# Protocole

## Matériel :

* Logiciel Rstudio
* Logiciel Excel

## Méthode

**1.** Calculer le pourcentage d’individus identifier jusqu’à l’espèce par parcelle afin de conserver uniquement les parcelles avec un taux d’identification supérieur à 80%. Effectuer cette étape pour les 2 réseaux TmFO et OEFo.

Remarque : Déjà effectué pour le réseau TmFO.

* **Packages utilisés :** plyr ; httr
* **Données utilisées :** variables « site », « plot », « species » des données brutes
* **Techniques utilisées :**
* Calculer le nombre d’individus non identifiés (NA, sp dans la colonne species) grâce à la fonction « count » du package « plyr »
* Calculer le nombre d’individus total
* Utiliser l’équation pour obtenir le pourcentage d’individus identifiés par parcelle
* **Forme des résultats attendus :** Obtention d’un tableau CSV

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **site** | **plot** | **nb tot ind** | **nb ind no id** | **% ind id** |
|  |  |  |  |  |

**2.** Ouvrir la base de données et former les deux bases de données des 2 réseaux.

Remarque : Utilisation du script de Camille Piponiot

* **Packages utilisés :** data.table ; plyr ; httr ; Biomass ; stringdist
* **Données utilisées :** Données brutes TmFO et OEFo :
* Tableaux CSV pour chaque site de forêts de production TmFO
* Tableaux Excel pour chaque site de forêts de production d’OEFo
* Tableaux Excel regroupant tous les sites de forêts secondaires d’OEFo
* **Techniques utilisées :**
* Réseau TmFO :
* Division des colonnes volume\_year et DBH\_year en colonnes year, DBH, Volume pour les parcelles dont les tables sont configurées comme cela.
* Enlever les mesures dupliquées.
* Création d’un tableau tree\_data qui rassemble les données pour chaque arbre : site, parcelle, sous-parcelle, id de l’arbre (site\_plot\_subplot\_treenumb), nom scientifique (genre et sp), statut d’exploitation (true/false), coordonnée x, coordonnée y
* Création d’un tableau data qui sert pour le DBH. Il rassemble Id de l’arbre, l’année de mesure et le DBH
* Ajouter la colonne « status », y inscrire 1 si l’arbre est toujours vivant l’année de mesure sinon 0
* Effectuer une correction taxonomique
* Création d’un tableau plot\_data qui renseigne sur le traitement de chaque sous-parcelle : site, parcelle, sous-parcelle, traitement sylvicole
* Assembler les variables utiles (colonnes) des différents tableaux dans un nouveau dataframe et les lignes avec des NA dans la colonne « species »
* Réseau OEFo :
* Ouvrir les tables en format csv sur R studio
* Enlever les mesures dupliquées
* Effectuer une correction taxonomique avec le package Biomass
* Convertir le DBH étant en mm en cm
* Correction du DBH si nécessaire : prendre comme seuil max 250 cm pour vérifier qu’il n’y a pas d’oublis de virgules
* Ajouter la colonne « status », y inscrire 1 si l’arbre est toujours vivant l’année de mesure sinon 0
* Enlever les variables (colonnes) inutiles et les lignes avec des NA dans la colonne « species »
* **Forme des résultats attendus :**

Obtention de 3 bases de données sous forme de tableau CSV :

* On obtient 2 bases de données des parcelles en forêt de production, une pour le réseau TmFO et une pour le réseau OEFo :

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **site** | **plot** | **subplot** | **treat** | **plot.size** | **idtree** | **year** | **genus** | **species** | **name** | **Coord x** | **Coord y** | **logged** | **dbh (cm)** | **status** |
|  |  |  | ***RIL, CL, silv, silv+, ctrl*** |  | ***site\_plot\_***  ***subplot\_***  ***numtree*** |  |  |  | ***Genus sp*** |  |  | ***False/true*** |  | ***1/0*** |

* On obtient une 3ème base de données des parcelles en forêt secondaires d’OEFo :

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **site** | **plot** | **subplot** | **treat** | **plot.size** | **idtree** | **year** | **genus** | **species** | **name** | **Coord x** | **Coord y** | **dbh (cm)** | **status** |
|  |  |  | ***operated*** |  | ***site\_plot\_***  ***subplot\_***  ***numtree*** |  |  |  | ***Genus sp*** |  |  |  | ***1/0*** |

Remarques :

Pour chaque base, on a une ligne pour chaque année de mesure d’un arbre.

Les coordonnées x et y sont respectivement la longitude et la latitude.

**3.** Calculer pour chaque parcelle, le nombre d’individus par espèce et par année.

* **Packages utilisés :** plyr ; httr ; data.table
* **Données utilisées :** Les variables « site, plot, idtree, year, name (genus species) » des bases de données csv obtenues dans l’étape précédente
* **Techniques utilisées :**
* Calculer pour chaque année de mesure d’une mesure, le nombre d’individus échantillonnés par espèces grâce à la fonction « count » du package « plyr »
* Rentrer les résultats dans un tableau
* **Forme des résultats attendus :** Une matrice contenant les abondances des espèces dont les lignes sont les espèces, les colonnes sont les différentes années de mesure d’une parcelle donnée :

|  |  |  |  |
| --- | --- | --- | --- |
|  | Plot (year 1) | Plot (year 2) | Plot (year 3) |
| Sp1 |  |  |  |
| Sp2 |  |  |  |
| Sp3 |  |  |  |

**4.** Calculer l’évolution des diversités alpha et béta dans le temps après perturbation :

* **Packages utilisés :** Entropart (Marcon & Hérault, 2015)
* **Données utilisées :** La matrice obtenue dans l’étape précédente contenant les abondances des espèces chaque année pour chaque parcelle.

On aura aussi besoin d’un vecteur contenant les poids de la parcelle à chaque année de mesure. Ici, le poids sera égal pour chaque année de mesure (=1) :

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Plot (year 1) | Plot (year 2) | Plot (year 3) | Plot (year 4) | … |
| weights | 1 | 1 | 1 | 1 | … |

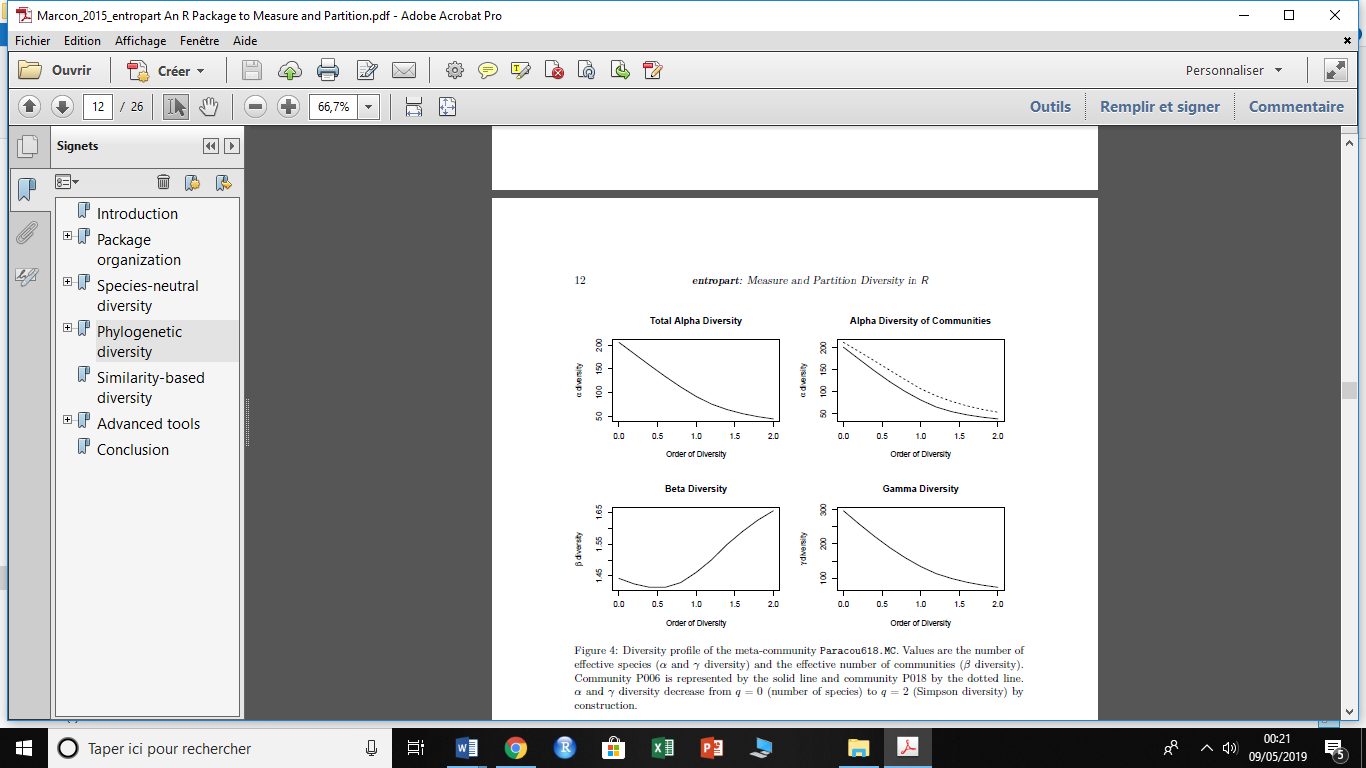
* On adapte le package à nos données : Métacommunauté = parcelle ; Communauté = Parcelle à un instant t (temps en année)
* **Techniques utilisées :**
* Calcul de l’entropie HCDT α et β pour chaque parcelle au cours du temps et chaque indice de diversité (Richesse spécifique ⬄ ordre de diversité q=0, Shannon ⬄ q = 1, Simpson ⬄ q=2). Pour cela, on va utiliser la fonction « Divpart » qui calcule toutes les diversités en une seule fois, càd la diversité alpha de chaque communauté, les diversités alpha, béta et gamma de la métacommunauté. On peut aussi utiliser la fonction « DivEst » qui a la même fonction mais elle estime les intervalles de confiances des diversités afin de réduire le biais d’espèces non vues.

La diversité alpha étant le nombre d’espèces présentes sur la parcelle, elle se calcule pour chaque année. A contrario, la diversité béta étant le taux de remplacement des espèces dans le temps, elle se calcule entre 2 années.

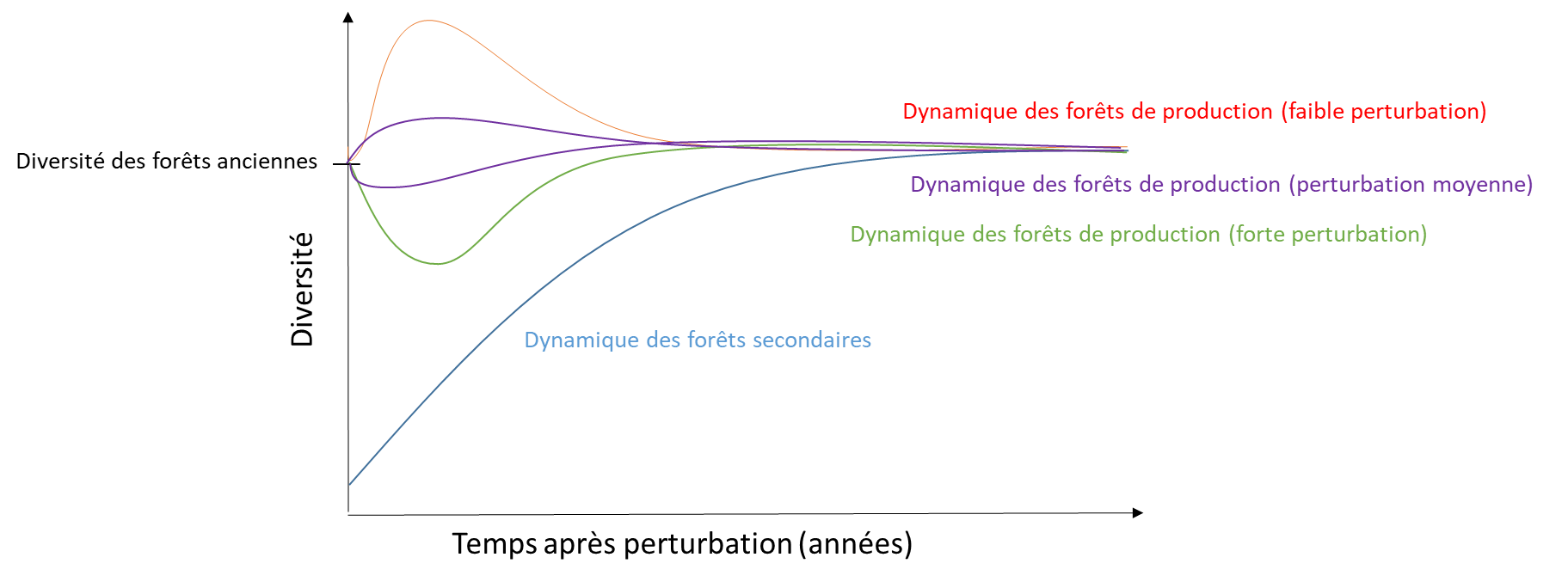
* Calculer et tracer le profil de diversité grâce à la fonction « DivProfile » afin de voir pour quel ordre de diversité (représentant les indices de diversité), la diversité est la plus élevée.
* **Forme des résultats attendus :** Obtention de deux tableaux CSV, un par diversité (alpha et béta) pour chaque parcelle et chaque indice (richesse spécifique, shannon et simpson)

|  |  |
| --- | --- |
| **Year of measurement** | **Diversity α** |
| Year 1 |  |
| Year 2 |  |
| Year 3 |  |

|  |  |
| --- | --- |
|  | **Diversity β** |
| Between years 1 and 2 |  |
| Between years 2 and 3 |  |
| Between years 3 and 4 |  |

Ainsi que les profils de diversité sous forme de graphique :

**4.** Construire les trajectoires de biodiversité = Constructions graphiques de l’évolution des diversités α et β dans le temps après perturbation en fonction des parcelles et des indices de diversité (la richesse spécifique, l’indice de Shannon et l’indice de Simpson).

* **Packages utilisés :** ggplot
* **Données utilisées :** Le tableau csv obtenu lors de l’étape 3 avec dans les colonnes, la variable densité α ou β par année et en ligne les parcelles
* **Techniques utilisées :** Grâce à la fonction plot (y ~ x), construire un graphique diversité (α ou β) en fonction du temps (années) après perturbation
* **Forme des résultats attendus :** Graphiques des trajectoires de biodiversité α et β pour chaque parcelle et chaque indice de diversité

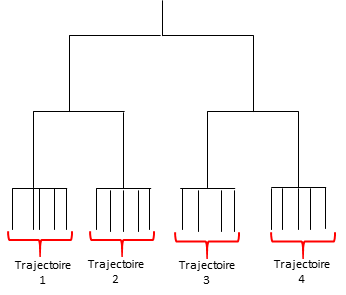
**5.** Regrouper les parcelles par pattern afin d’avoir les différents types de trajectoires représentés dans notre gamme de parcelles.

* **Packages utilisés :** Aucun
* **Données utilisées :** Matrice CSV contenant les indices alpha ou béta dont les lignes sont les parcelles et les colonnes sont les années de mesure :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Value of diversity index for each year of measurement** | | | |
| **Plot** | Index 1 | Index 2 | Index 3 | … |
| P1 |  |  |  |  |
| P2 |  |  |  |  |
| … |  |  |  |  |

* **Technique utilisée :** La classification ascendante et hiérarchique :

A partir du tableau précédent, on va calculer, pour chaque parcelle, la distance entre les valeurs de l’indice d’une année sur l’autre. On va réunir les parcelles qui ont les mêmes distances entre les valeurs de l’indice le long de la trajectoire.

* **Forme des résultats attendus :** Obtention d’un dendrogramme des différents types de trajectoires :

**6.** Analyser l’influence des facteurs intensité de perturbation pour les forêts de production et usages passés des terres pour les forêts secondaires.

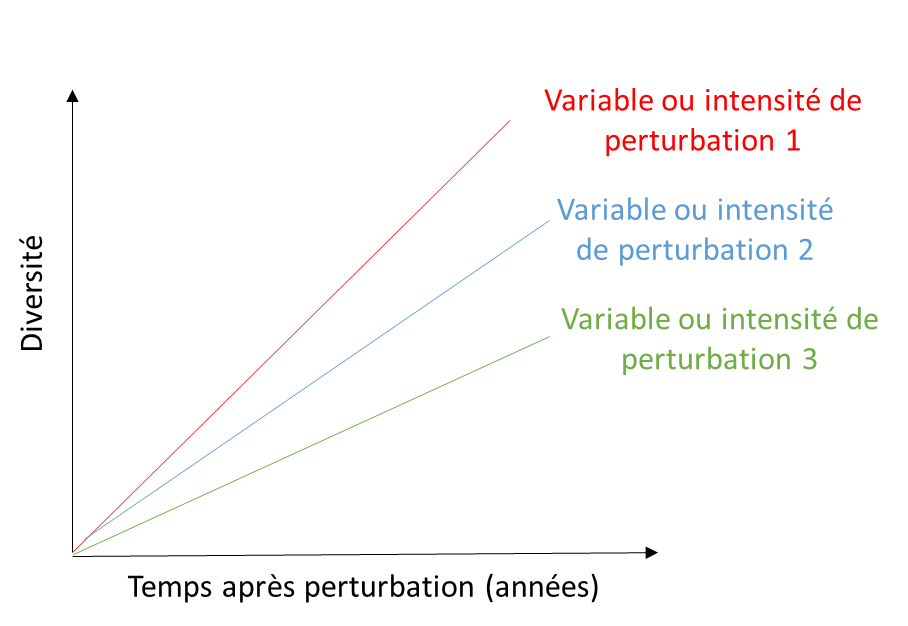
Remarque : Ce sera une perspective ou un résultat en fonction de l’avancée des étapes précédentes.

* **Packages utilisés :** Biomass, data\_table
* **Données utilisées :** variables des bases de données : site, plot, idtree, year, genus, species, name, coord x, coord y, dbh
* **Techniques utilisées :**
* Calcul de l’intensité de perturbation pour les bases de données des forêts de production TmFO et OEFo  (Réjou‐Méchain *et al.*, 2017):
* Trouver les coordonnées x et y des sites pour lesquels la valeur est manquante
* Retrouver la densité de bois de chaque espèce grâce à la fonction « getWoodDensity » grâce aux variables genre et espèce
* Calculer la biomasse (AGB) sans la hauteur grâce aux coordonnées x et y avec la fonction « computeAGB » avant et après exploitation
* Calculer la différence de biomasse avant et après exploitation et le convertir en pourcentage pour obtenir l’intensité de perturbation
* Création d’une nouvelle table :

|  |  |
| --- | --- |
| **Plot** | **Disturbance intensity (%)** |
| P1 |  |
| P2 |  |
| …. |  |

* Création d’une table avec la variable nominale « usage passé » pour les forêts secondaires OEFo :

|  |  |
| --- | --- |
| **Plot** | **past use** |
| P1 |  |
| P2 |  |
| … |  |

* Trouver s’il y a corrélation entre les types de trajectoire et les variables « intensité de perturbation » pour les forêts de production et « usage passé des terres » pour les forêts secondaires. Pour calculer la relation entre deux variables nominales (le type de trajectoire et l’usage passé), on fera le test chi² avec la fonction « chisq.test ». Pour calculer la relation entre une variable nominale (le type de trajectoire) et une variable continue (l’intensité de perturbation), on fera une anova avec la fonction « anova ».
* **Forme des résultats attendus :**