

# Données spatiales

## Données

Le fichier `bryo_belg.csv` est adapté des données de l'étude:

Neyens, T., Diggle, P.J., Faes, C., Beenaerts, N., Artois, T. et Giorgi, E. (2019) Mapping species richness using opportunistic samples: a case study on ground-floor bryophyte species richness in the Belgian province of Limburg. *Scientific Reports* 9, 19122. <https://doi.org/10.1038/s41598-019-55593-x>

Ce tableau de données indique la richesse spécifique des bryophytes au sol (*richness*) pour différents points d'échantillonnage de la province belge de Limbourg, avec leur position ( $x$ ,  $y$ ) en km, en plus de l'information sur la proportion de forêts (*forest*) et de milieux humides (*wetland*) dans une cellule de 1 km<sup>2</sup> contenant le point d'échantillonnage.

```
bryo_belg <- read.csv("../donnees/bryo_belg.csv")
head(bryo_belg)
```

##	richness	forest	wetland	x	y
## 1	9	0.2556721	0.5036614	228.9516	220.8869
## 2	6	0.6449114	0.1172068	227.6714	219.8613
## 3	5	0.5039905	0.6327003	228.8252	220.1073
## 4	3	0.5987329	0.2432942	229.2775	218.9035
## 5	2	0.7600775	0.1163538	209.2435	215.2414
## 6	10	0.6865434	0.0000000	210.4142	216.5579

## Ajustement d'un modèle géostatistique

Pour cet exercice, nous utiliserons la racine carrée de la richesse spécifique comme variable réponse. La transformation racine carrée permet souvent d'homogénéiser la variance des données de comptage afin d'y appliquer une régression linéaire.

*Note:* Pour modéliser directement les données de comptage, par exemple avec une distribution de Poisson, tout en incluant la dépendance spatiale, il faudrait passer par un modèle de vraisemblance personnalisé ou un modèle bayésien.

- Ajustez un modèle linéaire de la richesse spécifique transformée en fonction de la fraction de forêt et de milieux humides, sans tenir compte des corrélations spatiales. Quel est l'effet des deux prédicteurs selon ce modèle?
- Calculez le variogramme empirique des résidus du modèle en (a) et ajustez un modèle théorique pour ce variogramme avec les fonctions du package *gstat*. Parmi les modèles exponentiel, gaussien et sphérique, lequel convient le mieux à ces données?

*Note:* L'argument `cutoff` de la fonction `variogram` spécifie la distance maximale à laquelle le variogramme est calculé. Vous pouvez ajuster manuellement cette valeur pour vous assurer que le palier du variogramme soit atteint.

- Ré-ajustez le modèle linéaire en (a) avec la fonction `gls` du package *nlme*, puis ajoutez-y un terme de corrélation spatiale basé sur le meilleur modèle déterminé en (b). Comparez l'AIC des modèles avec et sans corrélation spatiale.

- d) Quel est l'effet de la fraction de forêts et de milieux humides selon le modèle en (c)? Expliquez les différences entre les conclusions de ce modèle et du modèle en (a).