



REPUBLIC ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de L'Enseignement Supérieur et de la Recherche Scientifique

Université Abdelhafid BOUSSOUF Mila



## Méthode d'étude et inventaire de la faune et la flore

•**Semestre** : 4

•2ème année Ecologie

Chargée de cours: Dr. BENSARKHI Zinette

Année universitaire: 2019/ 2020

## **Chapitre 4 : Collecte et analyses des données faunistiques et floristiques**

### **1. Présentation des données**

Quel que soit le type d'évaluation menée, il est essentiel de bien choisir les méthodes de collecte et d'analyse des données et de les appliquer correctement

Points principaux

1- Les méthodes de collecte et d'analyse de données doivent être choisies en fonction de l'évaluation concernée, de ses questions clés d'évaluation et des ressources disponibles.

2- Les méthodes de collecte et d'analyse de données doivent être sélectionnées de manière à assurer la complémentarité entre les points forts et les faiblesses des unes et des autres.

### **2. Préparation des données**

Après la collecte des données suit la préparation des données. La préparation des données, parfois appelée « **Pré-traitement** », est l'étape pendant laquelle les données brutes sont nettoyées et structurées en vue de l'étape suivante du traitement des données. Pendant cette phase de préparation, les données brutes sont vérifiées avec soin afin de déceler d'éventuelles erreurs. L'objectif est d'éliminer les données de mauvaise qualité (**redondantes, incomplètes ou incorrectes**) et de commencer à créer les données de haute qualité qui peuvent garantir la qualité de votre environnement de travail Exemple :

Données sur la faune et la flore d'une région ou d'un site c'est à dire tout dépend de l'échelle spatiale et temporelle (Diversité, richesse, distribution de la faune et la flore).

### **3. Importation des données**

Les données propres sont ensuite importées dans leur emplacement de destination (par exemple : un dossier qui contient un fichier Excel ou convertie vers un format supporté par cette destination). L'importation des données est la première étape au cours de laquelle les données brutes commencent à se transformer en information exploitable.

### **4. Traitement de données**

Le traitement des données est exécuté dès que celles-ci sont collectées, en vue de les traduire en information exploitable: Le traitement des données est généralement effectué par un spécialiste data scientist, biostatisticien, ...etc. (ou une équipe). Il est important qu'il soit

effectué correctement afin de ne pas impacter négativement le produit final ou la sortie des données.

Le traitement des données commence avec les données brutes propres :

Convertit les données sous une forme plus lisible (graphiques, documents de texte, ...etc.) en leur donnant le format et contexte nécessaires pour qu'elles puissent être interprétées par l'intéressé.

Exemple :

Des données sur la faune et la flore d'une région ou d'un site, c'est à dire tout dépend de l'échelle spatiale et temporelle (Diversité, richesse, distribution de la faune et la flore).

Les objectifs de cette étude est d'inventorier (richesse spécifique) la faune ou la flore avec les abondances et la densité qui existe dans un site bien déterminé, bien-sûr toutes les communautés (faune et flore) ont des attributs ou des caractéristiques qui diffèrent de ceux des composants qui composent la communauté et qui n'ont de sens que par rapport à l'assemblage collectif ou à la communauté.

Ces attributs sont:

- Nombre d'espèces (Richesse)
- Abondance des espèces et densité

1- **Abondance** : L'abondance des espèces est généralement basée sur les nombres d'individus par espèce, ou sur des variables telles que le pourcentage de couverture ou la biomasse des espèces.

2- **Densité des espèces**: le nombre d'individus d'une espèce collectée dans une zone totale particulière. Pour les échantillons de quadra ou d'autres méthodes d'échantillonnage d'une zone fixe, la densité des espèces est exprimée en unités d'espèces par zone spécifiée. exemple : 10 ind/m<sup>2</sup>, 100/quadra de 10 m<sup>2</sup>...etc.

3- **Estimation de la richesse spécifique :**

a- **Indice de Shannon**

$$H' = - \sum_{i=1}^s P_i \log p_i$$

Avec  $p_i$  = abondance proportionnelle ou pourcentage d'importance de l'espèce, se calcule ainsi :  $p_i = n_i/N$  ;  $S$  = nombre total d'espèces ;

$n_i$  = nombre d'individus d'une espèce dans l'échantillon ;

$N$  = nombre total d'individus de toutes les espèces dans l'échantillon.

Ainsi, la valeur de  $H'$  dépend du nombre d'espèces présentes, de leurs proportions relatives, de la taille de l'échantillon ( $N$ ) et de la base de logarithme. Le choix de la base de logarithme est arbitraire (Valiela, 1995) mais, lorsqu'on compare des indices, la base utilisée doit être mentionnée et être la même.

La dominance marquée d'une espèce révèle une faible diversité, alors que la co-dominance de plusieurs espèces révèle une grande diversité. Puisque l'équation est un estimateur biaisé (Valiela, 1995).

Cet indice permet de quantifier l'hétérogénéité de la biodiversité d'un milieu d'étude et donc d'observer une évolution au cours du temps. Cet indice varie toujours de 0 à  $\ln S$  (ou  $\log S$  ou  $\log_2 S$ , selon le choix de la base du logarithme)

**b- Indice d'équitable**

c- C'est le rapport de la diversité observée à une distribution de fréquence des espèces complètement égale (échelle de 0-1), peut être quantifiée séparément à l'aide de l'indice Shannon-Wiener comme étant :

**d-  $J' = H'/H'max$**

e- où  $H'$  est la diversité spécifique observée et  $H'max$  est le logarithme du nombre total d'espèces ( $S$ ) dans l'échantillon (Gray et al., 1992).

f- Par exemple, deux espèces présentant **50** individus chacune représenteraient une équitabilité ou une égalité complète d'une valeur de **1**. Deux espèces, dont la première est représentée par **un** individu, et la deuxième, par **99**, obtiendraient seulement un résultat de 0,08.

- **Traitements statistiques et méthodes multi-variées par l'identification des groupements d'espèce :**

#### **A. Historique :**

Les bases théoriques de ces méthodes sont anciennes et sont principalement issues de “psychomètres” américains : Spearman (1904) et Thurstone (1931, 1947) pour l'Analyse en Facteurs, Hotelling (1935) pour l'Analyse en Composantes Principales et l'Analyse Canonique, Hirschfeld (1935) et Guttman (1941, 1959) pour l'Analyse des Correspondances. Pratiquement, leur emploi ne s'est généralisé qu'avec la diffusion des moyens de calcul dans le courant des années 60. Sous l'appellation “**Multivariate Analysis**” elles poursuivent des objectifs sensiblement différents à ceux qui apparaîtront en France. Un individu ou unité statistique n'y est souvent considéré que pour l'information qu'il apporte sur la connaissance des liaisons entre variables au sein d'un échantillon statistique dont la distribution est le plus souvent soumise à des hypothèses de normalité.

En France, l'expression “Analyse des Données” recouvre les techniques ayant pour objectif la description statistique des grands tableaux ( $n$  lignes, où  $n$  varie de quelques dizaines à quelques milliers,  $p$  colonnes, où  $p$  varie de quelques unités à quelques dizaines). Ces méthodes se caractérisent par une utilisation intensive de l'ordinateur, leur objectif exploratoire et une absence quasi systématique d'hypothèses de nature probabiliste au profit des propriétés et résultats de géométrie euclidienne. Elles insistent sur les représentations graphiques en particulier de celles des individus qui sont considérés au même titre que les variables.

Depuis la fin des années 1970, de nombreux travaux ont permis de rapprocher ou concilier les deux points de vue en introduisant, dans des espaces multidimensionnels appropriés, les outils probabilistes et la notion de modèle, usuelle en statistique inférentielle. Les techniques se sont ainsi enrichies de notions telles que l'estimation, la convergence, la stabilité des résultats, le choix de critères. . . L'objectif essentiel de ces méthodes est l'aide à la compréhension de volumes de données souvent considérables. Réduction de dimension, représentation graphique optimale, recherche de facteurs ou variables latentes... sont des formulations équivalentes.

## **B. Méthodes**

Les méthodes de Statistique exploratoire multidimensionnelle se classifient selon leur objectif (**réduction de dimension** ou **classification**) et le type des données à analyser (quantitatives et/ou qualitatives) :

### ➤ **Description et réduction de dimension (méthodes factorielles) :**

1. Analyse en Composantes Principales (p variables quantitatives).
2. Analyse Factorielle Discriminante (p variables quantitatives, 1 variable qualitative).
3. Analyse Factorielle des Correspondances simple (2 variables qualitatives) et Multiple (p variables qualitatives).
4. Analyse Canonique (p et q variables quantitatives).
5. Multi-dimensional Scaling (M.D.S.) ou positionnement multidimensionnel ou analyse factorielle d'un tableau de distances. Toutes les précédentes méthodes sont basées sur des outils classiques de géométrie euclidienne qui sont développés dans les rappels et compléments d'algèbre linéaire.
6. Non negative Matrix Factorisation ou NMF.

### ➤ **Méthodes de classification :**

1. Classification ascendante hiérarchique,
2. Algorithmes de réallocation dynamique,
3. Cartes de Kohonen (réseaux de neurones)

## **5. Sortie et interprétation des données**

Lors de l'étape de sortie/interprétation, les données deviennent exploitables par tous, y compris ceux qui n'ont pas les compétences d'un spécialiste. Elles sont converties, deviennent lisibles et sont généralement présentées sous forme de graphiques...etc.