

Data Break 7: Kriging the Meuse River

BIOS 737 Fall 2012

Fall 2012

What do we have? (Reminder from Data Break 6)

- ▶ `library(gstat)`
- ▶ Data included in `gstat` library.
- ▶ `data(meuse)` makes the data available.
- ▶ Soil samples from flood plain of Meuse River, near village of Stein (Belgium).
- ▶ Geostatistical data: locations \mathbf{s} and attributes $Z(\mathbf{s})$.

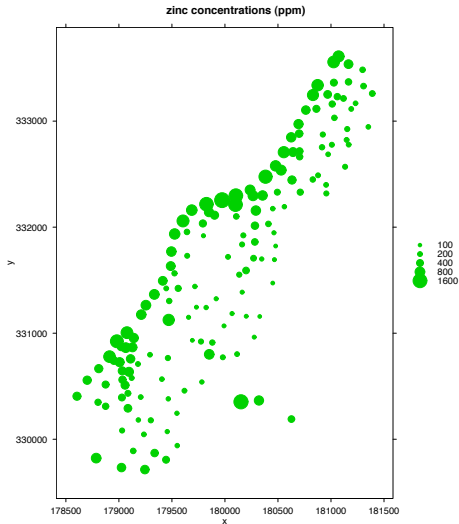
What do we want?

- ▶ Predictions of zinc across the entire flood plain.
- ▶ Mean square prediction error at each point.

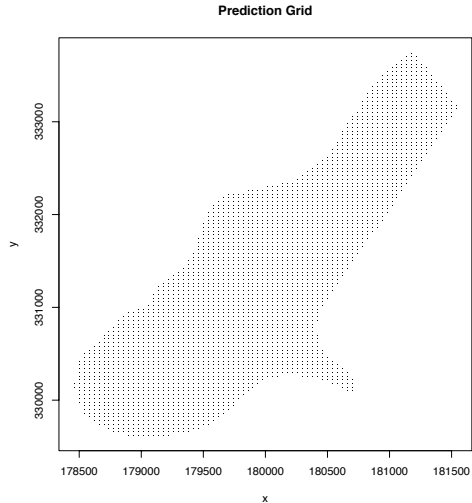
Kriging steps

- ▶ Empirical semivariogram.
- ▶ Fit theoretical semivariogram to empirical semivariogram values.
- ▶ Krige with best fitting semivariogram values.

What do we have? Bubble plot: Zinc



What do we want? Predictions here.



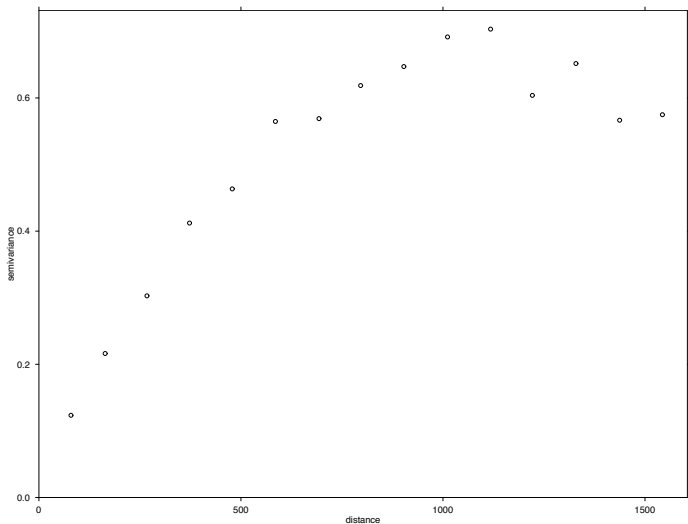
Step 1: Data Gaussian?

- ▶ Recall from Data Break 6, $\log(\text{zinc})$ closer to Gaussian than zinc.
- ▶ We will kriging the $\log(\text{zinc})$ values.
- ▶ (Note: biased prediction of $\exp(\log(\text{zinc})) = \text{zinc}$, but we ignore this in the example).

Step 2: Empirical Semivariogram

```
logzinc.var = variogram(log(zinc) 1,  
  loc=~x+y,  
  data=meuse)  
plot(logzinc.var$dist,  
  logzinc.var$gamma,  
  xlab="distance",  
  ylab="semivariance")
```


What do we want? Predictions here.



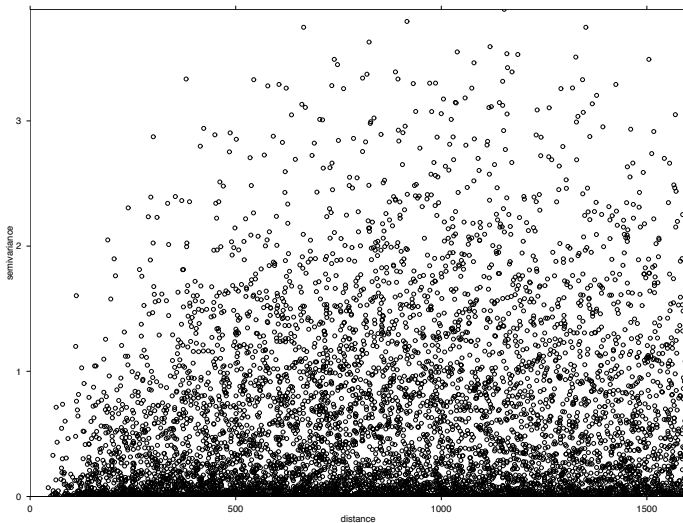
Some other cool plots

- ▶ The *variogram cloud* consists of the plot of all contrasts $Z(\mathbf{s}_i) - Z(\mathbf{s}_j)$ plotted versus $\|\mathbf{s}_i - \mathbf{s}_j\|$

```
plot(variogram(log(zinc)~1,  
              loc=~x+y,  
              data=meuse, cloud=TRUE))
```

- ▶ Also some need interactive commands here....

Variogram cloud

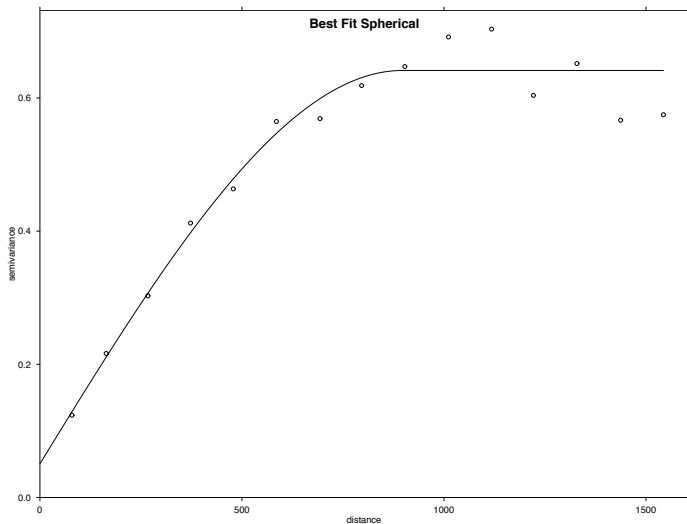


Step 2: Fit Theoretical Semivariograms

- ▶ `fit.variogram` function in `gstat`
- ▶ Not entirely clear what optimization rule, based on practical experience rather than theoretical motivation.
- ▶ To get best fitting *spherical* semivariogram...

```
model.1 = fit.variogram(logzinc.var,  
                        vgm(psill=1,model="Sph",  
                           range=300,nugget=1))  
plot(logzinc.var, model=model.1)  
title("Best Fit Spherical")
```

Best spherical

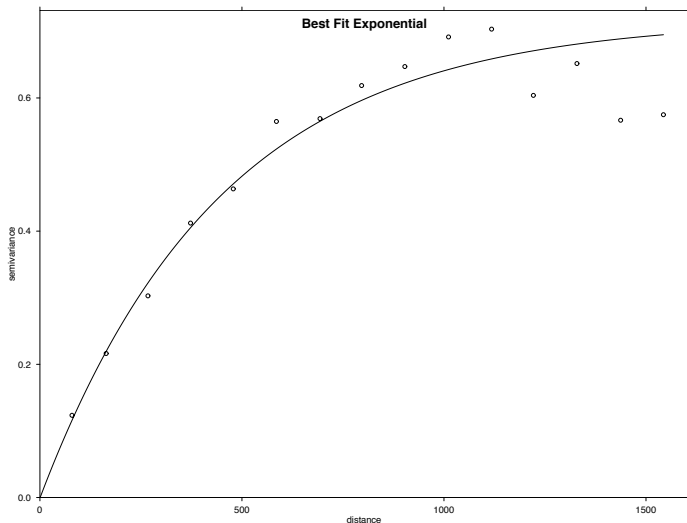


Step 2: Fit Theoretical Semivariograms

- ▶ To get best fitting *exponential* semivariogram...

```
model.1 <- fit.variogram(logzinc.var,  
                        vgm(psill=1,model="Exp",  
                           range=300,nugget=1))  
plot(logzinc.var, model=model.1)  
title("Best Fit Exponential")
```

Best exponential

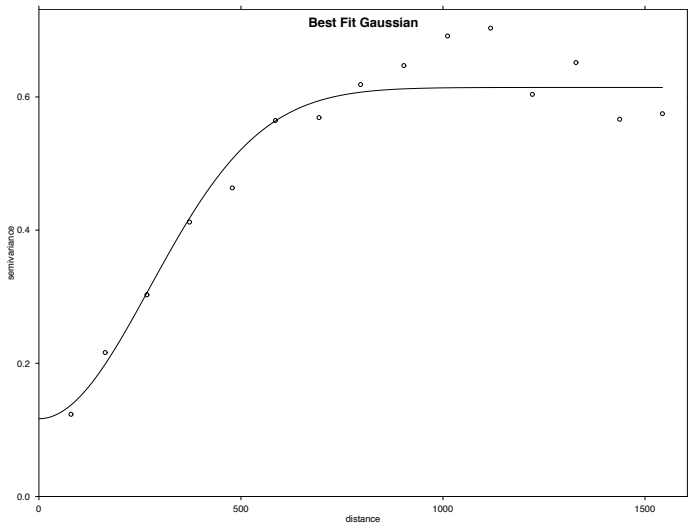


Step 2: Fit Theoretical Semivariograms

- To get best fitting *Gaussian* semivariogram...

```
model.1 <- fit.variogram(logzinc.var,  
                        vgm(psill=1,model="Gau",  
                           range=300,nugget=1))  
plot(logzinc.var, model=model.1)  
title("Best Fit Gaussian")
```


Best Gaussian

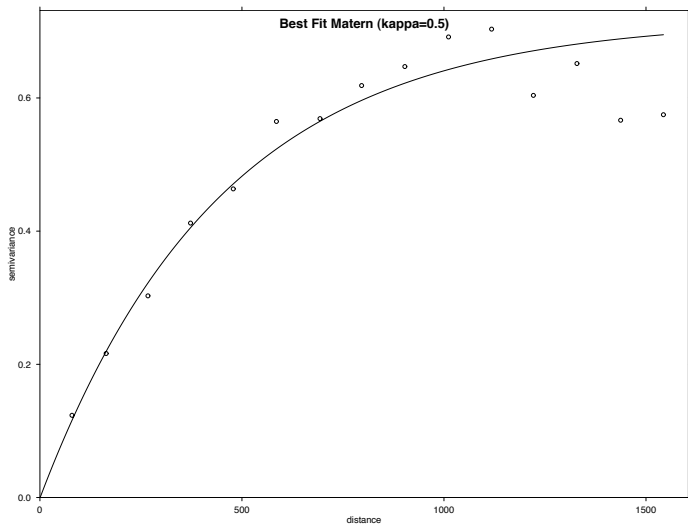


Step 2: Fit Theoretical Semivariograms

- To get best fitting *Matern* semivariogram...

```
model.1 <- fit.variogram(logzinc.var,  
                        vgm(psill=1,model="Mat",  
                           range=300,nugget=1))  
plot(logzinc.var, model=model.1)  
title("Best Fit Matern (kappa=0.25)")
```

Best Matern



What's best?

- ▶ We'll use the spherical theoretical semivariogram.
- ▶ (Skipping formal AIC or other comparison).
- ▶ What are the best fit values?

```
> model.1
```

```
model psill range
```

```
1 Nug 0.0506555 0.0000
```

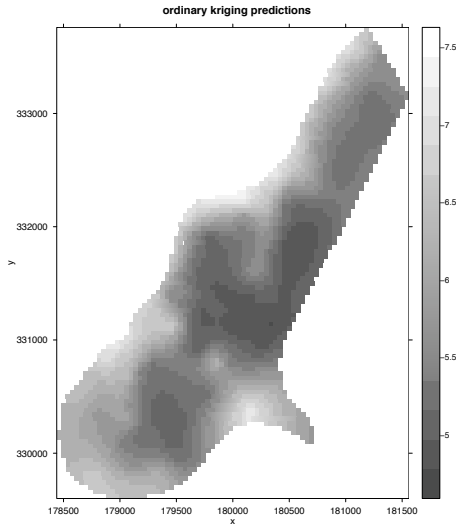
```
2 Sph 0.5906009 896.9702
```

- ▶ Nugget: 0.051, Partial sill: 0.591, Range: 896.97

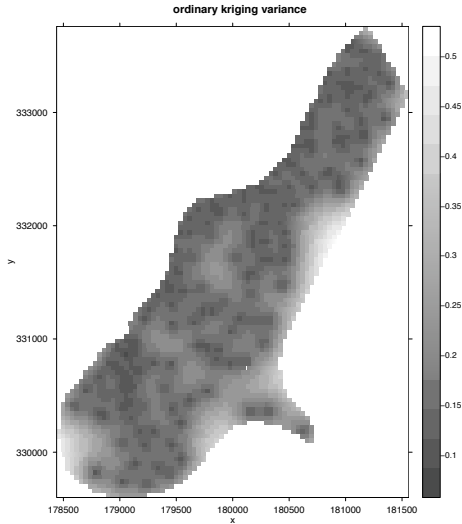
Kriging

```
# define the theoretical semivariogram
m = vgm(psil1=0.591,"Sph",
        range=896.97,nugget=0.051)
# ordinary kriging:
logzinc.krige = krige(log(zinc)~1,
                      ~x+y, model = m, data = meuse,
                      newd = meuse.grid)
# make 'levelplot'
levelplot(var1.pred~x+y,
          data=logzinc.krige,
          aspect = mapasp(logzinc.krige),
          main = "ordinary kriging predictions")
```

Prediction levels

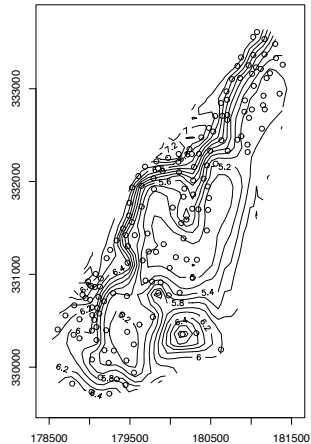
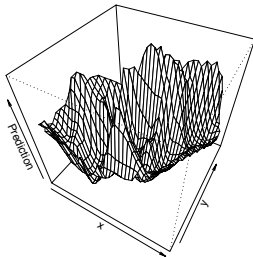


Kriging variance (MSPE)



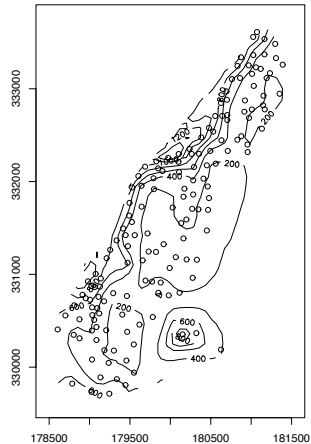
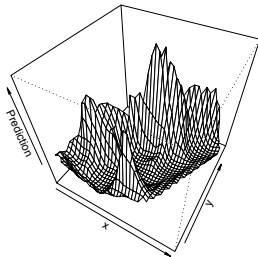
Prediction levels (perspective plot)

Prediction (log zinc)



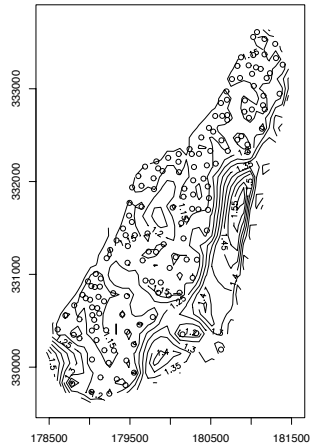
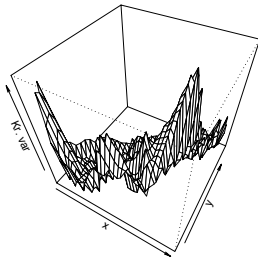
Prediction levels (zinc)

Prediction (zinc)



Kriging variance (zinc)

Kriging variance surface



- ▶ MSPE highest where fewest observations.
- ▶ Back-transforming to original zinc levels gives idea of prediction, but is no longer BLUP (we lose “L” and “U”).
- ▶ gstat does much more complicated analyses too.

Summary of geostatistical methods

- ▶ We've really only seen the whirlwind tour of the basic elements.
- ▶ Semivariogram.
- ▶ Ordinary kriging.
- ▶ Many other options too.
- ▶ Questions?

- ▶ Repeat this analysis for the Smoky Mountain pH data.
- ▶ Should be able to use the data break code to do this.
- ▶ Due 12/3, graphs, output.

Project 2

- ▶ Write up results of HW 4 as a report with Introduction, Methods, Results, and Discussion sections.
- ▶ Due 12/10, 5pm.