Spatio–temporal methods in environmental epidemiology

Lectures 9-11

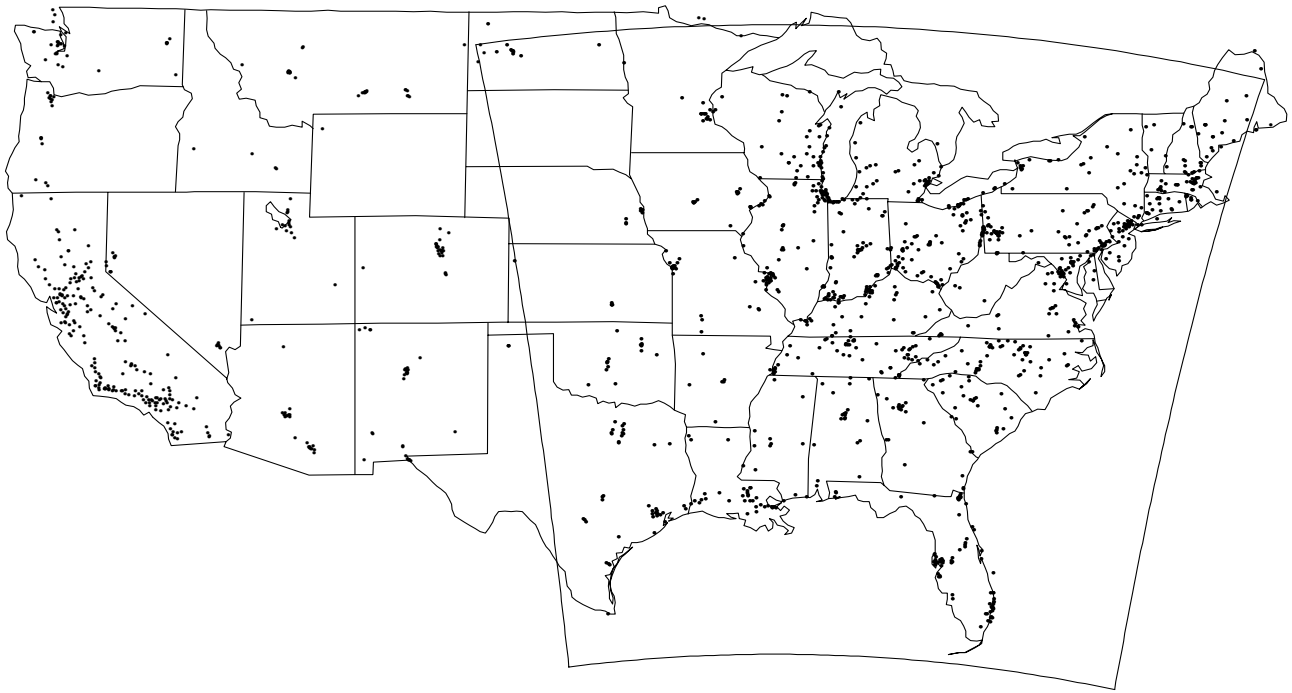
Overall outline

1. Temporal processes
2. Spatial processes or “fields”
   - Point referenced data
   - Area data
   - Point process data
3. Spatio–temporal processes

Spatial processes

Point referenced data

Example: US Ozone monitoring sites



Random field

a random process $Y(s),\ s \in D \subset R^d$ **for some** d

Usually $d = 2$. Interest focuses on stochastic inter-site spatial dependence (correlation in the case of Gaussian fields) between $Y(s_1)$ and $Y(s_2)$.

Why?

Correlation: is your enemy!

Suppose $Y(s_i) = \mu + W(s_i),\ i = 1, \ldots, p$ where for any two sites $corr[W(s_1), W(s_2)] = 0.97$. A naive statistician might take

$$\bar{Y} \pm 1.96 \frac{s}{\sqrt{p}}$$

as a 95% CI. But strong correlation effectively reduces the sample size to $p = 1$. It makes the CI much larger. Of particular concern in spatial regression where $Y = X\beta + \epsilon$ where Y is sample of measurements made at various locations in a random field.

Correlation: is your friend!

Strong intersite correlation enables strength to be “borrowed”. Measurements at a few sites can be used to predict the rest.

Bad and good correlation has thus led to an explosion of interest in stochastic models for random fields.

Some history about that interest

The foundations of geostatistics. D. G. Krige and G. Matheron formulated the theory of geostatistics and kriging in the 1950's [1].

Mining Industry.

- Ore deposits often highly concentrated in veins.
- Data collected through "core samples" at locations or sites at a fixed time.
- Data exhibit spatial correlation.

[1] Cressie, N. (1990). The Origins of Kriging. *Mathematical Geology*. **22**(3), pp. 239-252

Interest focussed on prediction at unsampled sites.

- Krige primarily interested in predicting block ore grades (areal block averages) from samples taken at points .
- Krige recognized spatial correlation of the measured values &
- predicted areal gold concentrations in South Africa using a lot of data with strong positive correlation[2].

[2] Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 52, 119–139

- Matheron first to publish a detailed exposition on geostatistics and kriging [3].
- G. Matheron called optimal prediction of responses at unsampled locations “kriging” to honor of D. G. Krige’s contributions.
- B. Matern as early in 1960, also derive the concept of spatial variation [4].
- Idea actually goes back a long way to at least Kolmogorov [5].

[3] Matheron, G. (1963). Principles of Geostatistics. *Economic Geology*. 58, pp. 1246-1266

[4] English version: Matern, B. (1986). *Spatial Variation*. Springer, New York

[5] Kolmogorov, A. N. (1941), Interpolirovanie i ekstrapolirovanie statsionarnykh sluchainyk posledovatel’ nostei (Interpolated and extrapolated staionary random sequences), *Izvestya Akademiya Nauk SSSR, Seriya Matematicheskaya*, **5**, 3–14

Geostatistics overview

A two step modelling approach is taken in the analysis of spatially correlated data:

1. **Modelling spatial variation:**
   - Develop a model for the correlation structure.
   - Primary method for determining spatial correlation uses the “variogram”.
   - Expresses observation differences between sites as function of their intersite separation index.
   - Immense literature on this topic.

2. Prediction:

- Use model to predict responses at unsampled locations or at areal units.
- Standard technique: kriging. Incorporates variogram to find optimal unbiased predictor (BLUP) for unsampled site.
- Has advantages over other interpolation methods e.g. inverse distance weighting.
  1. Standard interpolation techniques fail to exploit spatial correlation
  2. Kriging comes with predictive (95% say) interval to assess prediction accuracy.

Geostatistics: Study of spatial data where the spatial correlation is modeled through the variogram. Focus of geostatistical analyses: prediction of responses at unsampled sites using *distribution-free* methodology.

Model-based Geostatistics: Geostatistical analyses assuming the observed spatial data follow a specific sampling distribution (for example, normality is the most common). This assumption permits building several statistical models for which formal tests to choose between exist.

Primary modeling decision: must choose between

- a parametric family of distributions from which the responses arise or
- taking a distribution-free approach.

Avoiding the specification of a sampling distribution: has advantages, notably sidesteps limitations in resulting statistical inference.

Model-based approach: Gaussian random field traditionally assumed, partly for convenience and partly for the flexibility that this family offers.

Gaussian limitations: Many environmental responses are binary or positive with right-skewed distributions (such as water pollution concentrations). Then log-normality often assumed by default. But other transformations to normality can be employed.

More general linear models: (GLMs): can be used in model-based geostatistical analyses.

Spatial modelling: start with EDA! Case study 1

The Maas or Meuse: major European river. Rises in France. Flows through Belgium & the Netherlands. Draining into North Sea. ITotal length of 925 km. Has been monitored over time.

Various chemical measurements are stored in a dataset found in the gstat package.

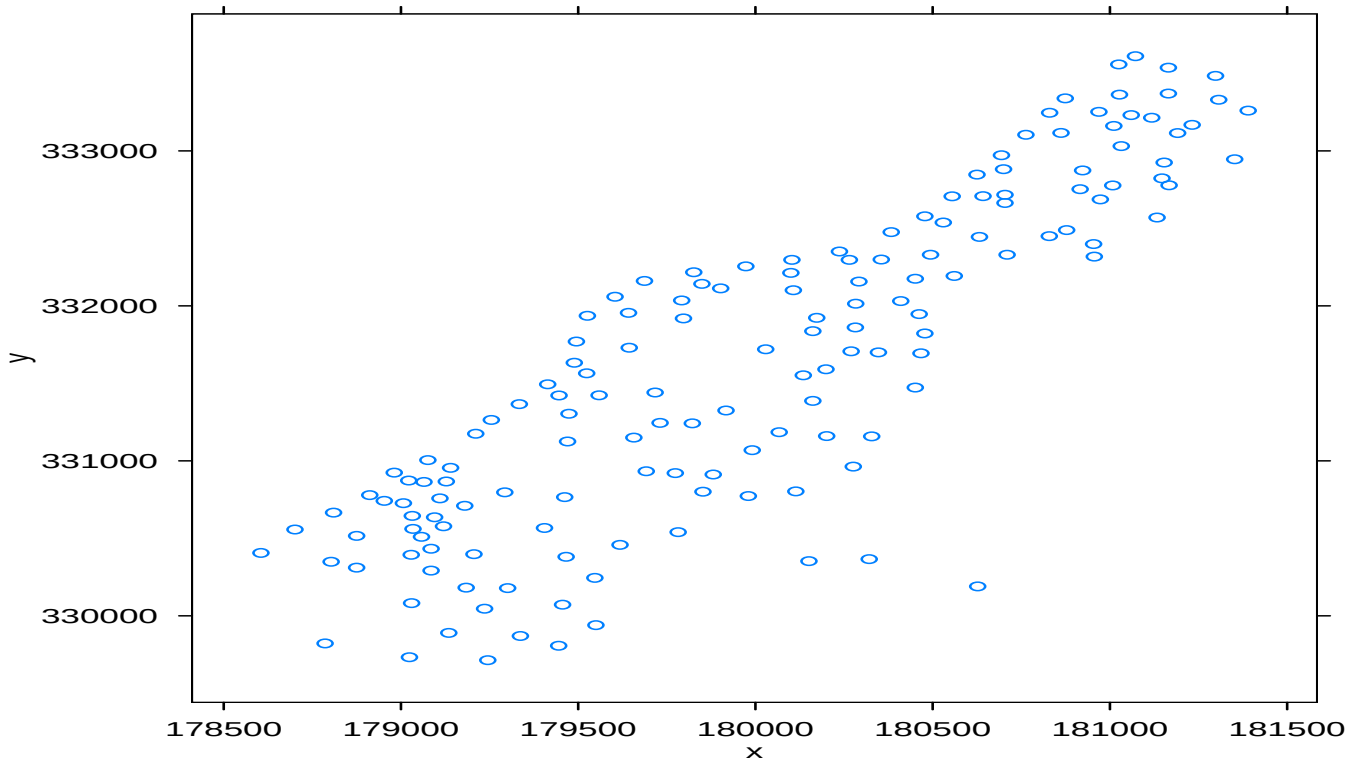
```
> library(gstat)
> data(meuse)
> str(meuse)
```

G

```
'data.frame':   155 obs. of  13 variables:
 $ x       : num  181072 181025 181165 181298 181307 ...
 $ y       : num  333611 333558 333537 333484 333330 ...
 $ cadmium: num  11.7 8.6 6.5 2.6 2.8 3 3.2 2.8 2.4 1.6 ...
 $ copper : num  85 81 68 81 48 61 31 29 37 24 ...
 $ lead   : num  299 277 199 116 117 137 132 150 133 80 ...
 $ zinc   : num  1022 1141  640  257  269 ...
 $ elev   : num  7.91 6.98 7.80 7.66 7.48 ...
 $ dist   : num  50 30 150 270 380 470 240 120 240 420 ...
 $ om     : num  13.6 14 13 8 8.7 7.8 9.2 9.5 10.6 6.3 ...
 $ ffreq  : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1
 $ soil   : Factor w/ 3 levels "1","2","3": 1 1 1 2 2 2 2 1
 $ lime   : Factor w/ 2 levels "0","1": 2 2 2 1 1 1 1 1 1 1
 $ landuse: Factor w/ 15 levels "Aa","Ab","Ag",..: 4 4 4 11
```

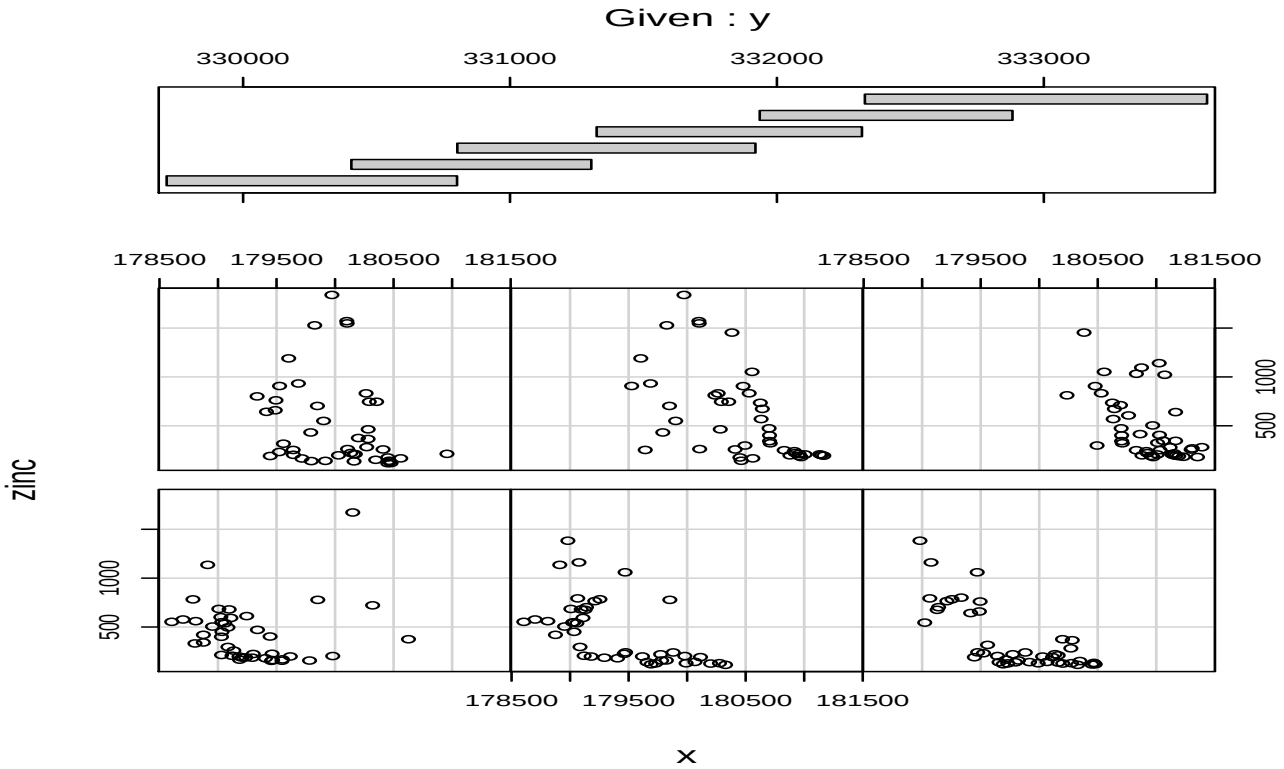
First of all: inspect sampling locations.

```
> print(xyplot(y ~ x, data = meuse))
```

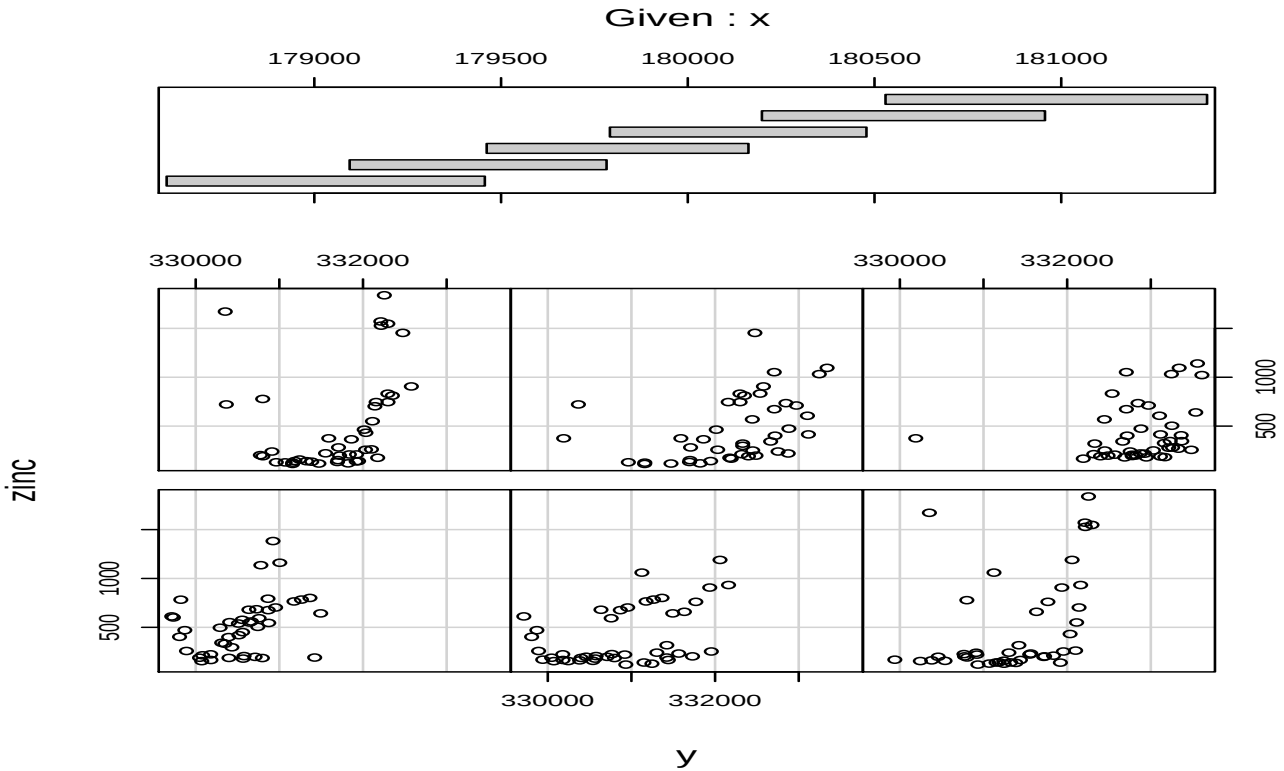


Ad hoc checks for non-stationarity can be done. Conditioning plots are one such approach.

```
> coplot(zinc ~ x | y, data = meuse)
```

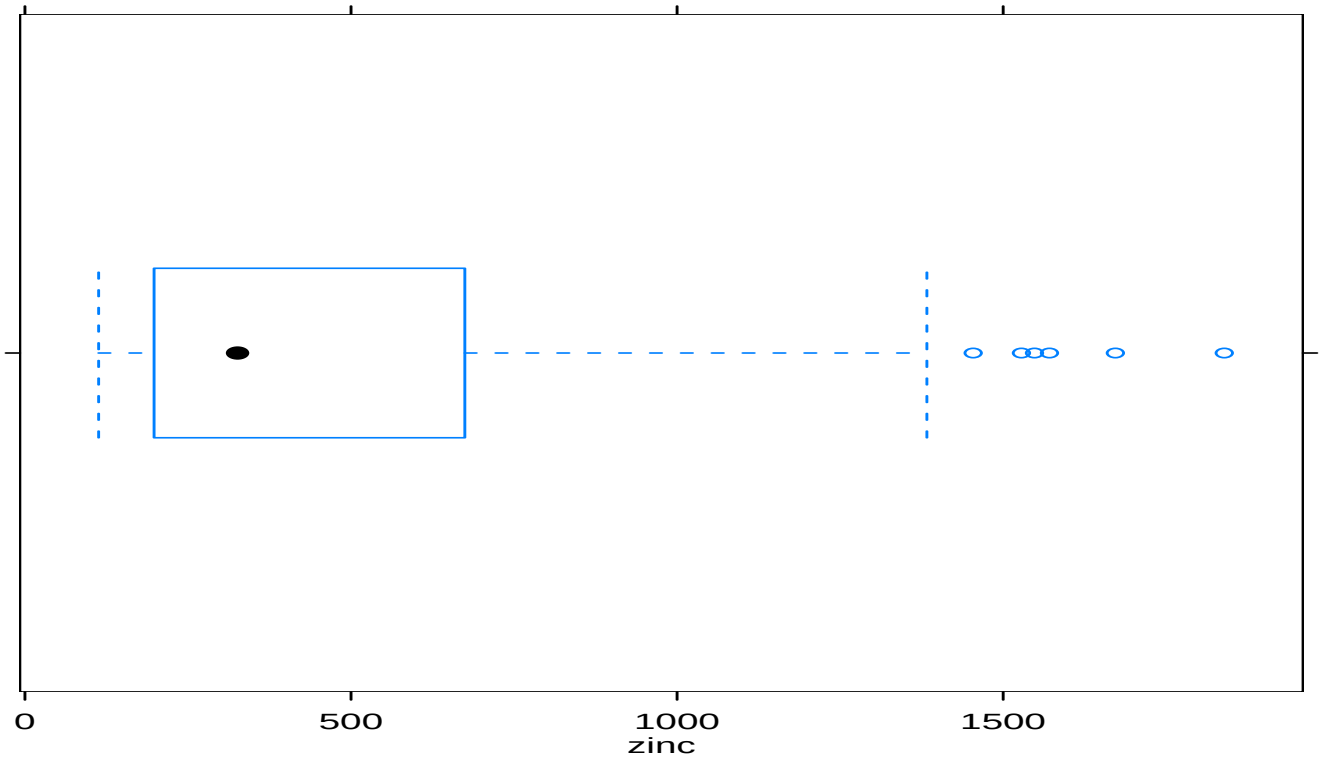


```
> coplot(zinc ~ y | x, data = meuse)
```



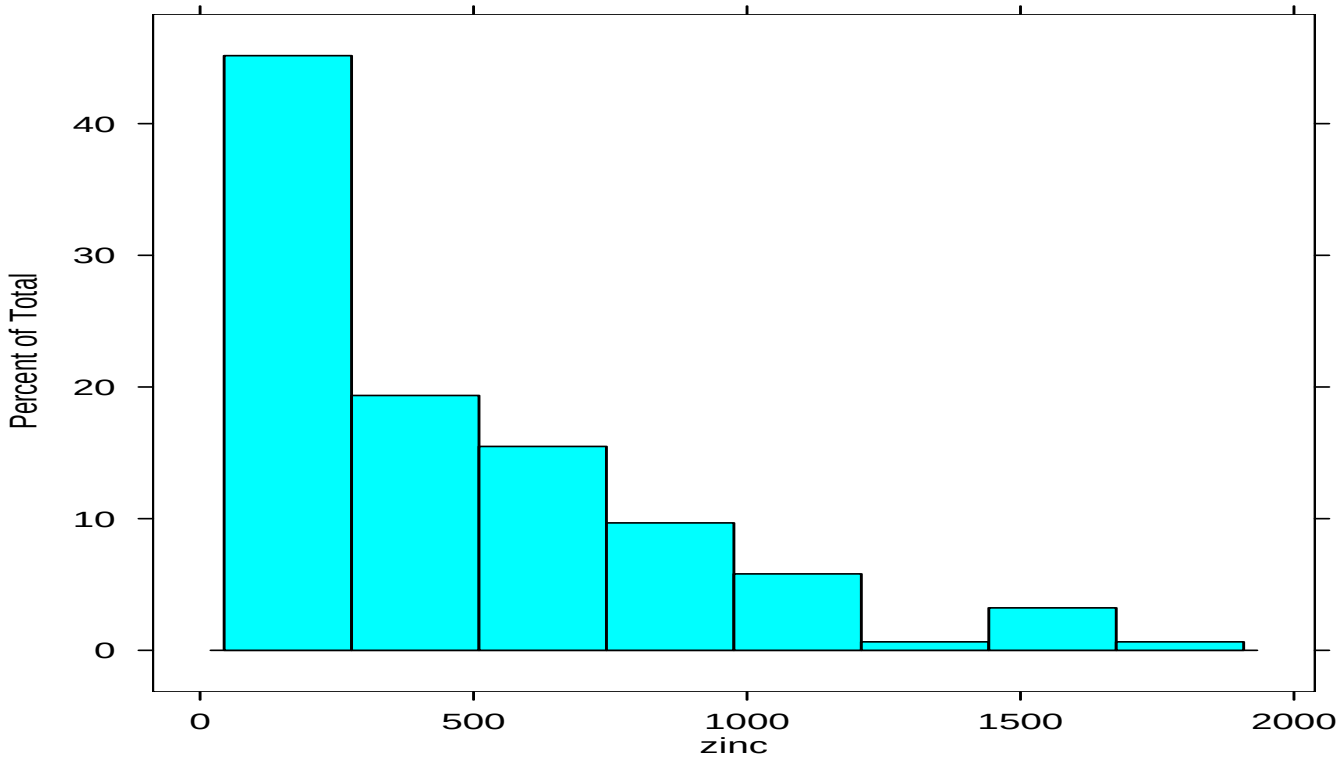
Next distributional checks, e.g. ”Box and Whisker plot”.

```
> print(bwplot(~zinc, data = meuse))
```



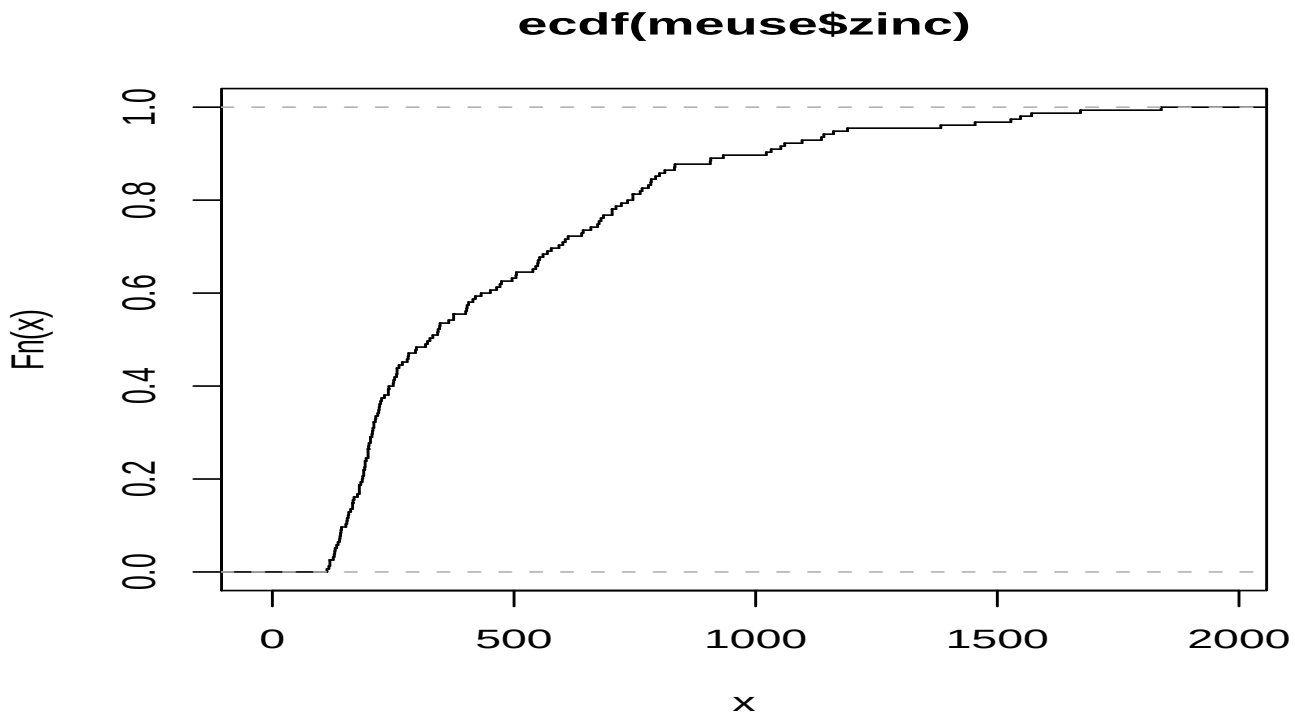
Or a histogram.

```
> print(histogram(~zinc, data = meuse))
```



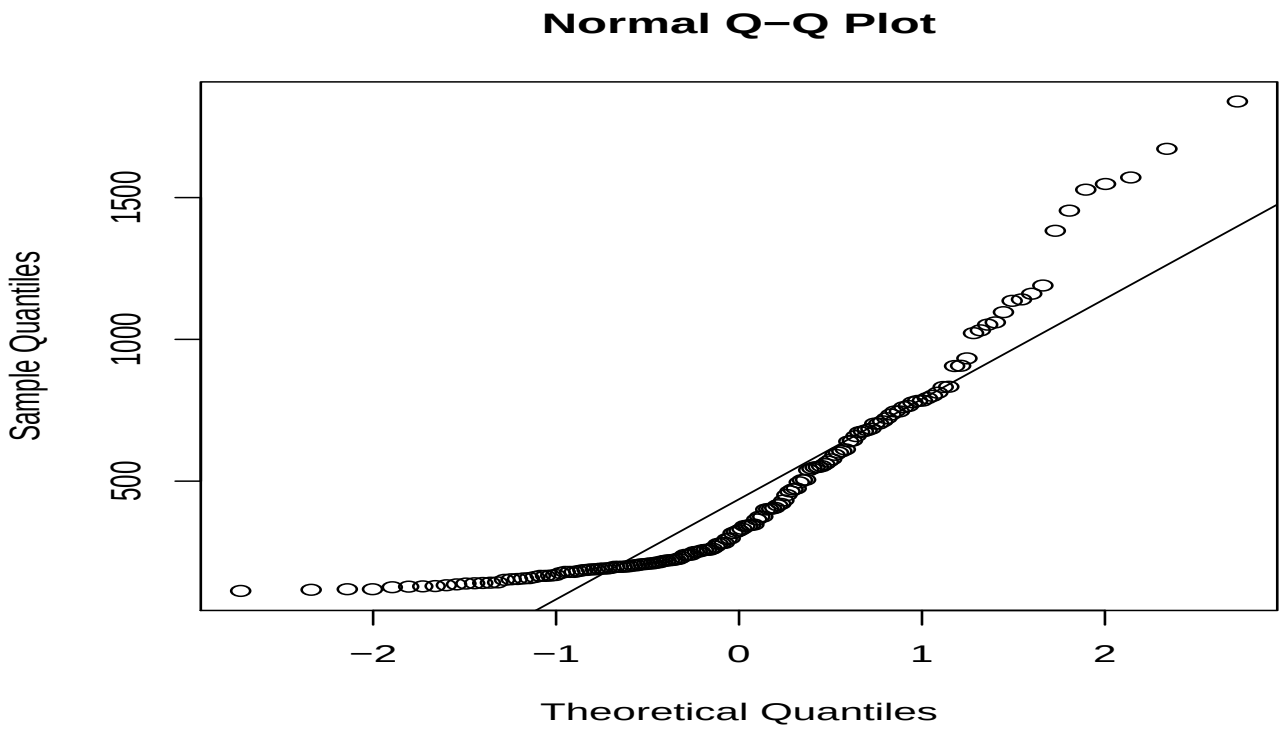
The empirical cdf

```
> library(stepfun)
> cdf.zinc <- ecdf(meuse$zinc)
> plot(cdf.zinc, verticals = T, do.points = F)
```



The q–q (quantile-quantile) plot

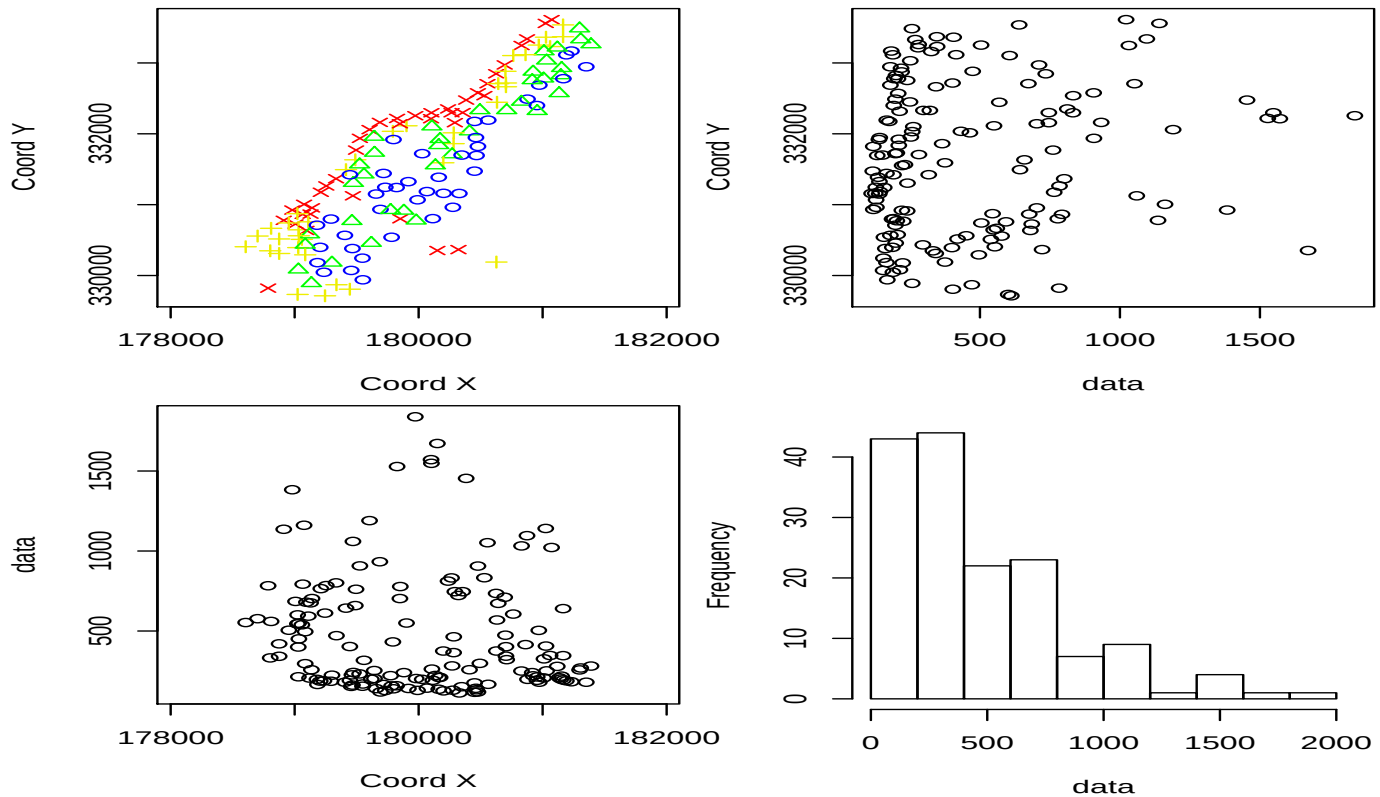
```
> qqnorm(meuse$zinc)
> qqline(meuse$zinc)
```



geoR alternative to gstat has nice plot function. But data must be converted to a geodata object from the meuse dataset, a dataframe object.

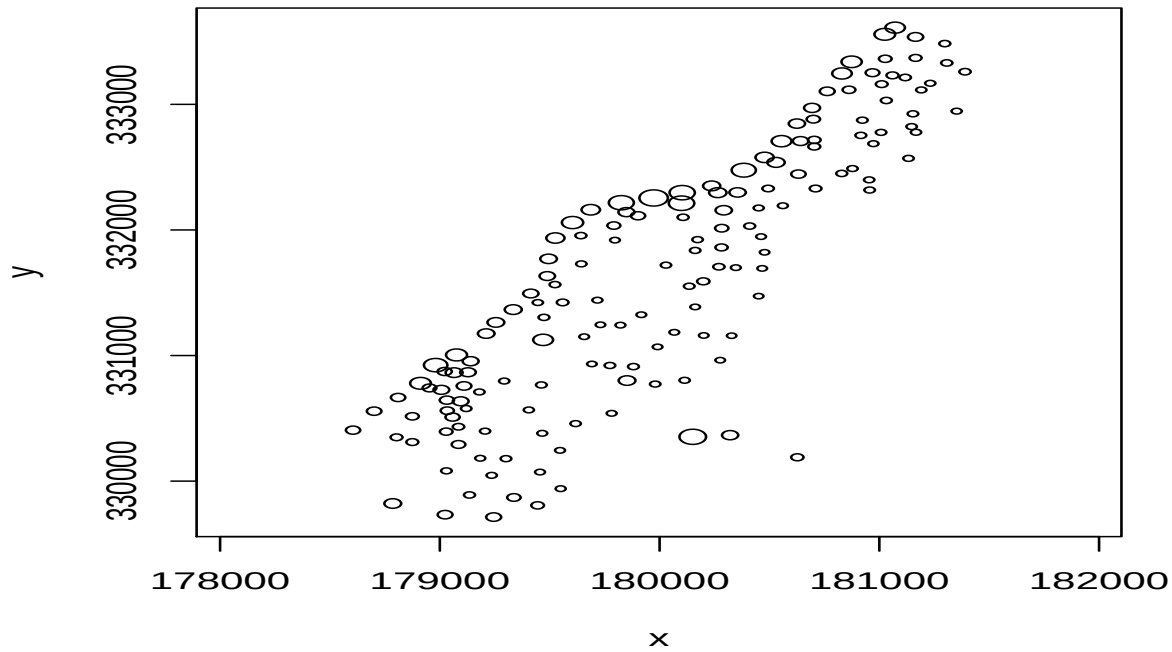
```
> library(geoR, warn = F)
> meuse.geo <- as.geodata(meuse, data.col = 6)

Loading required package: mva
```

```
> plot(meuse.geo)
```



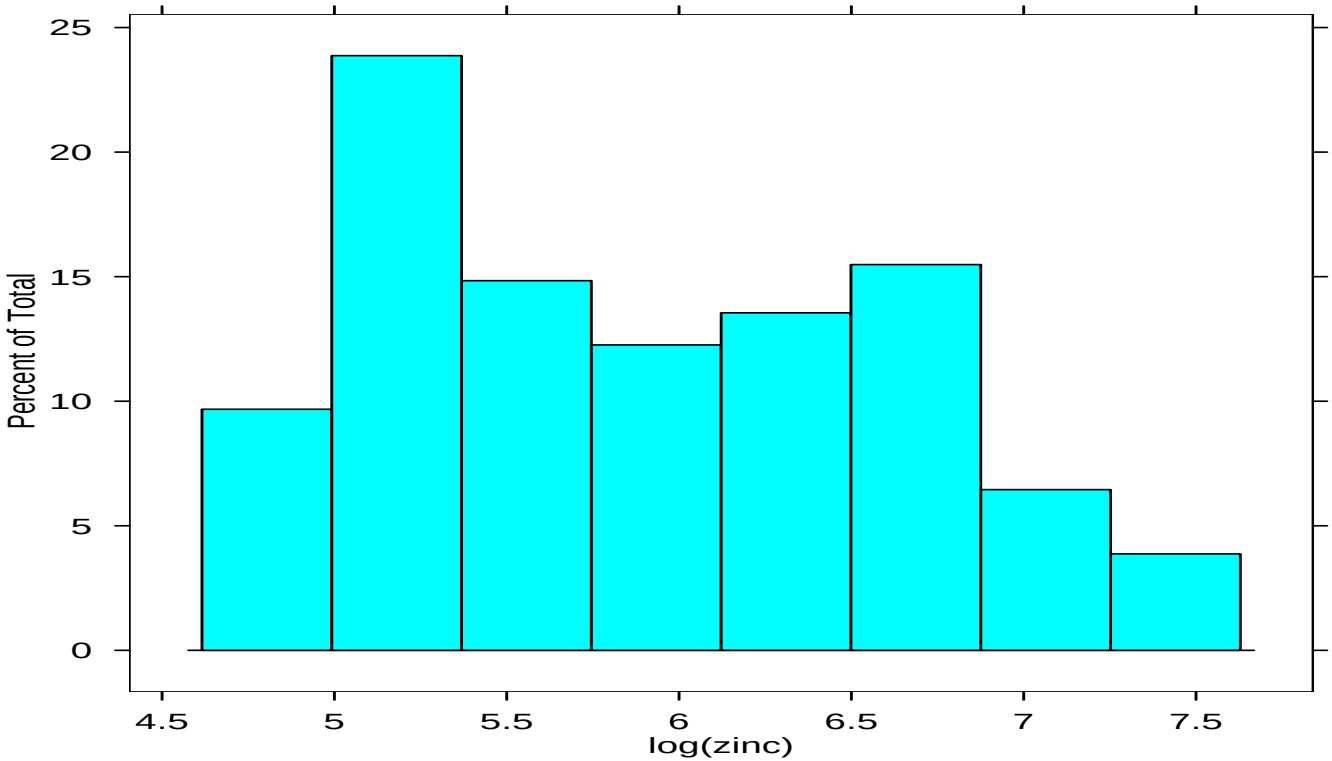
The geoR function `points.geodata()` checks for non-stationarity.

```
> points.geodata(meuse.geo)
```

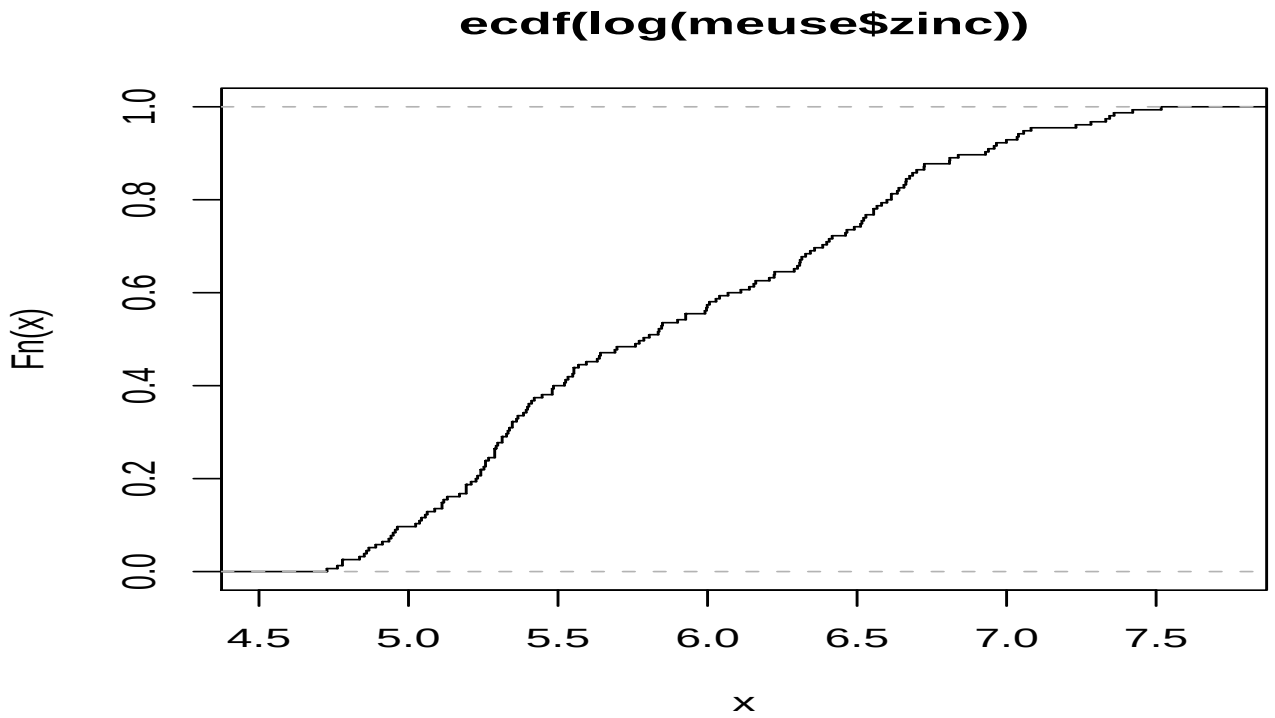


log transform the zinc data in the meuse dataset.

```
> print(histogram(~log(zinc), data = meuse))
```

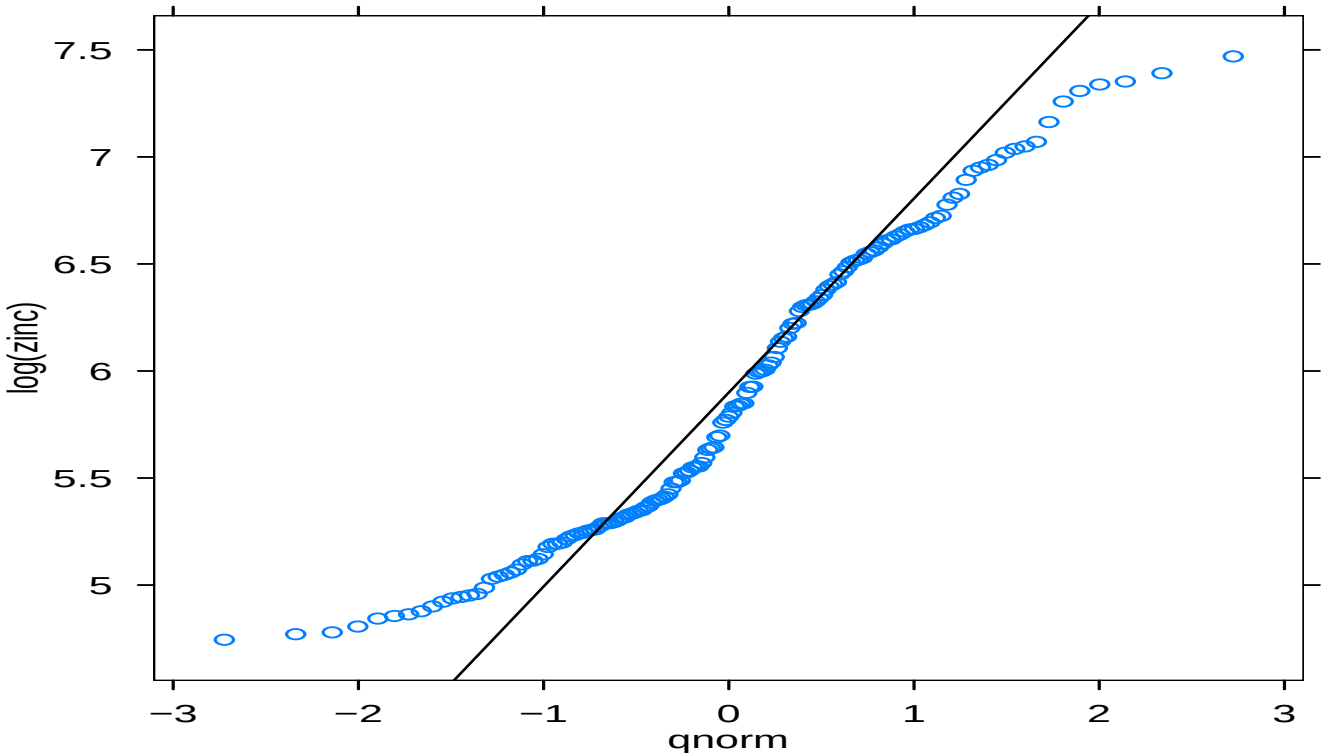


```
> cdf.zinc <- ecdf(log(meuse$zinc))
> plot(cdf.zinc, verticals = T, do.points = F)
```

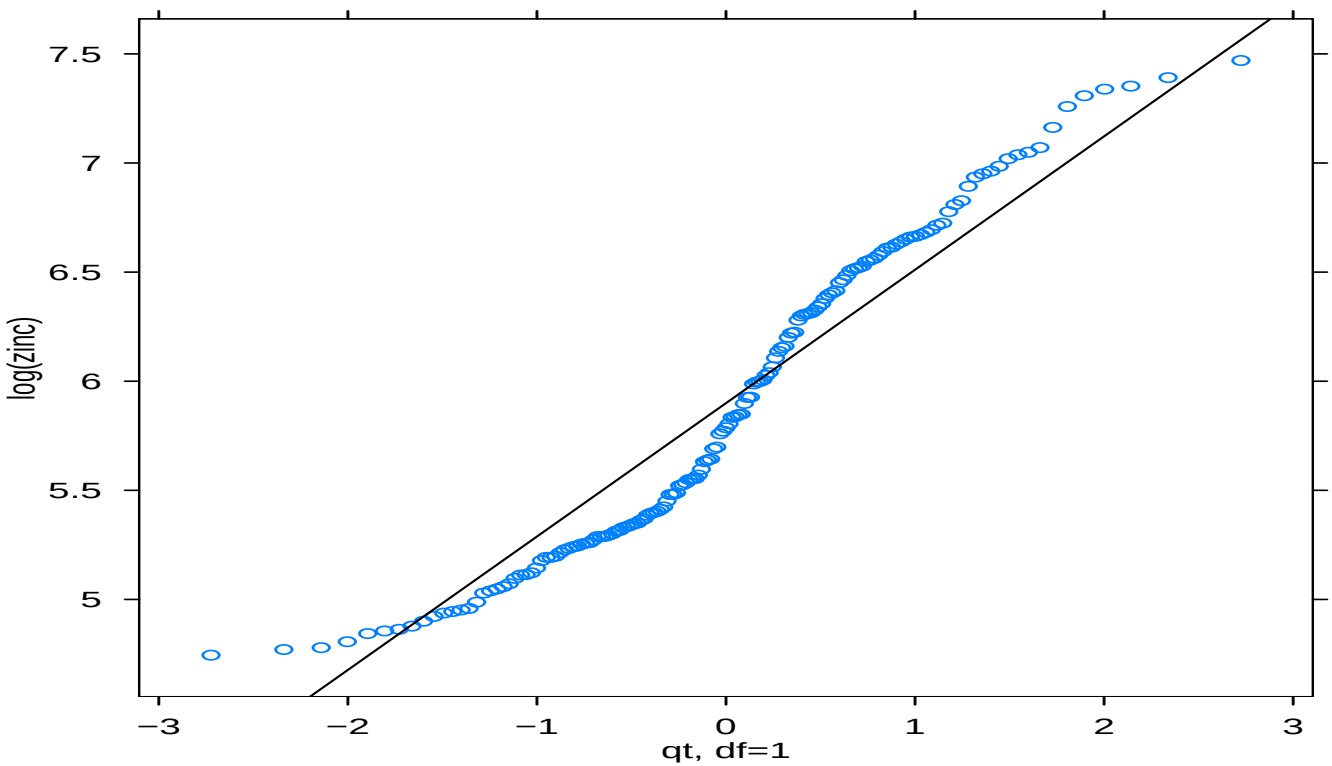


Instead of `qqnorm()` you can use the `qqmath()` from the lattice package.

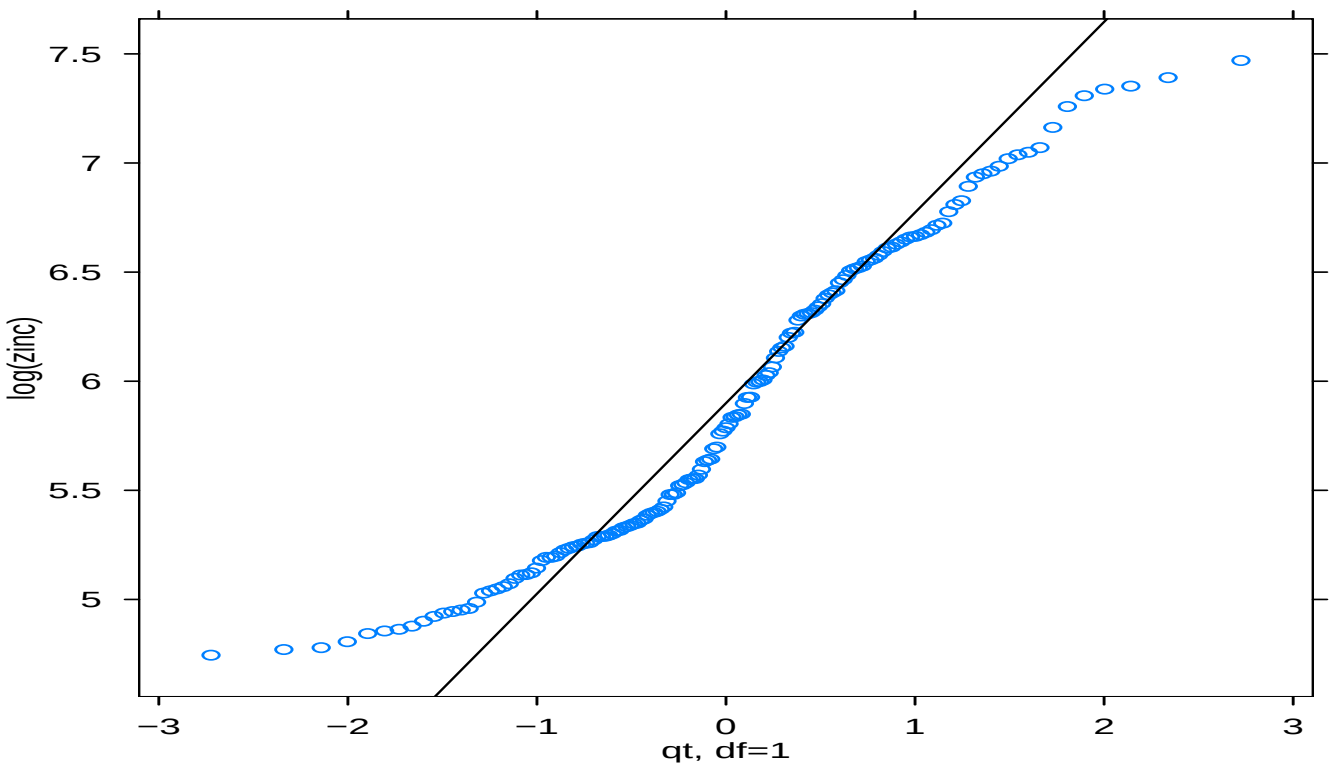
```
> print(qqmath(~log(zinc), data = meuse, panel = funct
+      panel.qqmath(x, ...)
+      panel.qqmathline(x, distribution = qnorm, ...)
+ }))
```



```
> print(qqmath(~log(zinc), data = meuse, panel = funct
+      panel.qqmath(x, ...)
+      panel.qqmathline(x, distribution = function(p) q
+          ...)
+ }, xlab = "qt, df=1"))
```



```
> print(qqmath(~log(zinc), data = meuse, panel = funct
+      panel.qqmath(x, ...)
+      panel.qqmathline(x, distribution = function(p) q
+          ...)
+}, xlab = "qt, df=10"))
```



CONCLUSION: Log zinc fitted reasonably well with a student t distribution - very heavy tails.

Some theory: Moments

$Y \sim F$: random vector field. Fixed time t omitted in sequel. s and x commonly used for spatial coordinates, e.g. (lat, long). We use s. For locations $\{s_1, \ldots, s_g\}$ for any g

$$F_{s_1,\ldots,s_g}(y_1, \ldots, y_g) \equiv P\{Y(s_1) \leq y_1, \ldots, Y(s_g) \leq y_g\}.$$

$F_{s_1,\ldots,s_g}(y)$ is joint distribution distribution (DF)

- **Moment** of k^{th}*-order:*

$$E[Y(s)]^k \equiv \int y^k dF_s(y)$$

- **Expectation**: If exists, defined as the 1^{st}-order moment for any s

$$\mu(s) \equiv E[Y(s)]$$

- **Variance:**

$$Var[Y(s)] \equiv E[Y(s) - \mu(s)]^2.$$

- **Covariance** between locations s_1 & s_2,

$$C(s_1, s_2) \equiv E[(Y(s_1) - \mu(s_1))(Y(s_2) - \mu(s_2))]$$

  - **NOTE:** $C(s_1, s_1) \equiv Var[Y(s_1)]$

Stationarity

An important concept in characterizing the random field Y

- **Strict stationarity**
  Y *strictly stationary* if:

$$F_{s_1,\ldots,s_n}(\mathbf{y}) = F_{s_1+h,\ldots,s_n+h}(\mathbf{y})$$

  for any vector h & an arbitrary n

- **Second–order stationarity**
  Y is *second-order stationary* if:

$$\begin{array}{lcccl} \mu(s) & = & E[Y(s)] & = & \mu \\ C(s, s+h) & = & C(s+h-s) & = & C(h) \end{array}$$

  - when $h = 0$: $Var[Y(s)] = C(s,s) = C(0)$
    ie. ***Mean, Variance do not depend on location***

Stationarity

- **Second–order stationarity - cont'd**
  - $C(h)$: *covariogram* (or *autocovariance* in time series)
  - Implies ***Intrinsic Stationarity*** *(weaker)*

$$\begin{aligned} Var[Y(s) - Y(s+h)] &= Var[Y(s)] + Var[Y(s+h)] \\ &\quad -2Cov[Y(s), Y(s+h)] \\ &= C(0) + C(0) - 2C(h) \\ &= 2[C(0) - C(h)]. \end{aligned}$$

or equivalently semi-variogram

$$\gamma(h) = C(0) - C(h).$$

Properties of C(h)

X second-order stationary process with covariance function $C(h)$.

- **Positive Definiteness (PD):** If $\Sigma = \{C(h_{ij})\}$ being covariance matrix of random vector $(Y(s_1), \ldots, Y(s_n))$ makes it PD implying for any vector a that:

$$\sum_i \sum_j a_i a_j C(h_{ij}) > 0$$

Exercise!

- **Anisotropy**: $C(h)$ - function of length & direction
- **Isotropy**: $C(h)$ - function only of length $|h|$

Variograms

Matheron supposed that at least for small $\mid h \mid$

$$E[Y(s+h) - Y(s)] = 0$$

would be reasonable assumption. He then defined the

- **Variogram:**

$$\begin{aligned} 2\gamma(h) &\equiv var[Y(s+h) - Y(s)] \\ &= E[Y(s+h) - Y(s) - (\mu(s+h) - \mu(s))]^2. \\ &= E[Y(s+h) - Y(s)]^2. \end{aligned}$$

- $\gamma(h)$ is called *semi-variogram*.

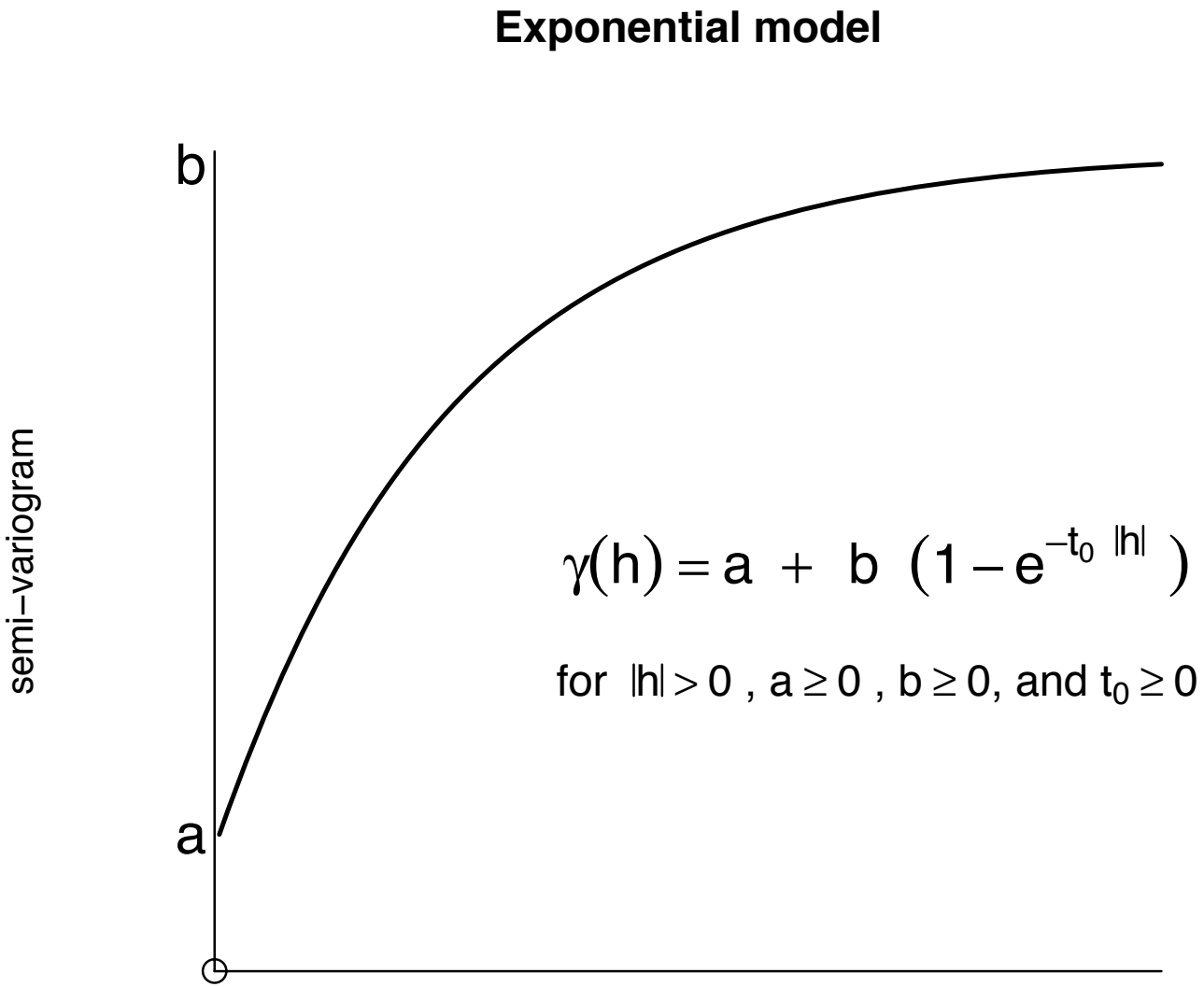
Isotropic Semi-Variogram Models

Second order stationarity implies $\gamma(h) = C(0) - C(h) \to \gamma(0) = 0$

- But often $\lim_{h \to 0} \gamma(h) \neq 0$. Discontinuity called ***nugget effect***.
- When $\gamma(h) \to B$ as $h \to \infty$, B called a ***sill***

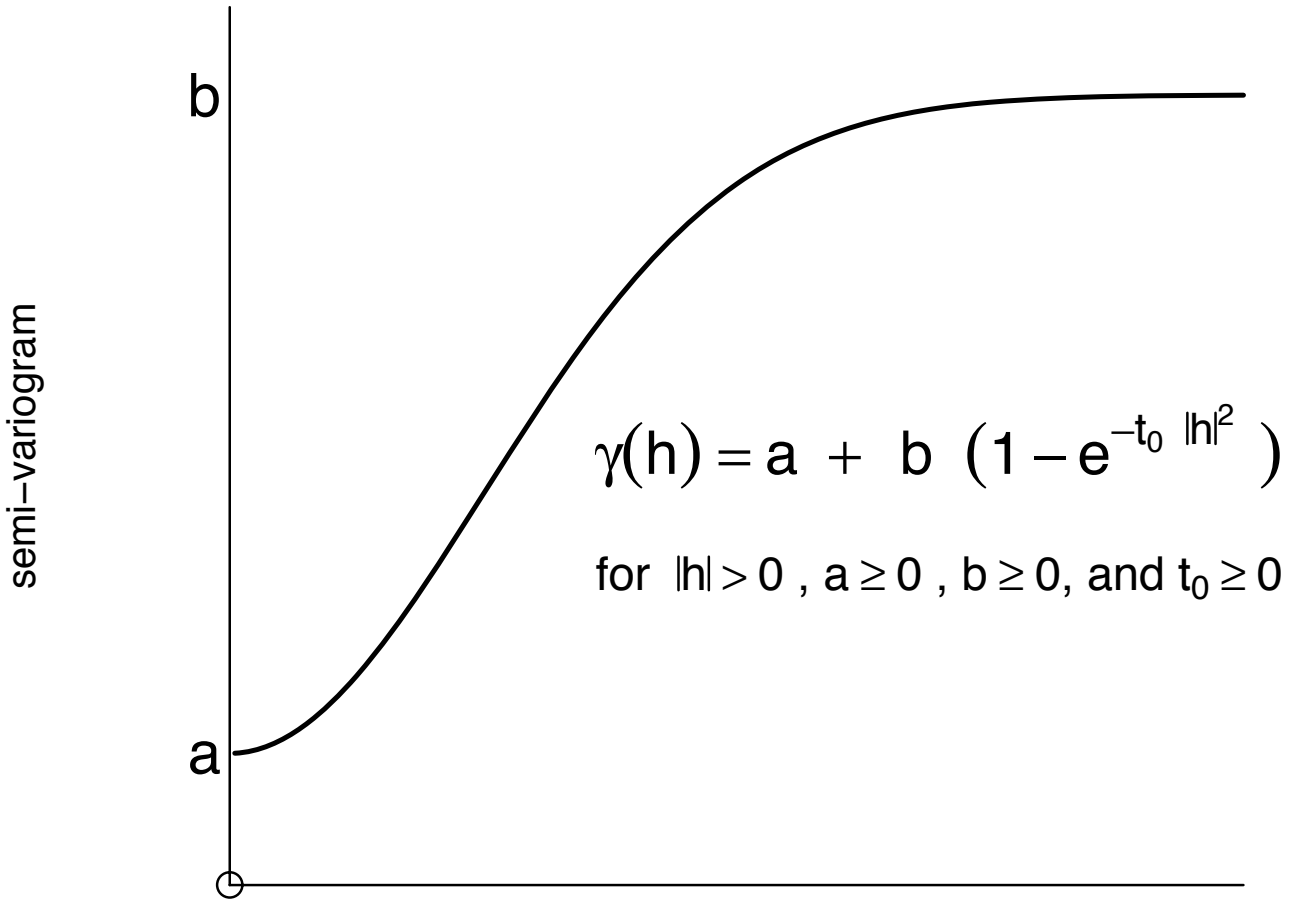
Note: Few functions satisfy positive definiteness condition - only certain ones (eg. variogram)

Common isotropic models

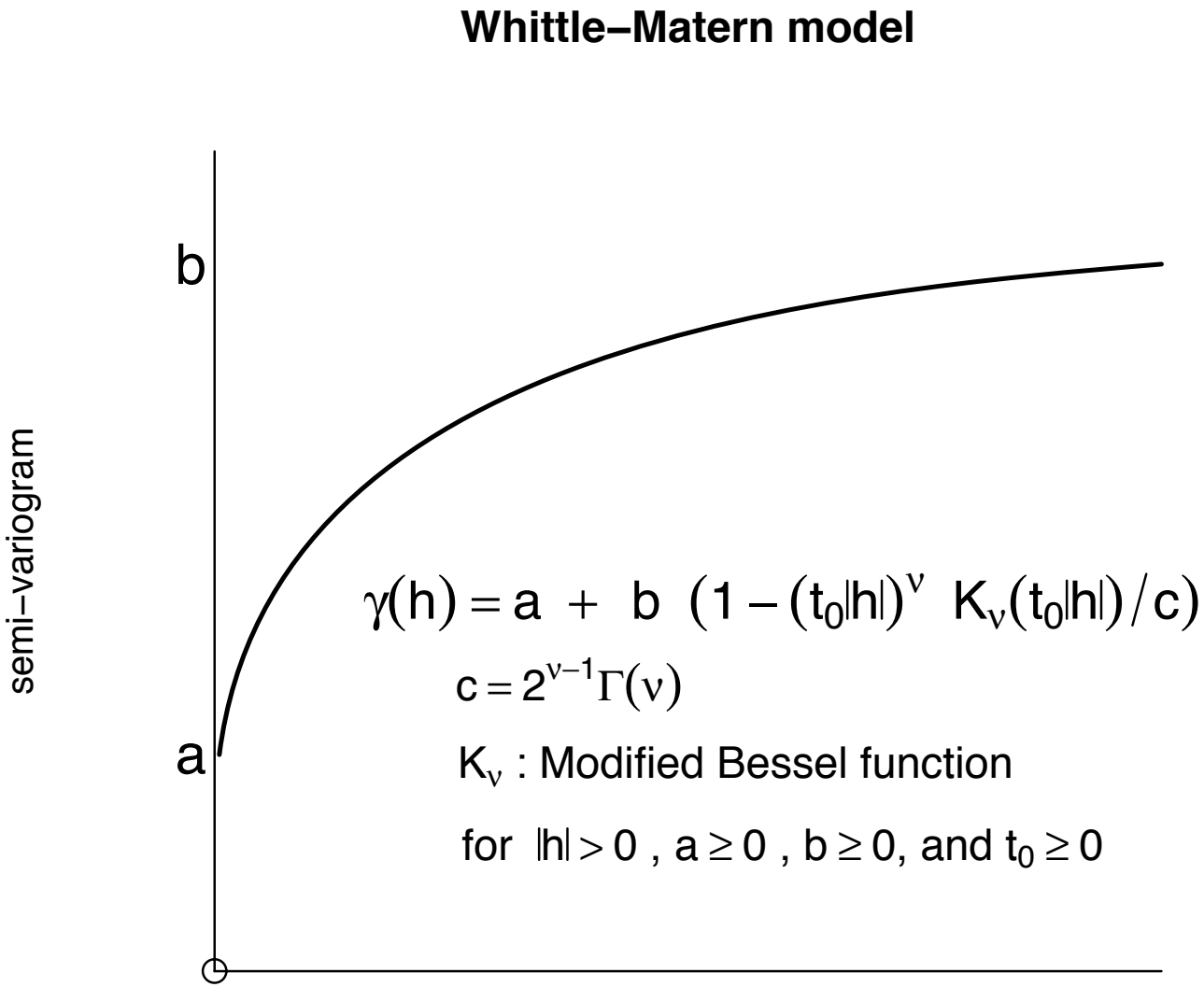


Common isotropic models

Gaussian model



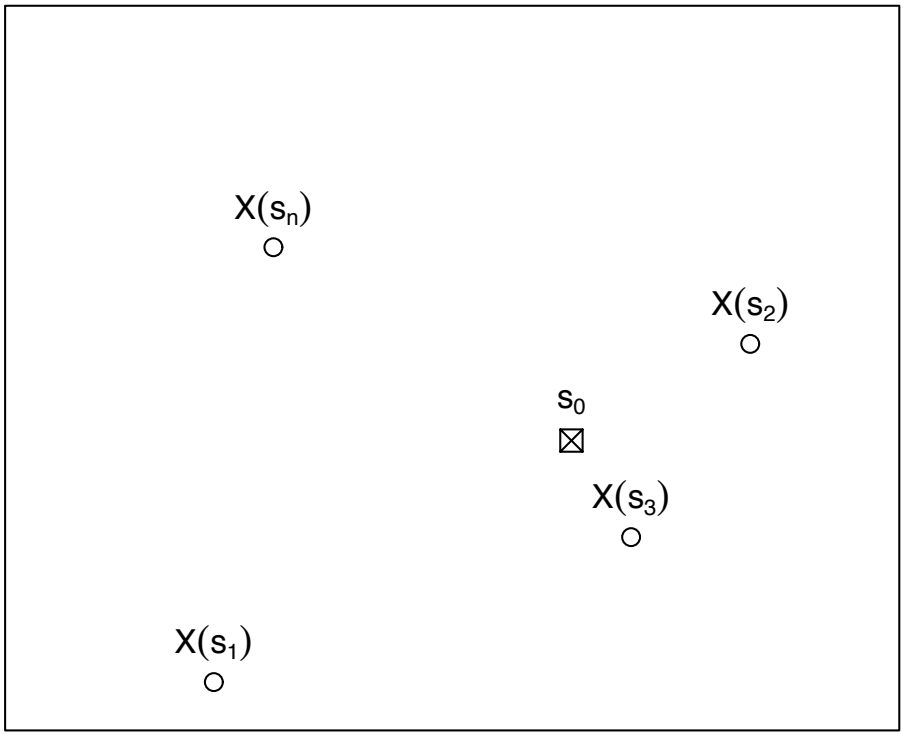
Common isotropic models



Spatial prediction

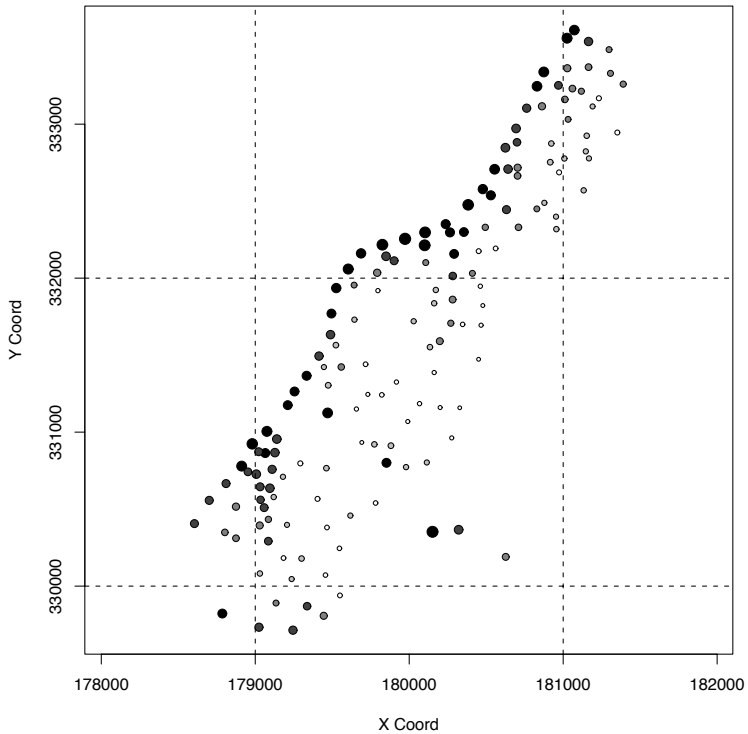
Typo: Change X to Y

Problem: Estimate at location s_0 given observed levels $X(s_i)$?



Case study 1: Zinc levels in the Netherlands

Values of log zinc at sampling locations. Mapping the basin would mean predicting unmeasured responses at other sites without measurements.



Ordinary Kriging

Goal:. Ignoring measurement error for simplicity predict $Y(s_0)$ given observations $y_1, \ldots, y_n$ at locations $s_1, \ldots, s_n$. Assumption

- Covariance structure known
- $Y(s) = \mu + W(s)$ & intrinsic stationary, ie.

$$\begin{aligned} E[Y(s)] &= \mu \\ Var[Y(s) - Y(s+h)] &= 2\gamma(|h|) \end{aligned}$$

- *Linear predictors:* $Y^*(s_0) = \sum_{i=1}^n \alpha_i Y(s_i)$

Reaching the goal: choose $\{\alpha\}$ to get unbiasedness & minimal prediction error $\sigma^2_{s_0} \equiv E\left[Y^*(s_0) - Y(s_0)\right]^2$

Result: Kriging predictor = best linear unbiased predictor (BLUP)

References: (?) & (?)

Ordinary Kriging system

- $E[Y^*(s_0)] = E\left[\sum_{i=1}^n \alpha_i Y s_i)\right] = \mu \sum_{i=1}^n \alpha_i \qquad (1)$
  implies $\sum_{i=1}^n \alpha_i = 1$.
- **Prediction error (Kriging variance). (Exercise: Verify)**

$$\begin{aligned}
\sigma^2_{s_0} &\equiv E\left[Y^*(s_0) - Y(s_0)\right]^2 = E\left[\sum_{i=1}^n \alpha_i [Y(s_i) - Y(s_0)]\right]^2 \\
&= \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j E[Y(s_i) - Y(s_j)]^2/2 \\
&\quad - \sum_{i=1}^n \alpha_i E[Y(s_i) - Y(s_0)]^2 \\
&= \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \gamma(|h_{ij}|) - 2\sum_{i=1}^n \alpha_i \gamma(|h_{i0}|) \qquad (2)
\end{aligned}$$

α's chosen to minimize (2) & satisfy (1)

- Solution for α's:

$$\begin{cases} \partial f/\partial \alpha_i &= 0 \quad i = 1, \ldots, n \\ \\ \partial f/\partial \lambda &= 0 \end{cases}$$

where $f(\alpha_1, \ldots, \alpha_n, \lambda) = \sigma^2_{s_0} + 2\lambda \left(\sum_{i=1}^n \alpha_i - 1\right)$

- $\Longrightarrow$ *ordinary Kriging system* (Exercise: Verify)

$$\begin{cases} \sum_{j=1}^n \alpha_j \gamma(|h_{ij}|) + \lambda &= \gamma(|h_{i0}|) \\ \\ \sum_{j=1}^n \alpha_j &= 1 \end{cases}$$

for $i = 1, \ldots, n$; h_{ij}: distance between s_i & s_j

Implementation in summary

- Select good semi-variogram model. Estimate $\hat{\gamma}(.)$ since it will not be known as assumed.
- Solve the *Kriging system* to obtain $\hat{\alpha}$'s

Resulting Kriging predictor & estimated **Kriging variance**

$$\begin{aligned} \hat{Y}^*(s_0) &= \sum_{i=1}^{n} \hat{\alpha}_i y_i \\ \hat{\sigma}^2_{s_0} &= \sum_{i=1}^{n}\sum_{j=1}^{n} \hat{\alpha}_i \hat{\alpha}_j \hat{\gamma}(|h_{ij}|) - \sum_{i=1}^{n} \hat{\alpha}_i \hat{\gamma}(|h_{i0}|) \end{aligned}$$

Remarks

- $Y \sim$ Gaussian implies 95% prediction interval:

$$[Y^*(s_0) - 1.96\sigma_{s_0}, Y^*(s_0) + 1.96\sigma_{s_0}]$$

- Kriging predictor is **exact interpolator**; (interpolator = observed value at that location)
- $\sigma^2_{s_0}$ is (Exercise: Verify)

$$\sigma^2_{s_0} = \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j C(s_i, s_j) - 2\sum_{i=1}^{n} \alpha_i C(s_i, s_0) + Var(Y(s_0))$$

- Stationarity required only because cannot otherwise estimate the covariance.

Universal Kriging

Random fields with non-constant means. Let $Y(s) = \mu(s) + W(s)$.

- . Here $W(s)$ is 2^{nd}-order stationary with mean $E[W(s)] = 0$
- $\mu(s) = \sum_{l=1}^{k} a_l f_l(s)$ $\{f_l(s), l = 1, \ldots, k\}$: known functions with parameters and $\{a_l\}$. Can be dummy variables.

Universal Kriging Estimator:

$$Y^*(s_0) = \sum_{i=1}^{n} \alpha_i Y(s_i)$$

Weights α's chosen to get unbiased estimate with smallest prediction error.

Derivation similar to ordinary Kriging.

- **Non-Bias Condition:** $E[Y^*(s_0)] = E[Y(s_0)]$, or

$$\mu(s_0) - \sum_{i=1}^{n} \alpha_i \mu(s_i) = 0$$

Equivalently $\sum_{l=1}^{k} a_l (f_l(s_0) - \sum_{i=1}^{n} \alpha_i f_l(s_i)) = 0$
Since a_l's are non zero, the condition becomes

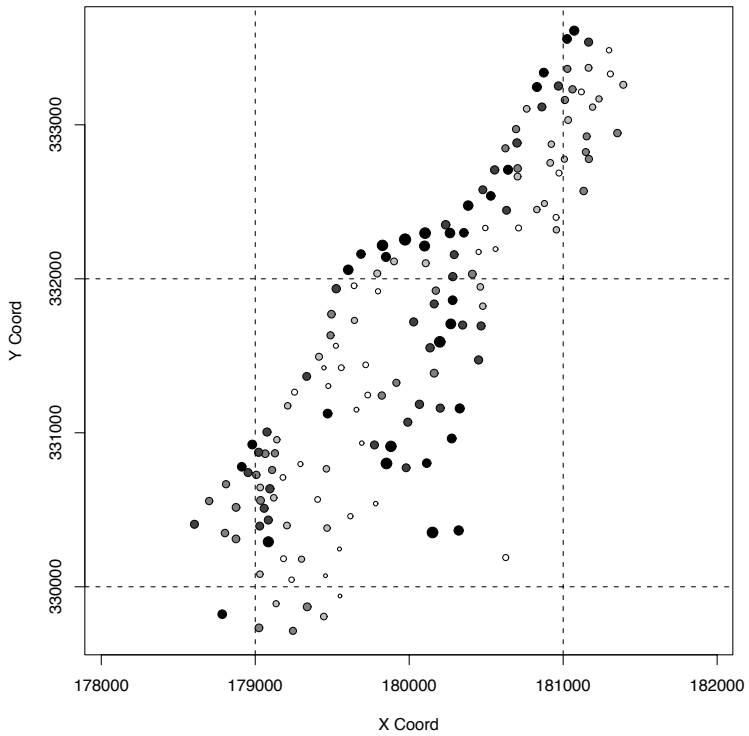
$$f_l(s_0) = \sum_{i=1}^{n} \alpha_i f_l(s_i) \quad \text{for } \; l = 1, \ldots, k \qquad (3)$$

- **Universal Kriging variance:** same form as (2). Hence α's chosen to minimize (2) & satisfy (3)

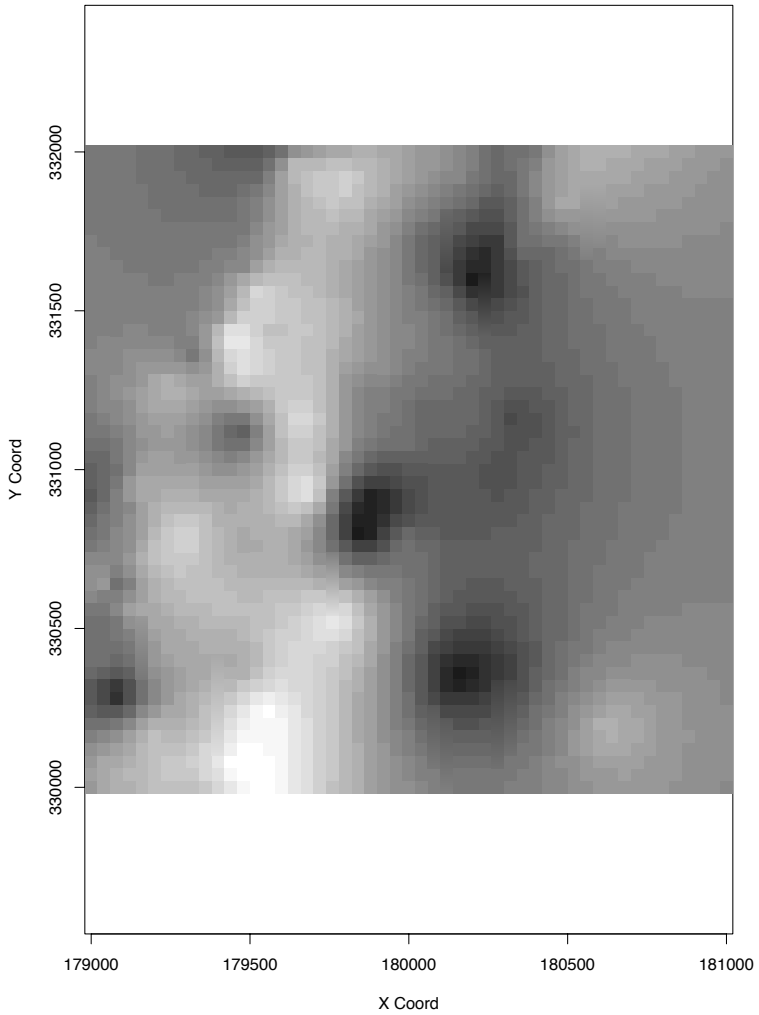
- Ordinary Kriging is a special case: e.g. $f_1 = 1$ & $f_2 = \ldots = f_l = 0$
- Like ordinary Kriging, covariance stationarity of residuals necessary

Case study 1 (cont'd).

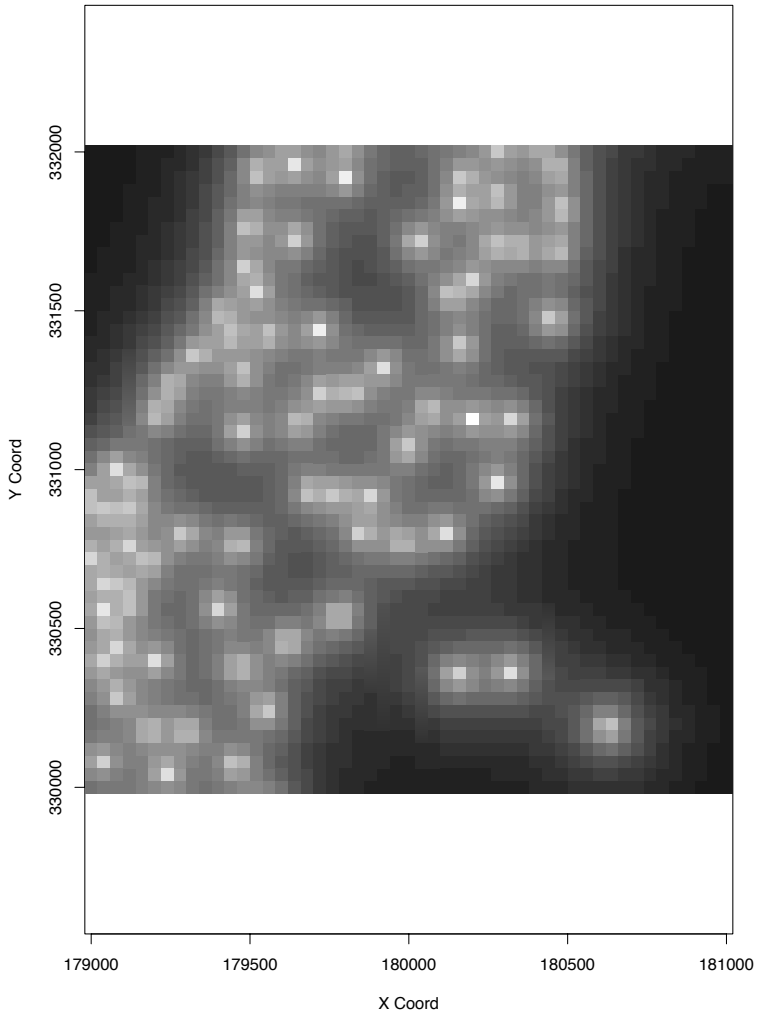
Values of log residuals, after detrending the data by removing effect of "distance from river" and "elevation" through universal kriging.



Predicted spatial residual surface.



Standard error of prediction of residual.



Fitting variogram models

Recommended steps (Webster & Oliver):

1. Plot empirical (experimental) variogram - smoothed or binned is best.
2. Find plausible variogram models – right shape for empirical (estimated) variogram
3. Fit each to estimated variogram but avoid manual fitting by eye.
4. Inspect plots estimated variogram together with estimated theoretical ones.
5. Choose one with best fit by LS or likelihood criterion.
6. Avoid “black-box” software fitting approach - thought may be needed!

Why model fitting difficult:

1. Getting $\mu(s)$ approx correct is hard but vital - failure here induces non-stationarity, bias in variogram estimation.
2. Even so anisotropic variation may well obtain.
3. Estimated variogram may contain too much point-to-point fluctuation.
4. Most models non-linear in one or more parameters.

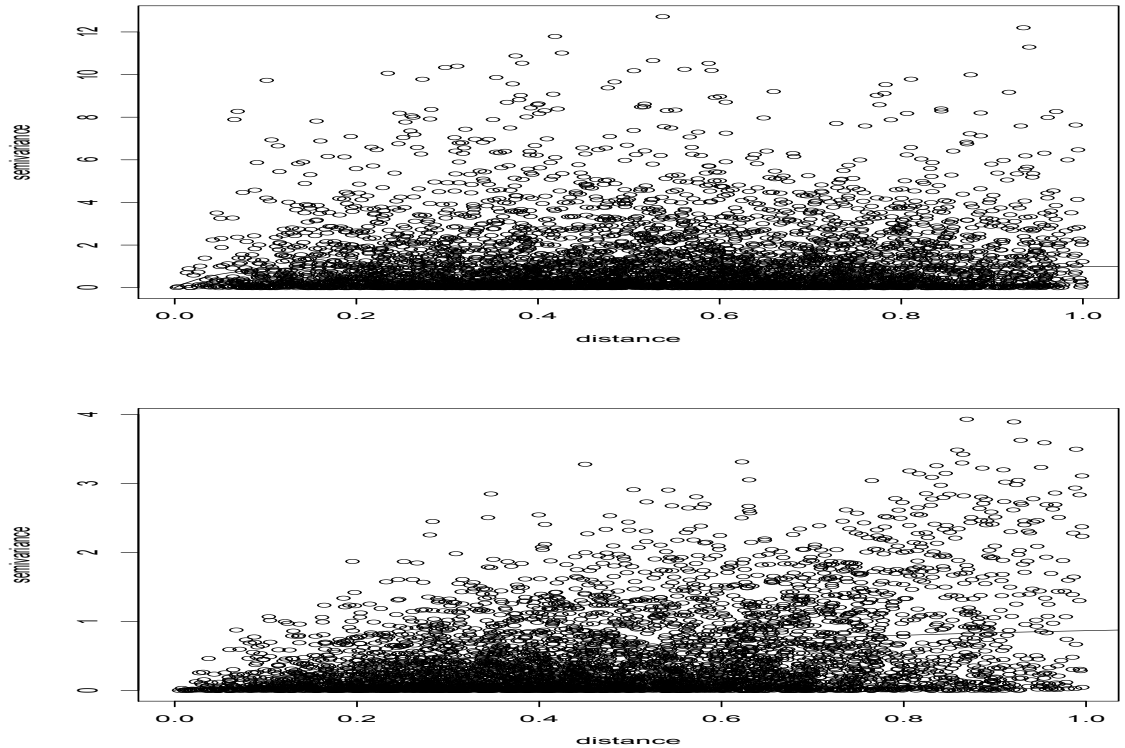
Fitting by eye difficult and unreliable.

Simulation study

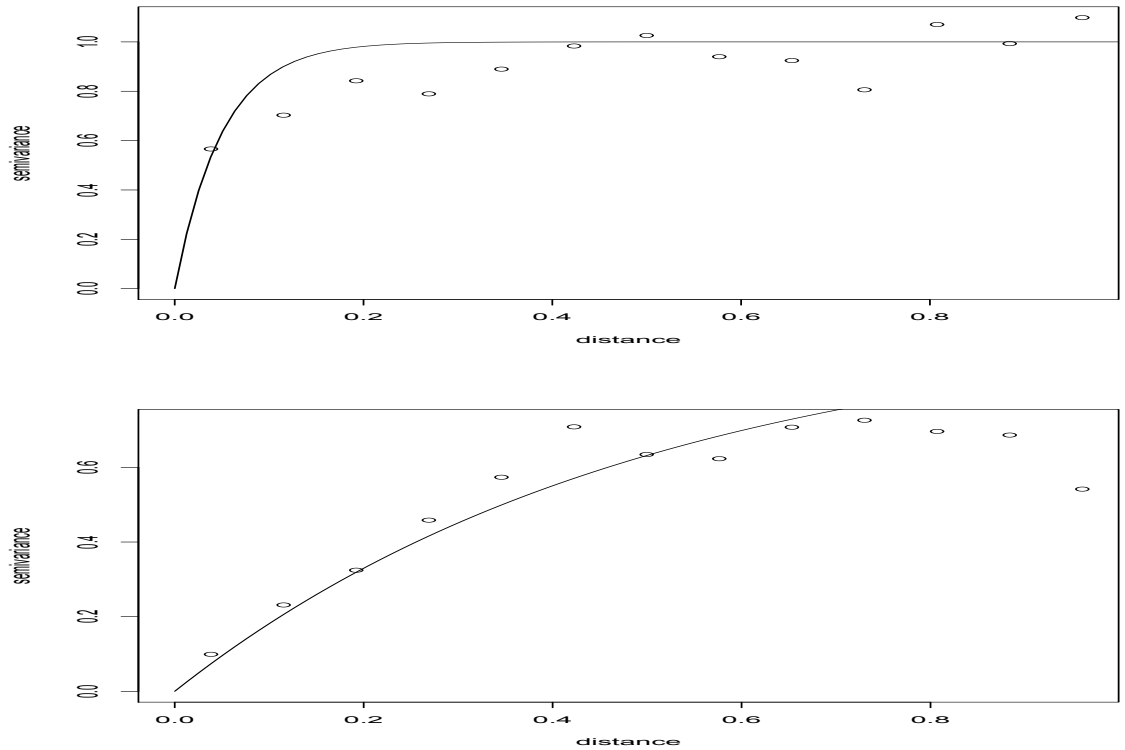
geoR provides a random field simulation function. Notice that we have used the Matern covariance function to generate the data with $\kappa = 0.5$ so it gives an exponential variogram. The range is $\phi = 0.05$ but this varies in the simulation study.

grf(n, grid = "irreg", nx, ny, xlims = c(0, 1),
ylims = c(0, 1), borders, nsim = 1,
cov.model = "matern", cov.pars = c(1,0.04) kappa = 0.5,
nugget = 0, lambda = 1, aniso.pars,
mean = 0, method, RF=TRUE, messages)

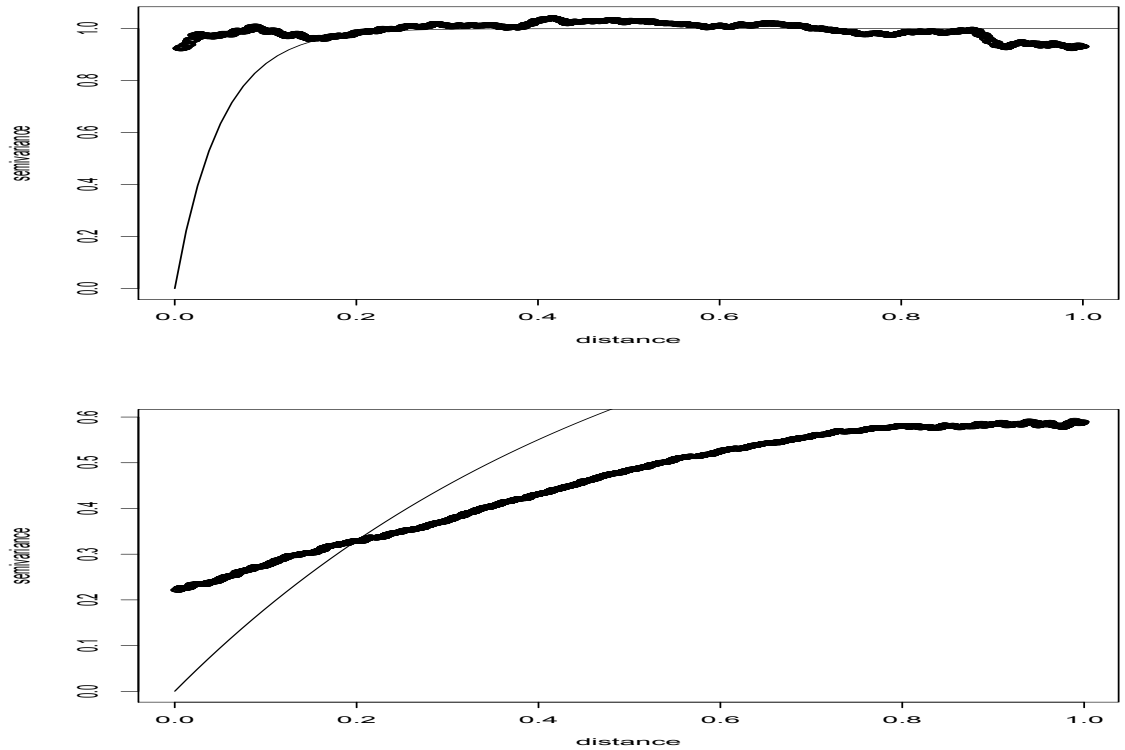
We begin with the variogram clouds for $\phi = 0.05, 0.50$.



We turn to bins or $\phi = 0.05, 0.50$.



We finish with smoothers $\phi = 0.05, 0.50$.



Variogram fitting strategies

First choose a parametric variogram family. Then use:

Least squares: We use four LS methods below, all of which fit to the binned variogram:

1. ordinary least squares
2. weighted least squared- bin counts; variances; Cressie weights.

Maximum likelihood: Needs to have a specified sampling distribution.

Bayes: Distributions put on the parameters.

Automatic fitting via LS

- Given a semivariogram $\hat{\gamma}(h)$ estimated at finite set of h's. May not be conditionally negative definite. So fitted parametric function $\gamma(h;\theta)$ preferable as estimate of $\hat{\gamma}(h)$.
- Example: exponential model with nugget effect. Then $\theta = (c_0, c, R)$, where c_0 is the nugget, c the sill, and R the range.
- Estimate of $\theta = \hat{\theta}$ chosen to make $\gamma(h;\theta)$ close to $\hat{\gamma}(h)$ in some sense. $\gamma_{\hat{\theta}}$ is vector of model estimates at h–values used to the estimate.

Three versions of weighted least-squares estimators follow.

LS estimators

Ordinary least squares: Choose θ to minimize

$$\big(\hat{\gamma} - \gamma_\theta\big)' \big(\hat{\gamma} - \gamma_\theta\big).$$

Ordinary LS immediately implementable by a nonlinear least squares procedure. But estimates $\hat{\gamma}(h)$ may vary a lot so assigning equal weights to all $\hat{\gamma}(h)$ unsatisfactory.

Number weighted least squares: Modification of equal weights scheme uses weights given by number of pairs in each bin as in second method above.

Choose θ to minimize

$$\left(\hat{\gamma} - \gamma_\theta\right)' M \left(\hat{\gamma} - \gamma_\theta\right),$$

where M is a diagonal matrix of the number of pairs of points in each bin.

Weighted least squares: Choose θ to minimize

$$\big(\hat{\gamma} - \gamma_\theta\big)' W_\theta \big(\hat{\gamma} - \gamma_\theta\big),$$

where W_θ is a diagonal matrix of the variances of the entries of γ_θ.

Generalized least squares: Choose θ to minimize

$$\big(\hat{\gamma} - \gamma_\theta\big)' V_\theta \big(\hat{\gamma} - \gamma_\theta\big),$$

where V_θ denotes the covariance matrix of γ_θ.

NOTES:

- The weighted and generalized least squares method require specification of the matrices W_θ and V_θ.
- Generalized LS is possible in principle, but complicated to implement.

Cressie (1985)[6] proposed compromise: use approximate weighted LS criterion.

Choose θ to minimize

$$\sum_j |m(h_j)| \left(\frac{\hat{\gamma}(h_j)}{\gamma_\theta(h_j)} - 1 \right)^2 .$$

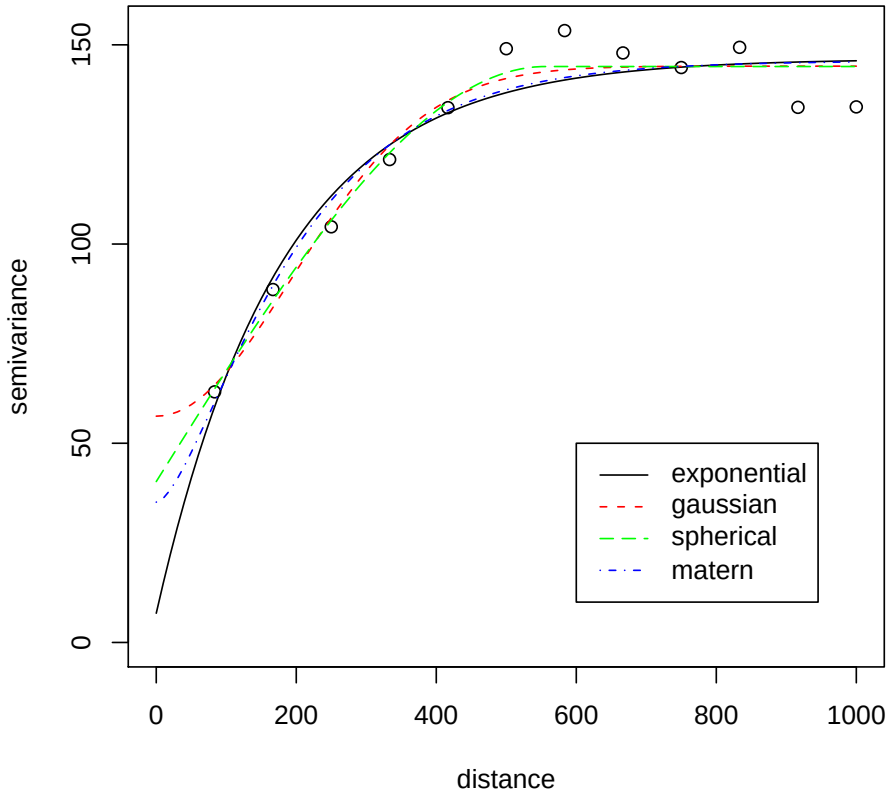
This is no more difficult to implement than ordinary least square.

[6] Cressie, N. (1985). Fitting variogram models by weighted least squares. *Mathematical Geology*,**17**, 563–586.

Case study 2: soil calcium in Brazil

These data consist of calcium content in soil from a region in Brazil. They are in the geoR library. For a description use > ?ca20 on the command line in R.

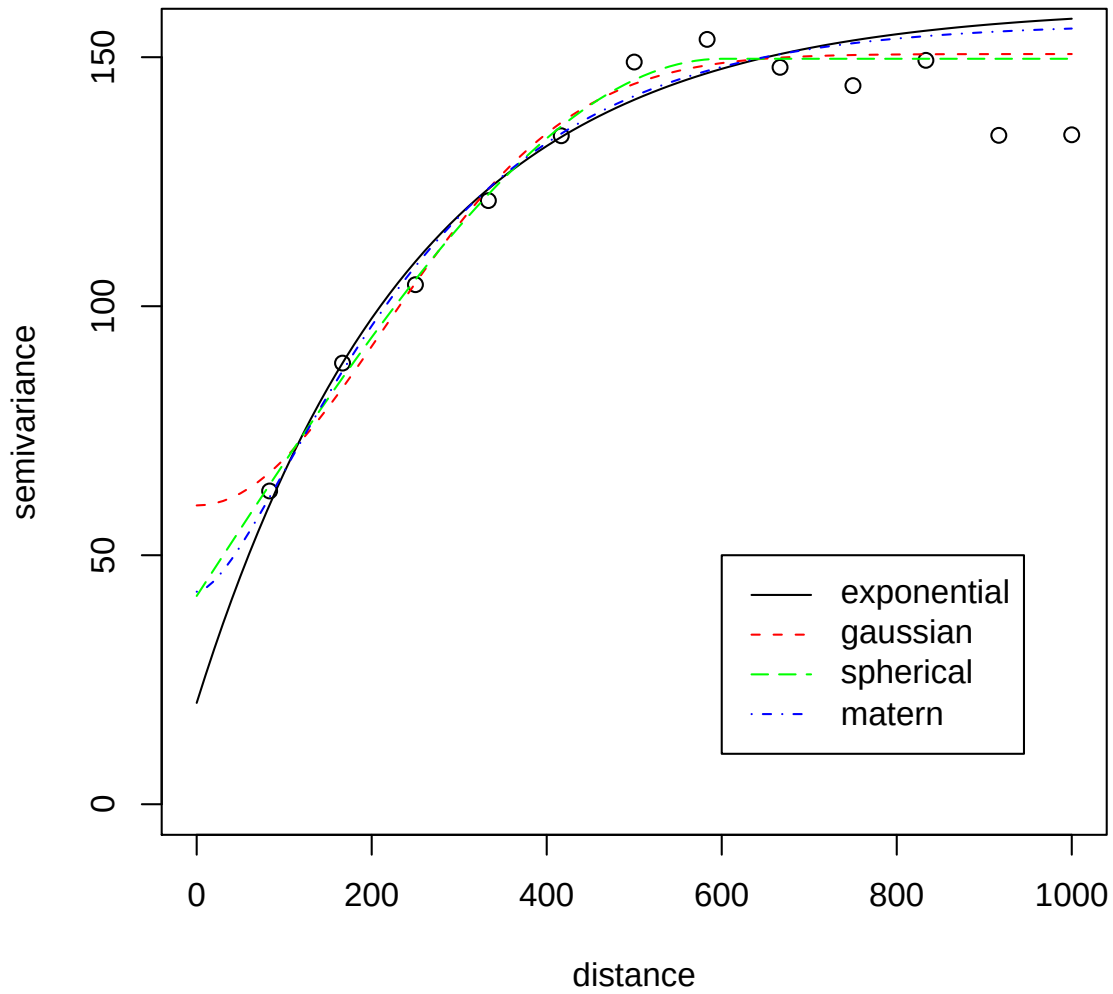
Fitting variograms by ordinary least squares.



Estimated parameters.

model	sill	range	nugget	RSS
exponential	139.1881	179.0273	7.3292	678.683
gaussian	87.8525	273.3551	56.7929	443.8803
spherical	104.1764	555.6048	40.3781	375.0742
matern	110.7297	135.1417	35.213	601.5659

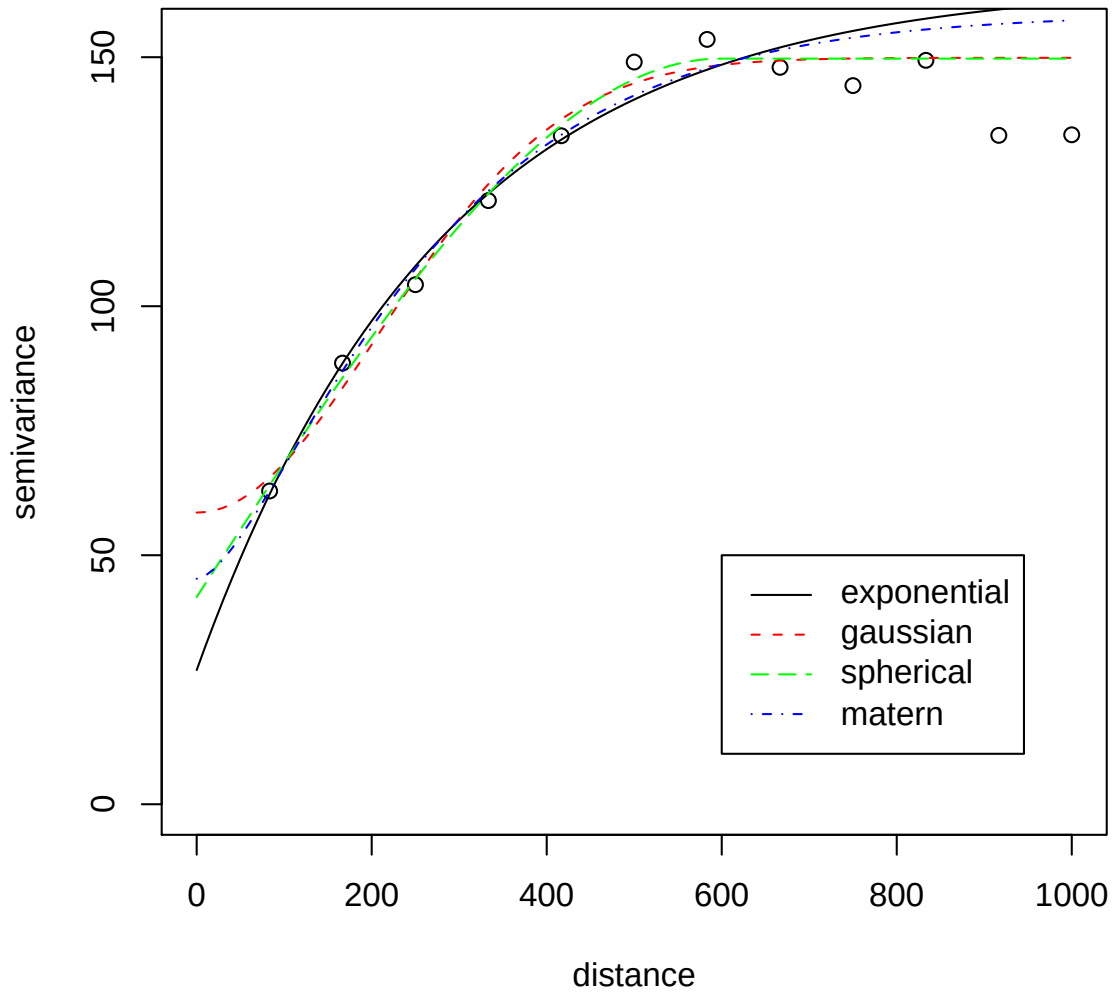
Number weighted least square fitting.



Estimated parameters.

model	sill	range	nugget	RSS
exponential	139.901	249.1264	20.3517	430123.7
gaussian	90.6371	303.4283	59.9994	251816.7
spherical	107.8644	599.9593	41.8199	149820.6
matern	114.1321	170.633	42.6684	355798.4

Cressie weighted LS



Estimated parameters.

model	sill	range	nugget	RSS
exponential	137.9816	281.9738	26.9473	21.3621
gaussian	91.3153	294.4789	58.5608	18.6959
spherical	108.1	598.2109	41.6151	8.8353
matern	113.4727	179.4307	45.275	17.6378

End of Case study #2

Estimated parameters –spherical model – different fitting methods.

method	sill	range	nugget	RSS
ordinary	104.1764	555.6048	40.3781	375.0742
number	107.8644	599.9593	41.8199	149820.6
cressie	108.1	598.2109	41.6151	8.8353

Likelihood approach

The likelihood approach is viewed as best method since points in the empirical variogram are highly correlated. Makes LS inefficient and misleading. geoR has that option.

Ensuring simple models

Adding parameters can always reduce residal sums of squares. But also need to minimize # of parameters. Distributional assumptions & Akaike Information Criterion (AIC) can do this:

$$\text{AIC} = -2\log(\text{maximized likelihood}) + 2(\text{number of parameters}),$$

AIC's variable part is estimated by

$$n\log(\text{RSS}) + 2p.$$

Here n = # of points, p= #of model parameter and RSS = residual sum of squares.

Cross-validation with kriging

Spatial prediction important goal of kriging. So choose model that does this best. How? By leave–one–out cross–validation.

1. Estimate variogram using sample data & fitted plausible models.
2. For each model, predict excluded Y's using kriging value there. Calculate kriging variance as well.

Diagnostics from results: *mean–deviation (ME)*; *mean–squared–deviation (MSE)*; *mean–squared–deviation–ratio (MSDR)* found from squared–errors & kriging variances, $\hat{\sigma}^2(s_i)$:

$$\text{ME} = \sum_{i=1}^{N} \mid y(s_i) - \hat{y}(s_i) \mid /N$$

$$\text{MSE} = \sum_{i=1}^{N} \mid y(s_i) - \hat{y}(s_i) \mid^2 /N$$

$$\text{MSDR} = \sum_{i=1}^{N} \frac{(y(s_i) - \hat{y}(x_i))^2}{\hat{\sigma}^2(s_i)} /N.$$

NOTES:

- ME should be close to 0, since kriging is an unbiased prediction method.
- MSE should be as small as possible.
- If the model is accurate then the MSDR should be close to 1.

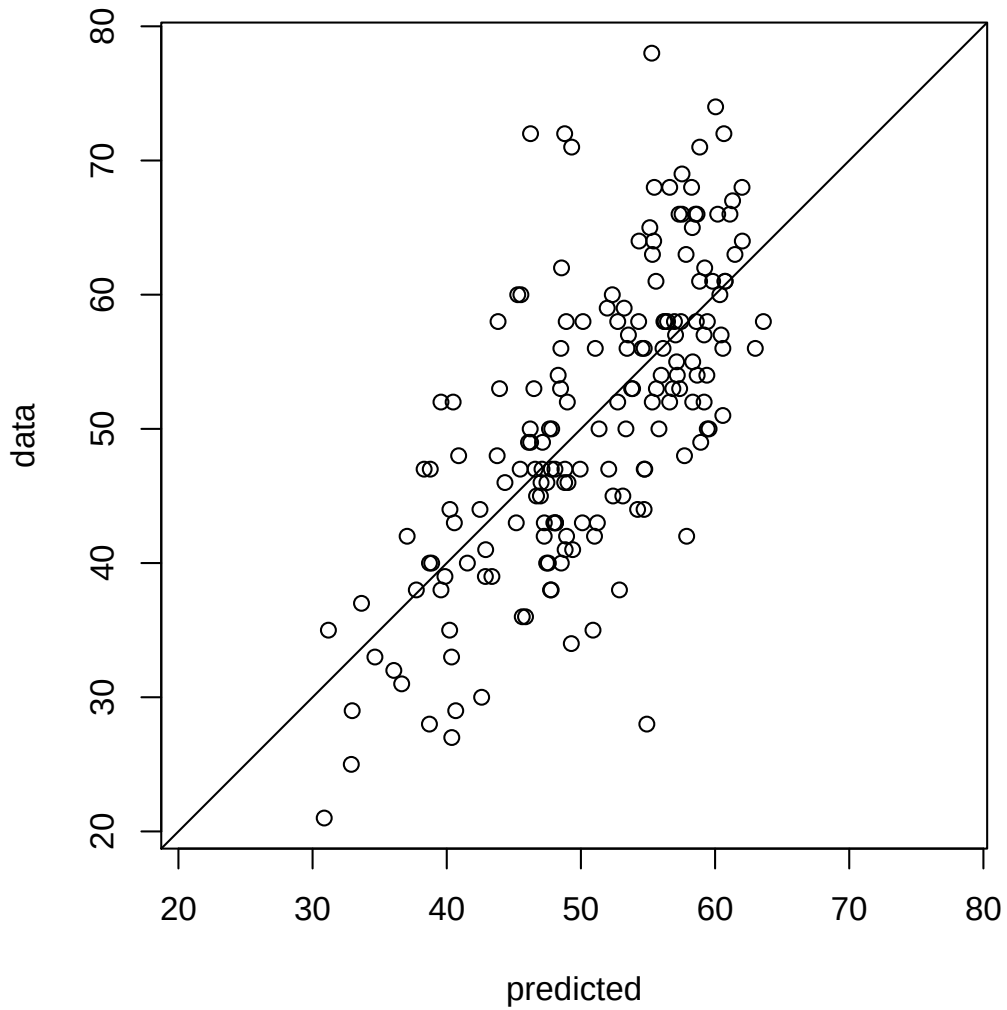
Case study 2 (cont'd).

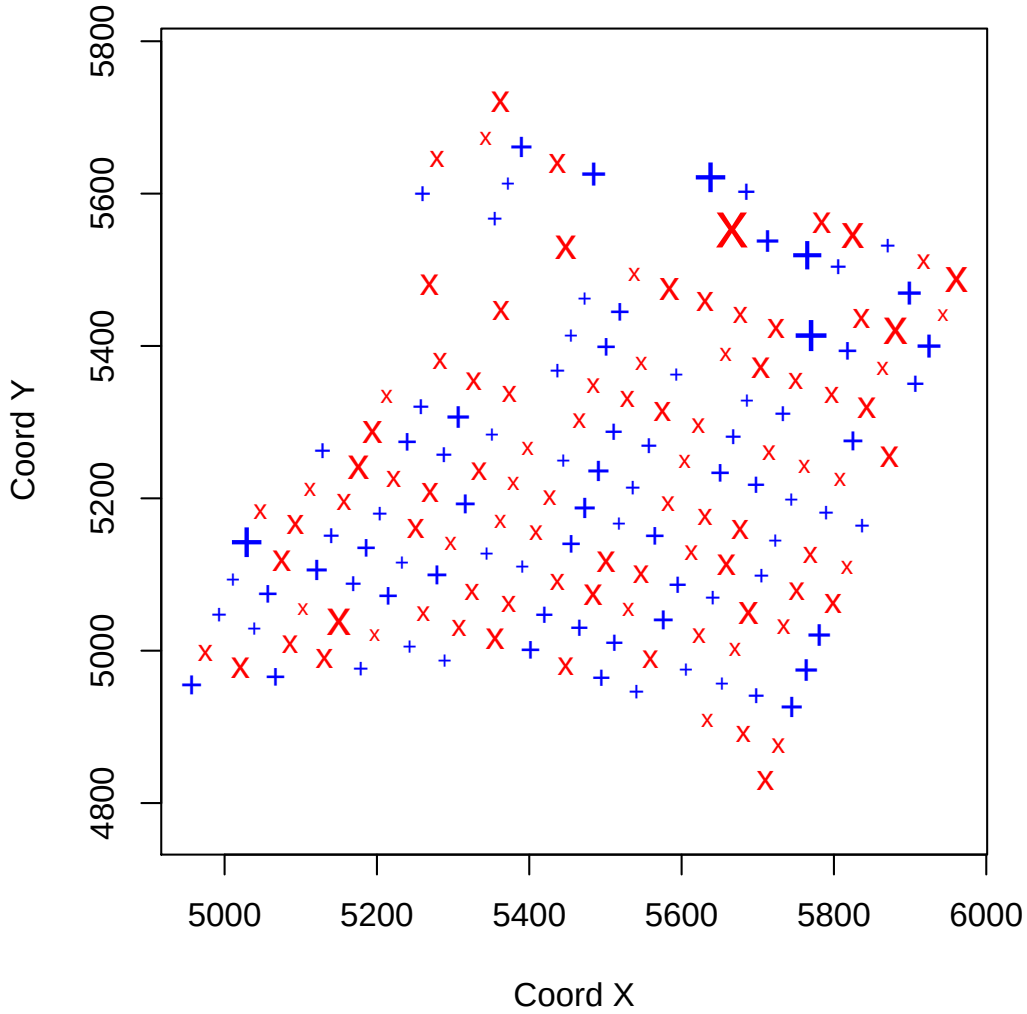
Do cross-validation for the four variogram models "exponential", "gaussian", "spherical", and "matern" on the ca20 data from the geoR package. Then calculate diagnostic indices.

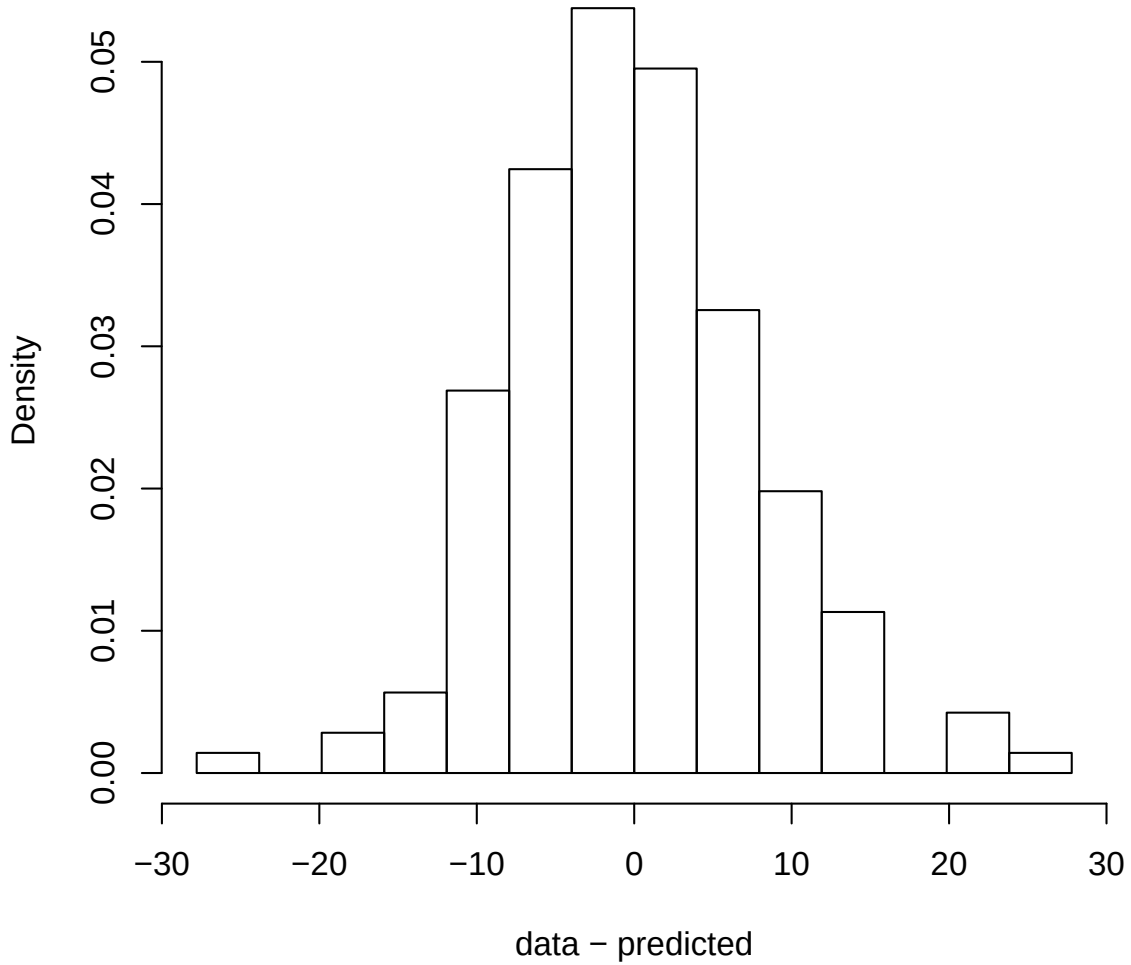
model	ME	MSE	MSDR
exponential	-0.008028705	60.94539	1.103823
gaussian	-0.007405837	69.02756	1.064712
spherical	-0.008975785	62.69338	1.022848
matern	-0.00870061	62.96571	1.057020

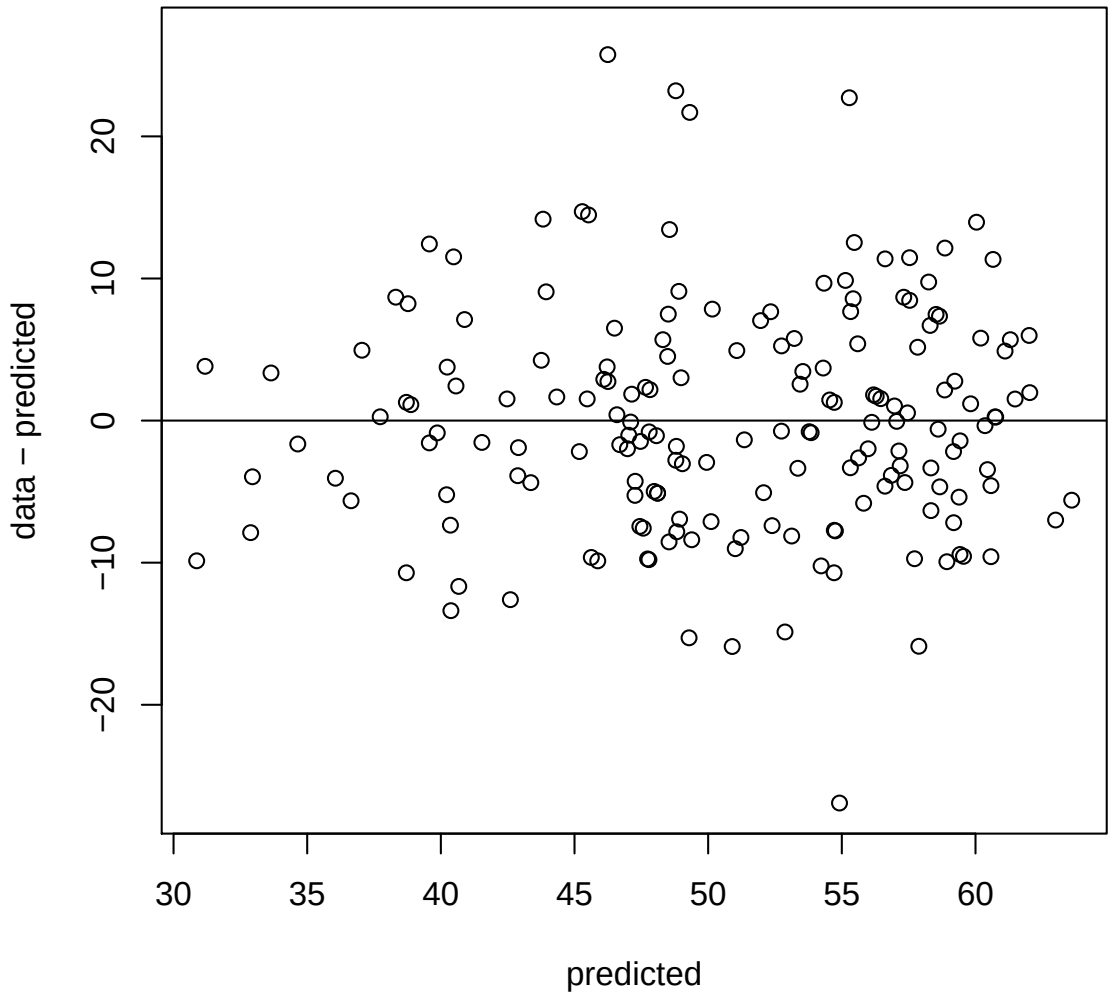
Some diagnostics plot.

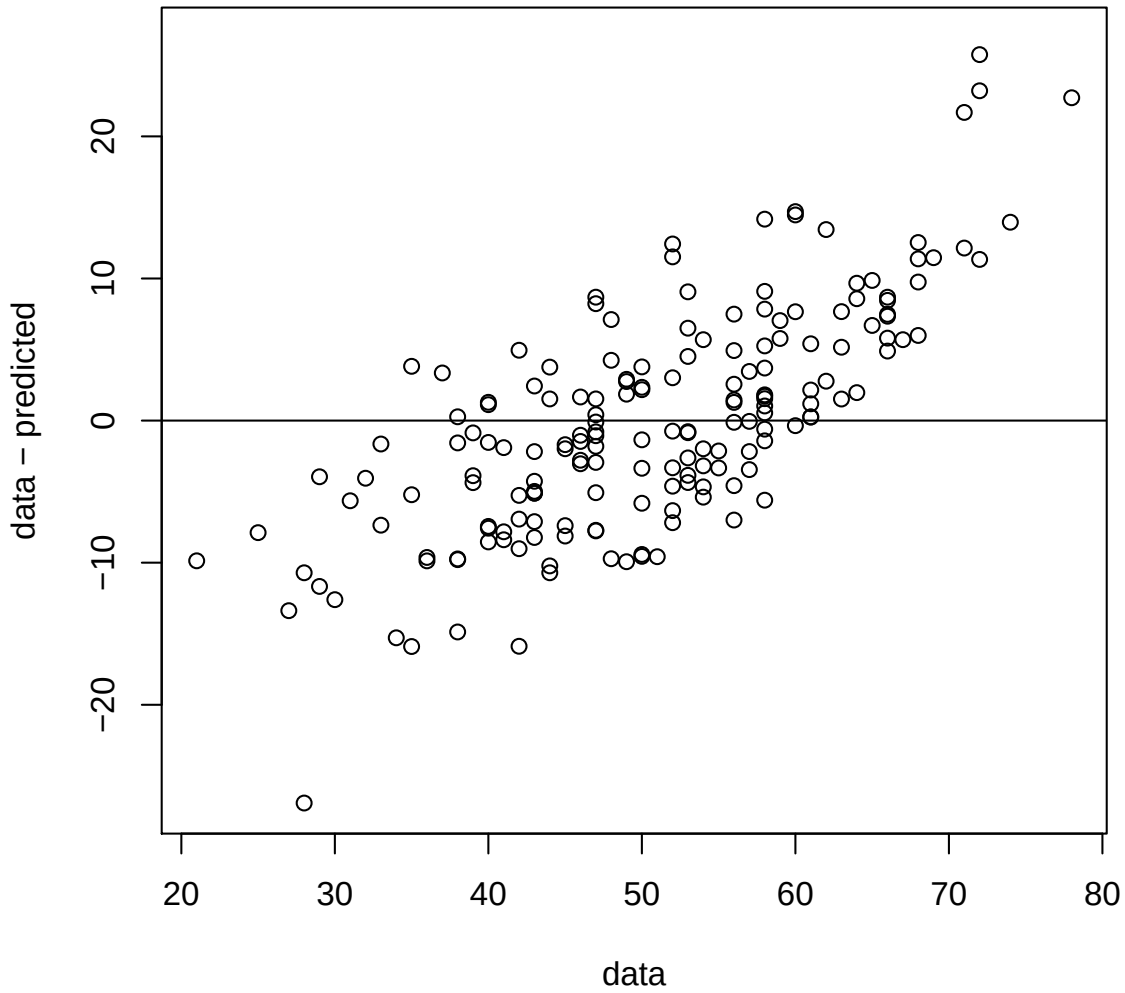
The spherical model seems to win also in the cross-validation competition. But diagnostic plots seen in the slides that follow can also be useful.

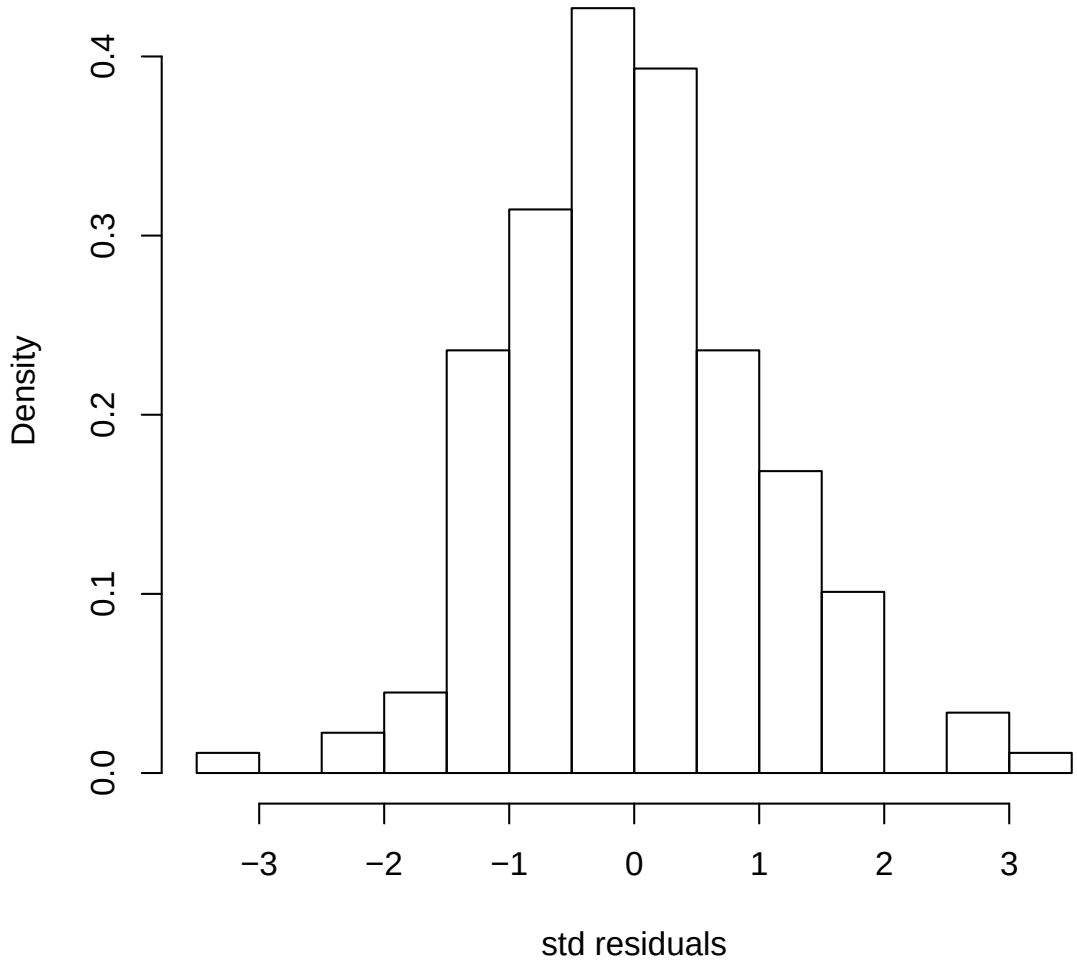

data
predicted
20
30
40
50
60
70
80

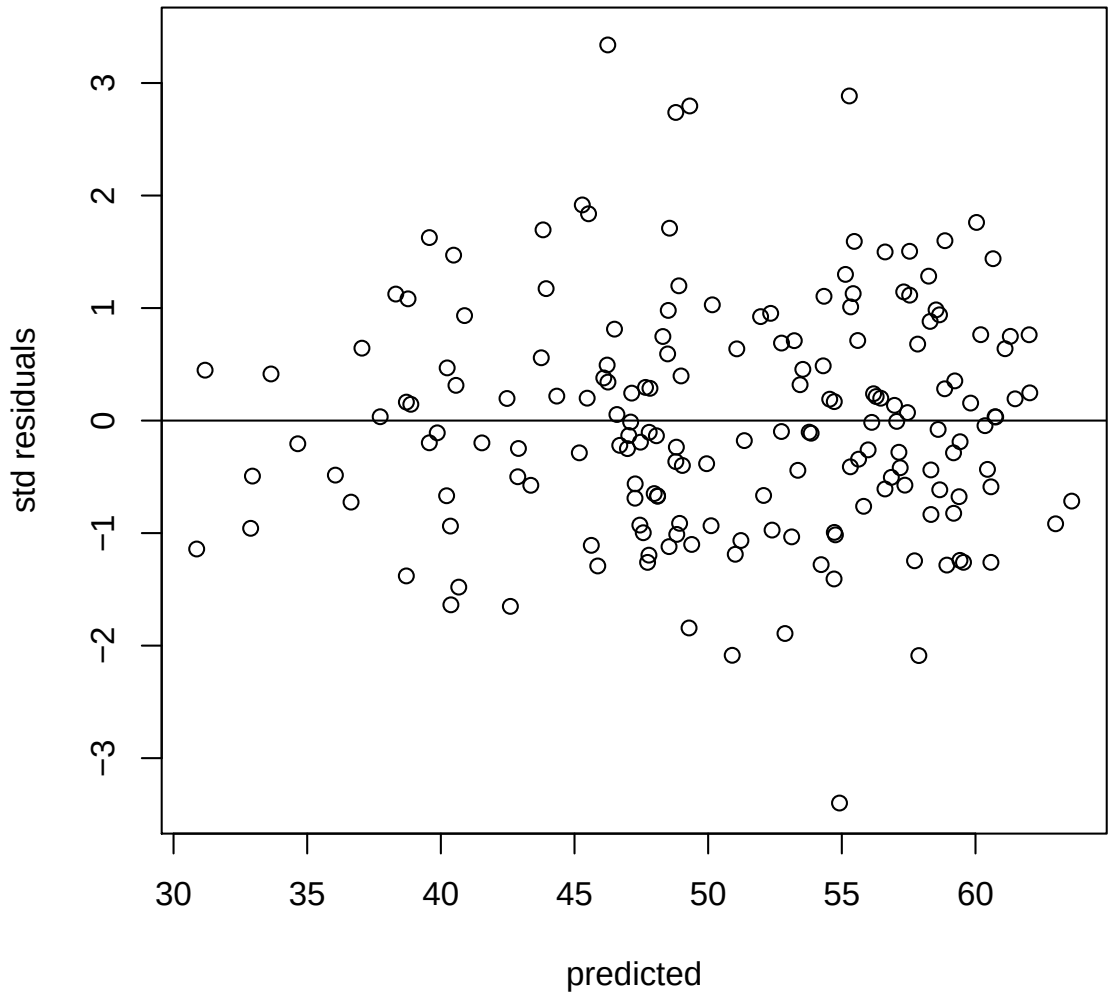

Coord Y
5800
5600
5400
5200
5000
4800
5000
5200
5400
5600
5800
6000
Coord X

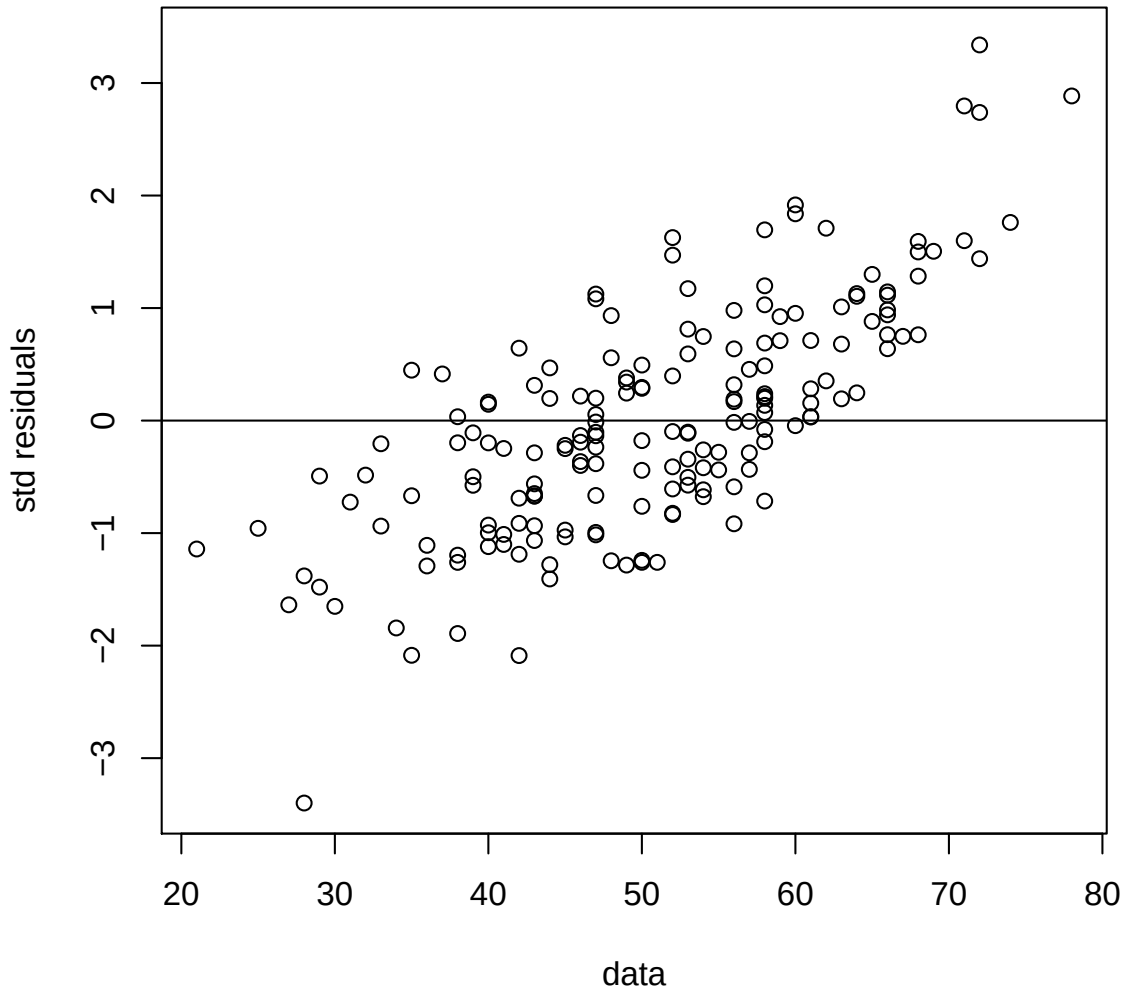

Density
0.00
0.01
0.02
0.03
0.04
0.05
−30
−20
−10
0
10
20
30
data − predicted


data − predicted
predicted
20
10
0
−10
−20
30
35
40
45
50
55
60


data – predicted
20
10
0
−10
−20
20
30
40
50
60
70
80
data


Density
0.0
0.1
0.2
0.3
0.4
−3
−2
−1
0
1
2
3
std residuals


std residuals
3
2
1
0
−1
−2
−3
30
35
40
45
50
55
60
predicted


std residuals
3
2
1
0
−1
−2
−3
20
30
40
50
60
70
80
data

What if process seems nonstationary?

Some options follow:

1. Change spatial mean: $\mu(s)$ will inevitably be mispecified as $\mu^*(s)$ so the residual is misspecified as $W^*(s) = Y(s) - \mu^*(s)$.

   Thus the calculated variogram will be nonstationary

$$E[W^*(s^1) - W^*(s^2)]^2 = E[W(s^1) - W(s^2)]^2 + [\{\mu^*(s^1) - \mu^*(s^1)\} - \{\mu^*(s^2) - \mu^*(s^2)\}]^2$$

What if process seems nonstationary?

2. Adopt nonstationary modeling approach, convolution approach:

   Represent the residual as

$$W(s) = \int K(s - s')W^*(s')ds'$$

   where W^* is stationary.

   NOTE: Allows only modest degree of nonstationarity.

What if process seems nonstationary?

3. Warping: The famous Sampson–Guttorp approach warps the geographic space into dispersion space so that strongly correlated sites are moved closet together, uncorrelated ones further apart.
4. Dimension expansion: Keep the geographic space as is but add additional dimensions.

3. Warping: the Sampson-Guttorp method

A nonparametric method to estimate spatial covariance structure without assuming stationarity (Sampson & Guttorp JASA,1992):

- Construct a $1-1$ non-linear smooth mapping
  $f(R^2 \rightarrow R^2)$: G-space to D-space
  G-space: geographic space with location s_i
  D-space: 'dispersion' space with location z_i
  where isotropy is reasonable
  ie. $z_i = f(s_i)$ or equivalently $s_i = f^{-1}(z_i)$
- Estimate (isotropic) semi-variogram, $\hat{\gamma}_D$, using D-distances (ie. between z_i) and the observed dispersion ($v_{ij} = 2 - 2\hat{corr}_{ij}$)

The SG-method

- Correlation c_{ij} between s_i and s_j, obtained by:
  - get D-distance, d_{ij} between z_i and z_j
  - evaluate $c_{ij} = 1 - \hat{\gamma}_D(d_{ij})$
- The SG-approach ensures the constructed correlation matrix, $\{c_{ij}\}$, is non-negative definite (ie. based on a variogram).
- Any estimate of variance field can be simply incorporated to obtain covariance matrix.

SG-method: Construction of f

A two-step procedure using the observed dispersion (v_{ij}):

- Using the multidimensional scaling to find a configuration of the locations, s_i, so that their new inter-distances are 'close' to the corresponding dispersions,
  ie.

$$min_{\delta} \sum_{i<j} \frac{(\delta(v_{ij}) - d_{ij})^2}{\sum d_{ij}^2}$$

  over all monotone functions

SG-method: Construction of f

- Fitting a thin-plate spline mapping, f, between new locations z_i and original locations s_i,
  ie.

$$f(s) = \alpha_0 + \alpha_1 s^{(1)} + \alpha_2 s^{(2)} + \sum_{i=1}^{n} \beta_i u_i(s)$$

where $u_i(s) = |s - s_i|^2 log|s - s_i|$

Find α's and β's by minimizing

$$\sum_{j=2}^{2} \sum_{i=1}^{n} (z_i^{(j)} - f_j(s_i^{(j)}))^2 + \lambda(J_2(f_1) + J_2(f_2))$$

Smoothing parameter $\lambda \to \infty$ leads to $\beta \to 0$

SG-method: Implementation

- Need to estimate λ in the construction of f
- Possibility: cross-validation to best estimate of dispersion
- A simpler approach is to ensure spatial 'integrity': longer inter-distance having smaller correlation (on the same direction)
  - Need input from user
  - Currently implemented in software - enviro.stat.ubc.ca. Download it!

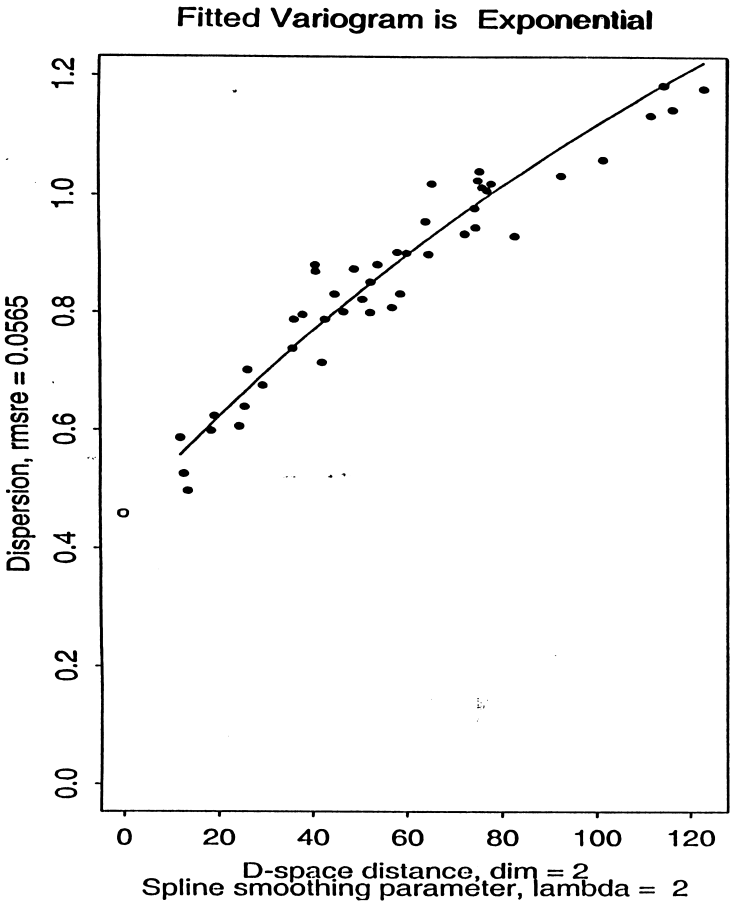
Example: Particulate matter in Vancouver

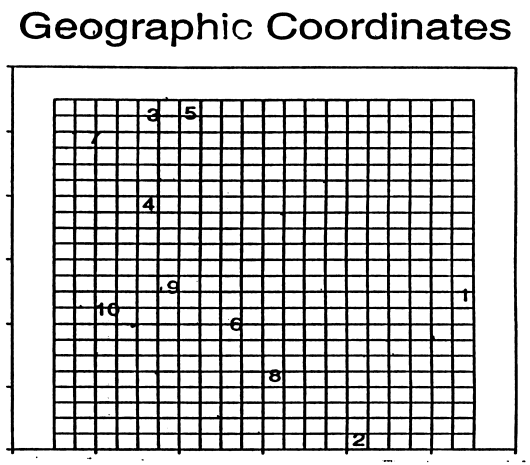
Small particulates, the size of those in cigarette smoke are nasty. They get deep into the lung to the gas exchange membrane where they can generate antiinflammatory mediators. These in turn affect the cardio–vascular system and cause heart problems.

PM10 are all up to 10 microns in size. PM2.5 is the fraction with the smallest sizes and are now of primary concern.

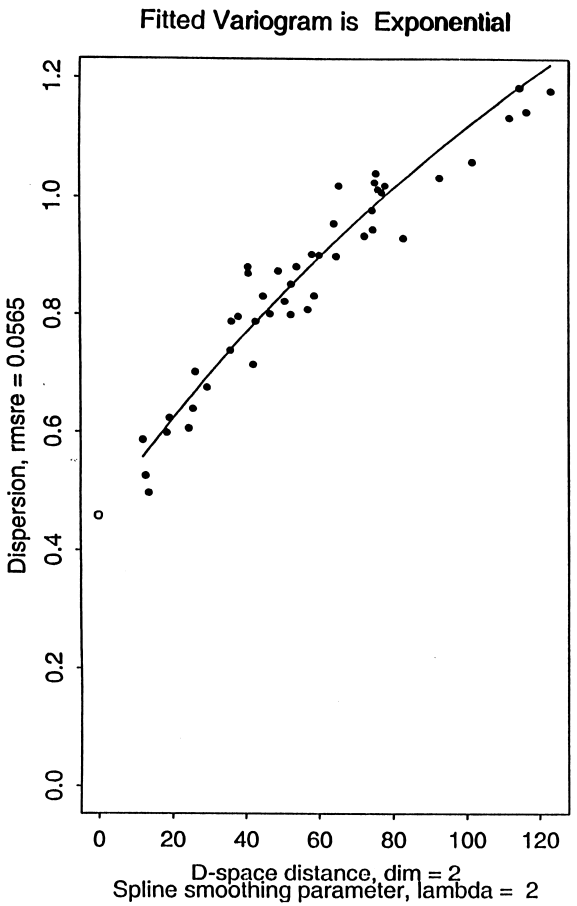
However the spatial field can be quite nonstationary since these particulates come from mobile and local sources.

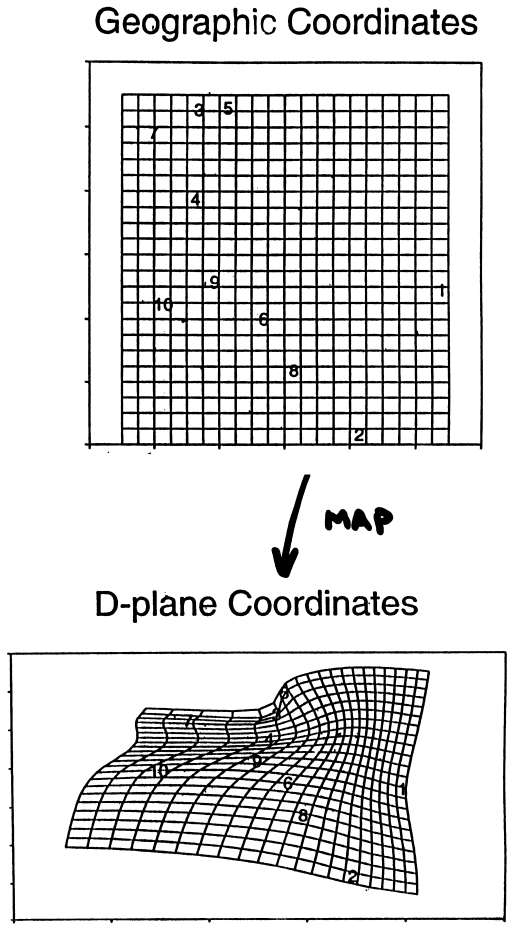
Hourly PM_{10} in Vancouver -1994-1999





Hourly PM_{10} in Vancouver -1994-1999





To be concluded in Lecture 12