

Analyse multivariée paramétrique à partir de matrice de distance

Abdoulaye Diabakhaté

5 juillet 2018

Plan de la présentation

- 1 Les données
- 2 Question
- 3 Description du problème
- 4 Inspection des variables
- 5 Méthode Adonis
- 6 Sélection de variable : bioenv

Les données de TARA

- 644 observations
- 15 variables :
 - Lat, Long, T, Sal, chl_a, O₂, NO_{3m}, NO₃, NO₂, NH₄, SSD, Phos, Si, depth, Fe
- 159 stations
- 6 fractions de taille :
 - 0 – 0.2, 0.22 – 3, 20 – 180, 5 – 20, 180 – 2000 et 0.8 – 5
 - Douzaine de matrices de distances pour chaque fraction

Question ?

La question que l'on se pose :

Quelle est l'influence des paramètres sur les matrices de distances ?

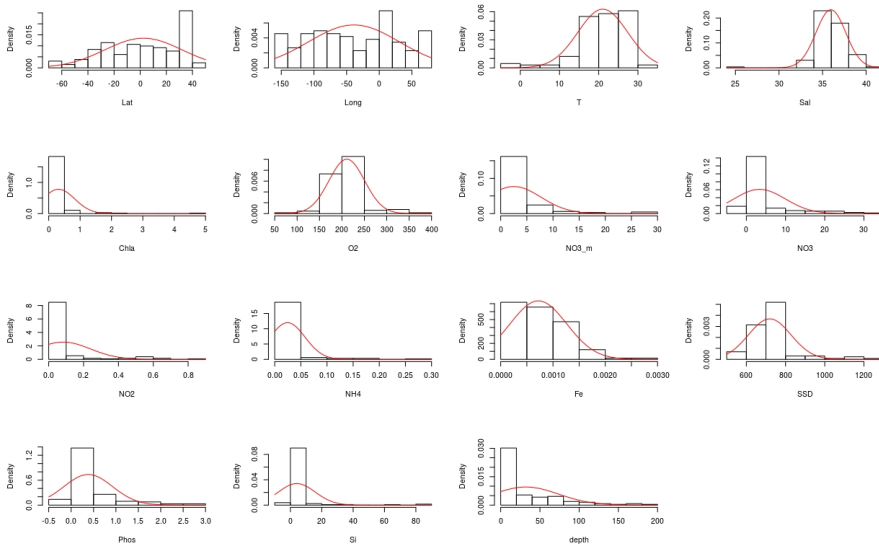
Description du problème

Notre modèle est :

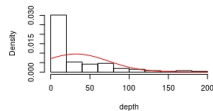
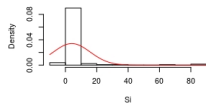
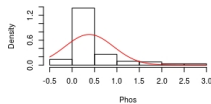
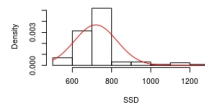
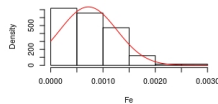
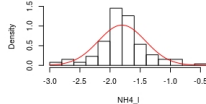
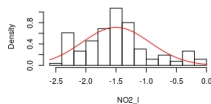
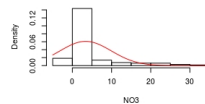
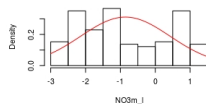
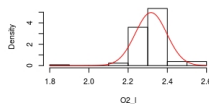
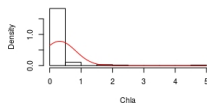
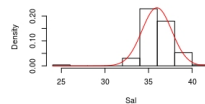
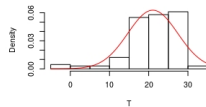
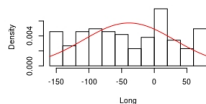
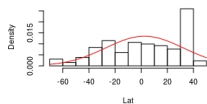
$$D_{ij} = |Y_i - Y_j| \text{ et } Y = X\beta + U$$

- $D = (D_{ij})$, matrice de distance
- X , matrice des covariables

Densité avant log-transformation



Densité après log-transformation



Résultat de la méthode Adonis



Best Subset Of Environmental Variables With Maximum (Rank) Correlation With Community Dissimilarities

- `vegdist`
- Sous-ensemble possibles de variables environnementales
- Distances Euclidiennes (`dist`)
- Corrélation
- $2^P - 1$

La corrélation est calculée entre les distances génétiques (par exemple `sorensen_prev`) et les distances environnementales et la méthode s'arrête quand rajouter des variables n'augmente plus la corrélation.

Résultat de la méthode bioenv

