

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 km² 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?

Solution

```
bryo_belg$sr <- sqrt(bryo_belg$richness)

bryo_lm <- lm(sr ~ forest + wetland, data = bryo_belg)
summary(bryo_lm)

##
## Call:
## lm(formula = sr ~ forest + wetland, data = bryo_belg)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8847 -0.4622  0.0545  0.4974  2.3116
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.34159    0.08369  27.981  < 2e-16 ***
## forest       1.11883    0.13925   8.034 9.74e-15 ***
## wetland     -0.59264    0.17216  -3.442 0.000635 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7095 on 417 degrees of freedom
## Multiple R-squared:  0.2231, Adjusted R-squared:  0.2193
## F-statistic: 59.86 on 2 and 417 DF,  p-value: < 2.2e-16
```

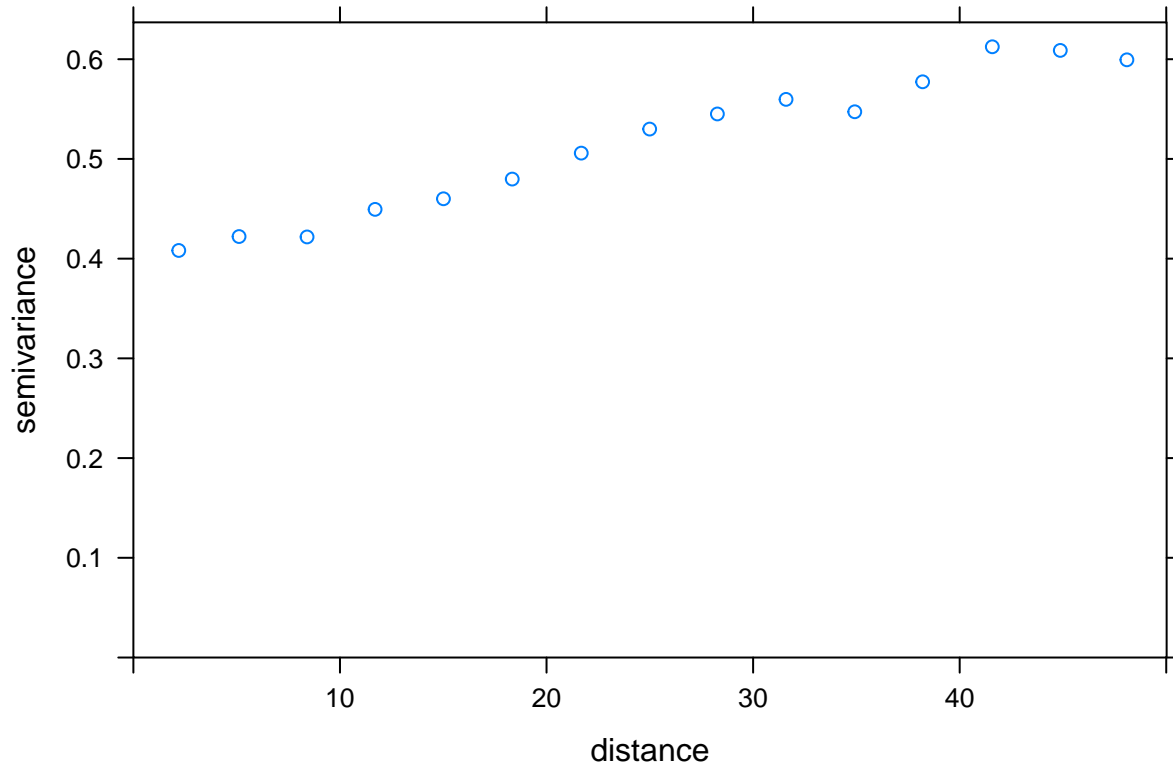
Il y a un effet positif de la fraction de forêts et négatif de la fraction de milieux humides; tous deux sont significatifs.

- b) Calculez le variogramme empirique des résidus du modèle en (a). Semble-t-il y avoir une corrélation spatiale entre les points?

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.

Solution

```
library(gstat)
vario <- variogram(sr ~ forest + wetland, locations = ~ x + y, data = bryo_belg,
                  cutoff = 50)
plot(vario)
```



Oui, la variance est plus faible pour de petites distances et semble augmenter jusqu'à environ 40 m.

- c) Ré-ajustez le modèle linéaire en (a) avec la fonction `gls` du package *nlme*, en essayant différents modèles de corrélation spatiale (exponentiel, Gaussien, sphérique). Comparez les modèles (incluant celui sans corrélation spatiale) avec l'AIC.

Solution

```
library(nlme)

bryo_exp <- gls(sr ~ forest + wetland, data = bryo_belg,
               correlation = corExp(form = ~ x + y, nugget = TRUE))

bryo_gaus <- gls(sr ~ forest + wetland, data = bryo_belg,
                 correlation = corGaus(form = ~ x + y, nugget = TRUE))

bryo_spher <- gls(sr ~ forest + wetland, data = bryo_belg,
                  correlation = corSpher(form = ~ x + y, nugget = TRUE))

rbind(AIC(bryo_lm), AIC(bryo_exp), AIC(bryo_gaus), AIC(bryo_spher))

##           [,1]
## [1,] 908.6358
## [2,] 867.8220
## [3,] 870.9592
## [4,] 866.9117
```

Le modèle sphérique a l'AIC le plus faible, suivi de près par le modèle exponentiel.

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

Solution

```
summary(bryo_spher)
```

```
## Generalized least squares fit by REML
##   Model: sr ~ forest + wetland
##   Data: bryo_belg
##           AIC      BIC    logLik
##   866.9117 891.1102 -427.4558
##
## Correlation Structure: Spherical spatial correlation
## Formula: ~x + y
## Parameter estimate(s):
##      range      nugget
## 43.1725287 0.6063165
##
## Coefficients:
##              Value Std.Error   t-value p-value
## (Intercept)  2.0368778 0.2481637  8.207800  0.000
## forest        0.6989832 0.1481690  4.717471  0.000
## wetland      -0.2441125 0.1809119 -1.349344  0.178
##
## Correlation:
##      (Intr) forest
## forest -0.251
## wetland -0.235  0.241
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -1.75204008 -0.06568726  0.61415414  1.15240184  3.23322262
##
## Residual standard error: 0.7998274
## Degrees of freedom: 420 total; 417 residual
```

La magnitude des coefficients est plus faible, leur erreur-type est plus élevée et l'effet des milieux humides n'est plus significatif. En raison des corrélations spatiales, nos différents points ne sont pas indépendants et une partie de l'effet originellement attribué aux variables prédictrices pourrait être une coïncidence due aux corrélations spatiales à la fois de la variable réponse et des variables prédictrices.