservation ne peut se réaliser sans effort; c'est pourquoi nous devons y consacrer une certaine quantité de ressources dans le but d'en assurer la pérennité. Évidemment, nos moyens limités ne nous permettent pas de tout préserver, d'où la nécessité de faire des choix et d'établir des priorités. Pour ce faire, nous devons être en mesure de comparer les différents projets entre eux, généralement sur la base d'un critère d'analyse avantage-coût.

Il est difficile de formuler des politiques de préservation en se basant sur un concept aussi vague que celui de la biodiversité. Nous avons donc besoin d'une définition opérationnelle de ce concept sur laquelle nous baser de façon à proposer des politiques efficaces.

Les écologistes ont développés différentes mesures qui fournissent des indications concernant la variété et l'intégrité de la biodiversité. Cependant, aucune d'entre elles prise isolément ne peut être utilisée pour décrire ce concept; c'est pourquoi le choix des mesures à utiliser devra dépendre de l'objectif recherché (Eiswerth et Haney, 2001). Généralement, les politiques de préservation sont formulées de telle sorte qu'on préserve un certain sous-ensemble de la biodiversité parmi tous les éléments mis à notre disposition.

Plusieurs approches ont été proposées dans le but de sélectionner des ensembles d'écosystèmes permettant de couvrir le plus grand nombre d'espèces possible. Citons par exemple les travaux de Margules, Nicholls et Pressey (1988) et de Church, Stoms et Davis (1996) qui utilisent de modèles de programmation numérique. L'idée derrière ces approches consiste généralement à déterminer le plus petit nombre de sites permettant de représenter le plus d'espèces possibles. Ce genre de modèle ne met pas directement l'emphase sur la détermination d'une valeur pour la biodiversité mais cherche plutôt à identifier un sous-ensemble représentatif d'une certaine région.

D'un point de vue économique, les principaux travaux concernant les mesures de la biodiversité en lien avec notre approche sont ceux de Solow, Polasky et Broadus (1993), Nehring et Puppe (2002) et plus particulièrement ceux de Weitzman (1992), Bossert, Pattanaik et Xu (2001) et Weikard (2002) que nous analyserons plus en profondeur. Ces différentes approches sont toutes construites sur la base de fonctions de diversité.

Dans ce mémoire, nous présentons quelques unes de ces fonctions en les appliquant au cas particulier de la biodiversité. Celles-ci mesurent chacune à leur façon la diversité comme étant le résultat d'une certaine forme d'agrégation des différences entre chacun des éléments d'un ensemble. Bien que nous envisageons d'appliquer ces fonctions au niveau de la biodiversité, elles peuvent être utilisées dans de multiples contextes; pour la diversité culturelle, linguistique ou artistique, en architecture, en sociologie, etc.

En fait, même en l'appliquant à notre sujet, ces différentes approches peuvent être employées à plusieurs niveaux. Weitzman (1992, 1993) mesure la valeur d'un ensemble d'espèces sur la base de leur diversité génétique alors que Weikard (2002) envisage plutôt de mesurer la diversité d'un ensemble d'écosystème en tenant compte des espèces qui les composent.

Pour des considérations pratiques, mais surtout parce qu'il s'agit d'une approche largement répandue chez les écologistes, nous porterons notre attention, tout comme Weikard (2002), sur les espèces pour définir les différentes unités de base de la biodiversité. En effet, elles sont à plusieurs égards de bons indicateurs, quoique incomplet, de la variété et de l'intégrité des écosystèmes. D'autre part, puisque la destruction des habitats représente une des principales cause de la détérioration de la biodiversité et puisqu'il est généralement admis que la meilleure façon de protéger les espèces consiste à préserver les écosystèmes, nous considérons ces derniers comme objets d'étude et nous les comparons entre eux sur la base des espèces qui les composent.

Pour les besoins de notre travail, nous définissons un écosystème comme étant une région géographiquement définie, offrant une certaine diversité biologique et constituant un système viable sans trop d'interaction avec les autres écosystèmes (Weikard, 2002). La diversité biologique, pour sa part, sera décrite uniquement en termes d'espèces. Évidemment, celles-ci peuvent être présentes dans un, plusieurs ou chacun des écosystèmes d'un ensemble. Une espèce est qualifiée d'endémique si elle n'est présente que dans un seul d'entre eux. Dans le

cas contraire, nous la considérons simplement comme étant commune à plusieurs écosystèmes.

Nous montrons qu'il est possible de construire des fonctions de diversité en se servant d'une approche basée sur un concept de différence entre écosystèmes. Pour être plus précis, étant donné une liste d'espèces bien définie, notre objectif consiste à proposer différentes fonctions de diversité à partir de la seule observation de la présence ou de l'absence de chacune des espèces dans chacun des écosystèmes d'un ensemble.

Les deux premières approches que nous présentons proposent un cadre d'analyse permettant de mesurer la diversité en se basant sur une notion de distance entre chacun des éléments d'un ensemble. Ces approches ne font que mentionner les caractéristiques que doivent respecter les mesures de distance sans invoquer la manière de les construire. Les chapitres III et V traitent de ce sujet. Nous introduisons au chapitre III la mesure de différence entre écosystème proposée par Weikard (2002). Le problème avec cette mesure provient du fait qu'elle ne respecte les caractéristiques d'une distance que dans un cas bien particulier. Un des apports de ce mémoire consiste justement à formuler des mesures de différence respectant ces caractéristiques en toute circonstance.

Nous faisons aussi ressortir un aspect intéressant de l'article de Weikard (2002) à savoir, qu'il est possible de formuler une mesure de différence entre écosystème autrement que sur la base de distance en se servant plutôt du concept d'endémisme. Il est intéressant de constater que les trois catégories de fonctions que nous présentons permettent chacune à leur façon de mesurer la diversité à partir d'une même information de base.

La première partie de ce mémoire présente le cadre général permettant de construire diverses fonctions de diversité en se basant sur des mesures de distance entre chaque paire d'écosystème d'un ensemble. Le chapitre I décrit l'approche hiérarchique inspirée de Weitzman (1992) alors que le deuxième chapitre traite d'une approche dite lexicographique (Bossert et al., 2001). Dans le troisième chapitre, nous décrivons notre façon de modéliser la biodiversité ainsi que l'approche basée sur le concept d'endémisme telle que proposée par Weikard (2002). Le chapitre IV présente certaines caractéristiques ainsi que quelques politiques de préservation sous-jacentes à ces différentes fonctions. Avant de conclure, nous présentons quelques extensions et limites au concept de fonction de distance et de diversité.

CHAPITRE I

LE CADRE GÉNÉRAL DE LA THÉORIE DE LA DIVERSITÉ : L'APPROCHE HIÉRARCHIQUE

Ce chapitre présente le cadre général de la théorie de la diversité tel que formulé par Weitzman (1992). Nous décrivons l'approche utilisée par ce dernier pour définir une fonction de diversité construite sur la base de mesures de distance entre chaque paire d'éléments d'un ensemble. Avant d'aller plus loin, présentons brièvement quelques uns des concepts de base utilisés en écologie pour mesurer la diversité d'un ensemble d'écosystème. L'idée derrière chacune des fonctions présentées dans ce mémoire consiste justement à décrire chacune à leur manière ces différents concepts.

Un des points de départ généralement proposé par les écologistes consiste souvent à représenter la diversité d'un ensemble d'écosystème à l'aide des mesures de diversité α , β et γ (voir Vane-Wright, Humphries et Williams (1991), Maignan et al. (2003) et Armsworth, Kendall et Davis (2004)).

- α représente la diversité d'un écosystème pris isolément. Ce type de diversité fait souvent référence à ce que les écologistes appellent la notion de richesse, qui peut se définir comme étant le nombre d'espèces différentes recensées dans un site. Cet indice est d'autant plus élevé que le nombre d'espèces y est grand. Cette notion est souvent utilisée comme une première approximation du niveau de la diversité biologique d'un écosystème.
- β mesure le niveau de disparité entre écosystèmes ou en d'autres termes, le degré d'hétérogénéité d'un ensemble. Ce concept permet de mettre l'emphase sur les groupes d'espèces endémiques. La valeur de cet indice est élevée si les écosystèmes ont un haut niveau d'endémisme alors qu'elle est faible s'ils sont composés sensiblement des mêmes espèces.

- γ fait référence à la diversité totale mesurée parmi tous les écosystèmes d'un ensemble. Elle peut être obtenue directement en y dénombrant le nombre d'espèces ou comme étant le résultat d'une certaine fonction des diversités α et β .

Comme nous venons de le mentionner, nous verrons qu'il est possible de construire différentes fonctions de diversité inspirées de ces trois concepts en se servant d'approches basées sur une notion de différence entre écosystèmes.

Dans le cas qui nous concerne, si nous voulons identifier un ensemble d'écosystèmes à préserver parmi ceux mis à notre disposition, nous devons tout d'abord être en mesure de définir une relation permettant de les comparer entre eux. Posons X pour identifier un ensemble d'écosystèmes. Ceux-ci peuvent être combinés entre eux de façon à former différents sous-ensembles, notés S . Ω représente l'ensemble des sous-ensembles S de X . Un décideur public s'intéressera donc à un de ces ensembles S , composé de n écosystèmes, sur lequel il a juridiction.

En partant de l'idée qu'il est souhaitable de préserver un maximum de biodiversité parmi les écosystèmes à notre disposition, Weikard (2002) suggère de comparer différents ensembles à l'aide de la relation de préférence suivante :

Pour tout $S \in \Omega$ et $S' \in \Omega$, $S \succeq S'$ si et seulement si S offre au moins autant de diversité que S' .

Pour être en mesure d'établir une telle relation, nous devons disposer d'un outil permettant de comparer ces différents ensembles. Weikard propose donc de définir une fonction de diversité de la manière suivante :

Une fonction de diversité est une fonction $V : \Omega \rightarrow \mathbb{R}_0^+$ telle que $V(S) \geq V(S')$ si et seulement si $S \succeq S'$ pour tout $S \in \Omega$ et $S' \in \Omega$.

Cette définition est plutôt vague mais elle nous permet d'identifier l'objectif recherché lors de l'élaboration d'un tel problème; à savoir, définir une fonction, notée $V(S)$,

permettant de mesurer la diversité de n'importe quel ensemble S de telle sorte qu'elle permet d'établir nos préférences.

Pour les besoins de notre analyse, une telle fonction devra dépendre du nombre d'écosystèmes inclus dans S , de leur diversité décrite en termes d'espèces ainsi que du niveau de différence entre chacun d'eux. En fait, la diversité de S devra être d'autant plus élevée que le nombre d'écosystème est grand, que ceux-ci sont composés d'un grand nombre d'espèces et que ces écosystèmes sont très diversifiés les uns par rapport aux autres. Sans faire l'inventaire de toutes les combinaisons possibles entre le nombre d'écosystèmes, leur diversité et leur niveau de différence, nous voyons clairement qu'une fonction inspirée de ces principes peut être utile pour comparer différents ensembles.

1.1 Mesure de diversité basée sur la notion de distances entre écosystèmes

Nous présentons ici le cadre d'analyse permettant de construire une fonction de diversité en se basant uniquement le niveau de différence entre toutes paires d'objets d'un ensemble. Pour définir ce concept de différence, Weitzman (1992) propose de se servir de la notion mathématique de distance. Ainsi, la différence entre toute paire d'objets peut être donnée par n'importe quelle mesure de distance $d : X \times X \rightarrow \mathbb{R}_0^+$ respectant :

$$d(x, x) = 0, \quad (1.1)$$

$$d(x, y) = d(y, x) \geq 0. \quad (1.2)$$

La fonction d représente une mesure de distance non-négative et symétrique exprimant le niveau de différence, ou de disparité, entre n'importe quel écosystème x et y appartenant à X . Nous allons voir aux chapitres III et V qu'il est possible de proposer différentes mesures de distances respectant les expressions (1.1) et (1.2). Notons par ailleurs que notre modélisation de la diversité au troisième chapitre fait en sorte que la notion de symétrie n'est pas toujours respectée. Nous reviendrons plus loin sur cet aspect.

Étant donné l'expression (1.2), le problème consiste à définir une mesure de la diversité d'un ensemble S en se servant des $n(n-1)/2$ mesures de distance de cet ensemble.

Cette mesure doit être construite de manière à ce qu'elle reflète l'apport de chaque écosystème par rapport à l'ensemble considéré. Il est donc nécessaire de proposer une mesure de la différence entre un écosystème et un ensemble construite à partir des distances préalablement définies. Ce concept sera décrit par une fonction $\delta : X \times \Omega \rightarrow \mathbb{R}_0^+$ représentant la diversité marginale d'un écosystème $x \in X \setminus S$ par rapport à S . Pour définir cette mesure, Weitzman (1992) propose d'utiliser la définition mathématique de la distance entre un objet et un ensemble :

$$\delta(x, S) = d(x, S) \equiv \min_{y \in S} d(x, y). \quad (1.3)$$

La diversité marginale de x par rapport à S est donnée par la distance minimale entre cet écosystème et l'ensemble considéré, c'est-à-dire la distance entre x et l'écosystème y qui lui est le plus semblable parmi ceux de S .

D'après ce que nous avons mentionné précédemment, la valeur d'un ensemble doit être le résultat d'une certaine forme d'agrégation des différences entre écosystèmes. En d'autre terme, une fonction de diversité doit être construite de manière à ce que la diversité augmente lorsqu'un écosystème x est ajouté à S et cette augmentation doit dépendre de la diversité marginale de x par rapport à l'ensemble considéré. En ce sens, Solow, Polasky et Broadus (1993) proposent trois caractéristiques essentielles que toute fonction doit respecter (voir aussi Sinclair-Desgagné (2005)). Celles-ci sont décrites par les axiomes suivants :

$$\text{MONOTONICITÉ : si } S' \text{ est un sous-ensemble de } S, \text{ alors } V(S') \leq V(S). \quad (1.4)$$

$$\text{ACUITÉ : } V(S \cup x) = V(S) \text{ si et seulement si } \delta(x, S) = 0. \quad (1.5)$$

$$\text{CONTINUITÉ : } V(S \cup x) \text{ doit être une fonction continue de } \delta(x, S). \quad (1.6)$$

Plusieurs mesures peuvent respecter ces axiomes. Un premier candidat consiste à poser, pour tout $x \in X \setminus S$ et pour tout $S \in \Omega$, la mesure de diversité suivante :

$$V(S \cup x) = V(S) + \delta(x, S), \quad (1.7)$$

La valeur d'un ensemble S auquel nous ajoutons un écosystème x est donnée par la valeur de S additionné de la diversité marginale de x par rapport à S . L'expression (1.7) est construite par récurrence en commençant par les plus petits ensembles S jusqu'aux plus grands. Il s'agit donc de calculer la valeur de chaque écosystème pris isolément, puis de chaque paire en se servant des résultats obtenus à l'étape précédente, puis de chaque triplet et ainsi de suite jusqu'à ce que le processus récursif soit complété. Évidemment, pour que la valeur de cette mesure soit unique, il est nécessaire que la valeur obtenue lors de chacune des étapes intermédiaires le soit aussi.

Puisque la résolution de cet algorithme débute par le calcul de la valeur de chaque écosystème pris isolément, nous devons préalablement imposer pour tout $x \in X$, la condition initiale suivante :

$$V(x) \equiv d_0 \quad (1.8)$$

où d_0 est une certaine constante non-négative. Cette condition nous dit simplement que la valeur d'un écosystème peut être donnée par n'importe quelle valeur supérieure ou égale à zéro, laquelle doit être identique pour tout écosystème.

En calculant la mesure de diversité inspirée des expressions (1.7), (1.8) et (1.3), on se rend compte qu'une telle approche ne permet pas toujours d'obtenir une valeur unique, ce qui implique que cette mesure n'est pas toujours bien définie. C'est ce que nous illustrons à l'aide de l'exemple suivant.

1.1.1 Exemple 1 : non-unicité de la mesure de diversité

Considérons le cas d'un ensemble $X = \{1, 2, 3\}$ où les distances sont données par $d(1, 2) = 1$, $d(1, 3) = 2$ et $d(2, 3) = 3$. Définissons la condition initiale de manière à ce que $V(x) = 5 \forall x \in X$. La première étape de l'algorithme consiste à calculer (1.7) à l'aide de (1.3) pour tout x et pour tout S tel que $|S| = 1$, c'est-à-dire pour $S = \{1\}$, $S = \{2\}$ et $S = \{3\}$. À cette étape, la valeur de la mesure de diversité est toujours bien définie en raison de l'unicité de la condition initiale et de la symétrie dans les distances. En effet, pour toute paire d'écosystème, nous avons $V(x \cup y) = V(x) + d(y, x)$ et $V(y \cup x) = V(y) + d(x, y)$. Or, l'unicité de la condition initiale implique que $V(x) = V(y)$ et la symétrie dans les distances implique que $d(x, y) = d(y, x)$ de telle sorte que la mesure offre une valeur unique pour tous S tel que $|S| = 1$. Dans le cas qui nous concerne, les valeurs de cette mesure sont données par :

$$V(\{1, 2\}) = 6, \quad V(\{1, 3\}) = 7 \quad \text{et} \quad V(\{2, 3\}) = 8.$$

Calculons maintenant la mesure de diversité pour $S = \{1, 2\}$, $S = \{1, 3\}$ et $S = \{2, 3\}$.

$$V(\{1, 2\} \cup 3) = V(\{1, 2\}) + d(3, \{1, 2\}) = 6 + 2 = 8$$

$$V(\{1, 3\} \cup 2) = V(\{1, 3\}) + d(2, \{1, 3\}) = 7 + 1 = 8$$

$$V(\{2, 3\} \cup 1) = V(\{2, 3\}) + d(1, \{2, 3\}) = 8 + 1 = 9$$

Nous voyons qu'il y a deux valeurs possibles pour mesurer la diversité de X . ■

Cet exemple illustre le fait que la mesure de diversité (1.7) n'est pas toujours bien définie lorsque $\delta(x, S)$ est donné par (1.3) et que $|S| \geq 2$. Généralement, la valeur de la diversité issue de cette approche dépend du sentier de calcul envisagé. Or, il existe une situation où cette mesure offre une solution unique pour toutes les combinaisons possibles, c'est-à-dire dans le cas particulier où les distances sont ultramétriques. Les distances d'un ensemble sont

dites ultramétriques si les distances de tous les triplets $\{x, y, z\}$ d'un ensemble respectent l'expression suivante :

$$\max \{d(x, y), d(x, z), d(y, z)\} = \text{med} \{d(x, y), d(x, z), d(y, z)\}, \quad (1.9)$$

ce qui signifie que la valeur maximale parmi les distances d'un triplet doit être égale à la valeur médiane; en d'autres termes, les deux plus grandes distances sont égales.¹ En considérant les distances de l'exemple 1, nous voyons clairement qu'elles ne respectent pas cette condition ce qui explique pourquoi la valeur de la mesure n'est pas unique. Par contre, lorsque les distances sont ultramétriques, la mesure décrite précédemment est toujours bien définie; c'est-à-dire qu'elle devient une fonction au sens où elle offre une valeur unique peu importe le sentier de calcul envisagé. Or, concrètement, il est plutôt rare de retrouver des situations où les distances sont exprimées sous une forme ultramétrique. Il serait donc intéressant de disposer d'une mesure offrant une solution unique dans le cas général.²

1.2 Fonction de diversité

Pour définir une mesure applicable dans le cas général, Weitzman (1992) propose de construire une fonction de diversité respectant une version faible de l'expression (1.7), définie par l'axiome suivant :

MONOTONICITÉ (DANS LES ÉCOSYSTÈMES) : *Lorsqu'un écosystème $x \in X \setminus S$ est ajouté à un ensemble $S \in \Omega$, la fonction de diversité doit respecter*

$$V(S \cup x) \geq V(S) + \delta(x, S), \quad (1.10)$$

où $\delta(x, S)$ est donné par (1.3).

¹ Nous disons aussi que les distances d'un ensemble sont ultramétriques si pour tout $x, y, z \in X$, $d(x, y) \leq \max \{d(x, z), d(y, z)\}$.

² Nous utilisons l'expression «dans le cas général» par opposition au cas particulier où les distances sont ultramétriques.

Cet axiome signifie que lorsqu'un écosystème x est ajouté à S , la valeur du nouvel ensemble doit augmenter d'au moins la distance entre x et l'écosystème qui lui est le plus semblable parmi ceux de S . Or, cet axiome n'est pas suffisant pour définir une seule et unique fonction de diversité. Pour combler cette lacune, Weitzman propose l'ajout d'un axiome supplémentaire que nous présentons dans ce qui suit.

Puisque dans le cas général la mesure de diversité n'offre pas la même valeur pour toutes les combinaisons possibles de $V(S) + \delta(x, S)$, il serait intéressant de faire en sorte que la valeur de la fonction de diversité soit donnée par au moins une de ces combinaisons. Cette intuition est décrite par l'axiome de liaison suivant :

LIAISON : Pour tout ensemble $S \in \Omega$ où $|S| \geq 2$, il existe au moins un écosystème $x \in S$ (appelé l'écosystème lien) pour qui

$$V(S) = V(S \setminus x) + \delta(x, S \setminus x) . \quad (1.11)$$

$V(S)$ est alors donnée par la valeur d'au moins un des ensembles S sans un des écosystème x auquel nous ajoutons la diversité marginale de cet écosystème par rapport à $S \setminus x$.

Dans le cas particulier où les distances sont ultramétriques, la valeur de la mesure (1.7)-(1.8) est la même pour toutes les combinaisons possibles de $V(S) + \delta(x, S)$, c'est-à-dire que tous les écosystèmes sont des écosystèmes liens au sens où chacun d'eux respecte l'axiome de liaison.

Dans le cas général, Weitzman (1992) démontre que la seule fonction de diversité respectant les axiomes (1.10) et (1.11) est donnée, pour tous $S \in \Omega$, par :

$$V(S) = \begin{cases} d_0 \\ \max_{x \in S} \{V(S \setminus x) + \delta(x, S \setminus x)\} \end{cases} \quad \text{si } |S| \begin{cases} = 1 \\ \geq 2 \end{cases} . \quad (1.12)$$

Cette expression consiste à modifier la mesure de diversité décrite précédemment pour n'en retenir que la valeur maximale parmi toutes les combinaisons possibles de

$V(S \setminus x) + \delta(x, S \setminus x)$. Par exemple, en appliquant (1.12) aux données de l'exemple 1, nous obtenons $V(X) = 9$.

Comme le mentionne Weitzman (1992), cette procédure est plutôt longue à réaliser puisqu'elle nécessite le calcul de toutes les combinaisons possibles de S et de x , ce qui peut devenir une lourde tâche lorsque l'ensemble considéré est très grand. Or, il est possible de formuler une fonction alternative, équivalente à (1.12), permettant de simplifier la résolution de la fonction de diversité en passant directement par le sentier de calcul maximal lors de chaque récurrence.

1.3 Fonction de diversité alternative

La section qui suit, inspirée de Roosen, Fadlaoui et Bertaglia (2003), présente une démarche relativement simple permettant d'obtenir cette nouvelle fonction en partant de l'expression (1.12) et des axiomes mentionnés précédemment.

Par l'axiome de liaison, nous savons qu'il existe au moins un écosystème (l'écosystème lien) pour qui l'axiome de monotonie dans les écosystèmes (expression (1.10)) est respecté avec égalité. Le point de départ de cette démarche provient du fait que cet écosystème lien est donné par un des deux écosystèmes à distance minimale de S ($|S| \geq 2$). Ces deux écosystèmes, notés $g(S)$ et $h(S)$, sont donc ceux qui respectent l'expression suivante :

$$d(g(S), h(S)) = \min_{\substack{x, y \in S \\ x \neq y}} d(x, y). \quad (1.13)$$

Pour déterminer quel est l'écosystème lien parmi la paire $(g(S), h(S))$, il s'agit de résoudre, en s'inspirant de (1.12), l'expression :

$$V(S) = \max \{V_g, V_h\} \quad (1.14)$$

où $V_g = V(S \setminus g(S)) + \delta(g(S), S \setminus g(S))$ et $V_h = V(S \setminus h(S)) + \delta(h(S), S \setminus h(S))$, de telle sorte que l'écosystème lien est celui qui satisfait :

$$\max \{V(S \setminus g(S)), V(S \setminus h(S))\}. \quad (1.15)^3$$

En se servant de l'axiome de liaison, il est possible de combiner (1.13) et (1.15), de façon à reformuler la fonction de diversité (1.12) de la manière suivante :

$$V(S) = \begin{cases} d_0 \\ d(g(S), h(S)) + \max \{V(S \setminus g(S)), V(S \setminus h(S))\} \end{cases} \text{ si } |S| \begin{cases} = 1 \\ \geq 2 \end{cases}. \quad (1.16)$$

La valeur de S ($|S| \geq 2$) est alors donnée par la distance minimale de cet ensemble auquel nous ajoutons la valeur maximale entre $S \setminus g(S)$ et $S \setminus h(S)$.

Nous présentons dans un premier temps l'algorithme sous-jacent à cette fonction. Celui-ci sera ensuite illustré à l'aide d'un exemple. Par la suite, nous analysons une caractéristique particulière de la fonction (1.16), à savoir qu'en partant des distances issues du cas général, il est possible de construire une matrice de distances ultramétriques en se servant de l'information obtenue lors de la résolution de l'algorithme. Or, il appert que dans le cas particulier où les distances sont ultramétriques, la diversité d'un ensemble peut être présentée graphiquement sous la forme d'une structure arborescente dite hiérarchique. Ces deux derniers points seront illustrés par les résultats de notre exemple.

³ Puisque $\delta(g(S), S \setminus g(S)) = \delta(h(S), S \setminus h(S)) = d(g(S), h(S))$, les diversités marginales de $g(S)$ et $h(S)$ n'interviennent pas dans l'identification de l'écosystème lien.

1.3.1 Algorithme sous jacent à la fonction de diversité

- 1) La première étape consiste à identifier les deux écosystèmes à distance minimale de l'ensemble S à l'aide de l'expression (1.13).
- 2) Il s'agit ensuite d'identifier l'écosystème lien parmi cette paire. Supposons par exemple que $g(S) = \arg \max \{V(S \setminus g(S)), V(S \setminus h(S))\}$ de telle sorte que la fonction (1.16) est donnée par $V(S) = d(g(S), h(S)) + V(S \setminus g(S))$.
- 3) Lorsque l'écosystème lien est identifié, il s'agit de le retirer de S et d'ajouter la valeur de la distance obtenue en 1) au calcul de $V(S)$. L'autre écosystème donné par $h(S)$ est appelé l'écosystème représentatif au sens où il représentera la paire $(g(S), h(S))$ lors des étapes subséquentes du processus récursif.
- 4) La résolution de l'algorithme se poursuit en reprenant pour $S' = S \setminus g(S)$, et pour tout ensemble subséquent, les étapes 1), 2) et 3).
- 5) Le processus récursif est terminé lorsqu'il ne reste plus qu'un seul élément. La valeur de la fonction de diversité est alors donnée par la somme des distances obtenues lors de chacune des récurrences auquel nous ajoutons finalement la valeur de la constante de normalisation d_0 .

1.3.2 Exemple 2 : calcul de la fonction de diversité $V(S)$

Cet exemple illustre l'algorithme décrit précédemment en se servant d'un ensemble $S = \{1, 2, 3, 4\}$. Supposons dans un premier temps que la constante de normalisation est donnée par $d_0 = 5$. La matrice qui suit présente les distances entre chaque paire d'écosystème de cet ensemble. Par exemple, l'élément situé à la première ligne, deuxième colonne représente la distance entre les écosystèmes 1 et 2, c'est-à-dire $d(1, 2) = 1$.

Tableau 1.1
Matrice des distances de l'ensemble S

	1	2	3	4
1	0	1	2	3
2	1	0	3	4
3	2	3	0	2
4	3	4	2	0

La première étape de l'algorithme consiste à identifier la distance minimale parmi celles de S . Celle-ci est donnée par $d(1,2)$. En insérant ce résultat dans (1.16), nous obtenons :

$$V(S) = d(1,2) + \max\{V(S \setminus 1), V(S \setminus 2)\}.$$

Il s'agit ensuite d'identifier quel est l'écosystème lien de la paire $(1,2)$ en déterminant la valeur maximale entre $V(S \setminus 1)$ et $V(S \setminus 2)$. La fonction de diversité de l'ensemble $S \setminus 1$ est donnée par :

$$V(S \setminus 1) = d(3,4) + \max\{V(\{2,4\}), V(\{2,3\})\}.$$

Puisque

$$V(\{2,4\}) = d(2,4) + \max\{V(2), V(4)\} = 4 + \max\{5,5\} = 9$$

et

$$V(\{2,3\}) = d(2,3) + \max\{V(2), V(3)\} = 3 + \max\{5,5\} = 8$$

nous obtenons :

$$V(S \setminus 1) = 2 + \max\{9,8\} = 11.$$

Pour l'ensemble $S \setminus 2$, il existe deux possibilités, donnant le même résultat. En appliquant le raisonnement précédemment, nous obtenons, dans le premier cas :

$$V(S \setminus 2) = d(1,3) + \max\{V(\{3,4\}), V(\{1,4\})\} = 2 + \max\{7,8\} = 10$$

et dans le second :

$$V(S \setminus 2) = d(3,4) + \max\{V(\{1,4\}), V(\{1,3\})\} = 2 + \max\{8,7\} = 10.$$

Puisque $V(S \setminus 1) > V(S \setminus 2)$, l'écosystème lien de S est donné par 1 alors que 2 est l'écosystème représentatif. Retirons 1 de S et retenons $d(1,2)$ pour le calcul de $V(S)$. L'étape 4) de l'algorithme suggère de reprendre ce processus pour l'ensemble $S \setminus 1$. En se servant des résultats obtenus précédemment, nous savons que :

$$V(S \setminus 1) = d(3,4) + \max\{V(\{2,4\}), V(\{2,3\})\}$$

où $V(\{2,4\}) > V(\{2,3\})$, de telle sorte que pour $S \setminus 1$, 3 joue le rôle de l'écosystème lien et 4 celui de l'écosystème représentatif. Retirons 3 et retenons $d(3,4)$ pour le calcul de $V(S)$. Il reste alors le sous-ensemble $\{2,4\}$ et la distance qui y est associée est $d(2,4)$. Le calcul de l'algorithme est maintenant terminé ce qui nous permet de déterminer la valeur de la fonction de diversité de la manière suivante :

$$V(S) = d(1,2) + d(3,4) + d(2,4) + d_0 = 1 + 2 + 4 + 5 = 12. \blacksquare$$

1.4 Transformation des distances en distances ultramétriques

Comme nous l'avons mentionné précédemment, il est possible de transformer les distances issues du cas général de façon à ce qu'elles deviennent ultramétriques. Nous disposons dans un premier temps de $n(n-1)/2$ distances pour mesurer la diversité d'un ensemble. Or, le calcul de la fonction (1.16) est donné par la somme des $n-1$ mesures de distances attachées aux différents écosystèmes liens. C'est donc dire qu'un certain nombre de distances ne sont pas utilisées dans le calcul de la fonction de diversité. L'idée consiste à les transformer en se servant de celles obtenues par la résolution de la fonction (1.16) et de la définition de l'ultramétrie, qui nous dit que les deux plus grandes distances d'un triplet sont égales. Il s'agit alors de passer en revue l'ensemble des triplets et de les compléter en se servant de cette définition.

Pour $\{1,2,4\}$, nous disposons des distances pour les paires $(1,2)$ et $(2,4)$. Pour que les distances de ce triplet soient ultramétriques, nous devons définir la distance entre 1

et 4 de telle sorte qu'elle soit égale à la valeur maximale entre $d(1,2)$ et $d(2,4)$, c'est-à-dire $d(1,4)=4$. En reprenant ce processus pour $\{2,3,4\}$, nous avons $d(2,4)=4$ et $d(3,4)=2$ d'où $d(2,3)=4$. Finalement, pour le triplet $\{1,2,3\}$, nous disposons de $d(1,2)=1$ et nous avons obtenu précédemment $d(2,3)=4$ ce qui fait en sorte que $d(1,3)=4$. La totalité des distances de S sont maintenant définies, ce qui donne la matrice des distances ultramétriques suivantes :

Tableau 1.2

Matrice des distances ultramétriques de l'ensemble S

	1	2	3	4
1	0	1	4	4
2	1	0	4	4
3	4	4	0	2
4	4	4	2	0

Puisque nous disposons de $n(n-1)/2$ mesures de distances différentes pour construire notre fonction et qu'il n'y a au minimum que $n-1$ distances qui ne changent pas, cette procédure peut modifier un certain nombre de distances pouvant aller jusqu'à $(n-1)(n-2)/2$. Notons que l'écart entre la matrice des distances originales et la matrice des distances ultramétriques transformée à l'aide de cette méthode est d'autant plus élevé que le nombre de distances modifiées est grand et que l'écart entre une distance modifiée et sa distance originale est élevé.

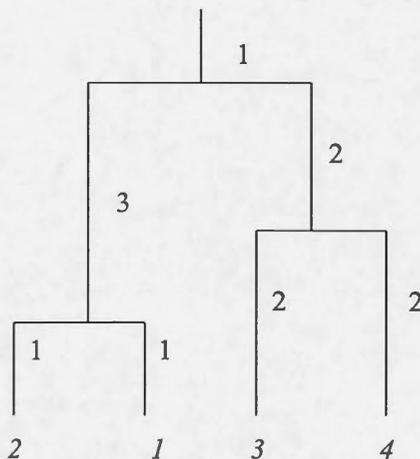
1.5 Représentation graphique d'un ensemble d'écosystèmes

Comme le mentionne Weitzman (1992, p.386), la fonction (1.16) représente la meilleure façon de regrouper les deux éléments les plus semblables lors de chacune des récurrences. Or, cette capacité de regrouper des paires d'écosystèmes fait en sorte qu'il est possible d'illustrer la diversité d'un ensemble sous la forme d'une structure arborescente dite hiérarchique. Cette représentation consiste à illustrer la valeur de la diversité par la longueur totale des branches verticales d'un arbre en se servant des distances obtenues via l'algorithme de la fonction (1.16). De façon à caractériser davantage la structure de cette arborescence, les écosystèmes devront être combinés entre eux de manière à ce que les écosystèmes représentatifs de chaque sous-ensemble soient placés le plus loin possible les uns des autres.

D'autre part, pour des raisons de cohérence, il est souhaitable de poser la constante de normalisation de telle sorte que $d_0 > \max_{x,y \in S} d(x,y)$. Ceci permettra de représenter chaque écosystème de façon à ce qu'ils soient tous liés les uns aux autres par une branche initiale commune.

Voyons maintenant comment construire cette représentation. En considérant les résultats de l'exemple 2, nous savons que les écosystèmes les plus semblables de S sont les écosystèmes 1 et 2. Ils seront donc placés l'un à côté de l'autre. Nous savons aussi que les écosystèmes les plus semblables du sous-ensemble $\{2,3,4\}$ sont donnés par 3 et 4 ce qui implique qu'eux aussi seront placés côte à côte. La dernière étape de l'algorithme lie 2 et 4. Or, puisque 2 est l'écosystème représentatif de S et que 4 est l'écosystème représentatif de $\{2,3,4\}$ et que nous devons les placer le plus loin possible les uns des autres, nous placerons 1 entre 2 et la paire (3,4) et 3 entre 4 et la paire (1,2) ce qui donne la représentation graphique suivante :

Figure 1.1
Représentation graphique de l'ensemble S



Cette capacité de représenter la diversité sous une forme arborescente hiérarchique permet d'illustrer de façon intuitive la structure de cet ensemble.⁴ Nous reviendrons plus loin sur l'interprétation que nous pouvons donner à cet arbre.

Le prochain chapitre présente une approche lexicographique, semblable à celle que nous venons de décrire, développée par Bossert, Pattanaik et Xu (2001).

⁴ Voir Weitzman (1992, p.368-372 et p.389-390) pour un traitement plus approfondi concernant la relation entre la représentation graphique, les distances ultramétriques et la fonction de diversité.

CHAPITRE II

L'APPROCHE LEXICOGRAPHIQUE

Bossert, Pattanaik et Xu (2001) présentent une approche alternative permettant de mesurer la diversité d'un ensemble en se basant uniquement sur des mesures de distances définies par les expressions (1.1) et (1.2). L'objectif de leur approche consiste à définir un vecteur $D(S)$ composé de n distances représentant la diversité d'un ensemble S . Le choix des distances incluses dans ce vecteur est déterminé sur la base d'un critère de classement lexicographique et de la mesure de diversité marginale (1.3).

Pour déterminer la diversité d'un ensemble à l'aide de cette approche, il s'agit de retirer chaque écosystème selon un certain processus récursif puis d'ajouter la distance minimale entre cet écosystème et l'ensemble considéré dans le vecteur $D(S)$. Un critère de classement lexicographique nous permettra de déterminer quel écosystème doit être éliminé lors de chaque récurrence.

2.1 Vecteur de distance $D(S)$

Posons tout d'abord, pour tout $x \in S$ et pour tout $S \in \Omega$, une fonction bijective $\sigma_x^S : \{1, \dots, |S|\} \rightarrow S$ de telle sorte qu'il est possible de classer les distances propres à chaque écosystèmes en ordre croissant, c'est-à-dire :

$$d(x, \sigma_x^S(1)) \leq \dots \leq d(x, \sigma_x^S(|S|)). \quad (2.1)$$

En se servant de (2.1), il est possible de construire des vecteurs de distance pour tout $x \in S$ en posant :

$$v_x(S) = (d(x, \sigma_x^S(1)), \dots, d(x, \sigma_x^S(|S|))). \quad (2.2)$$

Ces différents vecteurs seront utiles pour identifier quel écosystème devrait être retiré à chaque étape de l'algorithme.

L'idée derrière la construction du vecteur $D(S)$ consiste à retirer lors de chaque récurrence un des écosystèmes pour qui la distance avec le reste de l'ensemble est minimale. Or, ces distances sont symétriques ce qui implique qu'il en existe toujours au moins deux à distance minimale. Nous ne pouvons donc pas déterminer l'écosystème à retirer en se basant uniquement sur l'identification de ces distances. Il est toutefois possible de surmonter cette difficulté en se servant d'une approche lexicographique dite *leximin*.

Définissons une relation de préférence, notée \leq^S , permettant d'ordonner les écosystèmes de S de telle sorte que :

$$x \leq^S y \text{ si et seulement si } v_x(S) \leq_{lex}^S v_y(S) \quad (2.3)$$

où \leq_{lex}^S représente la relation de préférence lexicographique applicable aux vecteurs de distances (2.2). Définissons maintenant le critère de classement leximin permettant d'établir cette relation. Posons tout d'abord $v_x^j(S)$ et $v_x^k(S)$ pour définir les $j^{ième}$ et $k^{ième}$ éléments du vecteur $v_x(S)$. Ce critère s'énonce de la manière suivante (Knoblauch, 2000) :

$$v_x(S) <_{lex}^S v_y(S) \text{ si et seulement si il existe un certain élément } v_x^k(S) \\ \text{ où } k \leq |S| \text{ tel que } v_x^j(S) = v_y^j(S) \text{ pour tout } j < k \text{ et } v_x^k(S) < v_y^k(S). \quad (2.4)$$

Ce critère de classement procède en comparant tout d'abord, pour chaque paire d'écosystème, les premiers éléments des vecteurs de distances. Si le premier élément d'un des deux vecteurs est inférieur à l'autre, le vecteur lié à cet élément est considéré comme étant lexicographiquement inférieur à l'autre. S'ils sont égaux, nous n'avons qu'à reconsidérer cette démarche pour les éléments suivants et ce, jusqu'à ce que le $k^{\text{ième}}$ élément de l'un des deux vecteurs soit inférieur à l'autre.

Puisque le choix des écosystèmes à inclure dans le vecteur de distance $D(S)$ est le fruit d'un certain processus récursif, posons :

$$S_i = \begin{cases} S & \text{si } i = 1, \\ S_{i-1} \setminus \{s_{i-1}\} & \forall i \in \{2, \dots, |S|\} \end{cases} \quad (2.5)$$

et

$$s_i = \{x \in S_i \mid x \leq^{s_i} y \quad \forall y \in S_i\} \quad \forall i \in \{1, \dots, |S|\} \quad (2.6)$$

de façon à déterminer l'ordre d'apparition des écosystèmes dans $D(S)$. Ce vecteur est donné par :

$$D(S) = (\delta(s_1, S_1), \dots, \delta(s_{|S|}, S_{|S|})) \quad (2.7)$$

où, pour tout $i \in \{1, \dots, |S|\}$,

$$\delta(s_i, S_i) = \begin{cases} 0 & \text{si } S_i = \{s_i\} \\ \min \{d(s_i, y) \mid y \in S_i \setminus \{s_i\}\} & \text{si } S_i \neq \{s_i\} \end{cases} \quad (2.8)$$

de telle sorte que $\delta(s_1, S_1) \leq \dots \leq \delta(s_{|S|-1}, S_{|S|-1})$ et $\delta(s_{|S|}, S_{|S|}) = 0$.

Bossert, Pattanaik et Xu (2001) présentent les axiomes et caractéristiques propres au vecteur de distance $D(S)$. Or, un aspect intéressant lié à cette approche mérite d'être

mentionné, c'est-à-dire qu'il est possible d'agréger les distances du vecteur $D(S)$ de façon à obtenir une mesure de diversité comparable à celle proposée par Weitzman (1992).

2.2 Fonction de diversité issue du vecteur de distance

Définissons cette fonction, notée $U : \Omega \rightarrow \mathbb{R}_0^+$, de telle sorte que pour tout $S \in \Omega$,

$$U(S) = \sum_{i=1}^{|S|} \delta(s_i, S_i). \quad (2.9)$$

La seule différence entre les fonctions U et V provient du fait que l'approche hiérarchique est plus générale, en ce sens qu'elle considère une constante de normalisation d_0 pour déterminer la valeur de tout écosystème pris individuellement, laquelle n'est pas incluse dans la fonction proposée par l'approche lexicographique. Le théorème suivant, démontré par Bossert, Pattanaik et Xu (2001, p.9-11), présente pour tout $S \in \Omega$, l'équivalence entre les fonctions (2.9) et (1.12) (ou (1.16)) à une constante près.

THÉORÈME :
$$U(S) = V(S). \quad (2.10)$$

Pour être plus précis, $U(S) = V(S)$ seulement si $d_0 = 0$ sinon, $U(S) = V(S) - d_0$. Illustrons maintenant l'approche lexicographique en s'inspirant des données de l'exemple 2.

2.2.1 Exemple 3 : calcul de la fonction de diversité $U(S)$

L'objectif consiste dans un premier temps à construire le vecteur $D(S)$. Nous savons, par l'expression (2.5) que $S_1 = S$. La difficulté repose sur l'identification de l'écosystème x qui respectera la condition donnée par (2.6) pour définir s_1 . Construisons les vecteurs de distances pour chacun des écosystèmes de S_1 en se servant des expressions (2.1) et (2.2).

$$\begin{aligned} v_1(S_1) &= (0, 1, 2, 3) & v_2(S_1) &= (0, 1, 3, 4) \\ v_3(S_1) &= (0, 2, 2, 3) & v_4(S_1) &= (0, 2, 3, 4) \end{aligned}$$

En appliquant la relation de préférence (2.3) pour S_1 , nous obtenons $1 <^{S_1} 2 <^{S_1} 3 <^{S_1} 4$.

En effet, en se servant de (2.4) avec $v_1(S_1)$ et $v_2(S_1)$, nous vérifions pour $k = 3$ que :

$$v_1^1(S_1) = v_2^1(S_1) = 0, \quad v_1^2(S_1) = v_2^2(S_1) = 1 \quad \text{et} \quad v_1^3(S_1) = 2 < v_2^3(S_1) = 3,$$

ce qui implique que $v_1(S_1) <_{lex}^{S_1} v_2(S_1)$. De même, en comparant $v_1(S_1)$ et $v_x(S_1)$ pour $x = 3, 4$, nous vérifions pour $k = 2$ que :

$$v_1^1(S_1) = v_x^1(S_1) = 0 \quad \text{et} \quad v_1^2(S_1) = 1 < v_x^2(S_1) = 2,$$

ce qui implique que $v_1(S_1) <_{lex}^{S_1} v_x(S_1)$ pour $x = 3, 4$. Nous avons aussi identifié les relations liant les autres écosystèmes entre eux mais il est suffisant d'avoir déterminé que $v_1(S_1) <_{lex}^{S_1} v_x(S_1) \quad \forall x \in S_1, x \neq 1$ pour que l'expression (2.6) soit satisfaite pour $i = 1$. En posant $s_1 = 1$, il est possible de recommencer l'exercice précédent avec $S_2 = \{2, 3, 4\}$, ce qui donne en appliquant les expressions (2.1) et (2.2),

$$v_2(S_2) = (0, 3, 4), \quad v_3(S_2) = (0, 2, 3), \quad v_4(S_2) = (0, 2, 4).$$

La relation de préférence est alors donnée par $3 <^{S_2} 4 <^{S_2} 2$. Posons $s_2 = 3$ et reprenons ce processus pour $S_3 = \{2, 4\}$. Les vecteurs de distance de cet ensemble sont $v_2(S_3) = (0, 4)$ et $v_4(S_3) = (0, 4)$ ce qui donne $2 =^{S_3} 4$. Il y a donc deux

possibilités. Soit nous retirons 2 pour avoir $S_4 = \{4\}$ ou nous retirons 4 pour avoir $S_4 = \{2\}$. Dans le premier cas nous avons $s_3 = 2$ avec $v_4(S_4) = (0)$ et $s_4 = 4$ alors que l'autre possibilité donne $s_3 = 4$ avec $v_2(S_4) = (0)$ et $s_4 = 2$. Servons-nous du premier cas pour poursuivre le raisonnement.

En se servant de l'information obtenue lors de chaque étape de l'algorithme, il est possible d'identifier pour tout i , les diversités marginales telles que définit par l'expression (2.8). Le vecteur de distance est alors donné par :

$$D(S) = (d(1,2), d(3,4), d(2,4), d(4,4)) = (1,2,4,0).$$

En ayant construit ce vecteur, nous pouvons aussi calculer la valeur de la fonction de diversité qui y est associée. Celle-ci est donnée par :

$$U(S) = \sum_{i=1}^4 \delta(s_i, S_i) = 1+2+4+0 = 7. \blacksquare$$

Nous voyons facilement l'équivalence entre $U(S)$ et $V(S)$ à une transformation monotone près en comparant le résultat de cet exemple à celui de l'exemple 2.

Les deux approches que nous venons de présenter permettent de mesurer la diversité en se servant de distances entre chaque paire d'écosystème d'un ensemble. Le chapitre suivant présente notre façon de modéliser la biodiversité ainsi que l'approche basée sur le concept d'endémisme.

CHAPITRE III

MODÉLISATION DE LA DIVERSITÉ

Les deux premiers chapitres décrivent des approches permettant de mesurer la diversité d'un ensemble. Or, jusqu'à maintenant, nous n'avons rien mentionné concernant la construction des mesures de distances nécessaires au calcul de ces fonctions. Rappelons que ces distances, définies par les expressions (1.1) et (1.2), doivent à tout le moins être non-négatives et symétriques. Voici ce que nous proposons de faire.

Étant donné une liste d'espèces présente dans un ensemble d'écosystèmes, nous voulons construire une mesure de distance à partir de la seule observation de la présence ou de l'absence de chacune des espèces dans les différents écosystèmes constituant cet ensemble. Posons i , une espèce telle que $i = 1, \dots, m$ où $L = \{1, \dots, m\}$ représente la liste des espèces présentes dans S . Servons-nous d'une relation binaire de présence/absence pour établir la relation liant les espèces $i \in L$ aux écosystèmes $x \in S$. Celle-ci est donnée par :

$$\varepsilon_i^x = \begin{cases} 1 & \text{si } i \in x, \\ 0 & \text{sinon.} \end{cases} \quad (3.1)$$

C'est donc dire dans ce cas qu'un écosystème représente un certain sous-ensemble de L ($x \subseteq L$). Étant donné cette expression, il est possible d'illustrer la répartition des espèces parmi un ensemble d'écosystème à l'aide d'une matrice de dimension $n \times m$ où les lignes nous renseignent sur la composition des écosystèmes en termes d'espèces, alors que les colonnes décrivent la distribution des espèces parmi ces différents écosystèmes. Le tableau suivant présente les valeurs de ε_i^x pour l'ensemble $S = \{1, 2, 3, 4\}$ de l'exemple 2. À

l'aide des données de ce tableau, nous savons par exemple que l'écosystème 1 est composé des espèces 1 à 5 alors que l'espèce 3 est présente dans les écosystèmes 1, 2 et 4. Nous savons aussi que cet ensemble est composé de quatre écosystèmes et de neuf espèces différentes.

Tableau 3.1

Matrice des données relatives à ε_i^x

	1	2	3	4	5	6	7	8	9
1	1	1	1	1	1	0	0	0	0
2	1	1	1	0	1	0	0	1	0
3	1	1	0	1	0	1	1	0	0
4	0	0	1	1	0	1	1	0	1

En se servant de (3.1), la façon la plus simple de mesurer la valeur d'un écosystème consiste à faire la sommation des ε_i^x pour toutes espèces i faisant partie de x , c'est-à-dire :

$$v(x) = \sum_i \varepsilon_i^x. \quad (3.2)$$

$v(x)$ est alors donnée par le nombre d'espèces différentes recensées dans x . Cette expression illustre la notion de richesse généralement proposée par les écologistes. Dans le cas particulier où tous les écosystèmes sont composés d'un même nombre d'espèces, il est possible d'utiliser (3.2) pour représenter la constante de normalisation d_0 . Par exemple, en appliquant $v(x)$ aux données du Tableau 3.1, nous constatons qu'il y a cinq espèces différentes dans chacun des quatre écosystèmes, ce qui correspond à la valeur de la constante d_0 de l'exemple 2.

Maintenant, si nous voulons construire une mesure de la différence entre deux écosystèmes, il est nécessaire de proposer une mesure du nombre d'espèces détenues en commun par ces deux écosystèmes. Posons, pour tout $i \in L$ et pour tout $x, y \in S$:

$$\varepsilon_i^{xy} = \begin{cases} 1 & \text{si } \varepsilon_i^x = \varepsilon_i^y = 1 \\ 0 & \text{sinon} \end{cases} \quad (3.3)$$

et

$$c(x, y) = \sum_i \varepsilon_i^{xy}. \quad (3.4)$$

Par (3.3), nous savons que l'espèce i est commune à x et y si $\varepsilon_i^{xy} = 1$. Si elle n'est incluse que dans un des deux écosystèmes ou dans aucun des deux, alors $\varepsilon_i^{xy} = 0$. L'expression (3.4) représente donc le nombre d'espèces contenues dans x qui se retrouvent aussi dans y . Évidemment, $c(x, y)$ est symétrique et inférieur ou égal à $v(x)$ et $v(y)$ et ce, pour tout $x, y \in S$.

3.1 Mesure de différence de base

Weikard (2002) propose de mesurer la différence entre x et y ($\in S$) par le nombre d'espèces uniques à x lorsque celui-ci est comparé à y , c'est-à-dire :

$$d'(x, y) = v(x) - c(x, y). \quad (3.5)$$

L'expression (3.5) n'est symétrique que si la valeur de chaque écosystème, donnée par (3.2), est la même pour tous. En effet, puisque (3.4) est symétrique, il faut absolument que $v(x) = v(y)$ pour que $d'(x, y) = d'(y, x)$. Étant donné la nature des écosystèmes, il est possible d'envisager des situations où ceux-ci ne sont pas tous composés d'un même nombre d'espèces. Il serait donc intéressant de disposer de mesures de différence inspirées

des expressions (3.2) et (3.4) respectant la symétrie tout en permettant des situations où $v(x) \neq v(y)$ pour tout $x \in S, y \in S$. Mais avant de poursuivre, revenons brièvement sur l'interprétation de la Figure 1.1.

3.1.1 Interprétation de la Figure 1.1

Nous devons tout d'abord mentionner que les distances du Tableau 1.1 proviennent de l'application de l'expression (3.5) aux données issues du Tableau 3.1. Étant donné ces distances, le calcul de la fonction de diversité (1.16) tel que présenté dans l'exemple 2 permet de générer la représentation graphique de la Figure 1.1. Nous devons évidemment garder à l'esprit que les distances de cet ensemble ont subies les transformations nécessaires pour devenir ultramétriques. Puisque ces distances représentent le nombre d'espèces uniques à un certain écosystème lorsque celui-ci est comparé à un autre, il est possible d'interpréter la Figure 1.1 de la manière suivante.

Premièrement, les longueurs des branches terminales de l'arbre représentent le nombre d'espèces uniques à chaque écosystème. Nous voyons par exemple que deux espèces ne se retrouvent que dans 3 et nulle part ailleurs.

Dans un deuxième temps, les branches intermédiaires nous renseignent sur les groupes d'espèces communes à certains sous-ensembles. Selon la Figure 1.1, il y a trois espèces communes à la paire (1, 2) qui ne se retrouvent pas dans la paire (3, 4) et deux espèces qui sont communes à 3 et 4 qui ne se retrouvent ni dans 1, ni dans 2.

Finalement, la branche initiale nous renseigne sur le nombre d'espèces communes à tous les écosystèmes. Celle-ci suggère, dans le cas qui nous concerne, qu'une seule espèce est commune aux quatre écosystèmes de S .

D'autre part, nous avons mentionné à la section 1.5 que la valeur de la fonction de diversité est égale à la somme des branches verticales de l'arbre correspondant. En effet, en faisant la somme de la longueur de chacune des branches de la Figure 1.1, nous obtenons un résultat qui est égal à la valeur de la fonction de diversité issue de l'exemple 2.

Dans le même ordre d'idées, la valeur de chaque écosystème pris isolément est représenté graphiquement par la sommation des branches verticales qui leur sont propres, c'est-à-dire en additionnant les longueurs de chacune des branches d'un écosystème en partant de la branche terminale et en remontant jusqu'à la branche initiale.

Évidemment, toute cette interprétation n'est rigoureusement correcte que dans le cas particulier où les distances sont ultramétriques. Nous voyons facilement que cette interprétation n'est pas identique à celle que nous ferions à l'aide des données du Tableau 3.1, lesquelles représentent l'information brute décrivant la structure de la diversité de cet ensemble d'écosystème. Nous devons donc être prudents si nous voulons interpréter la diversité d'un ensemble de cette façon (voir Weitzman, 1992, p.389).

Poursuivons maintenant notre discussion concernant l'élaboration de mesures de différences symétriques.

3.2 Mesures de différences normalisées

Il existe un grand nombre d'approches permettant d'obtenir des mesures de distances non-négatives et symétriques. Nous présentons ici seulement deux mesures de façon à illustrer ce point.

Une manière relativement simple de réaliser cet objectif consiste premièrement à prendre pour chaque paire, la valeur moyenne du nombre d'espèces uniques à ces deux écosystèmes, ce qui donne :

$$\bar{d}(x, y) = v_{xy} - c(x, y) \quad (3.6)$$

où $v_{xy} = (v(x) + v(y)) / 2$. Cette mesure de distance s'interprète alors comme étant le nombre moyen d'espèces uniques à x et y .

Une deuxième façon de construire une mesure symétrique consiste à retenir pour chaque paire, la valeur minimale entre $d'(x, y)$ et $d'(y, x)$ ou, de façon équivalente :

$$d''(x, y) = \min \{v(x), v(y)\} - c(x, y). \quad (3.7)^5$$

Les expressions (3.6) et (3.7) respectent la notion de symétrie et deviennent équivalentes à (3.5) lorsque la valeur de chaque écosystème est la même pour tous.

Jusqu'à maintenant, nous avons mis l'emphase sur la construction de fonctions de diversité basées sur des mesures de distances entre toutes paires d'écosystèmes. Ces fonctions sont calculées en définissant la diversité marginale par la distance minimale entre un écosystème et un ensemble. Or, Weikard (2002) propose plutôt de considérer la diversité marginale comme étant le nombre d'espèces uniques à un écosystème. Cette approche ne nécessite donc plus l'utilisation de distances entre chaque paire. Voyons comment elle se présente.

3.3 Mesure basée sur le concept d'endémisme

Nous avons mentionné précédemment qu'une espèce est considérée comme endémique si, parmi un certain ensemble d'écosystème, elle ne se retrouve que dans un seul d'entre eux. La diversité marginale donnée par l'expression (1.3) ne correspond pas parfaitement à ce concept. En fait, elle exprime cette réalité seulement dans le cas particulier où les distances sont ultramétriques.

En prenant la définition de l'endémisme au sens littéral, Weikard (2002) propose d'utiliser la mesure de diversité initiale, c'est-à-dire :

$$V(S \cup x) = V(S) + \delta(x, S) \quad (3.8)$$

en se servant de (3.2) comme condition initiale et en définissant la diversité marginale de $x \in X \setminus S$ par rapport à $S \in \Omega$ à l'aide de l'expression suivante :

$$\delta(x, S) \equiv v(x) - c(x, S) \quad (3.9)$$

⁵ La valeur maximale est aussi envisageable, ce qui donnerait $d''(x, y) = \max \{v(x), v(y)\} - c(x, y)$.

où

$$c(x, S) = \sum_i \varepsilon_i^{xS} \quad (3.10)$$

et

$$\varepsilon_i^{xS} = \begin{cases} 1 & \text{si } \varepsilon_i^x = 1 \text{ et } \varepsilon_i^y = 1 \text{ pour au moins un } y \in S, \\ 0 & \text{sinon.} \end{cases} \quad (3.11)$$

Puisque (3.10) représente le nombre d'espèces communes à x et S , l'expression (3.9) exprime la diversité marginale comme étant le nombre d'espèces recensées dans x qui ne se retrouvent dans aucun des écosystèmes de S .

Cette façon d'aborder le problème, que nous nommerons l'approche basée sur le concept d'endémisme, permet de réaliser deux constats. Nous avons illustré précédemment que la mesure de diversité (1.7) n'offre pas toujours une valeur unique lorsque $\delta(x, S)$ est donnée par (1.3). Par contre, lorsque celle-ci est appliquée avec (3.9), la valeur obtenue est toujours bien définie peu importe le sentier de calcul envisagé. De plus, la valeur de la condition initiale n'a plus besoin d'être identique pour tout écosystème puisque (1.8) est remplacée par (3.2) qui nous dit seulement que la valeur d'un écosystème est donnée par le nombre d'espèces différentes qui y sont recensées. D'autre part, Weikard (2002, p.25) parvient à démontrer que la valeur de cette fonction de diversité est toujours égale au nombre total d'espèces recensées dans l'ensemble considéré. En appliquant par exemple cette approche aux données du Tableau 3.1, nous obtenons $V(S) = 9$ ce qui représente exactement le nombre d'espèces incluses dans L .

CHAPITRE IV

DISCUSSION

Nous présentons ici quelques caractéristiques intéressantes liées aux fonctions de diversité décrites précédemment. Nous faisons ensuite un bref survol de quelques politiques de préservation.

4.1 Caractéristiques des fonctions de diversité

Il est intéressant de constater que ces fonctions illustrent chacune à leur manière les trois types de diversité mentionnées dans l'introduction du chapitre I.

La diversité α représente la valeur de chaque écosystème pris isolément. Dans l'approche hiérarchique, celle-ci est donnée par (1.8), c'est-à-dire que la valeur de chaque écosystème pris isolément, peu importe la manière dont elle est modélisée, doit être la même pour tous. Cette expression ne permet donc pas de déterminer quel est l'écosystème le plus diversifié d'un ensemble. Par contre, l'approche basée sur le concept d'endémisme utilise (3.2) qui elle, permet d'identifier un tel écosystème puisque cette condition initiale n'implique pas nécessairement que chacun d'eux soit composé d'un même nombre d'espèces. Pour ce qui est de la fonction issue de l'approche lexicographique, celle-ci est construite seulement que sur la base d'un vecteur de distance qui ne nécessite pas de condition liée à la valeur des écosystèmes pris isolément.

La diversité β représente le niveau de disparité entre les écosystèmes d'un ensemble. Les fonctions de diversité marginale δ que nous avons décrites évoquent ce concept chacune à leur manière. Les expressions (1.3) et (2.8) sont équivalentes puisqu'elles

mesurent la diversité marginale comme étant la distance minimale entre un écosystème et un ensemble. L'approche basée sur le concept d'endémisme présente plutôt la diversité marginale d'un écosystème par le nombre d'espèces uniques à ce dernier (expression (3.9)).

La diversité γ illustre la diversité totale d'un ensemble. Celle-ci peut être obtenue par l'une ou l'autre des différentes approches que nous avons décrites. Lorsque les fonctions sont construites sur la base de distances entre écosystèmes, la diversité γ peut être donnée, dans le cas général, par les expressions (1.12), (1.16) ou (2.9) alors que les expressions (1.7)-(1.8)-(1.3) ne sont applicables que dans le cas particulier où les distances sont ultramétriques. Lorsque la fonction de diversité est basée sur le concept d'endémisme, la diversité γ est le résultat des expressions (3.8)-(3.9)-(3.2).

Comme nous pouvons le constater, les fonctions présentées ici utilisent différents outils pour définir la diversité d'un ensemble. Ces fonctions sont regroupées en deux principales catégories : celles basées sur le concept de distances (Weitzman, 1992; Bossert, Pattanaik et Xu, 2001) et celle basée sur le concept d'espèces endémiques (Weikard, 2002). Ces deux catégories de fonctions peuvent donc donner des résultats divergents même si elles sont construites à partir d'une information de base commune issue de la présence ou de l'absence de chaque espèce parmi les écosystèmes d'un ensemble.

4.2 Politiques de préservation

Ces différentes fonctions de diversité peuvent être utiles lors de l'élaboration de politiques visant la préservation des écosystèmes, lesquelles sont principalement inspirées de Weitzman (1992, p. 400-404).

Le tableau suivant présente les valeurs des trois fonctions de diversité pour tous les sous-ensembles $S' \subseteq S$, ($|S'| \geq 2$) de $S = \{1, 2, 3, 4\}$. Les fonctions (1.16), notée V et (2.9), notée U , sont calculées à l'aide des données de l'exemple 2. Celle basée sur le concept d'endémisme, notée V_e , est calculée en se servant des données du Tableau 3.1.

Tableau 4.1
Valeur des fonctions de diversité

	(1,2)	(1,3)	(1,4)	(2,3)	(2,4)	(3,4)	(S\4)	(S\3)	(S\2)	(S\1)	(S)
V	6	7	8	8	9	7	9	10	10	11	12
U	1	2	3	3	4	2	4	5	5	6	7
V_e	6	7	8	8	9	7	8	9	8	9	9

L'objectif derrière chacune de ces politiques consiste à maximiser la valeur des fonctions de diversité sujet à une contrainte budgétaire liée à la préservation de l'ensemble considéré. Comme le mentionne Weitzman (1992), l'aspect lié à la contrainte budgétaire est relativement facile à modéliser. Il s'agit simplement d'identifier le budget disponible à des fins de préservation ainsi qu'un estimé du coût lié à la conservation de chaque écosystème. Dans ce qui suit, les politiques de préservation sont énoncées sans égards aux contraintes budgétaires de façon à mettre l'emphase sur les propriétés des fonctions de diversité.

Une première recommandation que nous pouvons faire consiste à porter une attention toute particulière à la préservation des écosystèmes qui sont de façon non ambiguë plus distincts que tous les autres. En effet, si nous les sacrifions, la perte de diversité qui en résultera sera plus grande que pour tout autre écosystème.

En comparant les valeurs de V_e pour les ensembles composés de trois écosystèmes, nous voyons que cette fonction suggère de porter une attention particulière aux écosystèmes 2 et 4 puisque $V_e(S \setminus 2) = V_e(S \setminus 4) < V_e(S \setminus 1) = V_e(S \setminus 3) = V_e(S)$. Cela s'explique par le fait qu'ils comptent chacun une espèce endémique alors que les deux autres n'en possèdent pas. En effet, la valeur de V_e s'exprime directement en terme d'espèces. Nous pouvons donc affirmer que le retrait de 2 ou 4 fait en sorte que nous sacrifions assurément une espèce. Pour les fonctions V et U , nous voyons clairement que le retrait de 4 entraîne une perte de diversité plus grande que pour tout autre écosystème : $V(S \setminus 4) < V(S \setminus x)$, $\forall x \in S, x \neq 4$. Dans le cas qui nous concerne, la fonction V_e ne permet pas d'identifier de

façon non ambiguë un seul écosystème plus distinct que tous les autres puisqu'elle nous fait hésiter entre 2 et 4. Par contre, les trois fonctions identifient 4 comme étant un de ceux qui entraîne la plus grande perte de diversité. S'il y a donc un écosystème à préserver plus que tout autre, nous devrions privilégier l'écosystème 4 tout en portant une attention particulière à 2.

Dans le même ordre d'idée, si nous ne pouvons préserver que $n-1$ des n écosystèmes d'un ensemble, celui qui devrait être laissé de côté sera donné par un de ceux ayant la diversité marginale est la plus faible.

L'approche basée sur le concept d'endémisme suggère de retirer soit 1 ou 3 puisque leur diversité marginale respectives sont nulles; c'est-à-dire que $V_e(S \setminus 1) = V_e(S \setminus 3) = V_e(S)$. En d'autres termes, toutes les espèces de ces écosystèmes sont représentées au moins une fois dans les écosystèmes restants. Pour ce qui est des deux autres approches, nous voyons, par les distances du Tableau 1.1 que l'écosystème le moins diversifié devrait être 1 ou 2 puisque ce sont les deux écosystèmes à distance minimale. Par contre, en analysant les valeurs des fonctions V et U , nous voyons que le retrait de 1 entraîne une perte de diversité plus faible que pour tout autre écosystème: $V(S \setminus 1) > V(S \setminus x)$, $\forall x \in S, x \neq 1$. Ainsi, la fonction V_e nous fait hésiter entre 1 et 3 alors que V et U nous suggère plutôt de retirer 1 ou 2 avec une préférence pour 1. Étant donné ces résultats, si nous ne pouvons préserver que trois des quatre écosystèmes de S , il est préférable de se départir de 1.

Finalement, si nous ne pouvons préserver qu'une paire d'écosystème parmi toutes les paires d'un ensemble, il est souhaitable de préserver celle pour qui la valeur est maximale.

En analysant les valeurs des fonctions de diversité du Tableau 4.1, nous voyons qu'il est préférable de préserver la paire (2,4) puisque c'est elle qui maximise la valeur des fonctions de diversité pour toute paire d'écosystème et ce, peu importe la fonction utilisée. Qui plus est, la valeur de V_e pour cette paire est identique à la valeur de l'ensemble S . Autrement dit, toutes les espèces de S sont représentées dans ces deux écosystèmes. La paire (2,4) est donc le plus petit ensemble permettant de représenter la totalité des espèces recensées dans S .

CHAPITRE V

QUELQUES EXTENSIONS ET LIMITES

5.1 La diversité des écosystèmes en fonction des espèces ou des gènes ?

Dans les chapitres précédents, nous avons mesuré la diversité des écosystèmes en les comparant sur la base de leurs espèces respectives. En particulier, nous avons adapté l'approche de Weitzman (1992), qui mesure les différences entre espèces sur la base des gènes qu'elles renferment, à des situations où l'on désire mesurer les différences entre écosystèmes sur la base des espèces présentes dans ces derniers. Bien sûr il est intéressant, pour des raisons de simplicité, de mesurer la biodiversité comme nous l'avons fait. Mais pourquoi se limiter à cette relation entre les écosystèmes et les espèces ? Par exemple, du point de vue pharmaceutique, les espèces sont valorisées pour leur potentiel génétique dans la mesure où certaines d'entre elles peuvent être utiles pour le développement de traitements médicaux. Pourquoi alors ne voudrait-on pas mesurer la diversité des écosystèmes en fonction des gènes qui se retrouvent parmi les espèces qui y évoluent ? Là encore, il est possible d'adapter la méthode de Weitzman.

Weitzman mesure la biodiversité en se servant d'une fonction de diversité fondée sur des mesures de distance génétique entre espèces. Or, on peut se demander s'il est approprié d'aborder le problème de la mesure de la biodiversité sous cet angle sachant que la manière la plus efficace de protéger les espèces consiste à préserver les habitats dans lesquels elles évoluent. C'est pourquoi, dans les chapitres précédents, nous avons privilégié un cadre d'analyse où nous mettions l'accent sur les écosystèmes, qui eux, étaient comparés en fonction de leurs espèces. Cela étant, rien n'empêche de mesurer la diversité d'un écosystème sur une base génétique. Sa valeur réside alors dans les gènes apportés par

l'ensemble des espèces qui le composent. Si l'information génétique propre à chaque espèce est disponible, l'approche de Weitzman s'adapte de la manière suivante : on ne peut préserver un gène qu'en protégeant l'espèce qui l'abrite; et on ne peut protéger cette espèce qu'en préservant le milieu naturel qui l'héberge, lequel devient ainsi un réservoir de gènes. Il faut alors imaginer un écosystème comme étant une super-espèce composée de chacun des gènes de l'ensemble des espèces qui y sont présentes. Il s'agit alors d'appliquer la méthode de Weitzman à ces super-espèces. En fait, ce qui importe dans ce genre de modification de la méthode, c'est de bien identifier l'objet d'analyse ainsi que les caractéristiques pertinentes sur lesquelles il repose : ici, le gène est la caractéristique pertinente, mais c'est la réserve de gènes constituée par un écosystème particulier, et non pas l'espèce, qui constitue l'unité de base. Un écosystème est alors comparé à un autre écosystème sur la base des gènes que chacun abrite.

5.2 Les mesures de diversité des objets multidimensionnels

Jusqu'à maintenant, nous avons mesuré la diversité d'un ensemble d'objets en fonction d'une seule caractéristique⁶ et nous avons étudié des mesures fondées sur des notions de distance. La diversité d'un ensemble d'objets dépend plus généralement de plusieurs caractéristiques et également de la façon dont ces caractéristiques sont combinées. Par exemple, la valeur de la biodiversité peut dépendre non seulement des gènes présents, mais de la façon dont ces gènes apparaissent dans une espèce plutôt que dans une autre. Dans ce cas, la biodiversité dépendra non seulement de la liste des gènes présents, mais de la liste des espèces présentes. On peut aussi attribuer une valeur au nombre d'individus d'une espèce présents dans un site, par exemple si le risque d'extinction diminue avec le nombre d'individus.

Dans tous ces cas, la mesure de la diversité devient une question d'agrégation. L'agrégation pose le problème du niveau d'agrégation (espèce ou territoire?) et celui de l'agrégation des caractéristiques. Par exemple, peut-on faire la somme du nombre de gènes

⁶ On parle aussi à l'occasion d'attribut ou de descripteur.

ou est-il préférable de les regrouper par l'intermédiaire d'une fonction autre que la sommation?

Parmi les fonctions envisageables pour mesurer la diversité, la littérature que nous avons passée en revue privilégie celles qui sont basées sur des mesures de distance. Peut-on généraliser ces méthodes à des situations où la mesure est multidimensionnelle? Nous en discutons dans la section qui suit.

Plus généralement, si l'on ne se restreint pas à l'agrégation de mesures de distances s'ouvre un champ d'investigation immense qui dépasse largement le cadre de ce mémoire. Cependant, il est intéressant de souligner certaines implications des fonctions d'agrégation envisageables. Nous le ferons brièvement dans la section 5.4 en reprenant l'exemple de la super-espèce mentionnée à la section 5.1.

5.3 Les mesures de distance pour objets multidimensionnels

Les mesures de distances présentées jusqu'ici étaient construites en fonction d'une seule caractéristique. Par contre, rien dans le concept de distance ne requiert que les objets soient strictement unidimensionnels. D'ailleurs, les travaux de Nehring et Puppe (2002) présentent le cadre général d'une approche multi-attributs allant en ce sens.

Il peut être intéressant de mesurer la diversité d'un ensemble d'écosystèmes en les comparant non seulement sur la base de leurs espèces, mais aussi en fonction d'autres caractéristiques comme par exemple la fertilité du sol, le potentiel de séquestration du carbone, le volume de bois exploitable, etc. Dans ce cas, la difficulté repose sur l'élaboration d'une mesure de distance décrivant la différence entre chaque paire d'objets sur la base de leurs multiples attributs. Considérons le cas d'un ensemble où les objets sont composés de m caractéristiques. Par exemple, deux objets x et y peuvent être représentés par des vecteurs de caractéristiques composés respectivement des éléments (x_1, \dots, x_m) et (y_1, \dots, y_m) .⁷ Nous dirons alors que deux objets sont proches ou semblables si les vecteurs correspondants ont à peu près les mêmes valeurs.

⁷ Il peut y avoir des zéros pour certaines des caractéristiques si elles ne se retrouvent pas dans l'objet considéré.

Legendre et Legendre (1984) présentent un grand nombre de mesures de distance applicables à différents contextes. Parmi celles-ci, plusieurs d'entre elles peuvent être obtenues à l'aide d'une mesure générale appelée la distance de Minkowski : ⁸

$$d_{\lambda}(x, y) = \left(\sum_{i=1}^m |x_i - y_i|^{\lambda} \right)^{1/\lambda} \quad (5.1)$$

où $\lambda \geq 1$. Les principales valeurs utilisées pour le paramètre λ se résument généralement à 1, 2 et ∞ . Les valeurs de λ supérieures à deux ne sont habituellement pas utilisées car les puissances plus grandes que cet indice accordent un poids plus important aux caractéristiques ayant de grands écarts (voir Legendre et Legendre, 1984). Le tableau suivant présente les principales mesures issues de l'expression (5.1).

Tableau 5.1
Les principales mesures issues de la distance de Minkowski

$\lambda = 1$	$d_1(x, y) = \sum_{i=1}^m x_i - y_i $	Distance de Manhattan
$\lambda \rightarrow \infty$	$d_{\infty}(x, y) = \lim_{\lambda \rightarrow \infty} \left(\sum_{i=1}^m x_i - y_i ^{\lambda} \right)^{1/\lambda}$ $= \max_i x_i - y_i $	Distance de Chebyshev
$\lambda = 2$	$d_2(x, y) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2}$	Distance euclidienne

Lorsque $\lambda = 1$, nous avons la distance de Manhattan qui représente la somme des valeurs absolues des différences entre chacune des caractéristiques.

⁸ Les mesures issues de la distance de Minkowski respectent les axiomes de non-négativité et de symétrie.

La distance de Chebyshev est obtenue quand $\lambda \rightarrow \infty$; c'est la valeur maximale parmi les écarts en valeur absolue des m caractéristiques. Cela illustre, comme il a été mentionné précédemment (les puissances supérieures à deux accordent un poids plus important aux grands écarts), que lorsque λ tend vers l'infini, la valeur relative de l'écart maximal éclipse les valeurs des autres attributs à tel point qu'il devient le seul écart considéré.

Finalement, pour $\lambda = 2$, nous retrouvons la distance euclidienne qui se définit comme étant la racine de la somme des carrés des différences entre chacune des caractéristiques.

Le paramètre λ influence la forme de la fonction de distance. Il s'agit alors de justifier adéquatement le choix du paramètre à utiliser. Or, parmi ces mesures, la distance euclidienne est celle qui est la plus fréquemment utilisée. Il est donc important de souligner quelques-unes de ses propriétés et limites.⁹

Tout d'abord, sa valeur maximale n'est pas limitée. Elle dépend généralement du nombre de caractéristiques ainsi que de l'amplitude de leurs valeurs. Legendre et Legendre (1984) décrivent cette mesure de la manière suivante : «La distance euclidienne n'a pas de borne supérieure. Elle augmente à mesure que s'accroît le nombre de descripteurs, mais sa valeur dépend aussi de l'échelle de chacun des descripteurs, si bien qu'en changeant l'échelle de certains descripteurs, on peut obtenir des mesures de distance différentes qui ne sont pas monotones les unes aux autres.» L'ordonnement des distances peut donc différer selon les unités. La difficulté repose alors sur le choix de l'échelle de mesure à utiliser pour définir les valeurs à inclure dans les vecteurs propres à chaque objet.

D'autre part, le classement des distances pose problème dans la mesure où l'algorithme sous-jacent à la fonction de diversité ne passera pas nécessairement par le même sentier de calcul dépendamment du choix des unités. La différence entre les sentiers de calcul aura un impact sur le processus rendant les distances ultramétriques puisque ce processus ne transformera pas les mêmes distances selon les unités utilisées ce qui aura par le fait même des répercussions sur la représentation graphique de la structure de la diversité. L'exemple qui suit illustre les différents éléments de la discussion que nous venons d'aborder.

⁹ Les propriétés et limites de la distance euclidienne mentionnées ici s'appliquent aussi à la distance de Manhattan.

5.3.1 Exemple 4 : Modification de la structure de la diversité suite à un changement d'unité

La démarche que nous retenons ici consiste à calculer les distances entre différents objets multidimensionnels puis de recalculer ces mêmes distances en ayant préalablement modifié l'unité de mesure d'une des caractéristiques. L'objectif est de montrer qu'en plus d'altérer le classement des distances, le choix des unités peut avoir un impact sur la détermination des distances affectées par le processus de transformation rendant les distances ultramétriques. Nous montrons que ce sont certaines distances qui sont affectées lorsqu'elles sont construites selon certaines unités alors que ce sont d'autres distances qui sont affectées lorsqu'elles sont mesurées selon d'autres unités. Nous illustrons finalement l'effet que cela peut avoir sur la structure de la représentation graphique.

Le tableau 5.2 présente les valeurs des deux caractéristiques $\{1, 2\}$ appliquées à un ensemble de quatre objets $\{w, x, y, z\}$. Le tableau 5.3a) présente la matrice des distances euclidiennes obtenues à l'aide du tableau 5.2 alors que le tableau 5.3b) présente la matrice de ces mêmes distances lorsque les valeurs de la seconde caractéristique ont été décuplées.

Tableau 5.2

Valeurs de deux caractéristiques appliquées à quatre objets

	w	x	y	z
1	7	5	1	8
2	6	3	5	9

L'ordonnement des distances n'est pas le même selon l'unité de mesure utilisée. Pour le tableau 5.3a), $d_2(w, z) < d_2(w, x) < d_2(x, y) < d_2(w, y) < d_2(x, z) < d_2(y, z)$ alors que le tableau 5.3b) donne $d_2(w, y) < d_2(x, y) < d_2(w, z) < d_2(w, x) < d_2(y, z) < d_2(x, z)$. Il y a donc eu modification de leur classement suite à un changement d'échelle.

Tableau 5.3 a)

Matrice de distances initiale

	w	x	y	z
w	0	3.6	6.1	3.2
x	3.6	0	4.5	6.7
y	6.1	4.5	0	8.1
z	3.2	6.7	8.1	0

Tableau 5.3 b)

Matrice de distances modifiées

	w	x	y	z
w	0	30.1	11.7	30.0
x	30.1	0	20.4	60.1
y	11.7	20.4	0	40.6
z	30.0	60.1	40.6	0

En résolvant l'algorithme de la section 1.3.1 à l'aide des distances du tableau 5.3a), nous obtenons $V(\{\bullet\}) = d_2(w, z) + d_2(x, y) + d_2(y, z) + d_0$ alors que la fonction de diversité du tableau 5.3b) est donnée par $V(\{\bullet\}) = d_2(w, y) + d_2(w, z) + d_2(x, z) + d_0$. Le sentier de calcul n'est pas le même en raison de la différence de l'ordonnement des deux ensembles de distances. Appliquons maintenant le processus rendant les distances ultramétriques et traçons les représentations graphiques correspondantes.

Tableau 5.4 a)

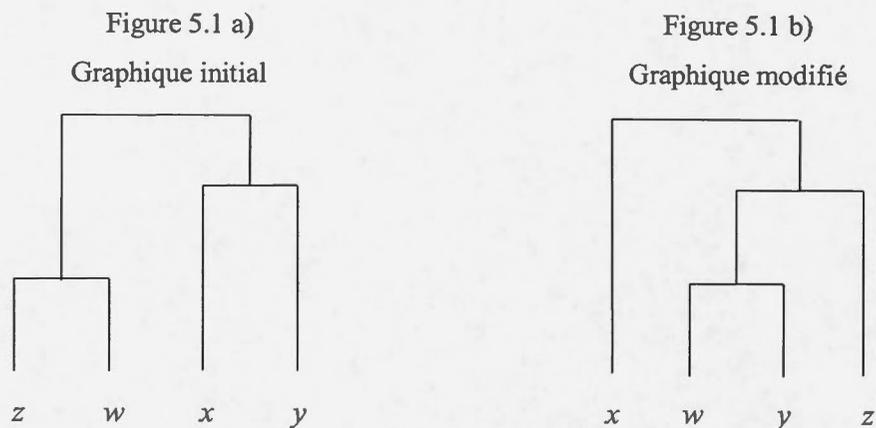
Matrice de distances ultramétriques

	w	x	y	z
w	0	8.1	8.1	3.2
x	8.1	0	4.5	8.1
y	8.1	4.5	0	8.1
z	3.2	8.1	8.1	0

Tableau 5.4 b)

Matrice de distances ultramétriques modifiées

	w	x	y	z
w	0	60.1	11.7	30.0
x	60.1	0	60.1	60.1
y	11.7	60.1	0	30.0
z	30.0	60.1	30.0	0



Nous constatons d'une part, que les distances qui ont été modifiées (en gras) ne sont pas nécessairement les mêmes dans les deux ensembles et d'autre part, que les arbres de diversité diffèrent l'un de l'autre. Dans la figure 5.1a), il y a deux sous arbres; un pour la paire $\{w, z\}$ et un autre pour $\{x, y\}$ alors que la figure 5.1b) n'en présente qu'un pour le triplet $\{w, y, z\}$ avec x dans une branche à part. De plus, dans la figure de gauche, y est un des deux objets ayant la plus grande diversité marginale alors que dans la figure de droite, il est un de ceux à distance minimale. ■

Les résultats de cet exemple s'expliquent par le fait que dans le cas d'objets multidimensionnels, les unités de mesures utilisées pour quantifier chacune des caractéristiques ont un impact sur l'ordonnancement des distances. Notons par ailleurs qu'il existe des mesures de distance qui ont l'avantage de ne pas augmenter en fonction du nombre de caractéristiques en plus d'être indépendantes des unités de mesure (voir Legendre et Legendre, 1984).

5.3.2 La distance euclidienne pondérée

Une autre particularité de la distance euclidienne provient du fait qu'elle accorde le même poids à chacune des caractéristiques. Une telle distance semble donc plutôt rigide dans la mesure où certains attributs peuvent être plus importants que d'autres. Il serait alors intéressant de disposer d'une mesure permettant de pondérer les valeurs que nous accordons à chacun d'entre eux. Si, par exemple, le nombre d'espèces est un élément important pour le chercheur alors que la fertilité du sol ne l'est pas vraiment, il est souhaitable que la mesure de distance et par le fait même, l'indice de diversité qui en résulte, soit plus sensible au nombre d'espèce qu'à la fertilité du sol. Une façon de résoudre ce problème consiste à introduire des coefficients pondérant les écarts des différents éléments à l'étude ce qui se traduit par la mesure de distance suivante :

$$d_p(x, y) = \sqrt{\sum_{i=1}^m \theta_i (x_i - y_i)^2} \quad (5.2)$$

où les coefficients θ_i , strictement positifs, pondèrent l'influence de la $i^{\text{ième}}$ caractéristique.¹⁰ La problématique réside alors dans la détermination de ces coefficients. En fait, le problème de la pondération s'apparente à celui du choix des unités puisqu'un changement d'échelle conduisant à multiplier par l les chiffres mesurant x_i et y_i équivaut à choisir $\theta_i = l$ au carré. Pour reprendre l'analogie de l'exemple 4, le calcul de la distance entre x et y où les valeurs de la seconde caractéristique mesurant x_2 et y_2 sont décuplées revient à poser $\theta_2 = 10^2 = 100$.

¹⁰ La distance euclidienne est un cas particulier de la distance pondérée où $\theta_i = 1$ pour tout i ; elle accorde une pondération unitaire à chacune des caractéristiques.

5.4 La diversité comme problème d'agrégation : l'exemple de la super-espèce

La discussion amorcée à la section 5.1 concernant les super-espèces consiste en fait en une question d'agrégation. Une super-espèce étant un agrégat composé de chacun des gènes de l'ensemble des espèces présentes dans un écosystème donné. Si l'on se concentre sur la valeur des espèces à des fins thérapeutiques, on doit retenir que leur valeur dépend des gènes qui peuvent être utilisés par l'industrie pharmaceutique. Chaque espèce i ($=1, \dots, m$) d'un écosystème x ($=1, \dots, n$) peut être décrite par une fonction similaire à une fonction d'utilité ou de production $f_x^i(x_1^i, \dots, x_l^i)$ où x_j^i représente le descripteur du $j^{\text{ième}}$ gène de la $i^{\text{ième}}$ espèce; l étant le nombre de gènes. Nous pouvons supposer sans perte de généralité que l est identique pour toutes les espèces en permettant que certaines variables x_j^i soient nulles.

Si l'on reste dans l'esprit de la méthode de Weitzman où la diversité est mesurée par une distance, l'agrégation des différentes espèces d'un écosystème x pour créer une super-espèce peut être décrite par la façon dont sont regroupées les caractéristiques de l'écosystème, ici la façon dont sont regroupés les gènes présents dans les différentes espèces évoluant dans l'écosystème. On mesurera toujours la biodiversité par une distance, mais ce sera la distance entre les agrégats ainsi constitués pour chaque super-espèce.

L'agrégation au niveau du site permet de représenter la contribution de ce dernier à la biodiversité par une fonction vectorielle $F_x(f_x^1(x_1^1, \dots, x_l^1), \dots, f_x^m(x_1^m, \dots, x_l^m))$. Là aussi, nous pouvons poser que m est identique pour tout écosystème en permettant que certaines fonctions f_x^i soient nulles; l'espèce i ne sera alors pas présente dans l'écosystème x . Plus généralement la contribution du site à la biodiversité peut être résumée par un vecteur de caractéristiques plus ou moins agrégées. Nous pouvons écrire :

$$F_x(f_x^1(x_1^1, \dots, x_l^1), \dots, f_x^m(x_1^m, \dots, x_l^m)) = F_x((x_1^1, \dots, x_l^1), \dots, (x_1^m, \dots, x_l^m)). \quad (5.3)$$

Le vecteur de dimension $l \times m$ du côté droit est alors agrégé en un vecteur de dimension m du côté gauche de l'expression. La répartition des gènes parmi les espèces est une caractéristique de la super-espèce.

On peut penser appliquer les diverses mesures de distance sur le vecteur «étendu» où les gènes sont distingués non seulement selon leur nature mais aussi selon leur hôte. Cette approche à l'avantage de la généralité, mais on peut se questionner sur son utilité étant donné que notre intérêt porte sur la présence de gènes dans un écosystème. Une agrégation de type

$$F_x \left(f_x^1(x_1^1, \dots, x_l^1), \dots, f_x^m(x_1^m, \dots, x_l^m) \right) = F_x \left(h(x_1^1, \dots, x_l^1), \dots, h(x_1^m, \dots, x_l^m) \right) \quad (5.4)$$

semble plus conforme aux attentes en la matière. Dans ce cas, la même fonction h est appliquée aux caractéristiques génétiques de chaque espèce pour en déterminer une caractéristique agrégée. Comme le vecteur de caractéristiques génétiques est différent pour chaque espèce, les valeurs que prend h sont différentes d'une caractéristique à l'autre.

La question de la mesure de diversité des écosystèmes en présence d'espèces-réservoirs de gènes se ramène au choix des fonctions agrégatives h et de la fonction de distance permettant d'apprécier la distance qui sépare le vecteur $F_x(h(x_1^1, \dots, x_l^1), \dots, h(x_1^m, \dots, x_l^m))$ d'un écosystème du vecteur de caractéristiques d'un autre écosystème.

La question de ces choix dépasse le cadre de ce mémoire. Nous pouvons cependant illustrer certaines de leurs conséquences. Nous avons vu précédemment que la distance de Minkowski permet de définir trois fonctions de distance : Manhattan ($\lambda=1$), Chebyshev ($\lambda \rightarrow \infty$) ou euclidienne ($\lambda=2$). Reste à définir la forme de la fonction h . Pour les fins de l'illustration, nous n'en considérerons que deux : la forme additive et la forme indicielle. Plus précisément :

$$\text{Forme additive : } h(x_1^j, \dots, x_l^j) = \sum_{i=1}^m x_j^i, \quad \forall j = 1, \dots, l,$$

$$F_x \left(f_x^1(x_1^1, \dots, x_l^1), \dots, f_x^m(x_1^m, \dots, x_l^m) \right) = F_x \left(\sum_{i=1}^m x_1^i, \dots, \sum_{i=1}^m x_l^i \right) \quad (5.5)$$

Forme indicielle : $h(x_j^1, \dots, x_j^m) = \min \left\{ 1, \sum_{i=1}^m x_j^i \right\}, \forall j = 1, \dots, l,$

$$F_x \left(f_x^1(x_1^1, \dots, x_l^1), \dots, f_x^m(x_1^m, \dots, x_l^m) \right) = F_x \left(\min \left\{ 1, \sum_{i=1}^m x_1^i \right\}, \dots, \min \left\{ 1, \sum_{i=1}^m x_l^i \right\} \right) \quad (5.6)$$

Dans le premier cas, la question de l'abondance des gènes parmi les espèces est prise en compte alors que dans le deuxième cas seule la présence du gène est retenue, peu importe que ce gène se retrouve parmi une ou plusieurs espèces.

Nous supposons dans ce qui suit que le descripteur x_j^i est une variable dichotomique égale à un si le gène j est présent dans l'espèce i et zéro sinon.¹¹ À partir de ces variables, nous obtenons une matrice de dimension $m \times l$ pour chaque écosystème décrivant la présence ou l'absence de chacun des l gènes de $i = 1, \dots, m$.

Pour illustrer notre propos, posons trois écosystèmes x , y et z contenant en tout sept espèces, certaines communes à plus d'un écosystème, d'autres étant propre à certains d'entre eux. L'écosystème x est composé des espèces $\{1, 2, 3\}$, y est composé de $\{3, 4, 5\}$ et z est composé de $\{2, 5, 6, 7\}$. Il y a six gènes dans ces sept espèces. Les matrices suivantes illustrent la répartition de ces gènes parmi les espèces de chaque écosystème.

Tableau 5.5 a)
Matrice des descripteurs x_j^i de l'écosystème x

x	1	2	3	4	5	6
1	1	1	1	1	1	0
2	1	1	1	0	0	0
3	1	0	0	1	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0
6	0	0	0	0	0	0
7	0	0	0	0	0	0

¹¹ L'espèce i ne fait pas partie de x si tous les descripteurs x_j^i pour $j = 1, \dots, l$ sont nuls.

Tableau 5.5 b)
Matrice des descripteurs y_j^i de l'écosystème y

y	1	2	3	4	5	6
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	1	0	0	1	0	0
4	0	0	1	1	1	0
5	1	0	1	1	1	0
6	0	0	0	0	0	0
7	0	0	0	0	0	0

Tableau 5.5 c)
Matrice des descripteurs z_j^i de l'écosystème z

z	1	2	3	4	5	6
1	0	0	0	0	0	0
2	1	1	1	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	1	0	1	1	1	0
6	0	0	1	0	0	1
7	0	0	1	0	1	1

Les vecteurs décrivant les super-espèces en fonction du nombre d'occurrence de chaque gène parmi les différentes espèces sont :

$$F_x(f_x^1(x_1^1, \dots, x_6^1), \dots, f_x^7(x_1^7, \dots, x_6^7)) = F_x\left(\sum_{i=1}^7 x_1^i, \dots, \sum_{i=1}^7 x_6^i\right) = F_x(3, 2, 2, 2, 1, 0),$$

$$F_y(f_y^1(y_1^1, \dots, y_6^1), \dots, f_y^7(y_1^7, \dots, y_6^7)) = F_y\left(\sum_{i=1}^7 y_1^i, \dots, \sum_{i=1}^7 y_6^i\right) = F_y(2, 0, 2, 3, 2, 0),$$

$$F_z(f_z^1(z_1^1, \dots, z_6^1), \dots, f_z^7(z_1^7, \dots, z_6^7)) = F_z\left(\sum_{i=1}^7 z_1^i, \dots, \sum_{i=1}^7 z_6^i\right) = F_z(2, 1, 4, 1, 2, 2).$$

En optant pour les vecteurs des super-espèces décrivant la présence/absence de chaque gène parmi les différents écosystèmes, nous obtenons :

$$F_x(f_x^1(x_1^1, \dots, x_6^1), \dots, f_x^7(x_1^7, \dots, x_6^7)) = F_x\left(\min\left\{1, \sum_{i=1}^7 x_1^i\right\}, \dots, \min\left\{1, \sum_{i=1}^7 x_6^i\right\}\right) = F_x(1, 1, 1, 1, 1, 0),$$

$$F_y(f_y^1(y_1^1, \dots, y_6^1), \dots, f_y^7(y_1^7, \dots, y_6^7)) = F_y\left(\min\left\{1, \sum_{i=1}^7 y_1^i\right\}, \dots, \min\left\{1, \sum_{i=1}^7 y_6^i\right\}\right) = F_y(1, 0, 1, 1, 1, 0),$$

$$F_z(f_z^1(z_1^1, \dots, z_6^1), \dots, f_z^7(z_1^7, \dots, z_6^7)) = F_z\left(\min\left\{1, \sum_{i=1}^7 z_1^i\right\}, \dots, \min\left\{1, \sum_{i=1}^7 z_6^i\right\}\right) = F_z(1, 1, 1, 1, 1, 1).$$

À partir de ces deux fonctions agrégatives, nous pouvons calculer les trois mesures de distances selon la valeur du paramètre λ pour la fonction de distance de Minkowski.

Tableau 5.6

Valeurs des distances issues de la distance de Minkowski
selon les formes additive et indicielle de h

	$\lambda = 1$	$\lambda = 2$	$\lambda \rightarrow \infty$
<i>h</i> additive			
$d_\lambda(F_x, F_y)$	5	2,6	2
$d_\lambda(F_x, F_z)$	8	3,5	2
$d_\lambda(F_y, F_z)$	7	3,6	2
<i>h</i> indicielle			
$d_\lambda(F_x, F_y)$	1	1	1
$d_\lambda(F_x, F_z)$	1	1	1
$d_\lambda(F_y, F_z)$	2	1,4	1

Nous constatons que le choix de la fonction agrégative h de même que le choix de la fonction de distance (représentée ici par le paramètre λ) détermine la distance entre les écosystèmes. Par exemple, pour la fonction agrégative basée sur l'occurrence des gènes (h

additive), les distances relatives sont grandement affectées par le choix du paramètre λ comme on peut le voir par les classements des distances suivants :

$$\begin{aligned}d_1(F_x, F_y) &< d_1(F_y, F_z) < d_1(F_x, F_z), \\d_2(F_x, F_y) &< d_2(F_x, F_z) < d_2(F_y, F_z), \\d_\infty(F_x, F_y) &= d_\infty(F_x, F_z) = d_\infty(F_y, F_z).\end{aligned}$$

Par ailleurs, si on prend une valeur particulière de λ , on constate que les distances varient considérablement selon le choix de la fonction agrégative h .

Le choix de l'approche utilisée ici pour construire les distances (euclidienne ou autre) peut avoir un impact sur l'ordonnement des distances et par le fait même sur le sentier de calcul issu de la fonction de diversité.

Comme on pouvait s'y attendre, la notion de diversité n'échappe pas aux problèmes d'agrégation que connaissent toutes les grandeurs multidimensionnelles. Le classement des écosystèmes peut différer selon le choix de la fonction h et du paramètre λ . Ce résultat ne peut être satisfaisant et justifie la poursuite des recherches méthodologiques sur la notion même de fonction de diversité.

CONCLUSION

Nous avons présenté dans ce mémoire un cadre d'analyse permettant de mesurer la biodiversité à l'aide de fonction de diversité. Puisque la biodiversité est un concept plutôt vague, les écosystèmes ont été considérés comme objet d'étude et nous les avons comparés entre eux sur la base des espèces qui les composent.

Nous avons par la suite établi certains liens entre les différentes fonctions, lesquelles décrivent chacune à leur manière la diversité comme étant le résultat d'une certaine forme d'agrégation des différences entre les écosystèmes d'un ensemble. Nous avons vu dans l'exemple 1 que la valeur de la mesure de diversité initiale n'est pas toujours unique lorsque la diversité marginale est construite sur la base de distance entre toute paire d'écosystèmes; en fait, celle-ci n'est bien définie que dans le cas particulier où les distances sont ultramétriques.

Nous avons ensuite présenté deux façons de contourner cette difficulté. La première, présentée dans les approches hiérarchique et lexicographique, consiste à modifier la forme de la mesure initiale à l'aide d'une approche axiomatique appropriée, tout en conservant la notion de diversité marginale construite à partir de mesures de distance. Il appert que ces deux approches sont toujours bien définies et équivalentes en toutes circonstances. Il est aussi possible de résoudre le problème de la non-unicité en se servant de l'approche basée sur le concept d'endémisme. Il s'agit alors d'utiliser la mesure de diversité initiale mais en modifiant la définition de la diversité marginale de façon à ce qu'elle exprime le nombre d'espèces endémique à l'écosystème considéré. Cette approche a la particularité de mesurer la diversité d'un ensemble d'écosystème de telle sorte que la valeur de la fonction soit égale au nombre d'espèces différentes qui y sont recensées.

Notre modélisation de la diversité nous a permis de présenter la mesure de différence entre toute paire d'écosystème proposée par Weikard (2002). Or, cette mesure n'est symétrique que lorsque tous les écosystèmes sont composés d'un même nombre d'espèces.

Pour contourner cette difficulté, nous avons proposé deux mesures de distance normalisées respectant la notion de symétrie en toutes circonstances.

Le chapitre IV a permis de faire le parallèle entre les différentes fonctions de diversité et les concepts de diversité α , β et γ généralement utilisés en écologie. Nous avons aussi vu que ces fonctions fournissent chacune à leur manière un cadre d'analyse permettant d'élaborer diverses politiques de préservation où l'objectif consiste à maximiser la valeur de la biodiversité sous contrainte budgétaire.

Notre façon de modéliser la biodiversité au chapitre III ne considère que la présence ou l'absence d'espèces parmi différents écosystèmes. Comme le suggère Weikard (2002), leur diversité peut se mesurer autrement qu'en termes d'espèces. Nous pouvons aussi la décrire en fonction d'une multitude de caractéristiques comme les biens et services fournis par chacun d'eux. De telles considérations ont été présentées dans le dernier chapitre en utilisant d'autres mesures de distance applicables à des objets multidimensionnels de façon à ce qu'elles reflètent les caractéristiques que nous souhaitons analyser.

Finalement, il serait intéressant que des notions comme l'abondance relative des espèces ou les interactions entre espèces et entre écosystèmes soient incluses dans l'élaboration d'un indice plus complet. Il serait aussi pertinent d'y incorporer une mesure de la perturbation d'un écosystème par rapport à son état naturel (voir Khazri et Lasserre (2006)). Un indice de biodiversité ne serait pas complet sans que nous y ajoutions une mesure de la valeur subjective que les agents économiques accordent à la biodiversité comme le propose Weitzman (1992, 1998).

BIBLIOGRAPHIE

- Armsworth, P. R., B. E. Kendall et F. W. Davis. 2004. "An introduction to biodiversity concepts for environmental economists". *Resource and Energy Economics*, vol.26, p.115-136.
- Barberà, S., W. Bossert et P. K. Pattanaik. 2001. "Ranking sets of objects". *Département de sciences économiques*, Université de Montréal, Cahier 2001-02, 57 p.
- Baumgärtner, S. 2006. "Measuring the diversity of what? And for what purpose? A conceptual comparison of ecological and economic biodiversity indices". *Department of Economics*, University of Heidelberg, 33 p.
- Bossert, W., P. K. Pattanaik et Y. Xu. 2001. "The measurement of diversity". *Département de sciences économiques*, Université de Montréal, Cahier 2001-17, 11 p.
- Church, R. L., D. M. Stoms et F. W. Davis. 1996. "Reserve selection as a maximal covering location problem". *Biological Conservation*, vol.76, p.105-112.
- deGroot, R. S., M. A. Wilson et R. M.J. Boumans. 2002. "A typology for the classification, description and valuation of ecosystem functions, goods and services". *Ecological Economics*, vol.41, p.393-408.
- Eiswerth, M. E. et J. C. Haney. 2001. "Maximising conserved biodiversity : why ecosystem indicators and thresholds matter". *Ecological Economics*, vol.38, p.259-274.
- Khazri, O. et P. Lasserre. 2006. "The natural state as a basis for measuring biodiversity". *Département des Sciences Économiques*, UQAM, 28 p.
- Knoblauch, V. 2000. "Lexicographic orders and preference representation". *Journal of Mathematical Economics*, vol.34, p.255-267.
- Legendre, L. et P. Legendre. 1984. *Écologie numérique*. *Les Presses de l'Université du Québec*, Québec.
- Maignan, C., G. Ottaviano, D. Pinelli et F. Rullani. 2003. "Bio-ecological diversity vs. socio-economic diversity : a comparison of existing measures". *Nota Di Lavoro*, 13, Fondazione Eni Enrico Mattei, 33 p.

- Margules, C. R., A. O. Nicholls et R. L. Pressey. 1988. "Selecting networks of reserves to maximise biological diversity". *Biological Conservation*, vol.43, p.63-76.
- Nehring, K., C. Puppe. 2002. "A theory of diversity". *Econometrica*, vol.70, p.1155-1198.
- Polasky, S., A. R. Solow. 1995. "On the value of a collection of species". *Journal of Environmental Economics and Management*, vol.29, p.298-303.
- Roosen, J., A. Fadlaoui et M. Bertaglia. 2003. "Economic evaluation and biodiversity conservation of animal genetic resources". *Department of Food Economics and Consumption Studies*, University of Kiel, FE Working Paper #0304, 73 p.
- Sinclair-Desgagné, B. 2005. "Analyse économique et préservation de la biodiversité". *Économie publique*, 16-2005/1, 16 p.
- Solow, A., S. Polasky et J. Broadus. 1993. "On the measurement of biological diversity". *Journal of Environmental Economics and Management*, vol.24, p.60-68.
- Underhill, L. G. 1994. "Optimal and suboptimal reserve selection algorithms". *Biological Conservation*, vol.70, p.85-87.
- Vane-Wright, R. I., C. J. Humphries et P. H. Williams. 1991. "What to protect? – Systematics and the agony of choices". *Biological Conservation*, vol.55, p.235-254.
- Weikard, H.-P. 1998. "On the measurement of diversity". *Institute of Public Economics*, University of Graz, Working Paper No. 9801, 12 p.
- Weikard, H.-P. 2002. "Diversity functions and the value of biodiversity". *Land Economics*, vol.78, p.20-27.
- Weitzman, M. L. 1992. "On Diversity". *The Quarterly Journal of Economics*, vol.107, p.363-405.
- Weitzman, M. L. 1993. "What to preserve? An application of diversity theory to crane conservation". *The Quarterly Journal of Economics*, vol.108, p.157-183.
- Weitzman, M. L. 1998. "The Noah's Ark Problem". *Econometrica*, vol.66, p.1279-1298.