Areal Data and Proximity

HES 505 Fall 2023: Session 20

Matt Williamson

Objectives

By the end of today you should be able to:

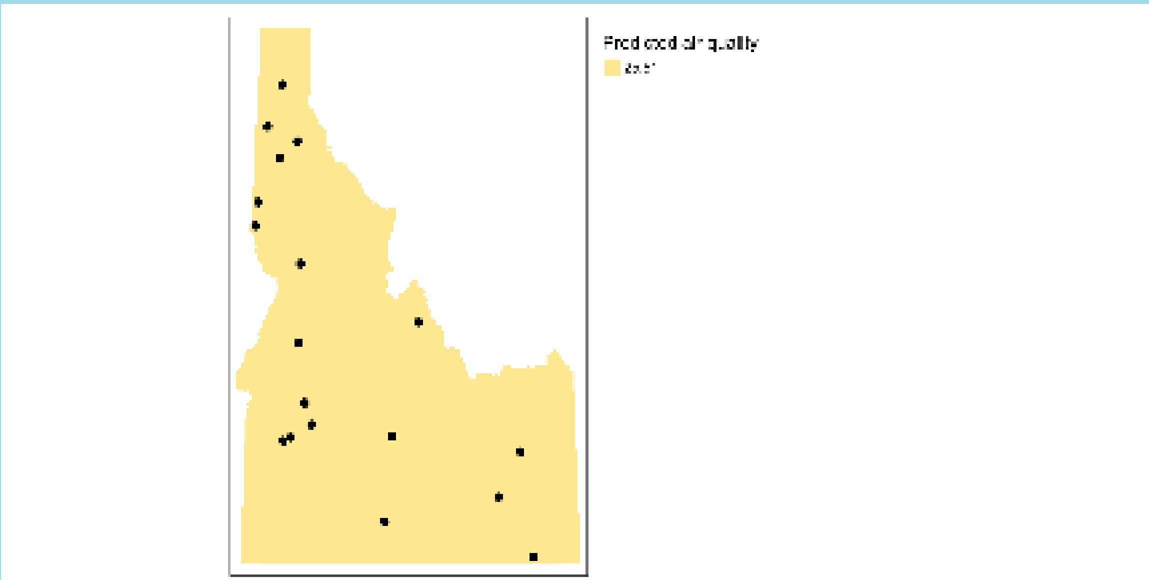
Statistical Interpolation

Statistical Interpolation

Trend Surface Modeling

- Basically a regression on the coordinates of your data points
- Coefficients apply to the coordinates and their interaction
- Relies on different functional forms

0th Order Trend Surface



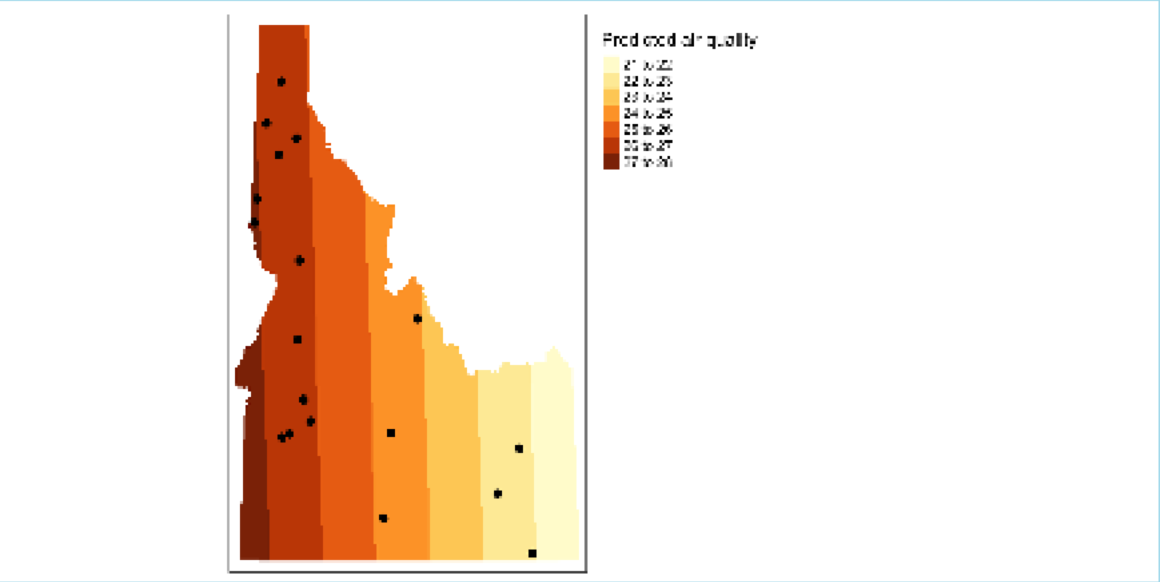
- Simplest form of trend surface
- Z = a where a is the mean value of air quality
- Result is a simple horizontal surface where all values are the same.

0th order trend surface

```
#set up interpolation grid
# Create an empty grid where n is the total number of cells
grd <- as.data.frame(spsample(as(id.cty, "Spatial"), "regular", n=20000))
names(grd)       <- c("X", "Y")
coordinates(grd) <- c("X", "Y")
gridded(grd)     <- TRUE  # Create SpatialPixel object
fullgrid(grd)    <- TRUE  # Create SpatialGrid object
proj4string(grd) <- proj4string(as(aq.sum, "Spatial"))
# Define the polynomial equation
f.0  <- as.formula(meanpm25 ~ 1)

# Run the regression model
lm.0 <- lm( f.0 , data=aq.sum)

# Use the regression model output to interpolate the surface
dat.0th <- SpatialGridDataFrame(grd, data.frame(var1.pred = predict(lm.0, n

# Convert to raster object to take advantage of rasterVis' imaging
```

1st Order Trend Surface

- Creates a slanted surface
- $Z = a + bX + cY$
- X and Y are the coordinate pairs

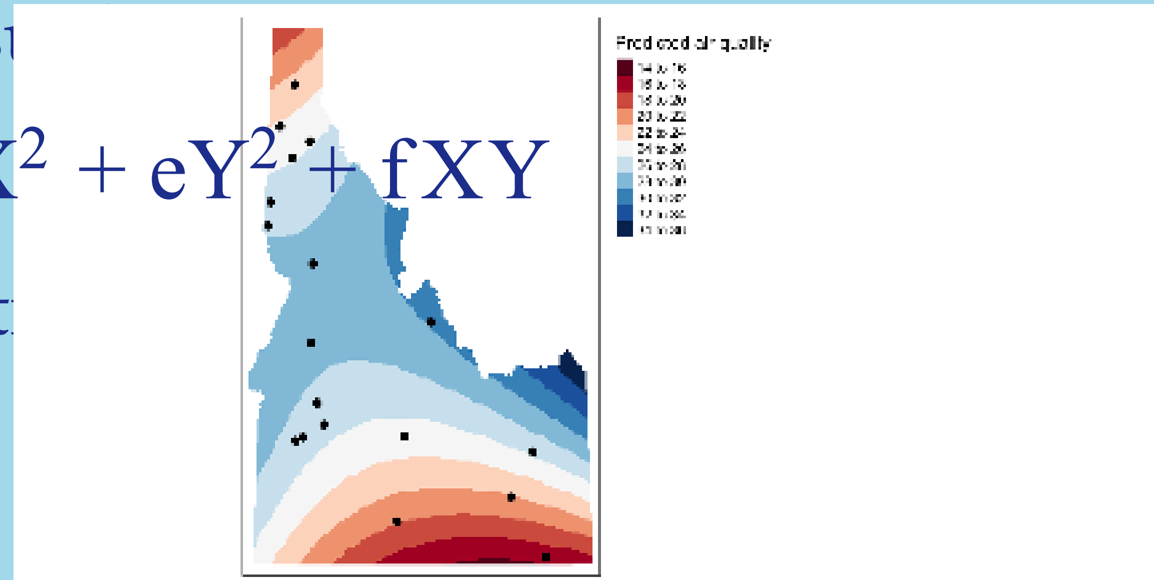


1st Order Trend Surface

```
# Define the polynomial equation
f.1  <- as.formula(meanpm25 ~ X + Y)

aq.sum$X <- st_coordinates(aq.sum)[,1]
aq.sum$Y <- st_coordinates(aq.sum)[,2]

# Run the regression model
lm.1 <- lm( f.1 , data=aq.sum)

# Use the regression model output to interpolate the surface
dat.1st <- SpatialGridDataFrame(grd, data.frame(var1.pred = predict(lm.1, n

# Convert to raster object to take advantage of rasterVis' imaging
# environment
r   <- rast(dat.1st)
r.m <- mask(r, st_as_sf(id.cty))
```

2nd Order Trend Surfaces

- Produces a parabolic su
- $Z = a + bX + cY + dX^2 + eY^2 + fXY$
- Highlights the interact
  directions



2nd Order Trend Surfaces

```
# Define the 1st order polynomial equation
f.2 <- as.formula(meanpm25 ~ X + Y + I(X*X)+I(Y*Y) + I(X*Y))

# Run the regression model
lm.2 <- lm( f.2, data=aq.sum)

# Use the regression model output to interpolate the surface
dat.2nd <- SpatialGridDataFrame(grd, data.frame(var1.pred = predict(lm.2, n

r   <- rast(dat.2nd)
r.m <- mask(r, st_as_sf(id.cty))

tm_shape(r.m) + tm_raster(n=10, palette="RdBu", title="Predicted air qualit
  tm_legend(legend.outside=TRUE)
```

Kriging

- Previous methods predict z as a (weighted) function of distance
- Treat the observations as perfect (no error)
- If we imagine that z is the outcome of some spatial process such that:

$$z(\mathbf{x}) = \mu(\mathbf{x}) + \epsilon(\mathbf{x})$$

then any observed value of z is some function of the process ($\mu(\mathbf{x})$) and some error ($\epsilon(\mathbf{x})$)

- Kriging exploits autocorrelation in $\epsilon(\mathbf{x})$ to identify the trend and interpolate accordingly

Autocorrelation

- **Correlation** the tendency for two variables to be related
- **Autocorrelation** the tendency for observations that are closer (in space or time) to be correlated
- **Positive autocorrelation** neighboring observations have ϵ with the same sign
- **Negative autocorrelation** neighboring observations have ϵ with a different sign (rare in geography)

Ordinary Kriging

- Assumes that the deterministic part of the process ($\mu(\mathbf{x})$) is an unknown constant (μ)

$$z(\mathbf{x}) = \mu + \epsilon(\mathbf{x})$$

Steps for Ordinary Kriging

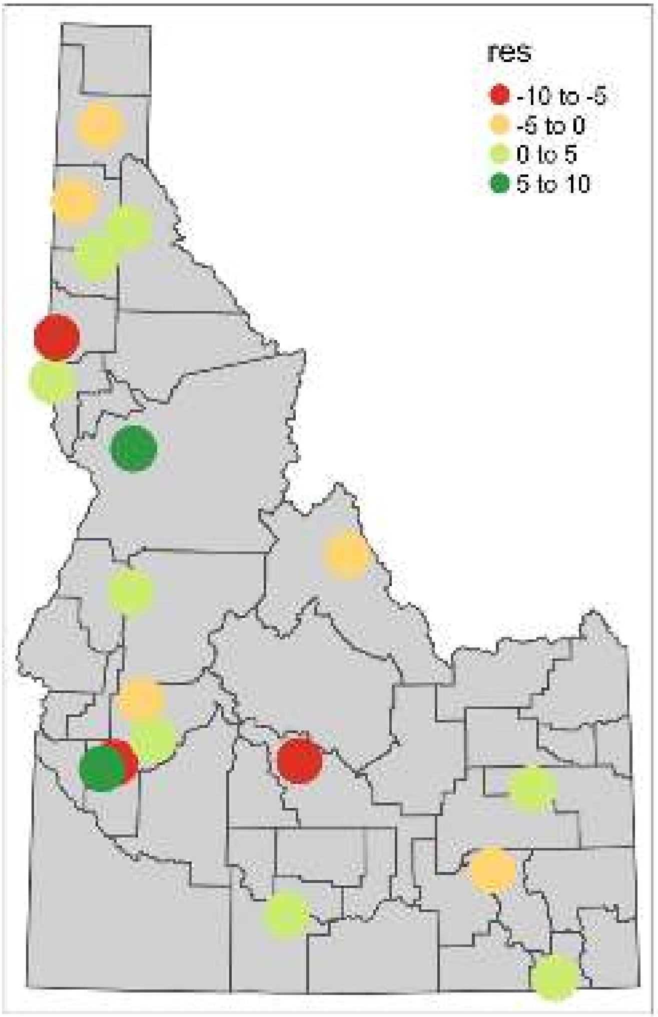
- Removing any **spatial trend** in the data (if present).
- Computing the **experimental variogram**, γ, which is a measure of spatial autocorrelation.
- Defining an **experimental variogram model** that best characterizes the spatial autocorrelation in the data.
- Interpolating the surface using the experimental variogram.
- Adding the kriged interpolated surface to the trend interpolated surface to produce the final output.

Removing Spatial Trend

- Mean and variance need to be constant across study area
- Trend surfaces indicate that is not the case
- Need to remove that trend

```
f.2 <- as.formula(meanpm25 ~ X + Y + I(X*X)+I(Y*Y) + I(X*Y))

# Run the regression model
lm.2 <- lm( f.2, data=aq.sum)

# Copy the residuals to the point object
aq.sum$res <- lm.2$residuals
```

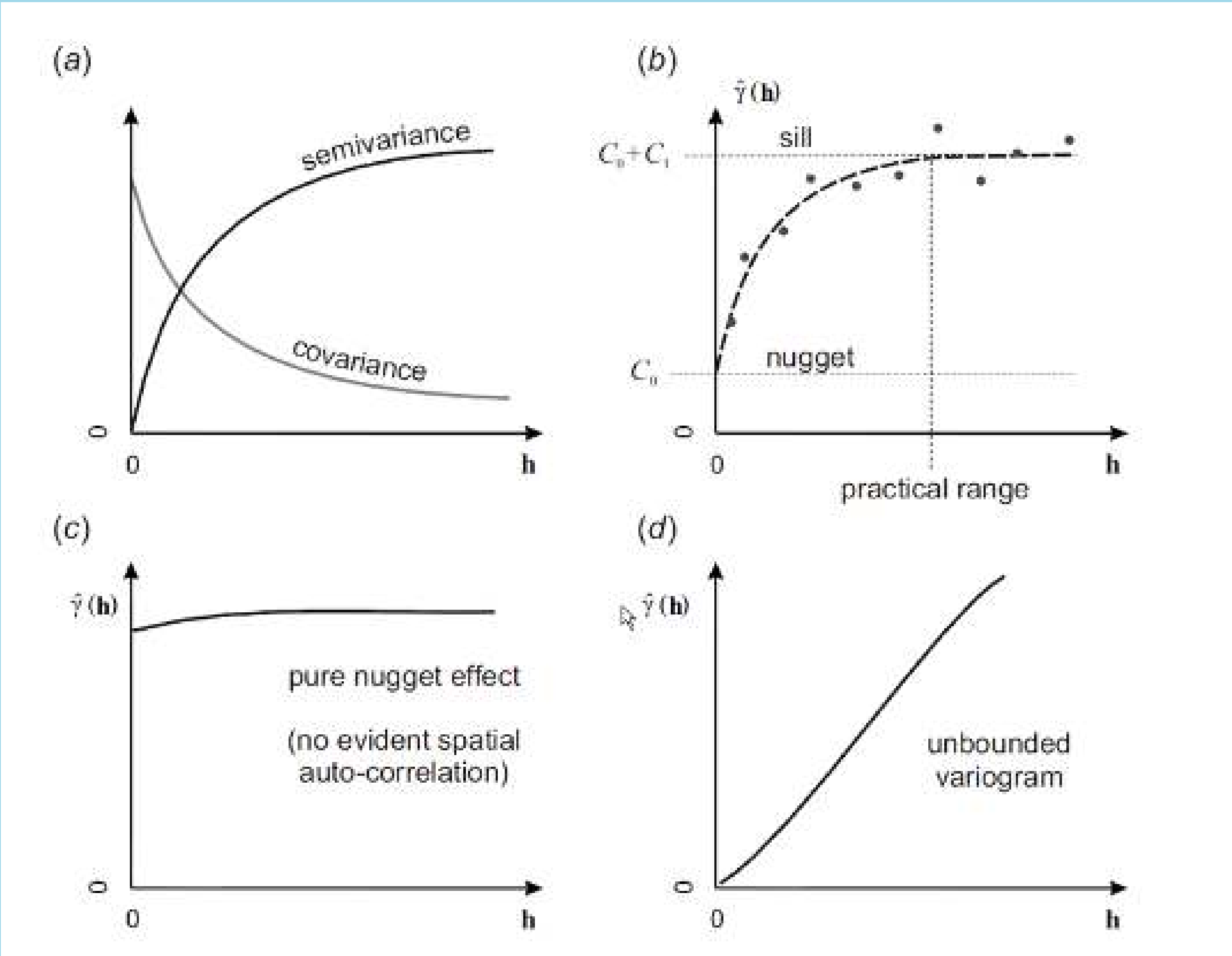
Removing the trend



Calculate the experimental variogram

- **nugget** - the proportion of semivariance that occurs at small distances
- **sill** - the maximum semivariance between pairs of observations
- **range** - the distance at which the **sill** occurs
- **experimental** vs. **fitted** variograms

A Note about Semivariograms



Fitted Semivariograms

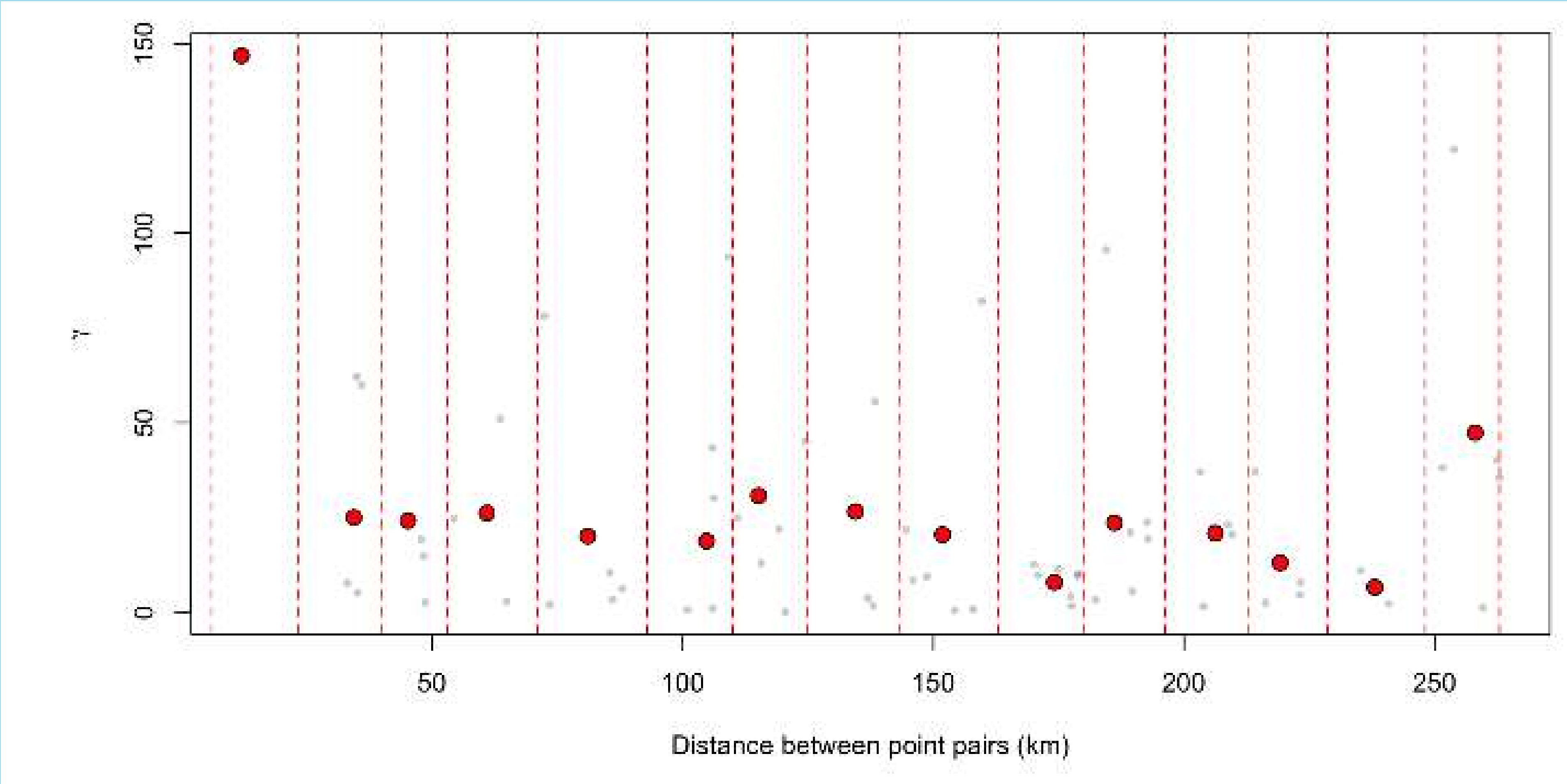
Calculate the experimental variogram

```
var.cld  <- gstat::variogram(res ~ 1, aq.sum, cloud = TRUE)
var.df  <- as.data.frame(var.cld)
index1  <- which(with(var.df, left==21 & right==2))

OP <- par( mar=c(4,6,1,1))
plot(var.cld$dist/1000 , var.cld$gamma, col="grey",
     xlab = "Distance between point pairs (km)",
     ylab = expression( frac((res[2] - res[1])^2 , 2)) )
```

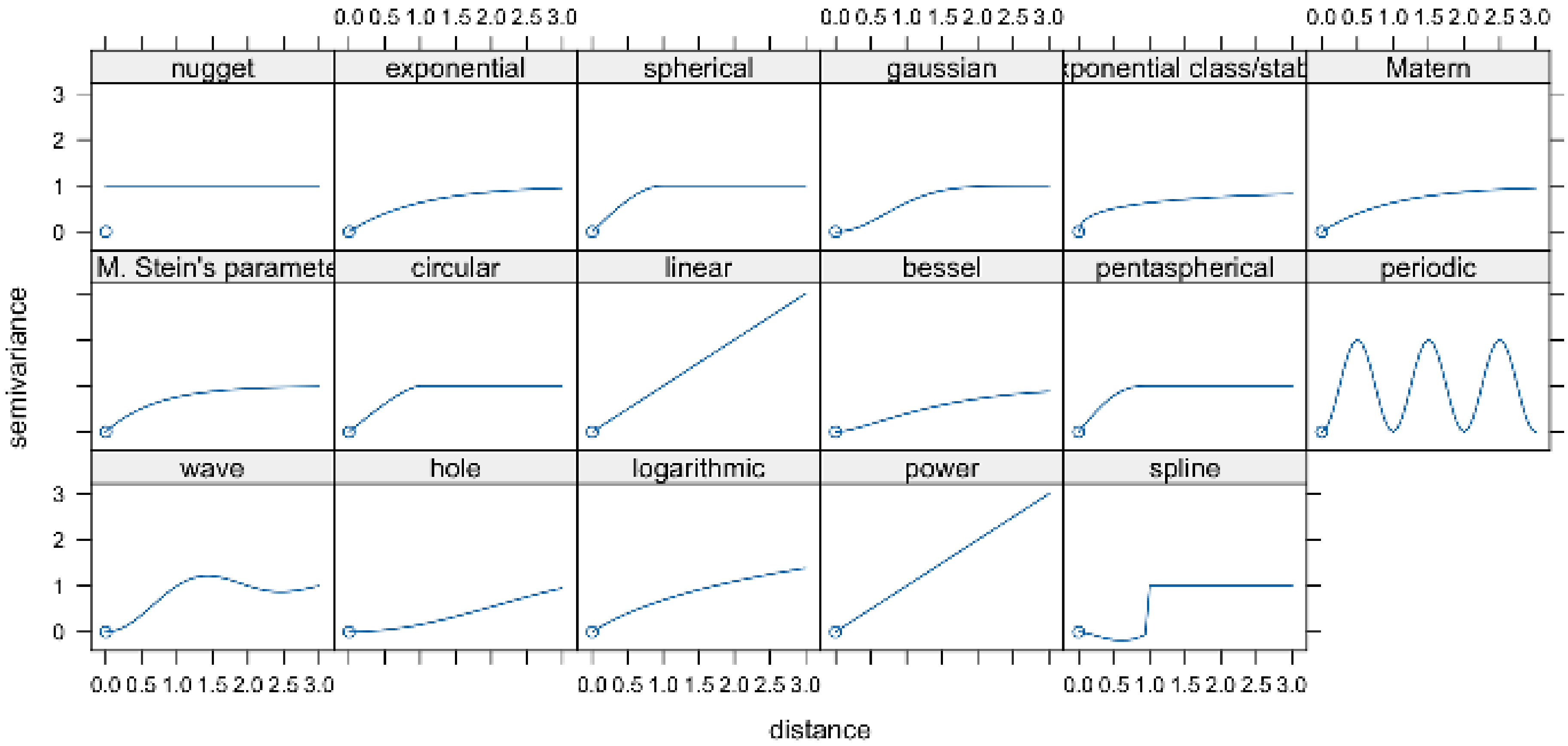
```
par(OP)
```

Simplifying the cloud plot

```r
# Compute the sample experimental variogram
var.smpl <- gstat::variogram(f.2, aq.sum, cloud = FALSE)

bins.ct <- c(0, var.smpl$dist , max(var.cld$dist) )
bins <- vector()
for (i in 1: (length(bins.ct) - 1) ){
  bins[i] <- mean(bins.ct[ seq(i,i+1, length.out=2)] )
}
bins[length(bins)] <- max(var.cld$dist)
var.bins <- findInterval(var.cld$dist, bins)

# Point data cloud with bin boundaries
OP <- par( mar = c(5,6,1,1))
plot(var.cld$gamma ~ eval(var.cld$dist/1000), col=rgb(0,0,0,0.2), pch=16, c
     xlab = "Distance between point pairs (km)",
     ylab = expression( gamma ) )
points( var.smpl$dist/1000, var.smpl$gamma, pch=21, col="black", bg="red",
abline(v=bins/1000, col="red", lty=2)
```

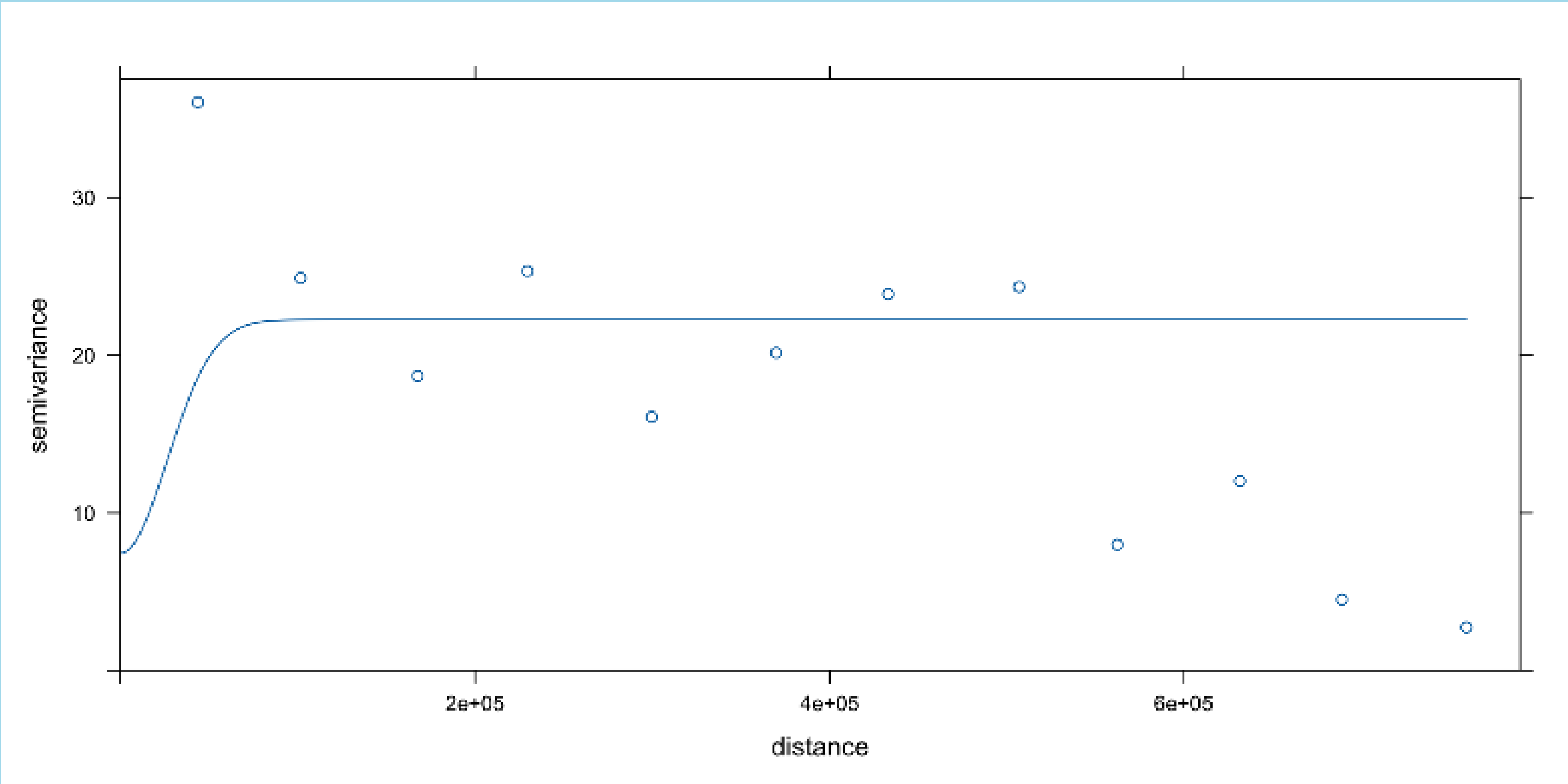

150
100
50
0
γ
50
100
150
200
250
Distance between point pairs (km)

```
par(OP)
```



nugget
exponential
spherical
gaussian
xponential class/stab
Matern
M. Stein's paramete
circular
linear
bessel
pentaspherical
periodic
wave
hole
logarithmic
power
spline
semivariance
distance
0.0 0.5 1.0 1.5 2.0 2.5 3.0
0
1
2
3

Looking at the sample Variogram

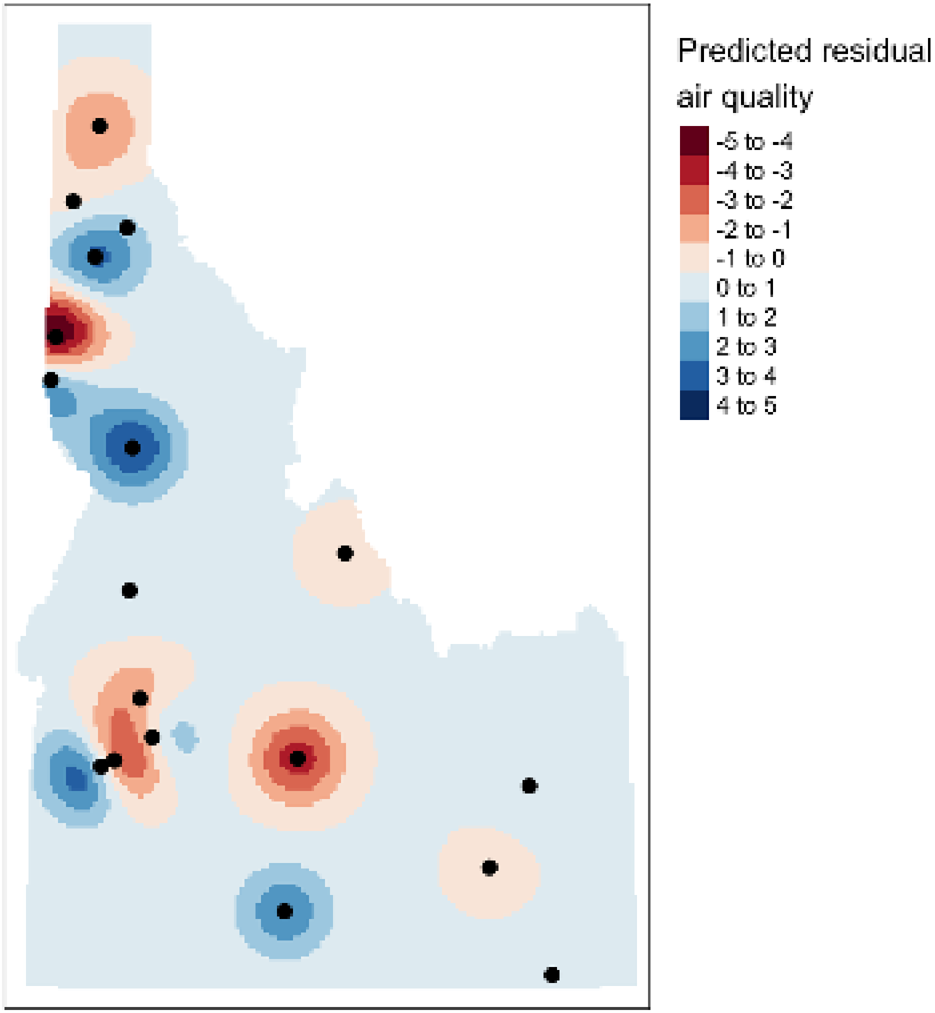


Estimating the sample variogram

```
var.smpl <- gstat::variogram(f.2, aq.sum, cloud = FALSE, cutoff = 1000000)


# Compute the variogram model by passing the nugget, sill and range values
# to fit.variogram() via the vgm() function.
dat.fit  <- gstat::fit.variogram(var.smpl, gstat::vgm(nugget = 12, range= 6
```

Ordinary Kriging

[using ordinary kriging]

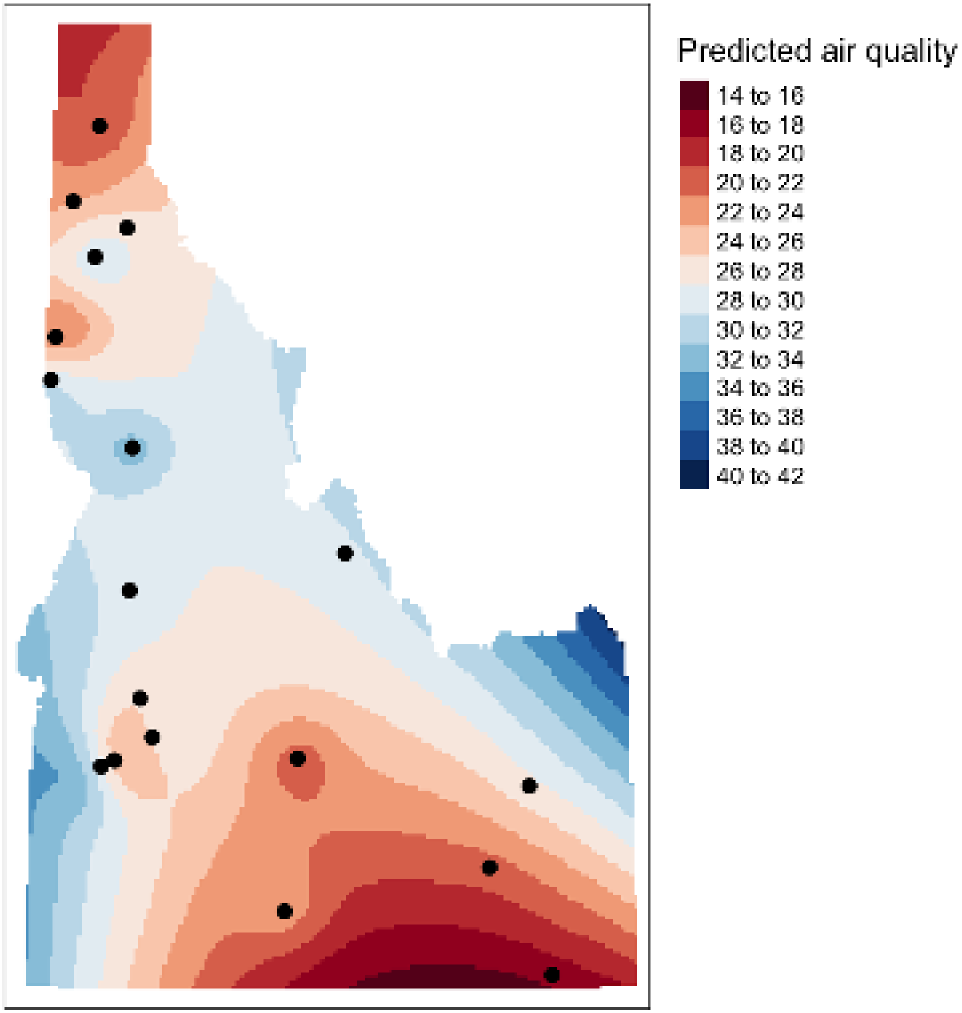


Ordinary Kriging

```
dat.krg <- gstat::krige( res~1, as(aq.sum, "Spatial"), grd, dat.fit)
```

Combining with the trend data

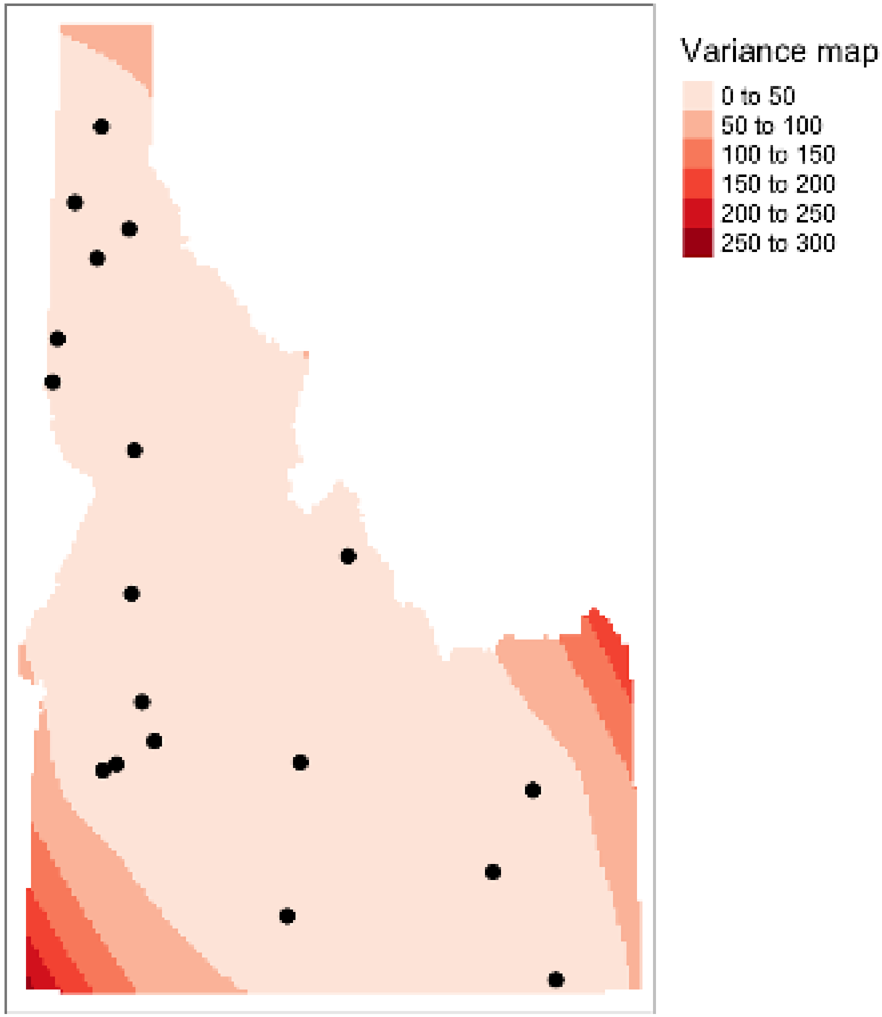
[using universal kriging]



Combining with the trend data

```
dat.krg <- gstat::krige( f.2, as(aq.sum, "Spatial"), grd, dat.fit)

r <- rast(dat.krg)$var1.pred
r.m <- mask(r, st_as_sf(id.cty))

# Plot the raster and the sampled points
tm_shape(r.m) + tm_raster(n=10, palette="RdBu", title="Predicted air qualit
  tm_legend(legend.outside=TRUE)
```

Visualizing Uncertainty



Universal Kriging

- Assumes that the deterministic part of the process ($\mu(\mathbf{x})$) is now a function of the location $\mathbf{x}$
- Could be the location or some other attribute
- Now y is a function of some aspect of x

```r
vu <- variogram(log(zinc)~elev, ~x+y, data=meuse)
mu <- fit.variogram(vu, vgm(1, "Sph", 300, 1))
gUK <- gstat(NULL, "log.zinc", log(zinc)~elev, meuse, locations=~x+y, model
names(r) <- "elev"
UK <- interpolate(r, gUK, debug.level=0)
```

Universal Kriging

Universal Kriging

```
vu <- variogram(log(zinc)~x + x^2 + y + y^2, ~x+y, data=meuse)
mu <- fit.variogram(vu, vgm(1, "Sph", 300, 1))
gUK <- gstat(NULL, "log.zinc", log(zinc)~x + x^2 + y + y^2, meuse, location
names(r) <- "elev"
UK <- interpolate(r, gUK, debug.level=0)
```

Universal Kriging

Co-Kriging

- relies on autocorrelation in $\epsilon_1(\mathbf{x})$ for z_1 AND cross correlation with other variables ($z_{2 \ldots j}$)
- Extending the ordinary kriging model gives:

$$z_1(\mathbf{x}) = \mu_1 + \epsilon_1(\mathbf{x})$$
$$z_2(\mathbf{x}) = \mu_2 + \epsilon_2(\mathbf{x})$$

* Note that there is autocorrelation within both z_1 and z_2 (because of the ϵ) and cross-correlation (because of the location, $\mathbf{x}$)

Co-Kriging

- Process is just a linked series of `gstat` calls

```
gCoK <- gstat(NULL, 'log.zinc', log(zinc)~1, meuse, locations=~x+y)
gCoK <- gstat(gCoK, 'elev', elev~1, meuse, locations=~x+y)
gCoK <- gstat(gCoK, 'cadmium', cadmium~1, meuse, locations=~x+y)
coV <- variogram(gCoK)
coV.fit <- fit.lmc(coV, gCoK, vgm(model='Sph', range=1000))

coK <- interpolate(r, coV.fit, debug.level=0)
```

Co-Kriging

Co-Kriging

A Note about Semivariograms