R pour le Géospatial

Modéliser

Etienne Racine

```r
library(tidyverse)
library(gstat)
library(mapview)
library(AmesHousing)
library(sf)
```

Outils spatiaux pour la Science des Données

- Préparation des données (feature engineering)
- Modélisation des dépendances spatiales

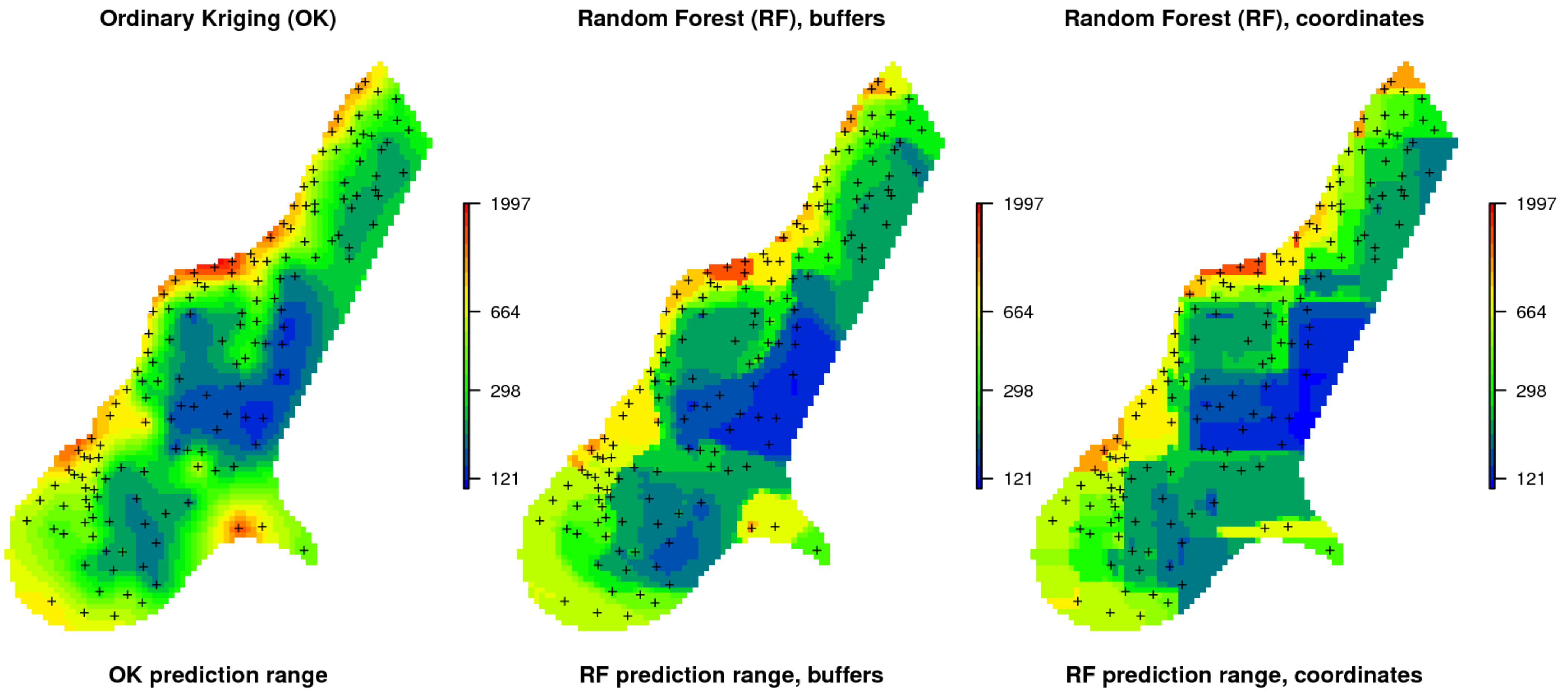
Dépendance spatiale

La relation entre les observations croît avec la proximité

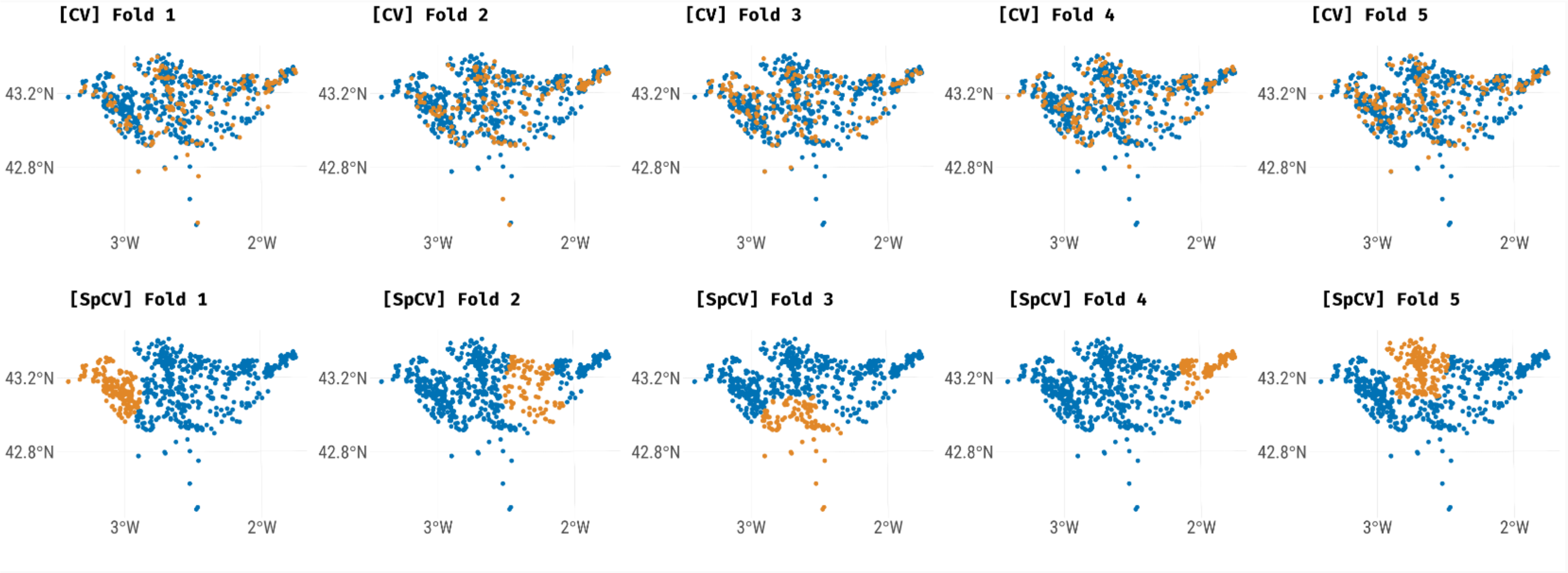
```
gstat::show.vgms()
```

Modélisation

Hengl et al. 2018 https://peerj.com/articles/5518



Méthodes de validation croisée



- Risque de surestimer la performance du modèle.
- https://mlr.mlr-org.com/

https://pjs-web.de/slides/papers/paper1/2018_06_Kolloquium.pdf

Distance

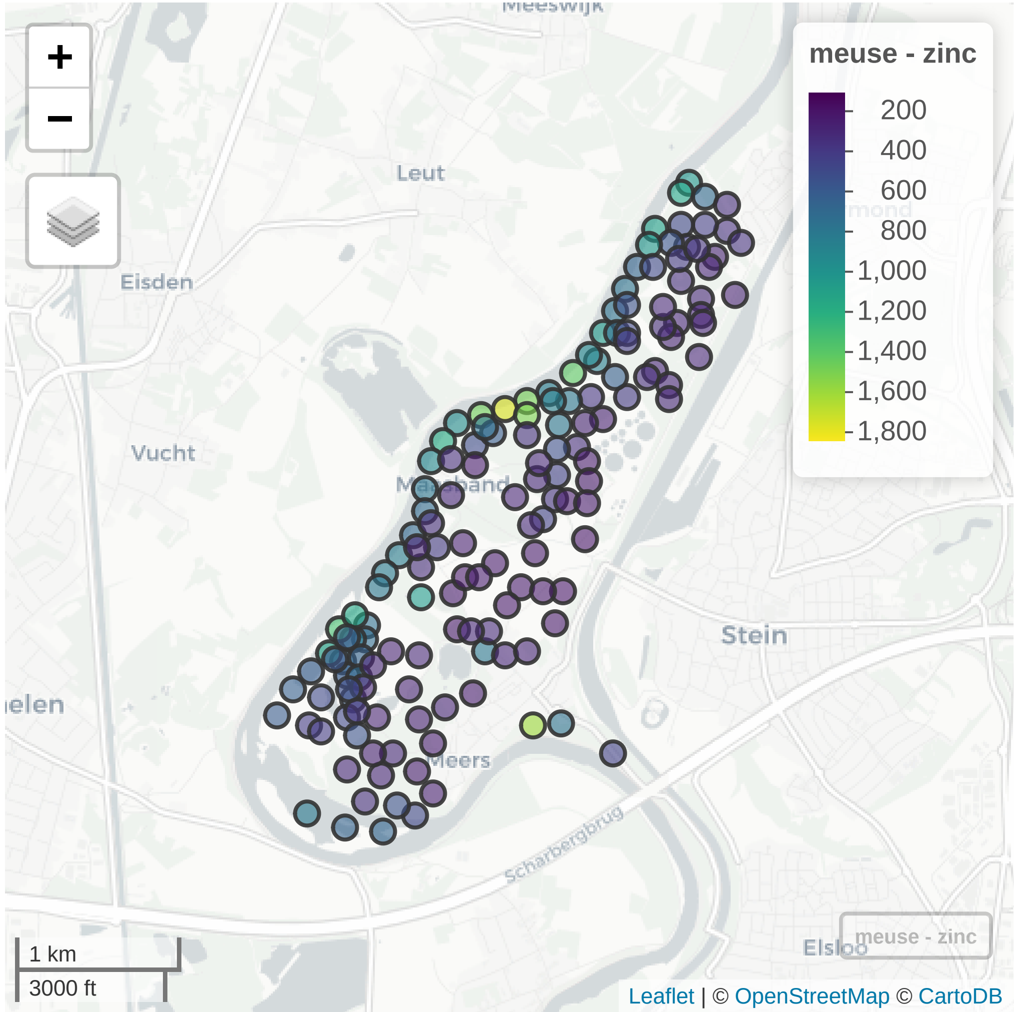
- Euclidienne
- Grand arc
- Trajet
- Interaction

Analyse exploratoire

```
meuse <- system.file("sqlite/meuse.sqlite", package = "sf") %>% read_sf()
```

```
## Simple feature collection with 155 features and 12 fields
## geometry type:  POINT
## dimension:      XY
## bbox:           xmin: 178605 ymin: 329714 xmax: 181390 ymax: 333611
## epsg (SRID):    28992
## proj4string:    +proj=sterea +lat_0=52.15616055555555 +lon_0=5.38763888888889 +k=0.9999079 +x_
## # A tibble: 155 x 13
##    cadmium copper  lead  zinc  elev    dist    om ffreq soil  lime  landuse
##      <dbl>  <dbl> <dbl> <dbl> <dbl>   <dbl> <dbl> <chr> <chr> <chr> <chr>
##  1    11.7     85   299  1022  7.91 0.00136  13.6 1     1     1     Ah
##  2     8.6     81   277  1141  6.98 0.0122   14   1     1     1     Ah
##  3     6.5     68   199   640  7.8  0.103    13   1     1     1     Ah
##  4     2.6     81   116   257  7.66 0.190     8   1     2     0     Ga
##  5     2.8     48   117   269  7.48 0.277     8.7 1     2     0     Ah
##  6     3       61   137   281  7.79 0.364     7.8 1     2     0     Ga
##  7     3.2     31   132   346  8.22 0.190     9.2 1     2     0     Ah
##  8     2.8     29   150   406  8.49 0.0922    9.5 1     1     0     Ab
##  9     2.4     37   133   347  8.67 0.185    10.6 1     1     0     Ab
## 10     1.6     24    80   183  9.05 0.310     6.3 1     2     0     W
## # … with 145 more rows, and 2 more variables: dist.m <dbl>,
```

```r
mapview(meuse, zcol = "zinc")
```



Variogram

```r
library(gstat)

vgm <- variogram(zinc ~ 1, meuse)
plot(vgm)
```

Modélisation

```
zinc_m <- lm(zinc ~ cadmium+copper+lead+elev+om+soil, data = meuse)

meuse <- meuse %>%
  mutate(zinc_pred = predict(zinc_m,
                             list(
                               cadmium = cadmium,
                               copper = copper,
                               lead = lead,
                               elev = elev,
                               om = om,
                               soil = soil
                               )),
         zinc_resid = zinc - zinc_pred)
```

Vérification des résidus

```
meuse_vgm <- variogram(zinc_resid ~ 1, meuse %>% na.omit)
plot(meuse_vgm)
```

À vous

- Quelle est la relation de dépendance spatiale dans les données `AmesHousing`

```
# install.packages("AmesHousing")
ames <- inner_join(
  AmesHousing::ames_geo,
  AmesHousing::ames_raw
  ) %>%
  st_as_sf(coords = c("Longitude", "Latitude"), crs = 4326)
```

```
## Joining, by = "PID"
```

- Ajustez un modèle puis vérifiez la dépendance des résidus
- Utilisez `osmdata` pour ajouter de nouvelles features (e.g. proximité d'un parc, distance à la route, restaurant, ...)

Ames

```
ames <- AmesHousing::ames_geo %>% inner_join(AmesHousing::ames_raw) %>%
  st_as_sf(coords = c("Longitude", "Latitude"), crs = 4326)
```

```
## Joining, by = "PID"
```

```
vgm <- variogram(SalePrice ~ 1, ames)
```