

# Aufgabenzettel 07

Gruppe 01

9.6.2020

Laden sie den Workspace yingtan\_20\_ueb3.Rdata sowie das Paket gstat und überführen sie das Objekt l1z in ein SpatialPointsDataFrame. Reproduzieren sie ihr Variogrammodell aus Übung 05.

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.0 --
## v ggplot2 3.3.0      v purrr  0.3.4
## v tibble  3.0.1      v dplyr  0.8.5
## v tidyr   1.0.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

load("data/yingtan_20_ueb3.RData")

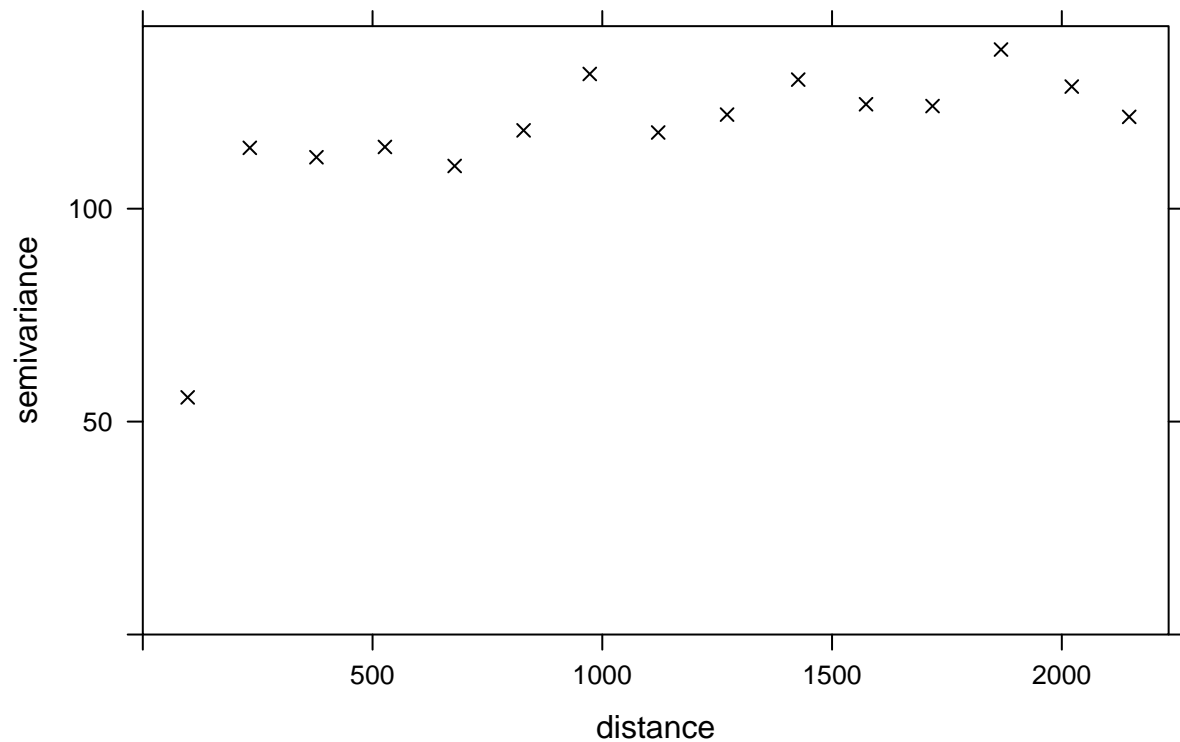
##SpatialPointsDataFrame##
library(sp)
SPDFl1z <- l1z
coordinates(SPDFl1z) <- ~ EAST + NORTH
proj4string(SPDFl1z) <- CRS("+proj=utm +zone=50 +ellps=WGS84 +datum=WGS84")

##Reproduktion des Variogrammodells##
#omnidirektionales empirisches Variogramm
library(gstat)
Ca <- SPDFl1z@data$Ca_exch

vario_omni_Ca <- variogram(Ca ~ EAST + NORTH,
                          data = SPDFl1z,
                          cutoff = 2202,
                          width = 150)

plot(vario_omni_Ca,
     main = "Omnidirektionales empirisches Variogramm",
     pch = 4,
     col = "black")
```

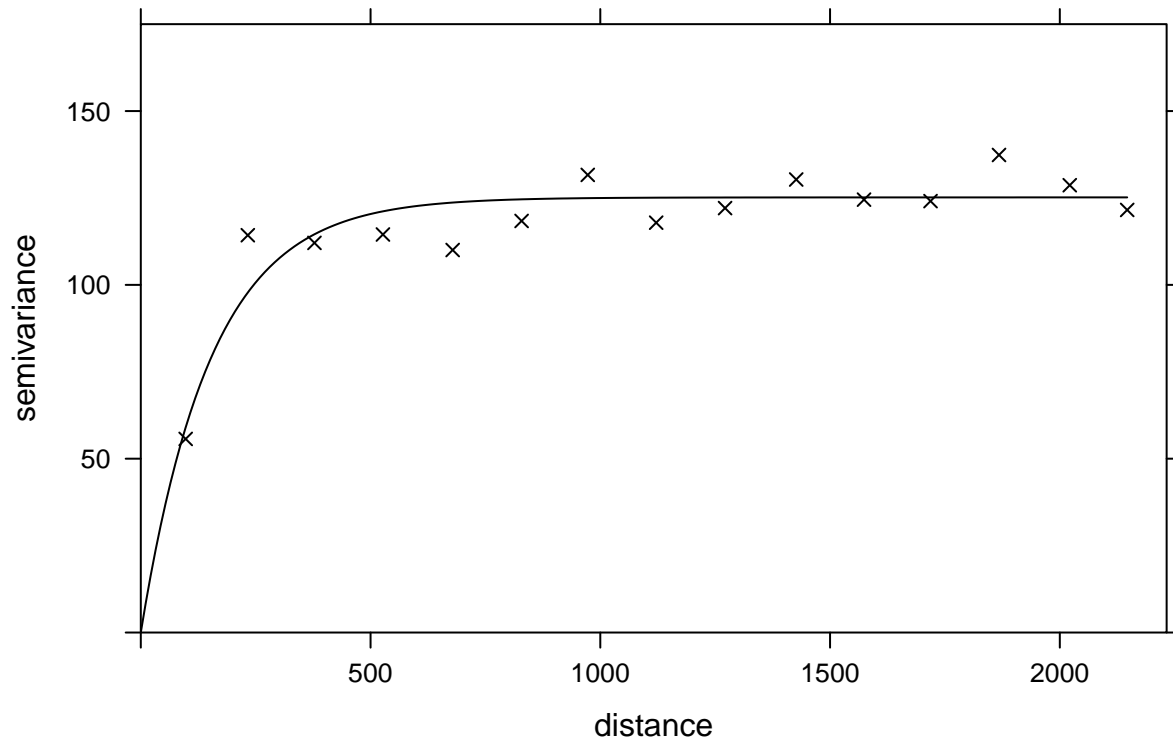
## Omnidirektionales empirisches Variogramm



```
#Modell zum Variogramm
vario_omni_Ca_fit <- fit.variogram(vario_omni_Ca,
                                  vgm(model = "Exp"))

plot(vario_omni_Ca,
     model = vario_omni_Ca_fit,
     cutoff = 2202,
     ylim = c(0, 175),
     pch = 4,
     col = "black",
     main = "Variogrammodell der austauschbaren Ca-Ionen")
```

## Variogrammmodell der austauschbaren Ca-Ionen



### ## Aufgabe 14 Leave-One-Out-Cross-Validation

Die Validierung der Ergebnisse ist ein wichtiger Schritt jeder Modellierung. Um erfolgreich und unabhängig validieren zu können, bedarf es Daten, die nicht in die Kalibrierung des Modells eingeflossen sind. Um den häufig ohnehin schon kleinen Datenpool durch eine Aufteilung in Kalibrierungs- und Validierungsdatensatz nicht noch weiter zu reduzieren wird bei geostatistischen Modellen häufig das LOOCV-Verfahren angewendet. Dabei wird nacheinander ein Probenstandort aus dem Modell entfernt und die Zielgröße für diesen Ort vorhergesagt; so lange, bis alle Beprobungspunkte einmal ausgeschlossen worden sind.

- a) Führen Sie mit der Methode `krige.cv` für die Ca-Ionen eine leave-one-out-cross-validation durch. Verwenden Sie das Variogrammmodell aus Aufg. 13 und notieren Sie ihre R-Syntax im Protokoll. (1 Punkt)

```
#Leave-one-out-Cross-Validation
LOOCV <- gstat::krige.cv(formula = Ca_exch ~ 1,
                        locations = SPDFljz,
                        model = vario_omni_Ca_fit)
```

- b) Vergleichen Sie die Struktur des mittels `krige.cv` generierten Objekts mit dem Ergebnis der `krige`-Funktion aus Aufg. 13. Welche Daten-Attribute sind hinzugekommen und wofür stehen sie? (1 Punkt)

```
summary(LOOCV)
```

```
## Object of class SpatialPointsDataFrame
## Coordinates:
##           min      max
## EAST    490441 493591
## NORTH  3121290 3125630
```

```
## Is projected: NA
## proj4string : [NA]
## Number of points: 335
## Data attributes:
##      var1.pred      var1.var      observed      residual
## Min.   : 6.981    Min.   : 1.726    Min.   : 3.772    Min.   : -23.80832
## 1st Qu.:17.229    1st Qu.: 27.727    1st Qu.:13.633    1st Qu.: -5.53713
## Median :20.594    Median : 81.690    Median :19.491    Median : -1.06772
## Mean   :21.862    Mean   : 64.167    Mean   :21.926    Mean   : 0.06368
## 3rd Qu.:26.059    3rd Qu.: 91.831    3rd Qu.:28.718    3rd Qu.: 5.06027
## Max.   :43.774    Max.   :124.508    Max.   :94.311    Max.   : 68.09722
##      zscore      fold
## Min.   : -3.074485    Min.   : 1.0
## 1st Qu.: -0.688066    1st Qu.: 84.5
## Median : -0.153649    Median :168.0
## Mean   : 0.006176    Mean   :168.0
## 3rd Qu.: 0.666500    3rd Qu.:251.5
## Max.   : 7.125407    Max.   :335.0
```

observed:

residual: Residuen als Differenz des errechneten Werts zum tatsächlichen Wert, der zur Überprüfung ausgelassen wurde.

zscore: Kriging Standard-Fehler, bei dem die Kriging-Varianz eine Rolle spielt. mean und variance sollten dicht an 0 und 1 liegen

fold: zeigt zu welchem "fold" jeder Datensatz gehört (Bivand, Pebesma, and Gomez-Rubio 2008)

- c) Wie sähe das Vorhersageergebnis aus, wenn der Probenstandort während der Kreuzvalidierung nicht ausgeschlossen werden würde? Was ergäbe sich konsequenterweise bei der Fehlerberechnung? (1 Punkt)

Dasselbe wie beim Ordinary Kriging???

## Aufgabe 15 Root-Mean-Squared-Error

Der RMSE gibt Auskunft darüber, wie nah das Modell an die bekannten, tatsächlich gemessenen Daten herankommt. Er ist ein Maß für die 'accuracy' des gewählten Prädiktionsverfahrens.

- a) Berechnen Sie den RMSE für die austauschbaren Ca-Ionen. (1 Punkt)

```
#Root-Mean-Square-Error Ca
rmse <- function(x,y) {
  sqrt(mean((x - y)^2))
}

rmse(x = Ca,
     y = LOOCV@data$residual)
```

```
## [1] 22.90189
```

- b) Was bedeuten die einzelnen Silben des Wortes RMSE und warum wird der Vorhersage-Fehler gerade so beschrieben? (2 Punkte)

RMSE steht für Root-Mean-Square-Error. Zur Fehlerberechnung des Modells wird die Wurzel aus der quadrierten mittleren Abweichung des berechneten Werts von dem tatsächlichen Wert berechnet.

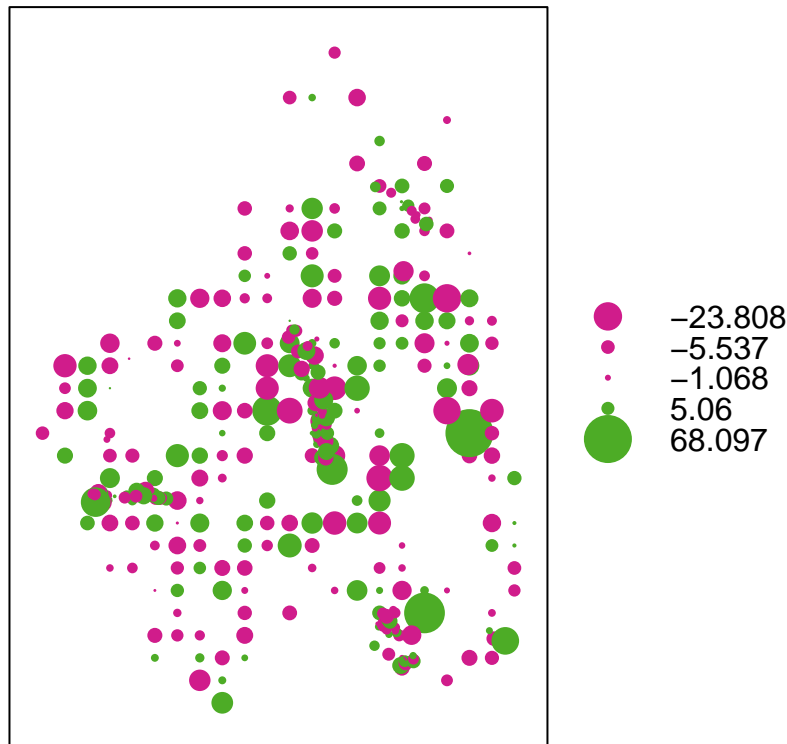
## Aufgabe 16 Grafische Validierung

Die Angabe eines Gesamtfehlers reicht nicht aus, um die Güte eines Modells hinreichend zu beschreiben. Eine Darstellung der Verteilung der Fehler im Raum ist ebenso nützlich wie die Betrachtung der Streuung im Werteraum.

- a) Erstellen Sie für ihre Modell-Residuen einen ansehnlichen Bubble-Plot und gehen sie der Frage nach, ob räumliche Muster erkennbar sind. (2 Punkte)

```
#Bubbleplot der LOOCV
library(lattice)
bubble(LOOCV, "residual",
       main = "Austauschbare Ca-Ionen: Residuen der LOOCV")
```

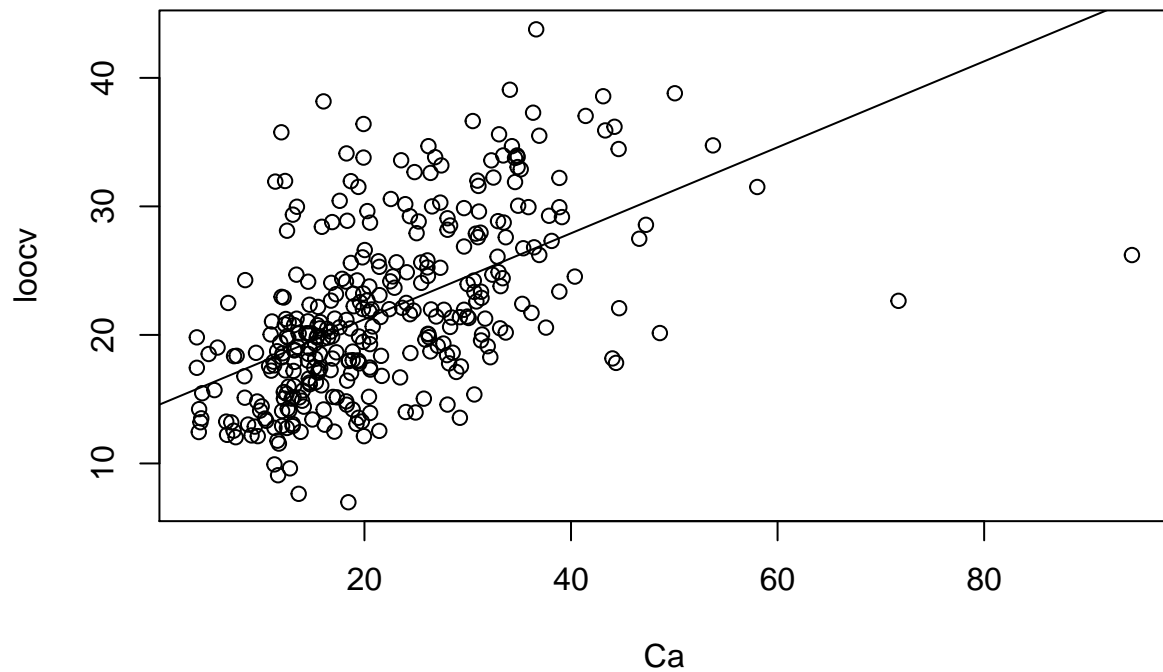
### Austauschbare Ca-Ionen: Residuen der LOOCV



- b) Plotten Sie die tatsächlichen Ca-Ionen-Konzentrationen gegen die vorhergesagten Werte. Ergänzen Sie eine Ausgleichsgerade mit der Steigung eins und einem Verlauf durch den Ursprung. (1 Punkt)

```
#library(ggplot2)
#DFljz <- as(SPDFljz, "data.frame")
#ggplot(DFljz, aes(x = Ca,
#                  y = leave_one_out_cv)) +
#  geom_smooth()

loocv <- LOOCV@data$var1.pred
plot(Ca, loocv)
abline(lm(loocv ~ Ca))
```



c) Bewerten Sie kurz das durchgeführte Interpolationsverfahren. Beziehen Sie sich auf den RMSE und ihre Diagnose-Plots. (2 Punkte)

Bivand, R. S., E. J. Pebesma, and Gomez-Rubio. 2008. *Applied Spatial Data Analysis with R*. New York, NY: Springer New York. <https://doi.org/10.1007/978-0-387-78171-6>.