

Model for spatial data with trend (Advanced)

Spatial Statistics

Madlene Nussbaum

m.nussbaum@uu.nl

Utrecht University, Department of Physical Geography

1 Model for Gaussian spatial data

Model for data: $Y_i = S(\mathbf{x}_i) + Z_i = \mu(\mathbf{x}_i) + E(\mathbf{x}_i) + Z_i$

- with Y_i : i^{th} datum; $S(\mathbf{x}_i)$: “signal” (true quantity) at location $\mathbf{x}_i$; $\mu(\mathbf{x}_i)$: trend
- $\{E(\mathbf{x}_i)\}$: a zero-mean Gaussian process, parametrized by covariance function $\gamma(\mathbf{h}; \theta)$ or variogram $V(\mathbf{h}; \theta)$
- Z_i : iid Gaussian measurement error with variance τ^2

Trend $\mu(\mathbf{x}_i)$ modeled by linear regression model with spatial covariates $d_k(\mathbf{x}_i)$

$$\mu(\mathbf{x}_i) = \sum_k d_k(\mathbf{x}_i) \beta_k = \mathbf{d}(\mathbf{x}_i)^T \boldsymbol{\beta}$$

Unknown elements of the model:

1. Structure and parameters $\boldsymbol{\beta}$ of the trend model
2. Covariance (or variogram) parameters θ
3. Nugget variance τ^2

2 Trend modelling

Ordinary least squares (OLS) trend estimation

Gaussian model in vector notation: $\mathbf{Y} = \mathbf{X}\beta + \mathbf{E} + \mathbf{Z}$

Estimation of trend parameters β by ordinary least squares:

$$\hat{\beta}_{\text{OLS}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$$

For spatially uncorrelated data ($\mathbf{E} = \mathbf{0}$; $\text{Cov}(\mathbf{Y}, \mathbf{Y}^\top) = \tau^2 \mathbf{I}$):

$$\hat{\beta}_{\text{OLS}} \sim \mathcal{N}(\beta, \tau^2 (\mathbf{X}^\top \mathbf{X})^{-1})$$

For spatially auto-correlated data $\text{Cov}(\mathbf{Y}, \mathbf{Y}^\top) = \text{Cov}(\mathbf{Z}, \mathbf{Z}^\top) + \text{Cov}(\mathbf{E}, \mathbf{E}^\top) = \mathbf{\Gamma}_\theta = \tau^2 \mathbf{I} + \mathbf{\Sigma}_\theta$

$$\hat{\beta}_{\text{OLS}} \sim \mathcal{N}(\beta, \tau^2 (\mathbf{X}^\top \mathbf{X})^{-1} + (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{\Sigma}_\theta \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1})$$

Conclusion: Ignoring auto-correlation: $\hat{\beta}_{\text{OLS}}$ unbiased, but the standard errors are too small. Tests based on OLS fit are biased!

Generalized least squares (GLS) trend estimation

Generalized least squares estimates:

$$\hat{\beta}_{\text{GLS}} = (\mathbf{X}^\top \mathbf{\Gamma}_\theta^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{\Gamma}_\theta^{-1} \mathbf{Y}$$

GLS = OLS with “orthogonalized” response and design matrix.

Sampling distribution:

$$\hat{\beta}_{\text{GLS}} \sim \mathcal{N}(\beta, (\mathbf{X}^\top \mathbf{\Gamma}_\theta^{-1} \mathbf{X})^{-1})$$

For spatially uncorrelated data ($\mathbf{\Gamma}_\theta = \tau^2 \mathbf{I}$):

$$\hat{\beta}_{\text{GLS}} = \hat{\beta}_{\text{OLS}}$$

$\hat{\beta}_{\text{GLS}}$ has the smallest standard errors among all linear estimators (Gauss-Markov theorem), hence it's the BLUE (Best Linear Unbiased Estimator).

$\hat{\beta}_{\text{GLS}}$ is the maximum likelihood estimate for Gaussian $\mathbf{Y}$.

Generalised least squares (GLS) is the method of choice for estimating coefficients of the trend model.

3 Computing the sample variogram

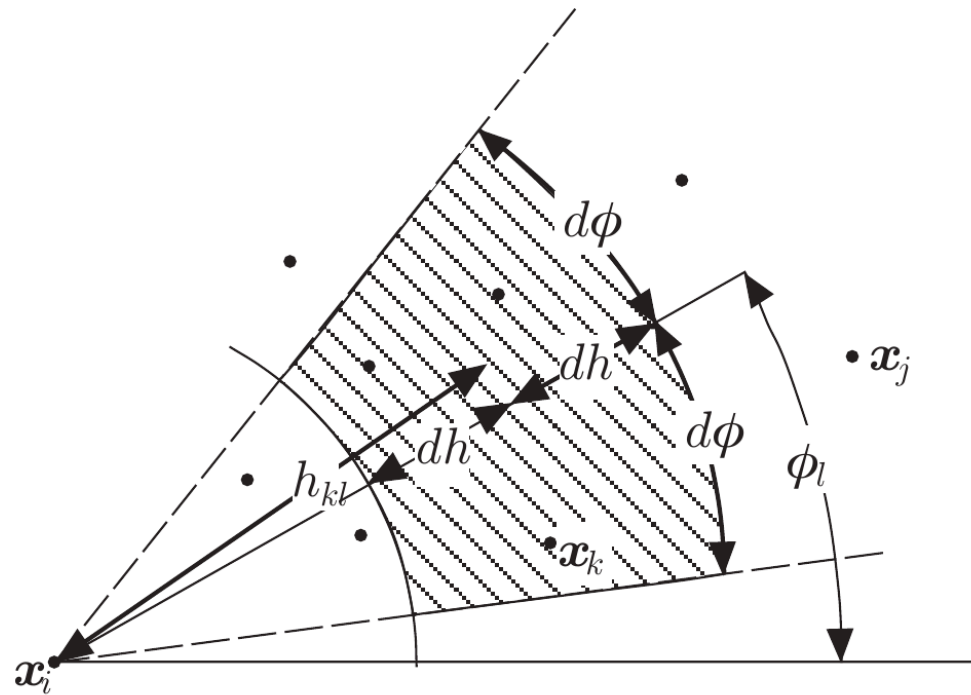
- Sample variogram (also experimental variogram) consists of semivariances at a finite set of discrete lags h .
- Underlying function is continuous for all h .
- Variogram modelling:
 1. Compute sample variogram
 2. Fit smooth function that describes principal features of semivariance sequence

4 Sample variogram of residuals

Extract residuals $\mathbf{R} = \mathbf{Y} - \mathbf{X}\hat{\beta}$ of the fitted linear model (or use data $\mathbf{Y}$ if the model has constant $\mu(\mathbf{x})$).

Choose bin width dh (and width of angular classes $d\phi$) to define the $(k, l)^{\text{th}}$ lag class, $\mathbf{h}_{kl}$, characterized by:

- Distance: $(h_k - dh, h_k + dh]$
- Angular class: $\phi_l - d\phi, \phi_l + d\phi$



Computing sample variogram – formally

Form all N_{kl} pairs (i, j) with $\mathbf{x}_i - \mathbf{x}_j \approx \mathbf{h}_{kl}$.

Compute for each lag class $\mathbf{h}_{kl}$ the semivariance:

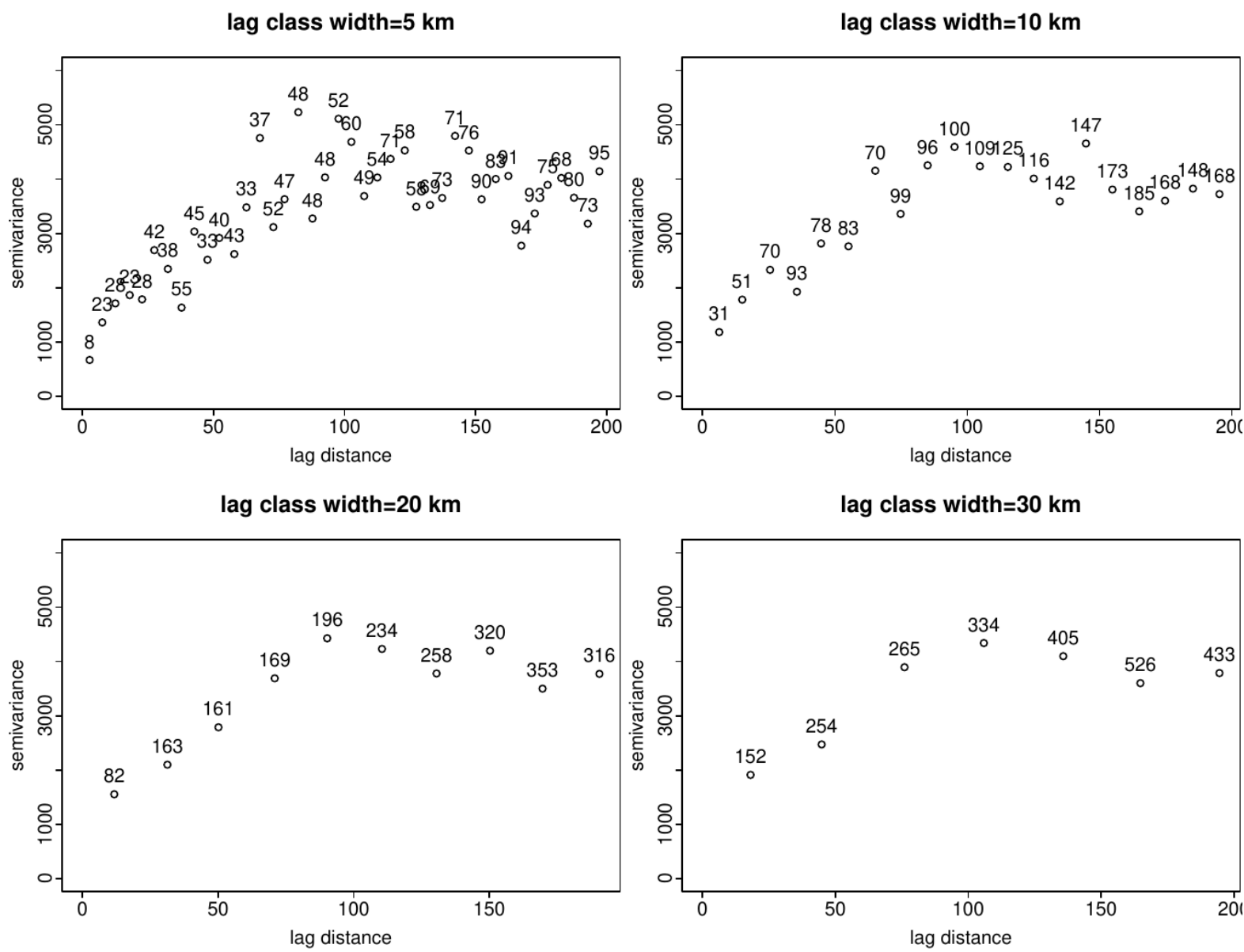
$$\hat{V}(\mathbf{h}_{kl}) = \frac{1}{2N_{kl}} \sum_{(i,j) \in \mathbf{h}_{kl}} [R(\mathbf{x}_i) - R(\mathbf{x}_j)]^2$$

Sample variogram: Plot of $\hat{V}(\mathbf{h}_{kl})$ vs. $\mathbf{h}_{kl}$.

Rules of thumb:

1. Ensure enough data pairs per lag class. Choose $d\mathbf{h}$ (and $d\phi$) such that $N_{kl} > 30 - 50$.
2. Only compute the variogram for half the size of the study area.
Largest $\mathbf{h}_{kl} \leq 0.5 \max(\mathbf{x}_i - \mathbf{x}_j)$.

Examples of different lags (Wolfcamp data)



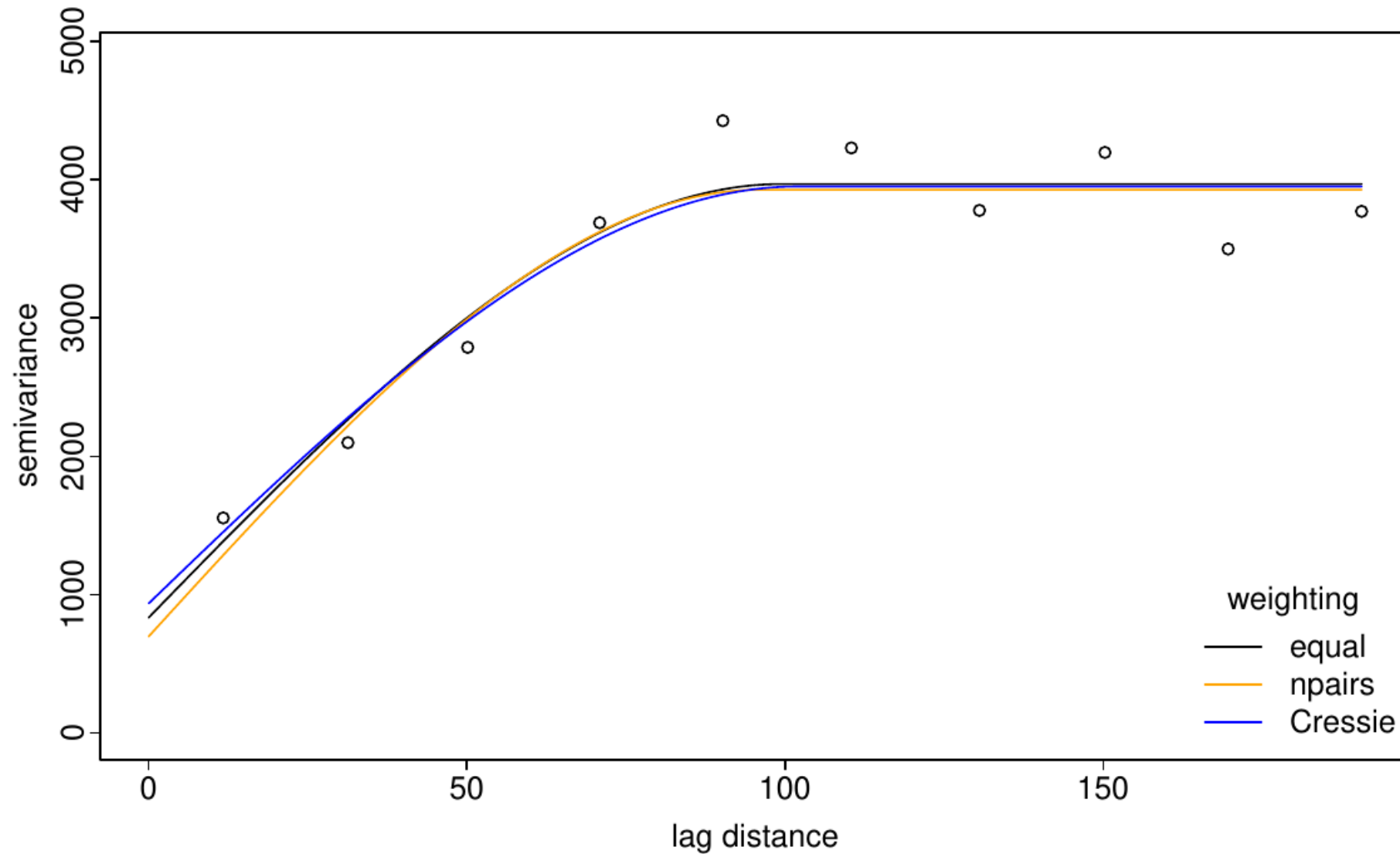
5 Fitting model to sample variogram

- **Semivariance** is required for arbitrary lag distances when computing predictions.
- **Smoothing** the sample variogram by fitting a parametric variogram function $V(\mathbf{h}, \theta)$.
- Choose a **variogram function** that approximates the shape of the sample variogram well, particularly close to the origin.
- **Fit parameters** θ by (weighted) least squares: *[Math Processing Error]*

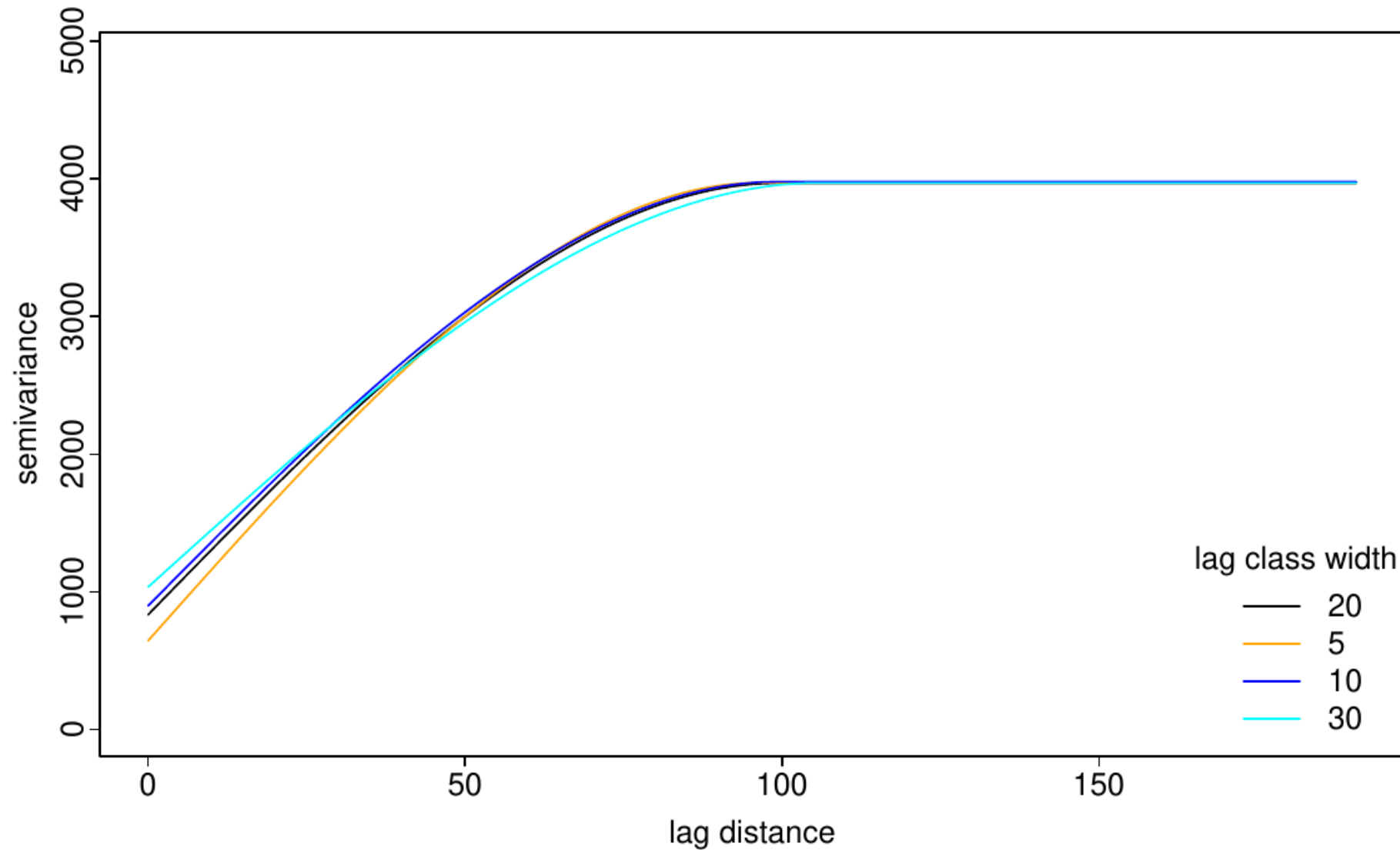
Options for weighing:

1. **Equal weights:** $w(\mathbf{h}_{kl}) = 1$
2. **By number of pairs:** $w(\mathbf{h}_{kl}) = N_{kl}$
3. **Cressie's weights:** $w(\mathbf{h}_{kl}) = \frac{N_{kl}}{V(\mathbf{h}_{kl}, \theta)^2}$

Fits with different weights (Wolfcamp data)



Fits with different lags (Wolfcamp data)



6 Problems with ad-hoc estimation

- **Subjective choice** of lag class width and weighting method for model fitting.
- **Estimates of semivariance** for different lag classes are mutually correlated; the choice of variogram function based on the sample variogram is problematic.
- The sample variogram is **susceptible to outliers**, hence robust estimators are preferred.
- Fitting a model function to the sample variogram requires **further subjective choices**.
- An ad-hoc approach provides **biased estimates** of the variogram of the underlying stochastic process if the trend is modeled.
- Therefore, the estimate of the variogram based on the sample variogram of OLS residuals is biased.
- Thus, estimate trend and variogram parameters **simultaneously** using maximum likelihood.

7 Maximum likelihood estimation

Maximum likelihood estimation (MLE) of parameters of Gaussian model for spatial data

- Principle of maximum likelihood estimation: find parameters that maximize joint probability for observed data
- Properties of maximum likelihood estimates: asymptotically unbiased and fully efficient; asymptotically normally distributed
- profile likelihood useful for exploring shape of likelihood surface and for computing confidence intervals based on likelihood ratio test

Maximum likelihood estimation

Consider a **Gaussian stochastic process** $\{Y(\mathbf{x})\}$ with a linear trend function.

Any arbitrary set of random variables $\mathbf{Y} = (Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))$ has a multivariate Gaussian distribution with expectation:

$$\mathbb{E}[\mathbf{Y}] = \mathbf{X}\beta$$

and covariance matrix:

$$\text{Cov}(\mathbf{Y}, \mathbf{Y}^T) = \mathbf{\Gamma}_\theta$$

The joint probability density for $\mathbf{Y}$ is given by:

[Math Processing Error]

Maximum likelihood estimation (cont.)

Unknown model parameters:

1. Regression coefficients β
2. Covariance (or variogram) parameters θ

The log-likelihood function (up to a constant) is given by:

$$L(\beta, \theta; \mathbf{y}) = -\frac{1}{2}\log(|\mathbf{\Gamma}_\theta|) - \frac{1}{2}\{\mathbf{y} - \mathbf{X}\beta\}^T \mathbf{\Gamma}_\theta^{-1} \{\mathbf{y} - \mathbf{X}\beta\}$$

For known variogram parameters θ maximum likelihood estimate for β equals **GLS** estimator:

$$\hat{\beta}_{\text{GLS}} = (\mathbf{X}^T \mathbf{\Gamma}_\theta^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{\Gamma}_\theta^{-1} \mathbf{Y}$$

- Plugging $\hat{\beta}_{\text{GLS}}$ into $L(\beta, \theta; \mathbf{y})$ gives the profile likelihood function for θ .
- No closed form expression for maximum likelihood estimate
- Maximize L numerically by a non-linear optimization method to find $\hat{\theta}$
- Numerical optimization requires initial values of θ
- Make sure numerical optimization converged!

Example: MaxLik estimates Wolfcamp data

```
1 library(gstat); library(georob)
2 data("wolfcamp"); d.w <- wolfcamp
3 coordinates(d.w) <- c("x", "y")
4 r.georob.ml <- georob(pressure~x+y, d.w,
5   locations=~x+y, variogram.model="RMspheric",
6   param=c(variance=3000, nugget=1000, scale=100),
7   tuning.psi=1000, control=control.georob(ml.method="ML"))
8 summary(r.georob.ml)
```

Call:georob(formula = pressure ~ x + y, data = d.w, locations = ~x +
y, variogram.model = "RMspheric", param = c(variance = 3000,
nugget = 1000, scale = 100), tuning.psi = 1000, control = control.georob(ml.method = "ML"))

Tuning constant: 1000

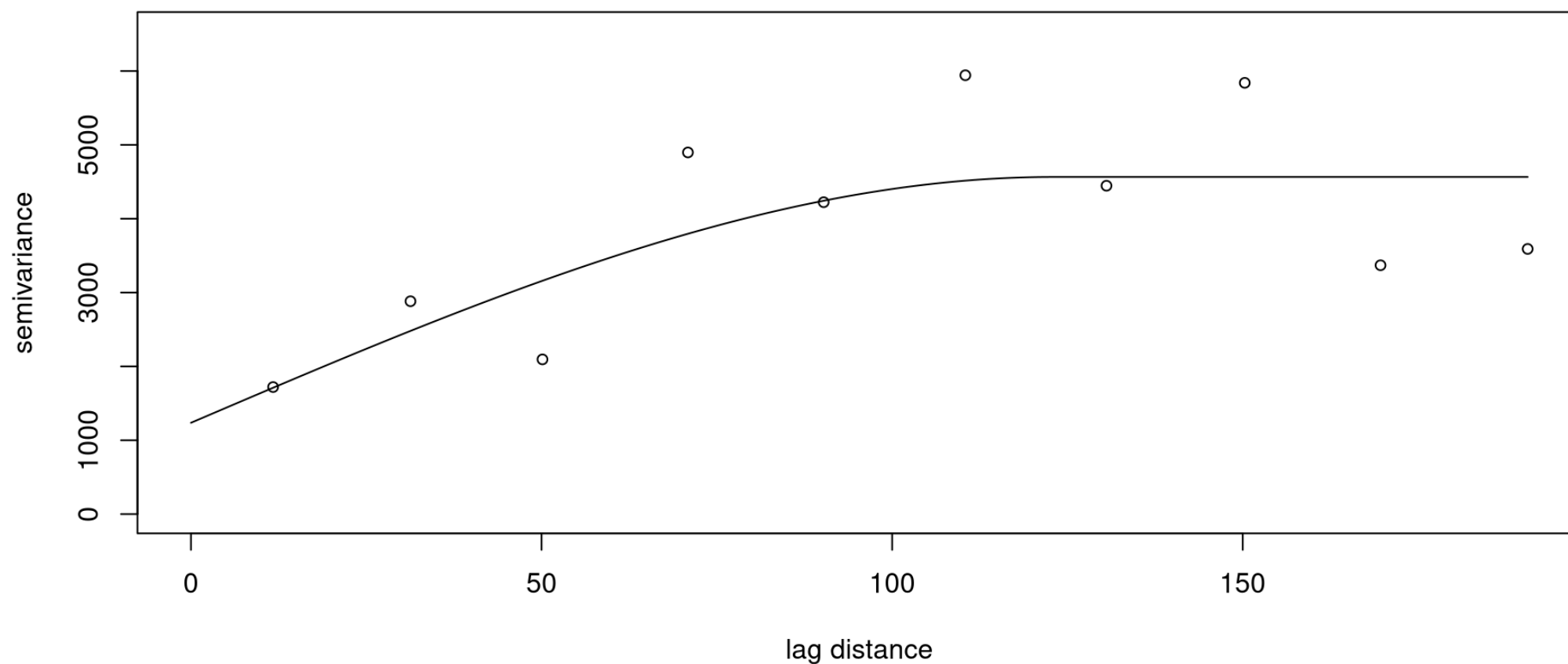
Convergence in 12 function and 7 Jacobian/gradient evaluations

Estimating equations (gradient)

	eta	scale
Gradient	: -8.880367e-05	7.654524e-04

Wolfcamp: MaxLik fitted variogram

```
1 plot(r.georob.ml, lag.dist.def=20, max.lag=200)
```



8 Restricted maximum likelihood

Equivalent number of independent observations of a sample of spatial data often much smaller than nominal sample size $\Rightarrow$ bias of maximum likelihood estimates of variance parameters important

The bias of MLEs of variogram parameters θ can be reduced by **restricted maximum likelihood estimation** (REML).

Principle of **REML**:

1. Form linear combinations $\mathbf{Z} = \mathbf{A}\mathbf{Y}$ of the data $\mathbf{Y}$ that have zero expectation (no longer depend on β):

$$\mathbb{E}[\mathbf{Z}] = \mathbf{A}\mathbf{X}\beta = \mathbf{0}$$

The matrix $\mathbf{A}$ must satisfy: $\mathbf{A}\mathbf{X} = \mathbf{0}$

$\mathbf{A}$ is non-unique with many possibilities

2. Estimate θ by maximizing the likelihood function for $n - p$ elements of $\mathbf{Z}$.

REML estimates Wolfcamp data

```
1 r.georob.reml <- georob(pressure~x+y, d.w,  
2   locations=~x+y, variogram.model="RMspheric",  
3   param=c(variance=3000, nugget=1000, scale=100),  
4   tuning.psi=1000)  
5 summary(r.georob.reml)
```

```
Call:georob(formula = pressure ~ x + y, data = d.w, locations = ~x +  
y, variogram.model = "RMspheric", param = c(variance = 3000,  
nugget = 1000, scale = 100), tuning.psi = 1000)
```

Tuning constant: 1000

Convergence in 6 function and 5 Jacobian/gradient evaluations

Estimating equations (gradient)

	eta	scale
Gradient	: -2.248651e-04	-1.070402e-01

Maximized restricted log-likelihood: -456.3802

Predicted latent variable (B):

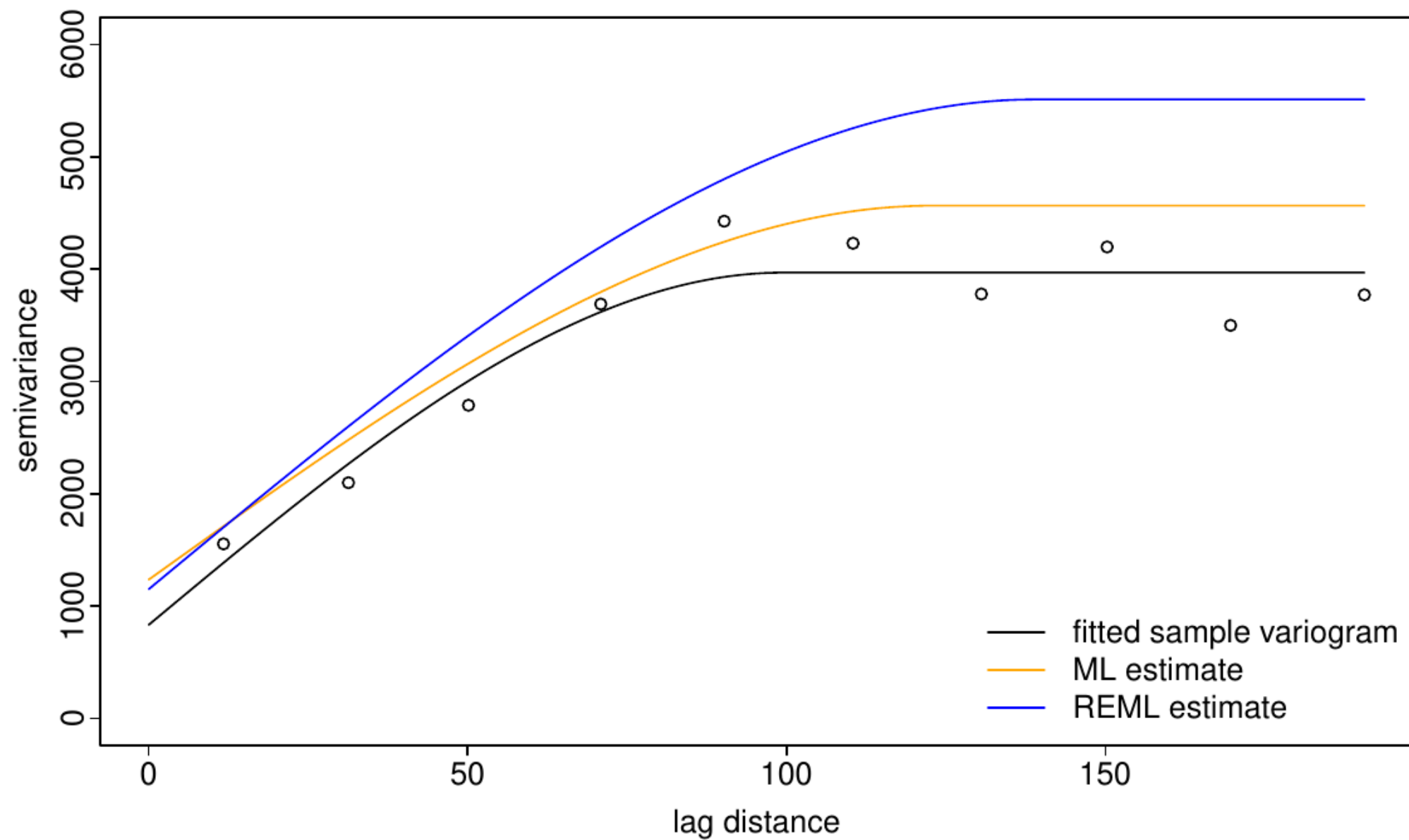
Min	1Q	Median	3Q	Max
-94.58	-60.99	-17.59	23.10	115.72

Residuals (epsilon):

Min	1Q	Median	3Q	Max
-59.148	-18.009	6.251	15.982	54.620

Standardized residuals:

Min	1Q	Median	3Q	Max
-2.4030	-0.7131	0.2282	0.6937	3.1932





**Utrecht
University**

Sharing science,
shaping tomorrow