STATISTICAL AND MACHINE LEARNING FOR BIG GEOSPATIAL DATA: Part I

Abhi Datta
Johns Hopkins University
Department of Biostatistics

Course outline

Part I: Introduction to geostatistics and spatial linear models

Part II: Random forests for geospatial data

Part III: Neural networks for geospatial data

Part IV: Software demonstration

Course outline

Part I: Introduction to geostatistics and spatial linear models

Part II: Random forests for geospatial data

- Part IV a: Software demonstration of random forests for spatial analysis in R

Part III: Neural networks for geospatial data

- Part IV b: Software demonstration of neural nets for spatial analysis in Python

Course materials available at …

Overview of Part I

Introduction to geostatistics

Exploratory data analysis

- Maps and variograms

Gaussian Processes (GP) and spatial linear regression

- Estimation and prediction (kriging)
- Spatial linear mixed effect models

Big spatial data

- Computing challenges
- Fast alternatives (Nearest Neighbor Gaussian Process)

What is spatial data

Any data with some geographical information

Common sources of spatial data:
- climatology, forestry, ecology, environmental health, disease epidemiology, real estate marketing

Other examples where spatial need not refer to space on earth:
- Neuroimaging (data for each voxel in the brain)
- Genetics (position along a chromosome)
- Spatial transcriptomics (gene expression on slides)

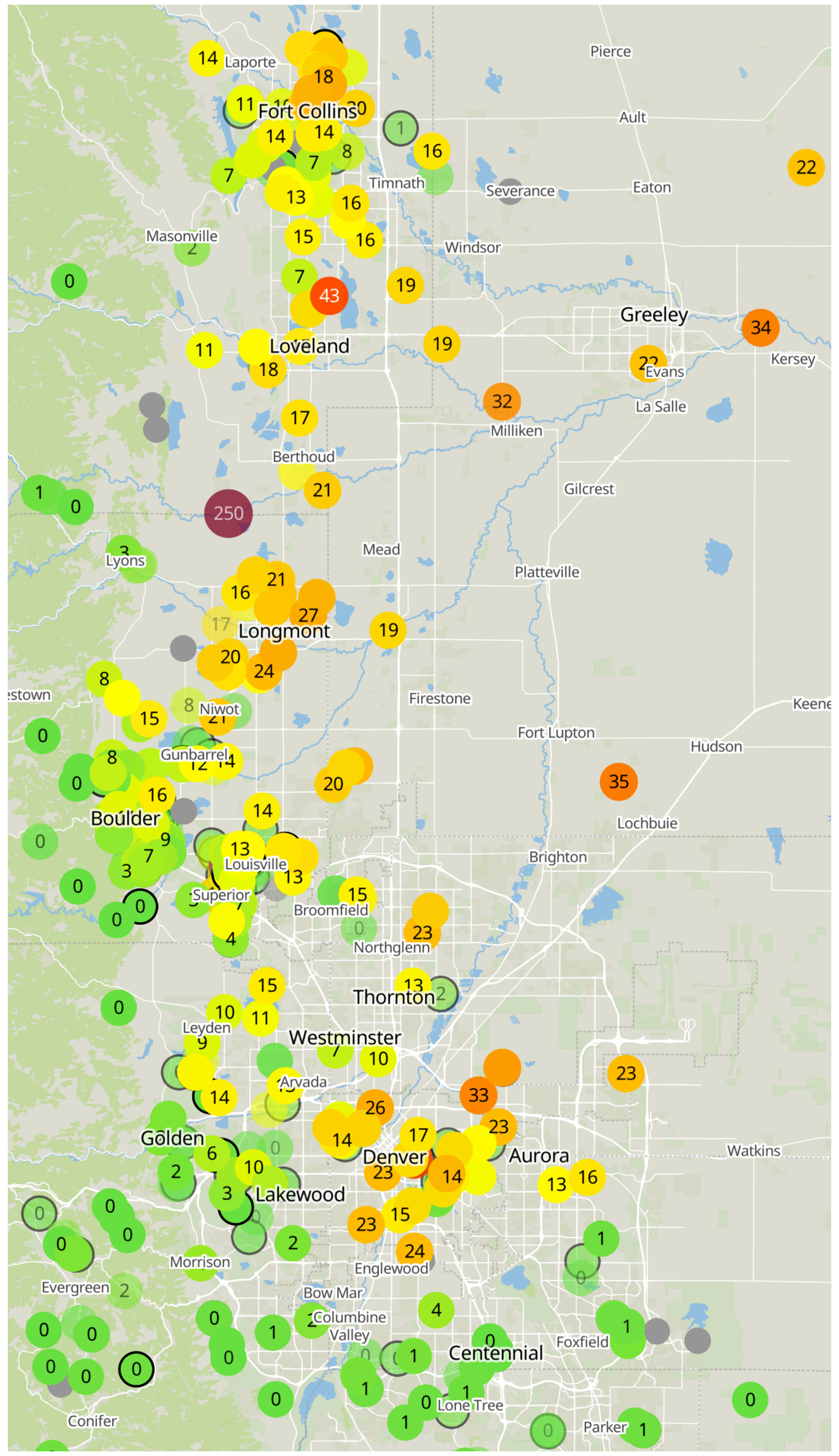
Geostatistics

Each observation (data unit) is associated with a geographical location (latitude-longitude)

Data represents a sample from a continuous spatial domain

Often displayed on a map

Referred to as geocoded/ geostatistical/ point referenced data

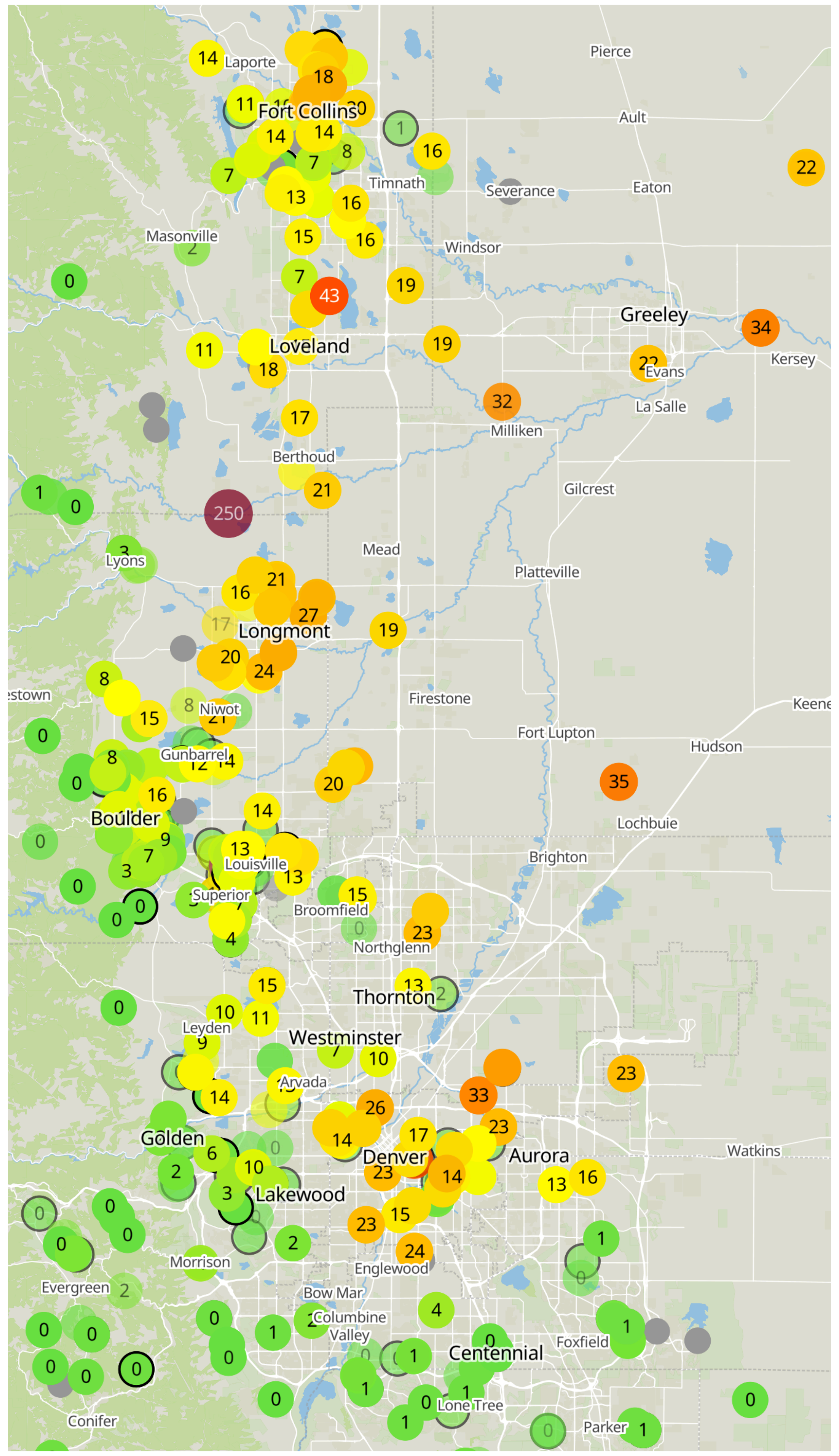


PM$_{2.5}$ ($\mu g/m^3$) in Colorado on Nov 12, 2024 from PurpleAir.com

Geostatistics

Point reference data:

- Data collected at locations $s_1, \ldots, s_n$
- $Y_i = Y(s_i)$: scalar response at location s_i



PM$_{2.5}$ ($\mu g/m^3$) in Colorado
on Nov 12, 2024 from PurpleAir.com

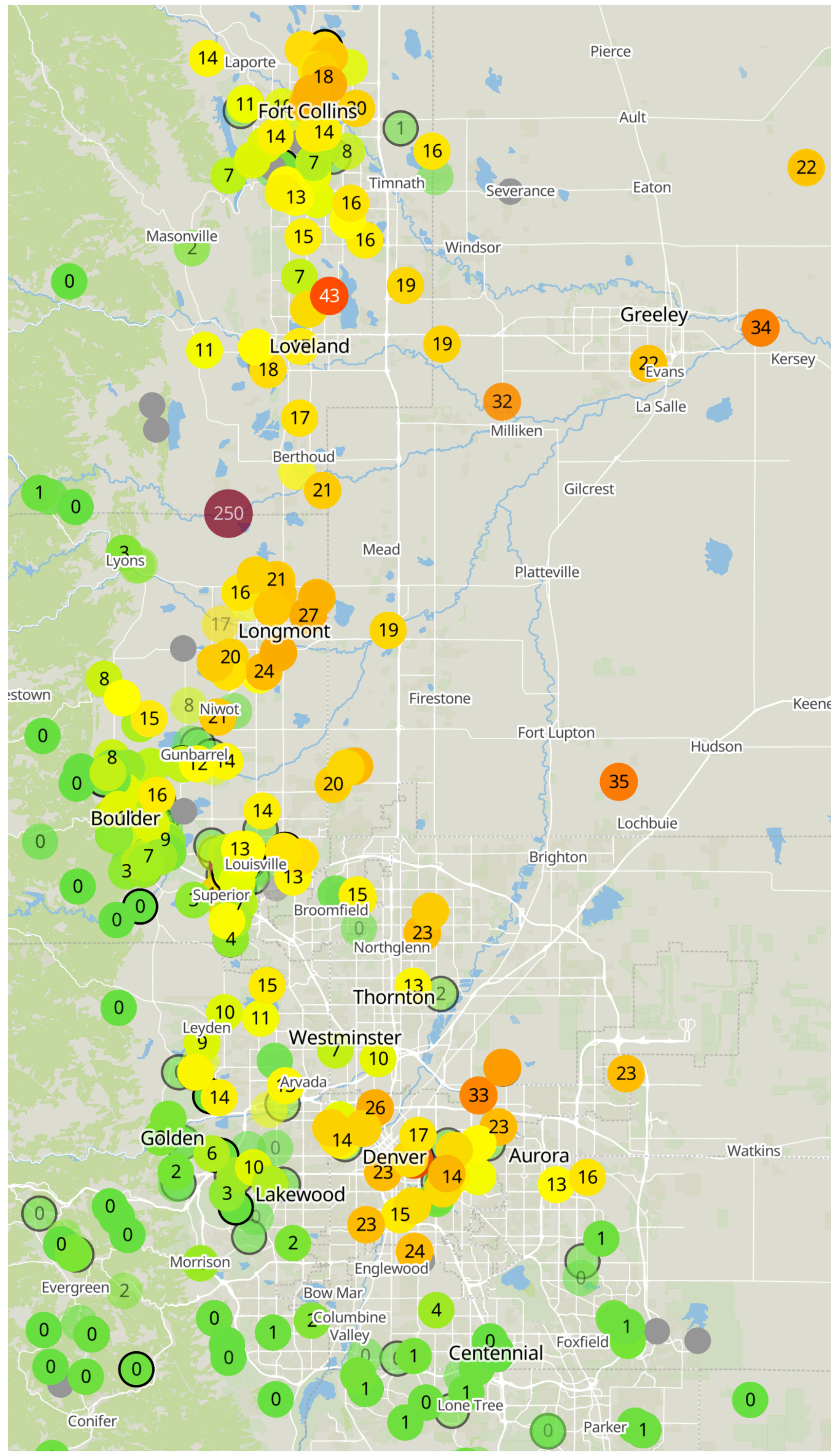
Geostatistics

Point reference data:

Data collected at locations $s_1, \ldots, s_n$

$Y_i = Y(s_i)$: scalar response at location s_i

$X_i = X(s_i)$: $d \times 1$ vector of covariates (explanatory variables)



PM$_{2.5}$ $(\mu g/m^3)$ in Colorado on Nov 12, 2024 from PurpleAir.com

Geostatistics

Point reference data:

Data collected at locations $s_1, \ldots, s_n$

$Y_i = Y(s_i)$: scalar response at location s_i

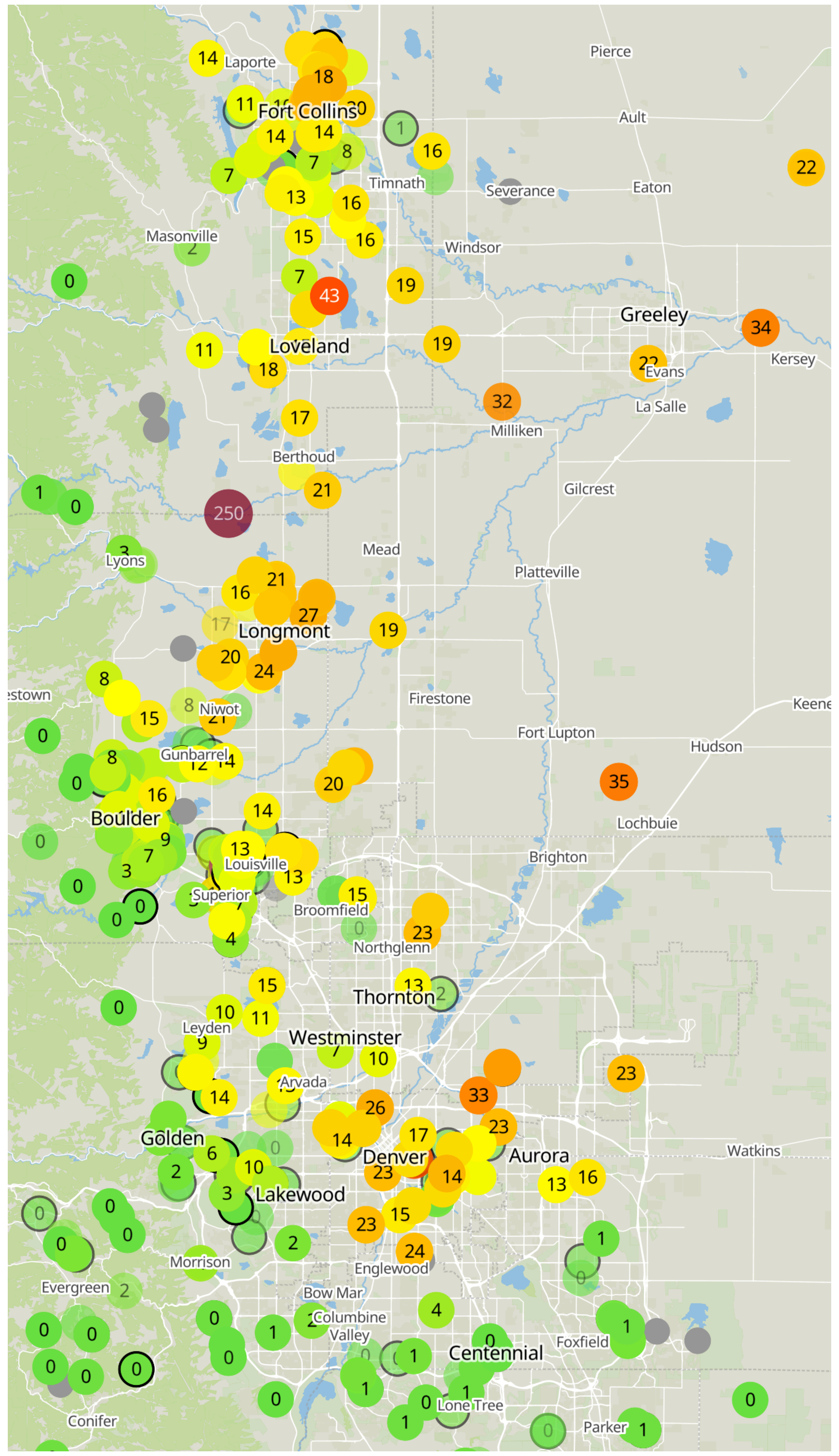
$X_i = X(s_i)$: $d \times 1$ vector of covariates (explanatory variables)

Objectives:

Predict Y at any location without data

Understand spatial patterns in Y

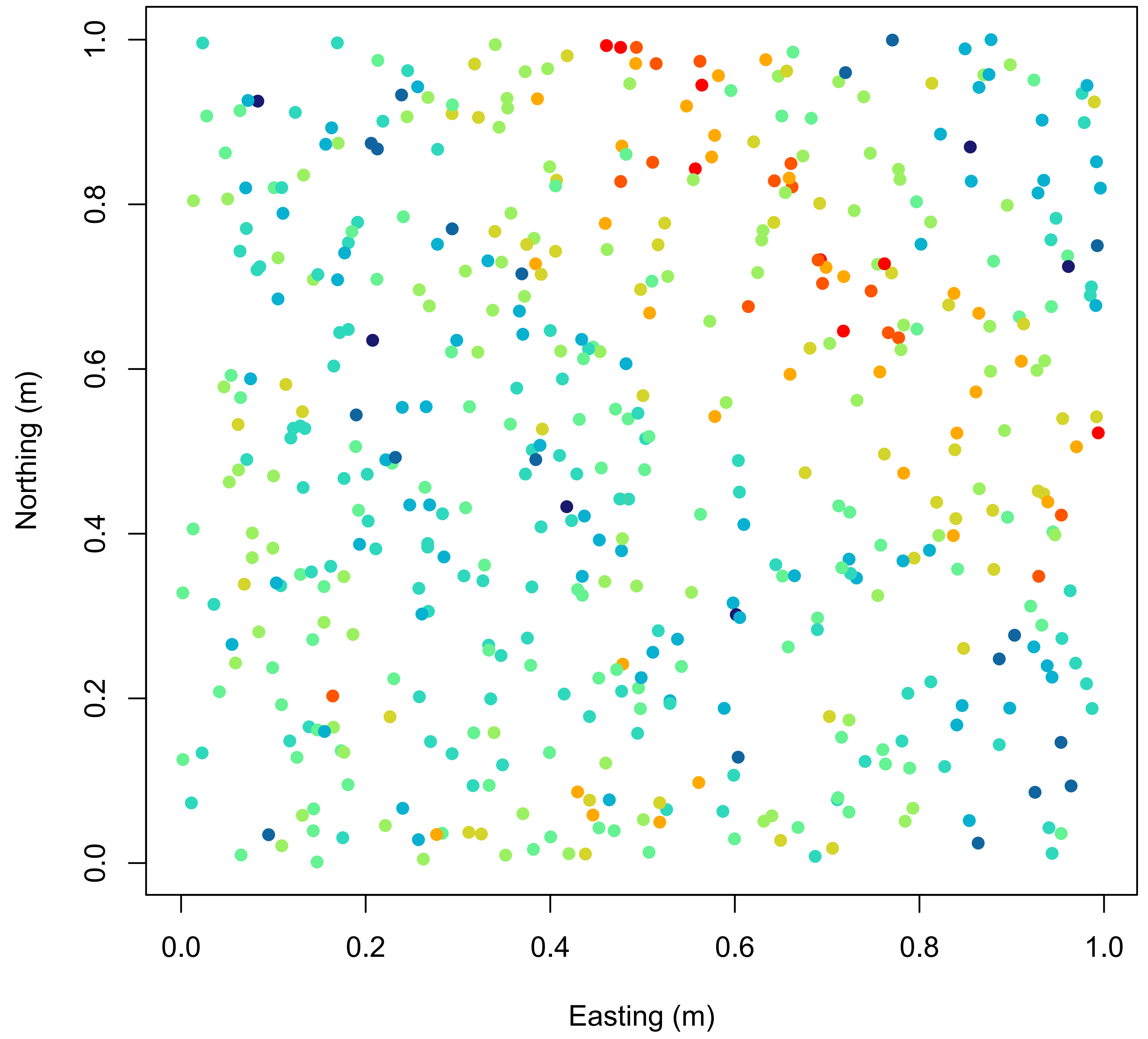
Understand relationship between X and Y



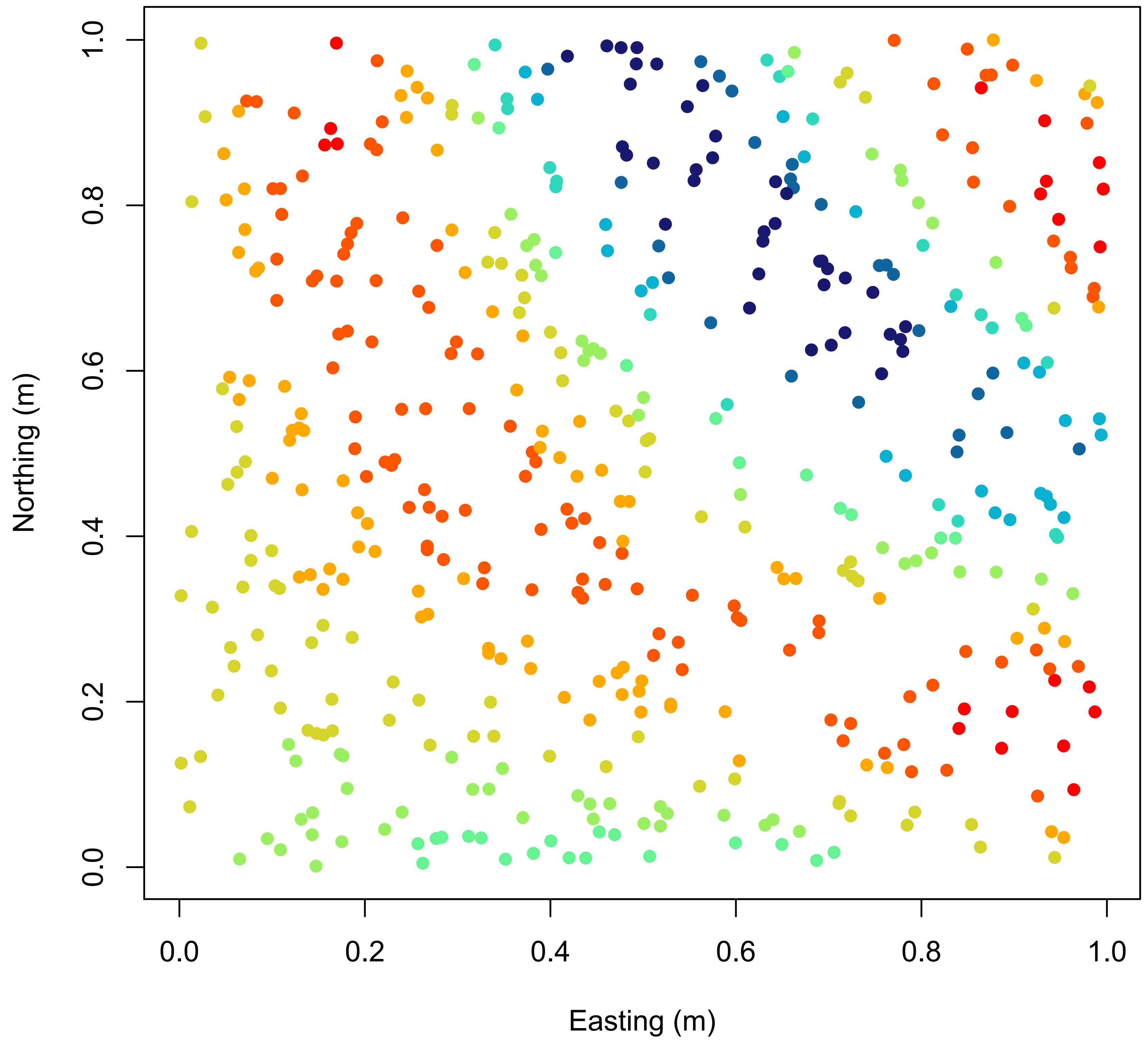
PM$_{2.5}$ ($\mu g/m^3$) in Colorado on Nov 12, 2024 from PurpleAir.com

Exploratory data analysis (EDA): Plotting the data

Point plots help to visualize the exact data where they are observed



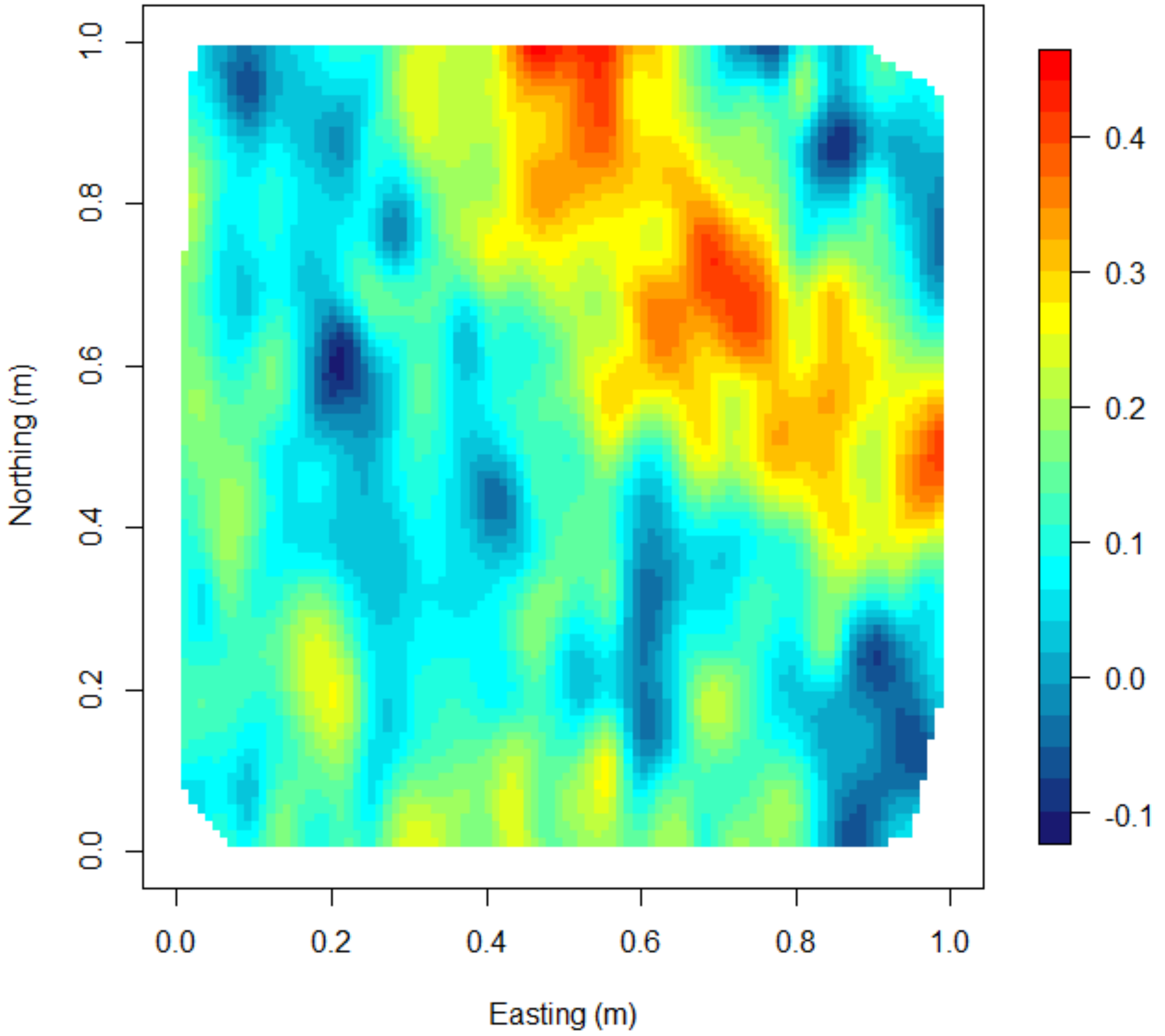
$Y(s_i)$



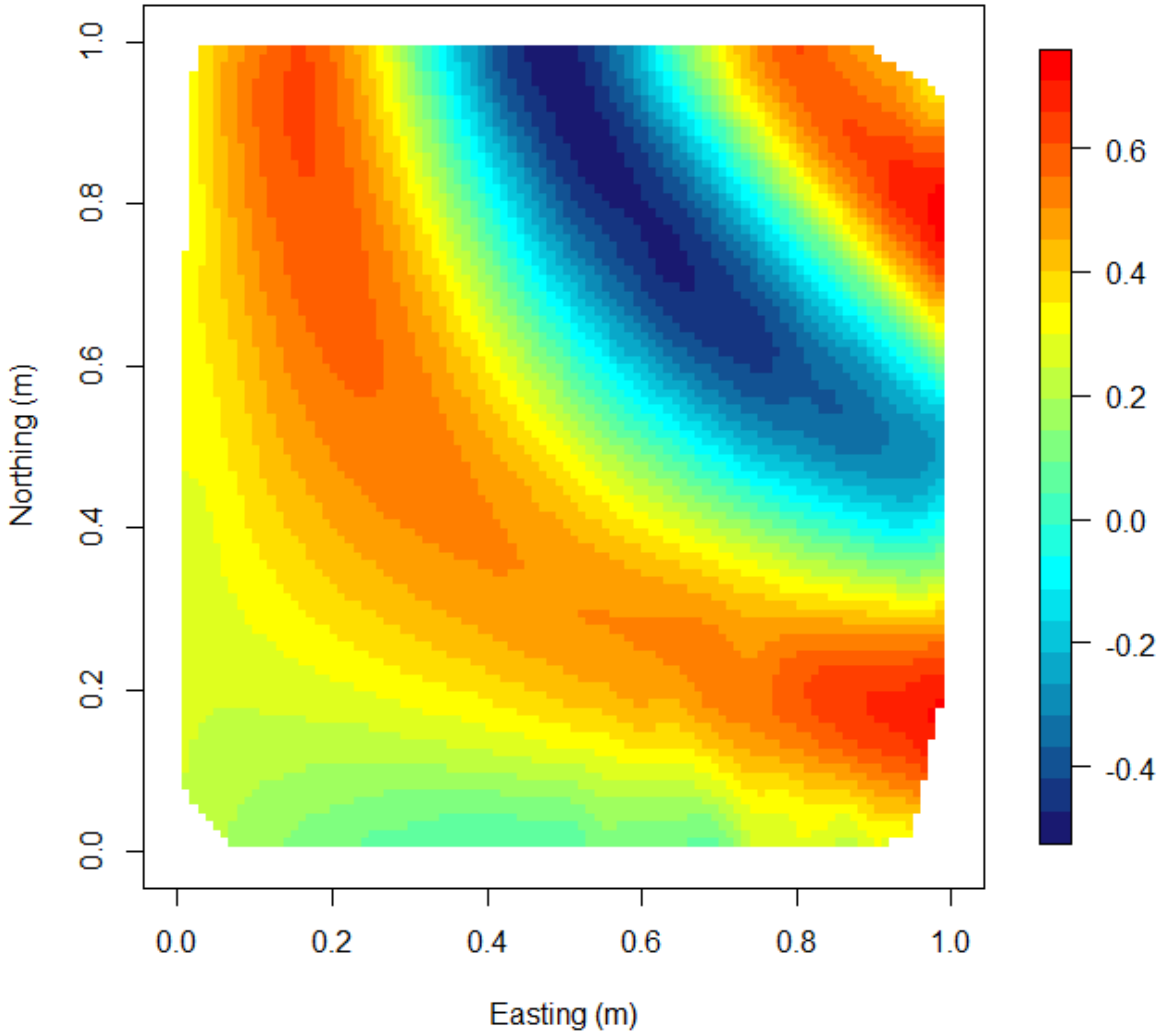
$X(s_i)$

Exploratory data analysis (EDA): Plotting the data

Surface plots (interpolated data) are often better help understand spatial patterns



$Y(s)$



$X(s)$

What's so special about spatial?

Linear regression model: $Y(s_i) = X(s_i)'\beta + \epsilon(s_i)$

$\epsilon(s_i)$ are iid $N(0,\tau^2)$ errors

$Y = (Y(s_1), Y(s_2), \ldots, Y(s_n))'; X = (X(s_1)', X(s_2)', \ldots, X(s_n)')'$

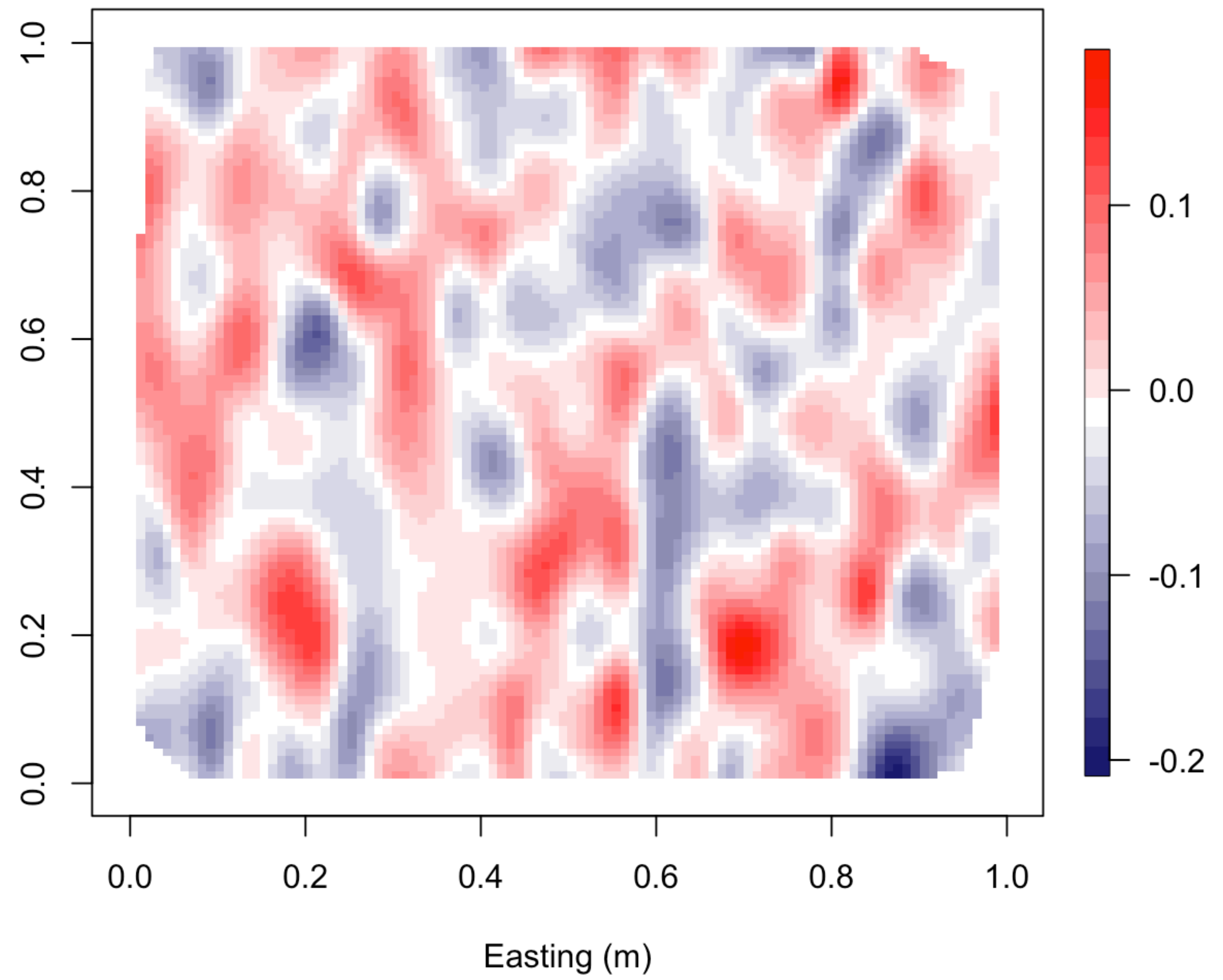
Inference: $\hat{\beta} = (X'X)^{-1}X'Y \sim N(\beta, \tau^2(X'X)^{-1})$

Prediction at new location s_0: $\widehat{Y(s_0)} = X(s_0)'\hat{\beta}$

Although the data is spatial, this is an ordinary linear regression model

Residual plots

Surface plots of residuals $y(s) - \hat{y}(s)$ help identify residual spatial patterns not explained by the covariates



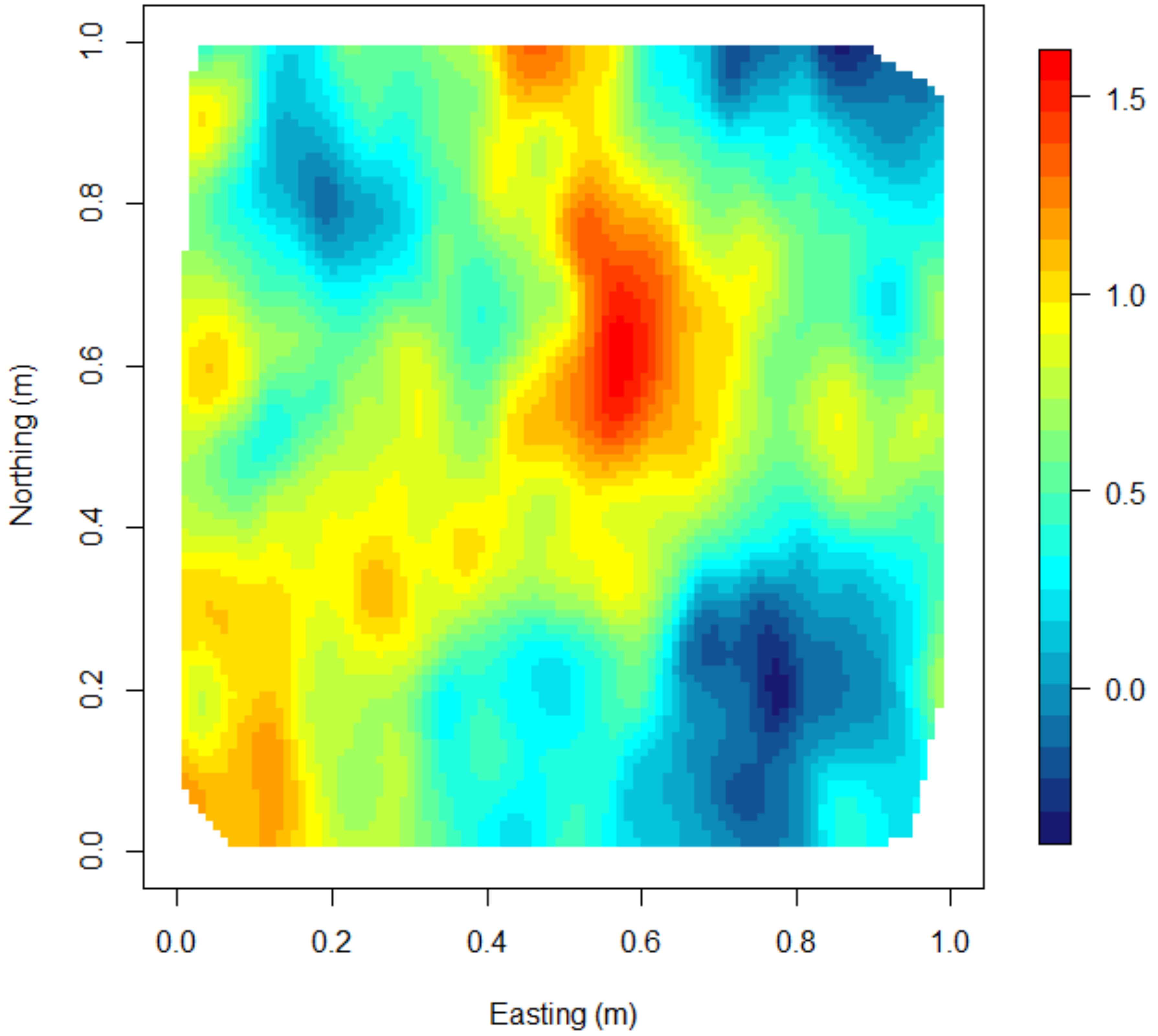
Surface plot of residuals

Surface plot of residuals not showing any large scale spatial patterns

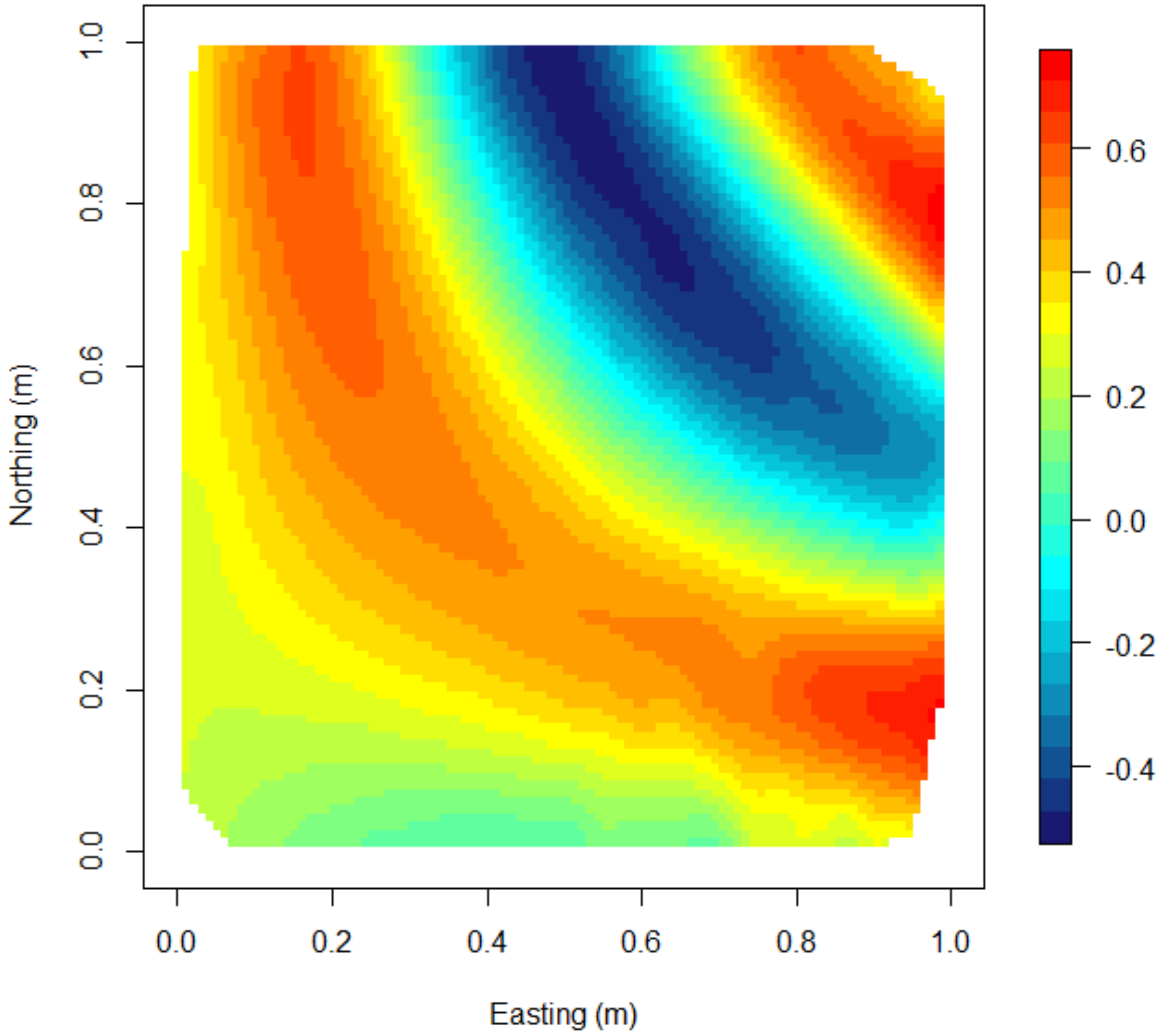
The covariate $X(s)$ seems to explain all spatial variation in the response $Y(s)$

When does such a non-spatial analysis suffice?

Another dataset



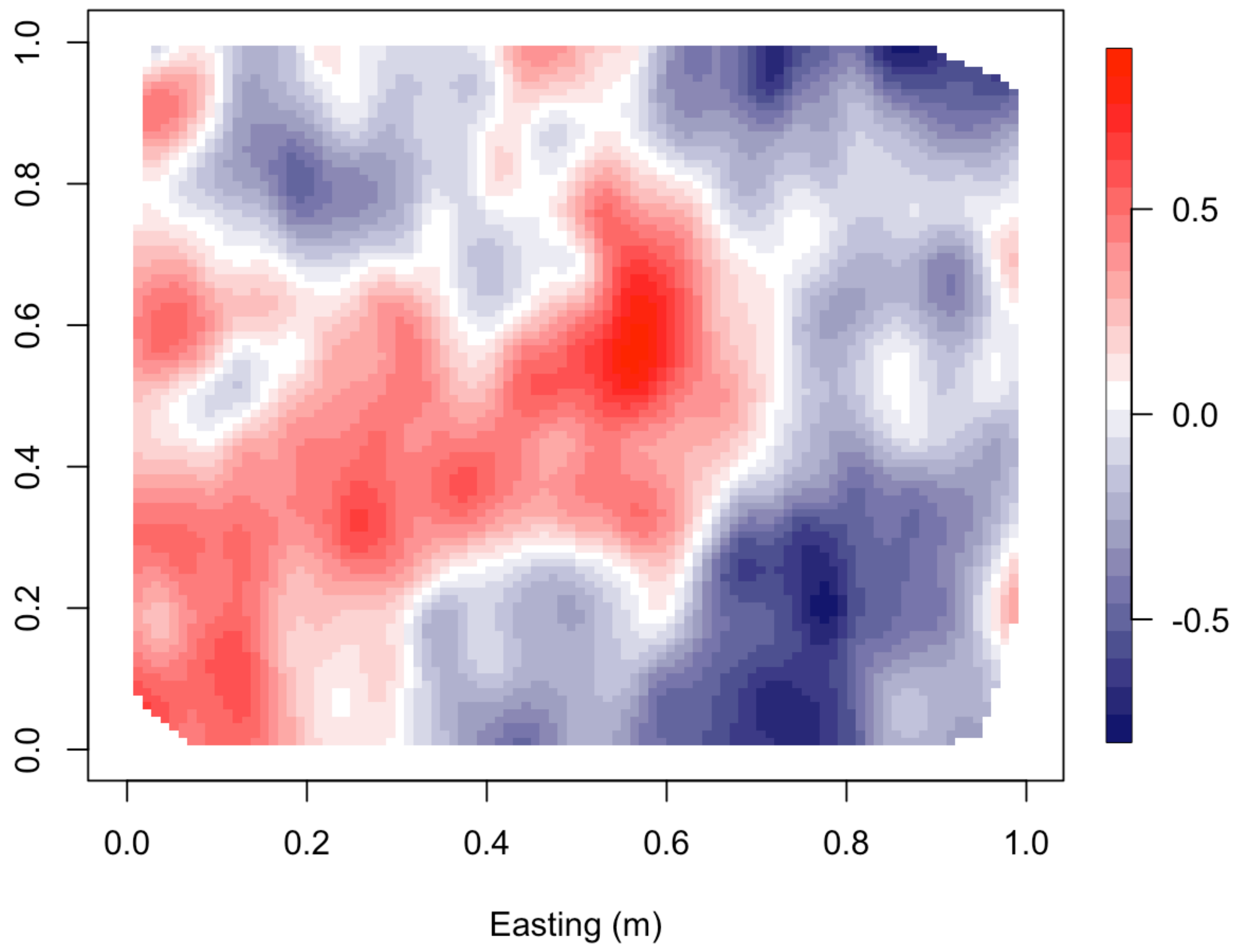
Dataset 2: $Y(s)$



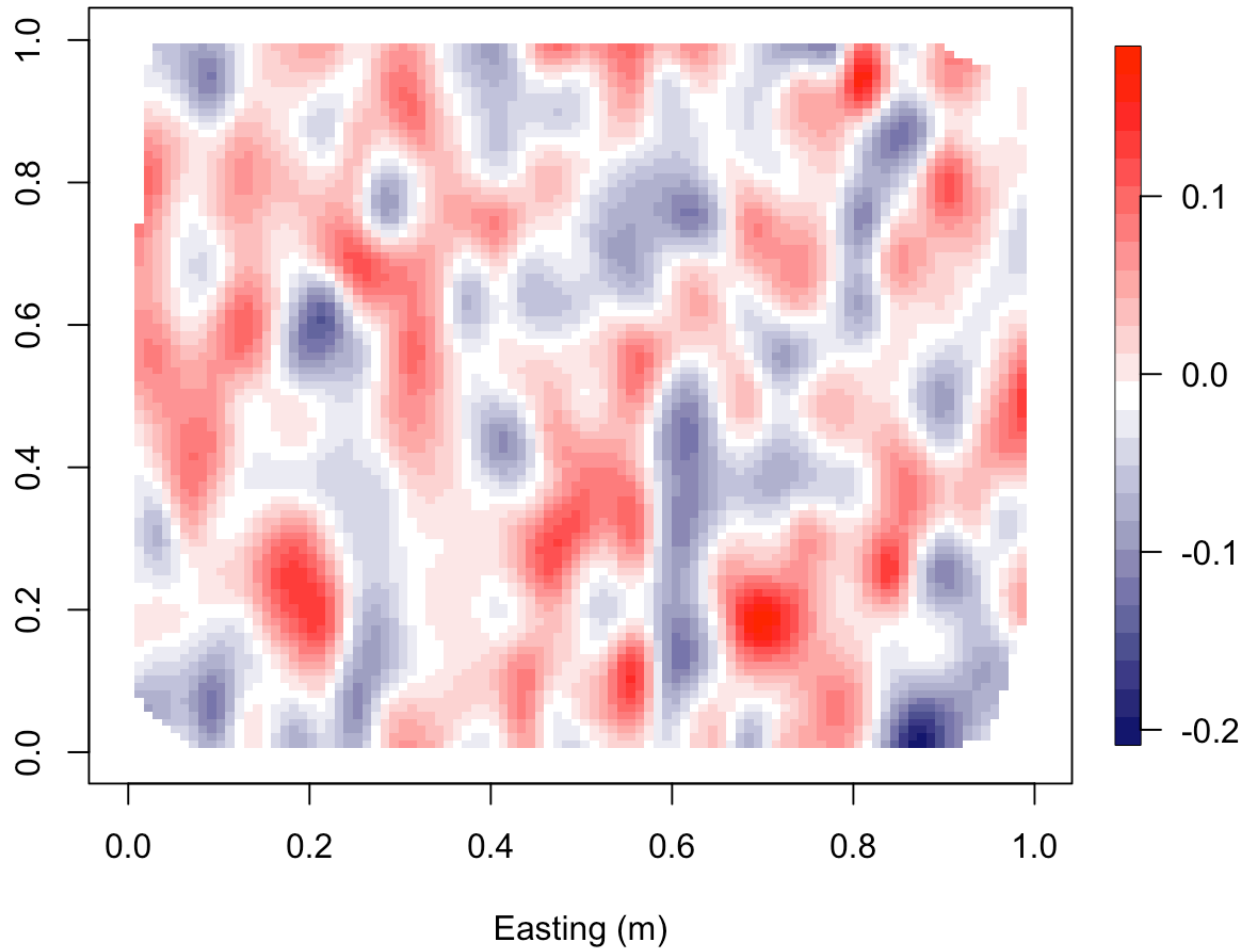
Same $X(s)$

Another dataset

Linear regression: $y(s_i) = \beta_0 + x(s_i)\beta_1 + \epsilon(s_i)$



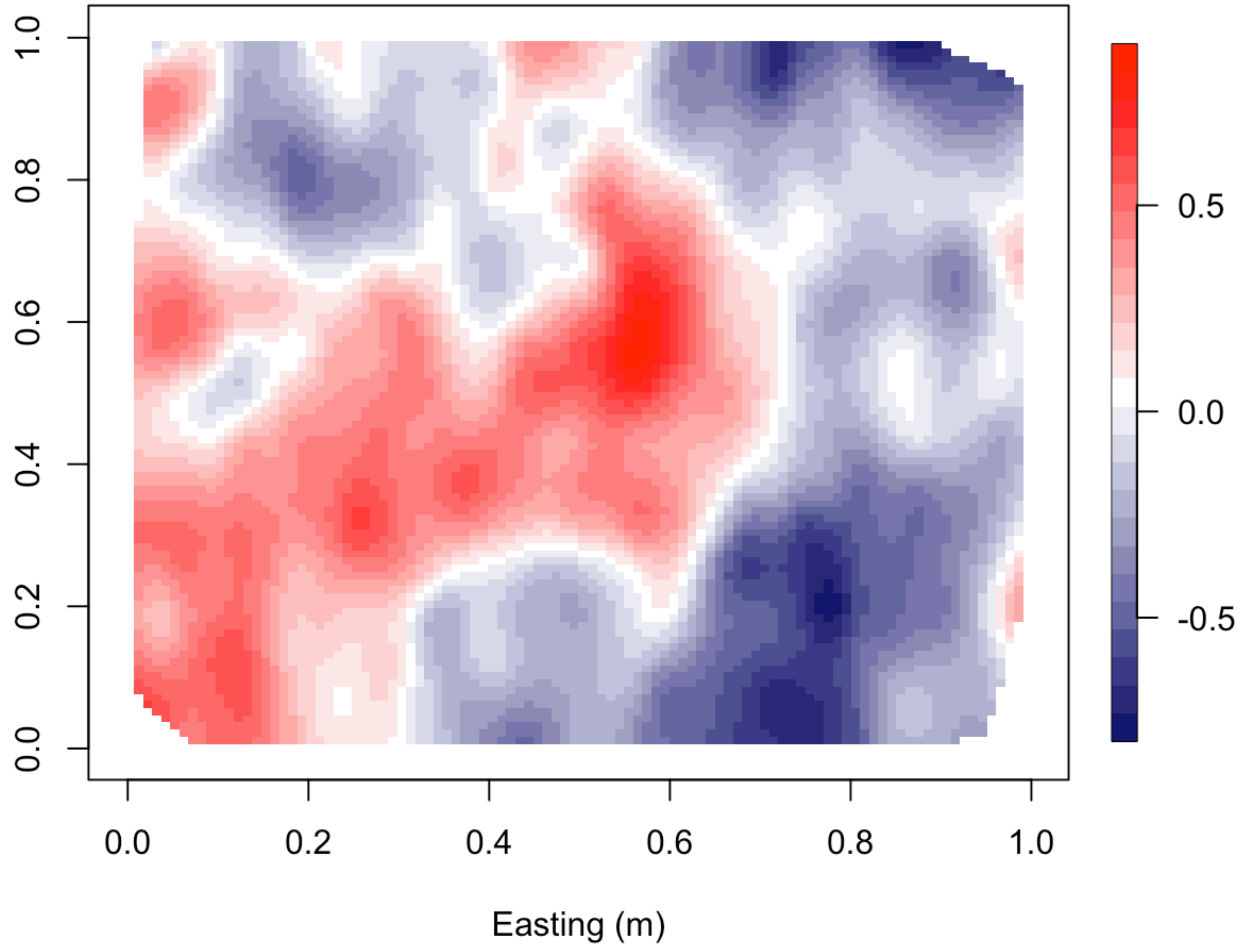
Dataset 2: Residual plot



Dataset 1: Residual plot

Another dataset

Linear regression: $y(s_i) = \beta_0 + x(s_i)\beta_1 + \epsilon(s_i)$



Dataset 2: Residual plot

Strong residual spatial pattern

The covariate $X(s)$ does not explain all spatial variation in the response $Y(s)$

Besides eyeballing residual surfaces, how to do more formal EDA to identify spatial pattern ?

Semi-Variogram

First Law of Geography: *"Everything is related to everything else, but near things are more related than distant things."* – Waldo Tobler

$Y(s_1)$ and $Y(s_2)$ should be more similar if s_1 is near s_2

$(Y(s_1) - Y(s_2))^2$ should be small when $||s_1 - s_2||$ is small and increase as $||s_1 - s_2||$ increases

Can this be formalized to identify spatial pattern in data?

Semi-Variogram

Empirical semi-variogram:

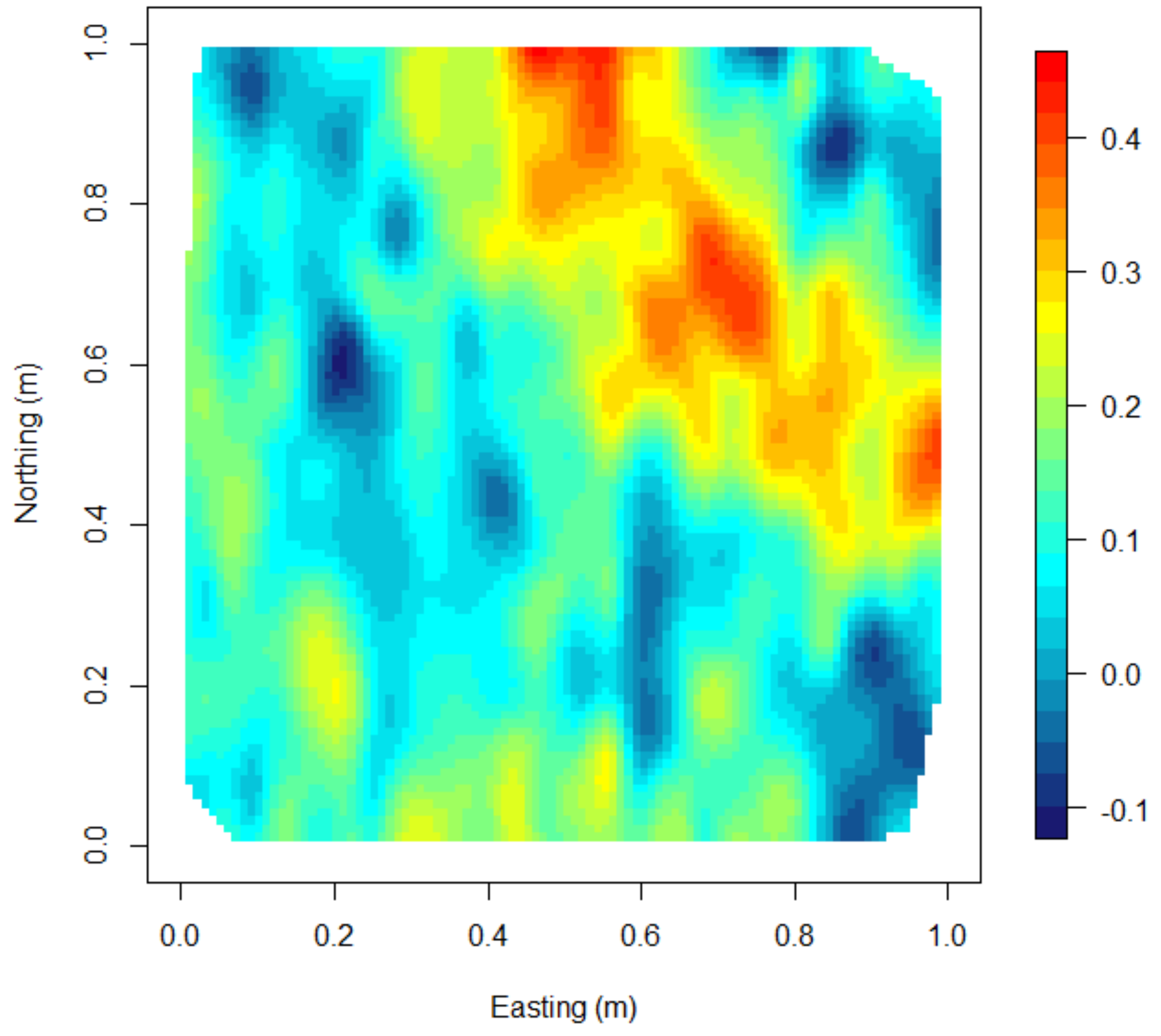
$\gamma(h) =$ Average of $(Y(s_1) - Y(s_2))^2$ for all pairs s_1, s_2 such that $s_1 - s_2 \approx h$

For spatial data, the $\gamma(h)$ is expected to roughly increase with the distance h

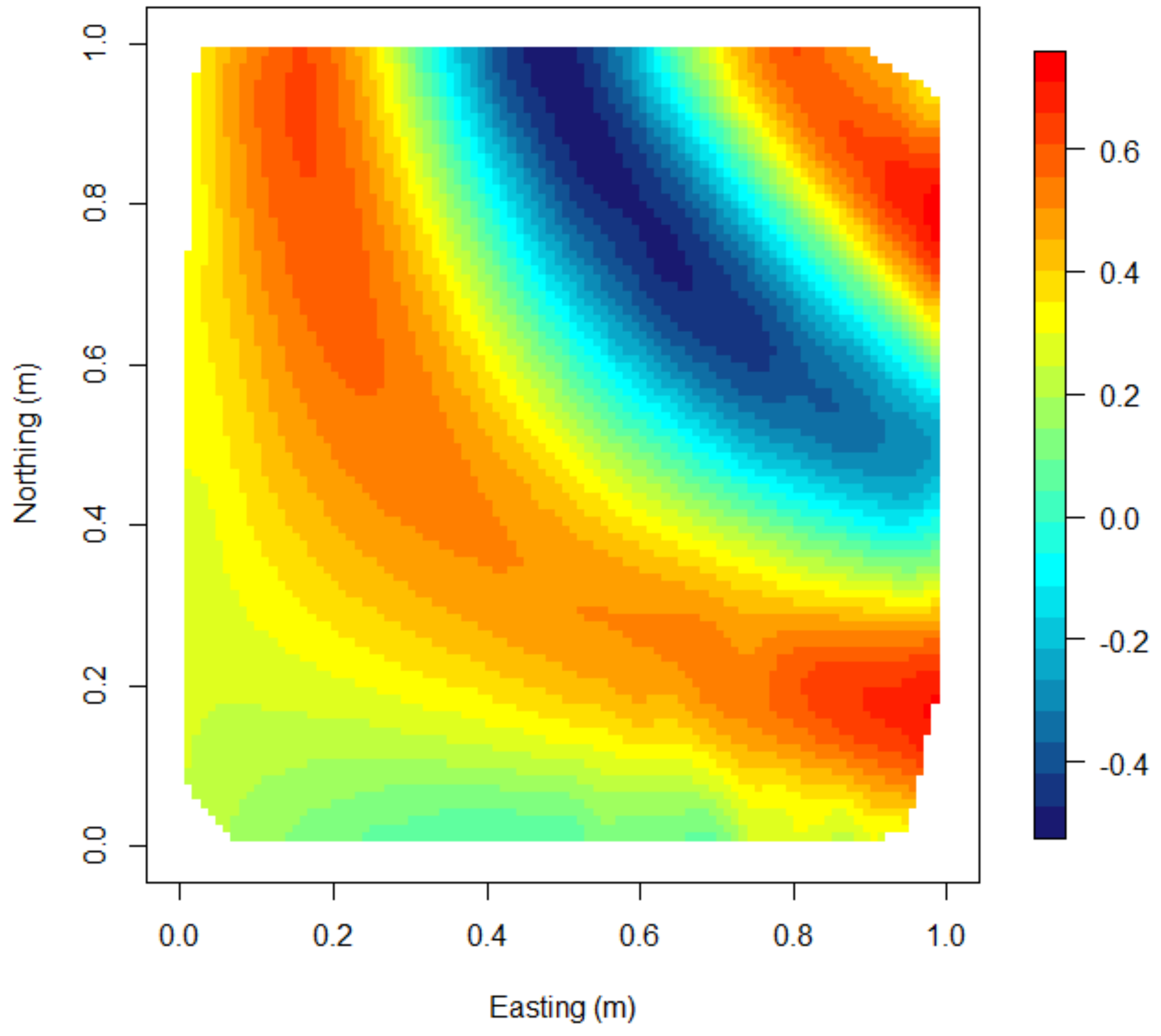
A flat semivariogram would suggest little spatial variation

variog command in the *geoR* R-package calculates empirical semivariograms

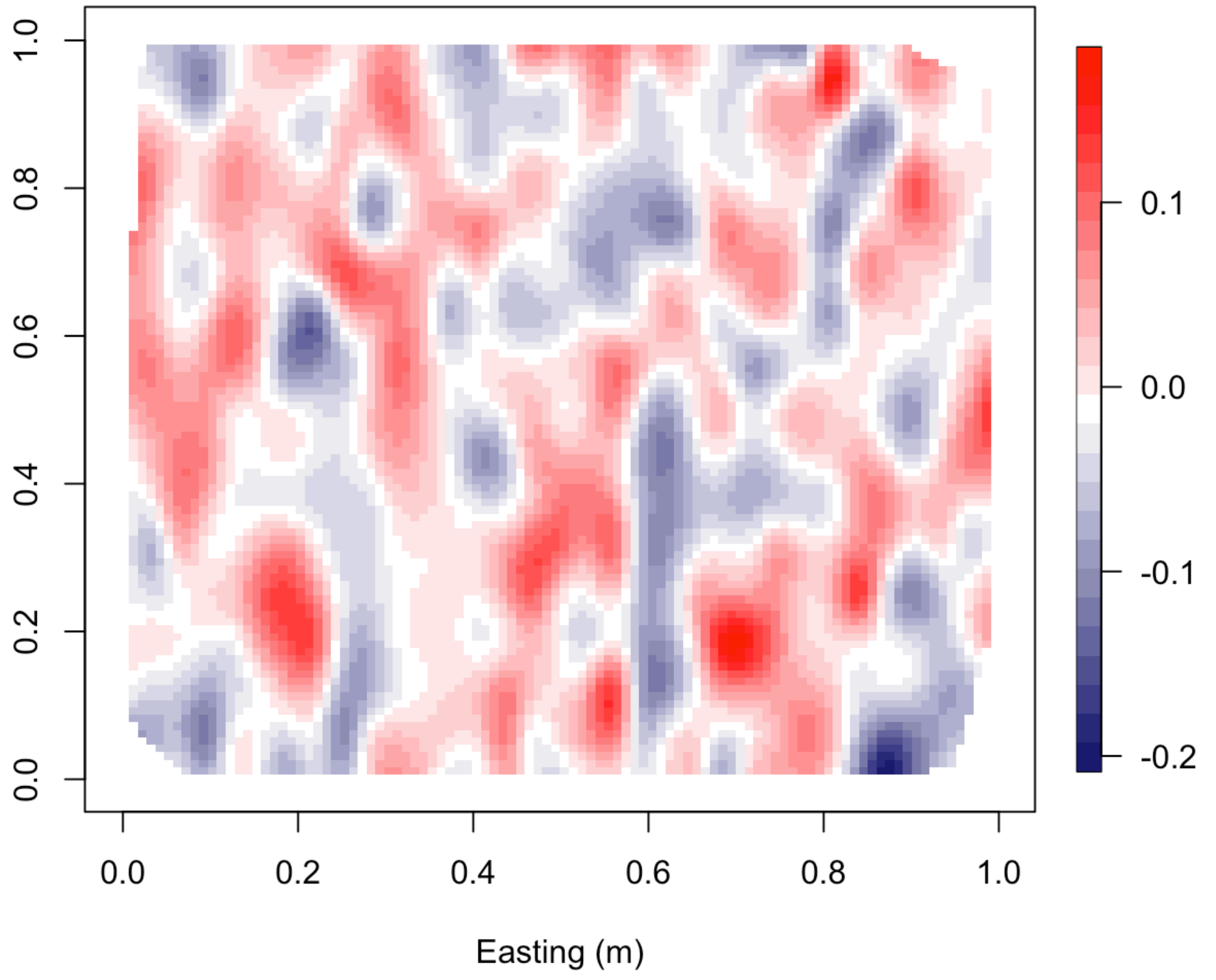
Dataset 1



$Y(s)$

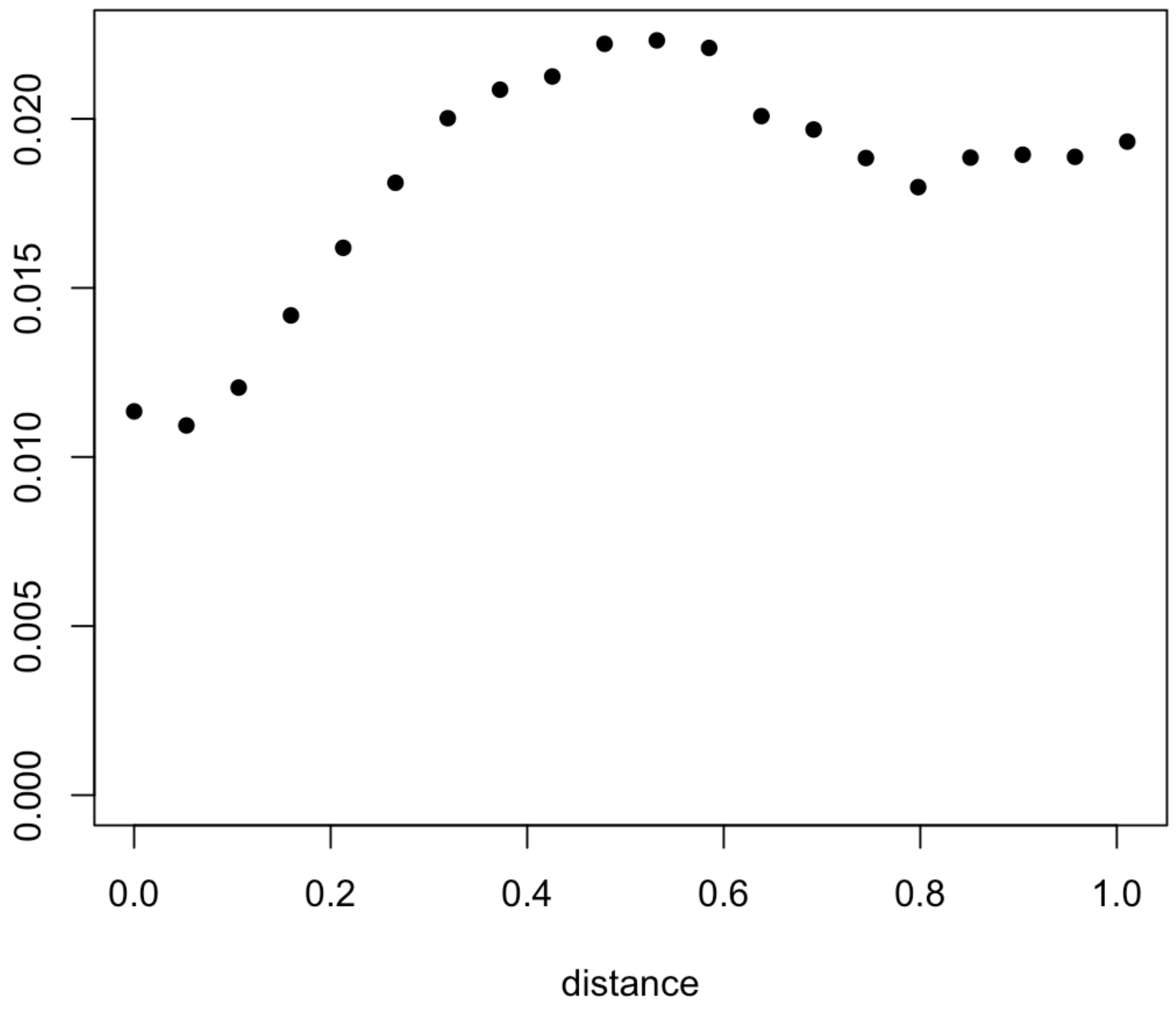


$X(s)$

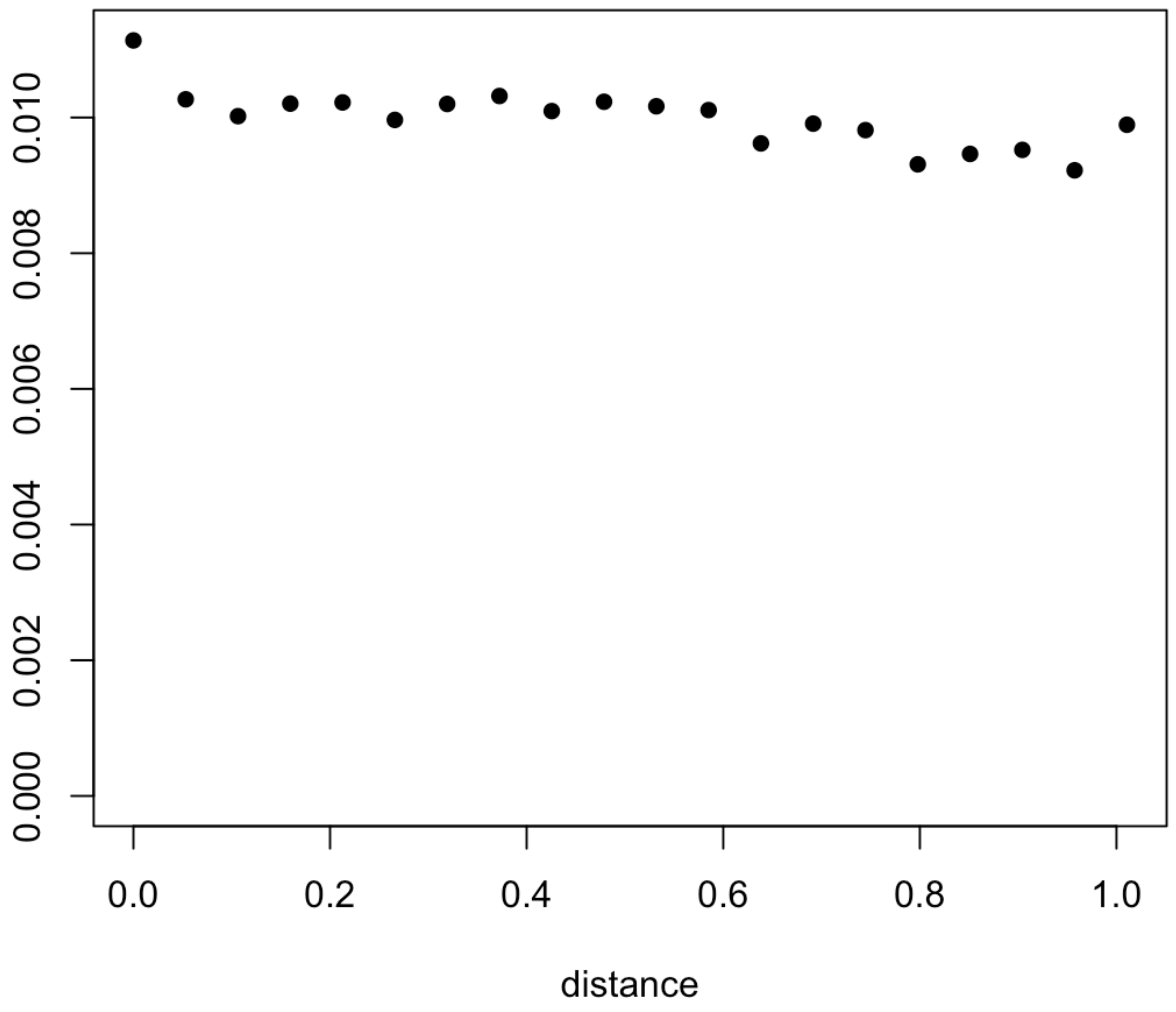


Residuals $Y(s) - \hat{\beta}X(s)$

Dataset 1



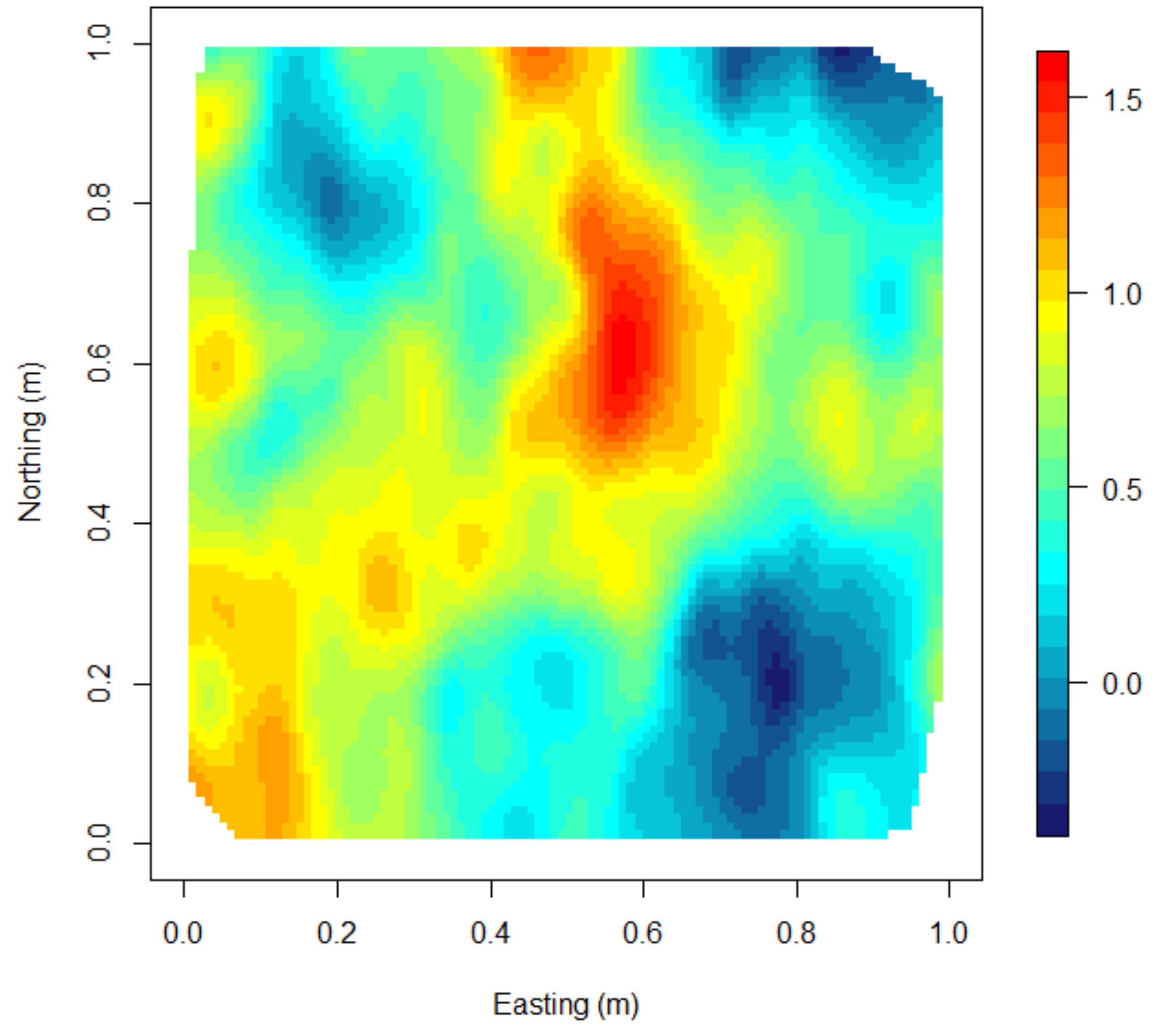
Variogram of $Y(s)$



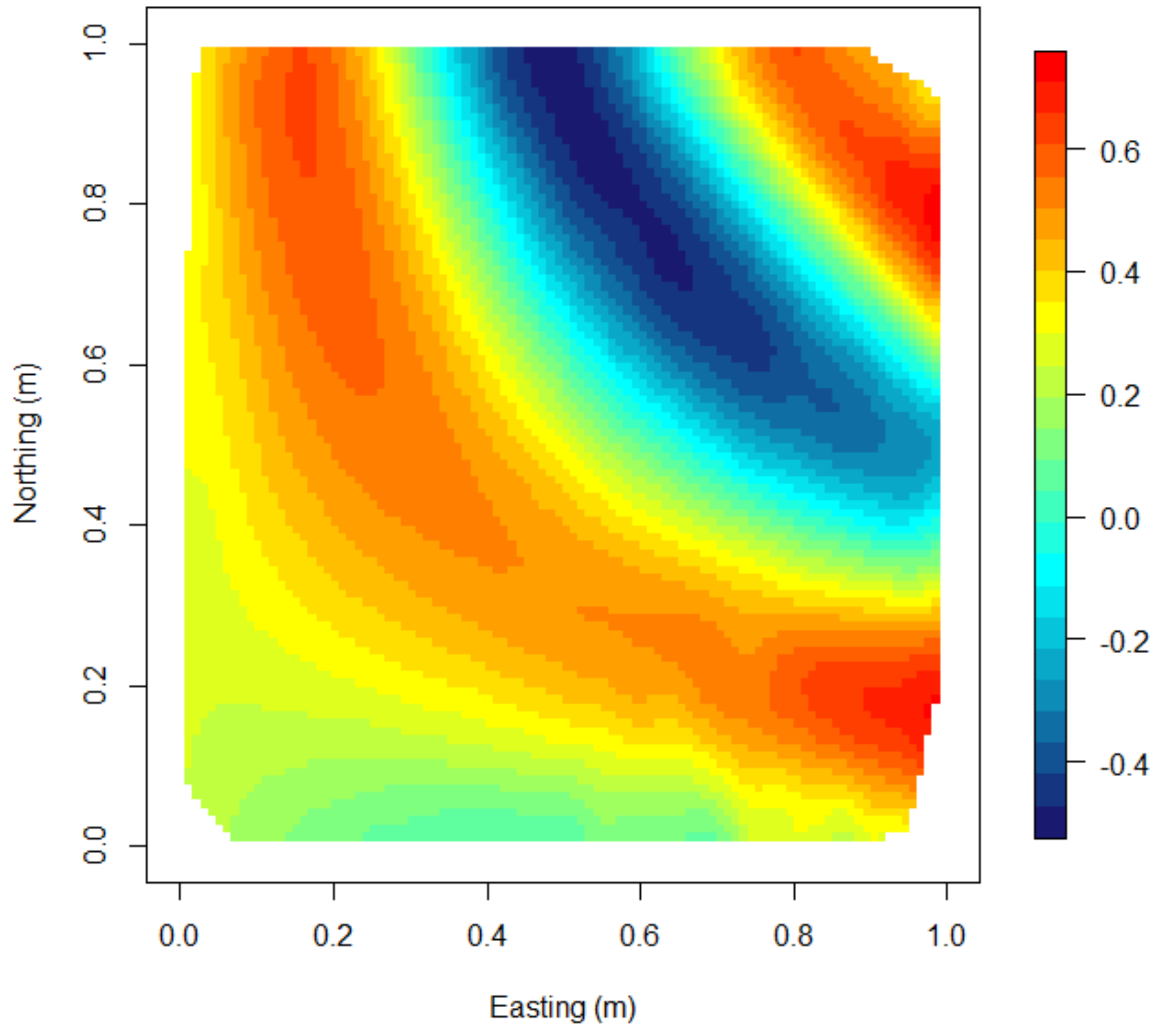
Variogram of residuals
$Y(s) - \hat{\beta}X(s)$

Variogram of residuals suggests very little spatial variation

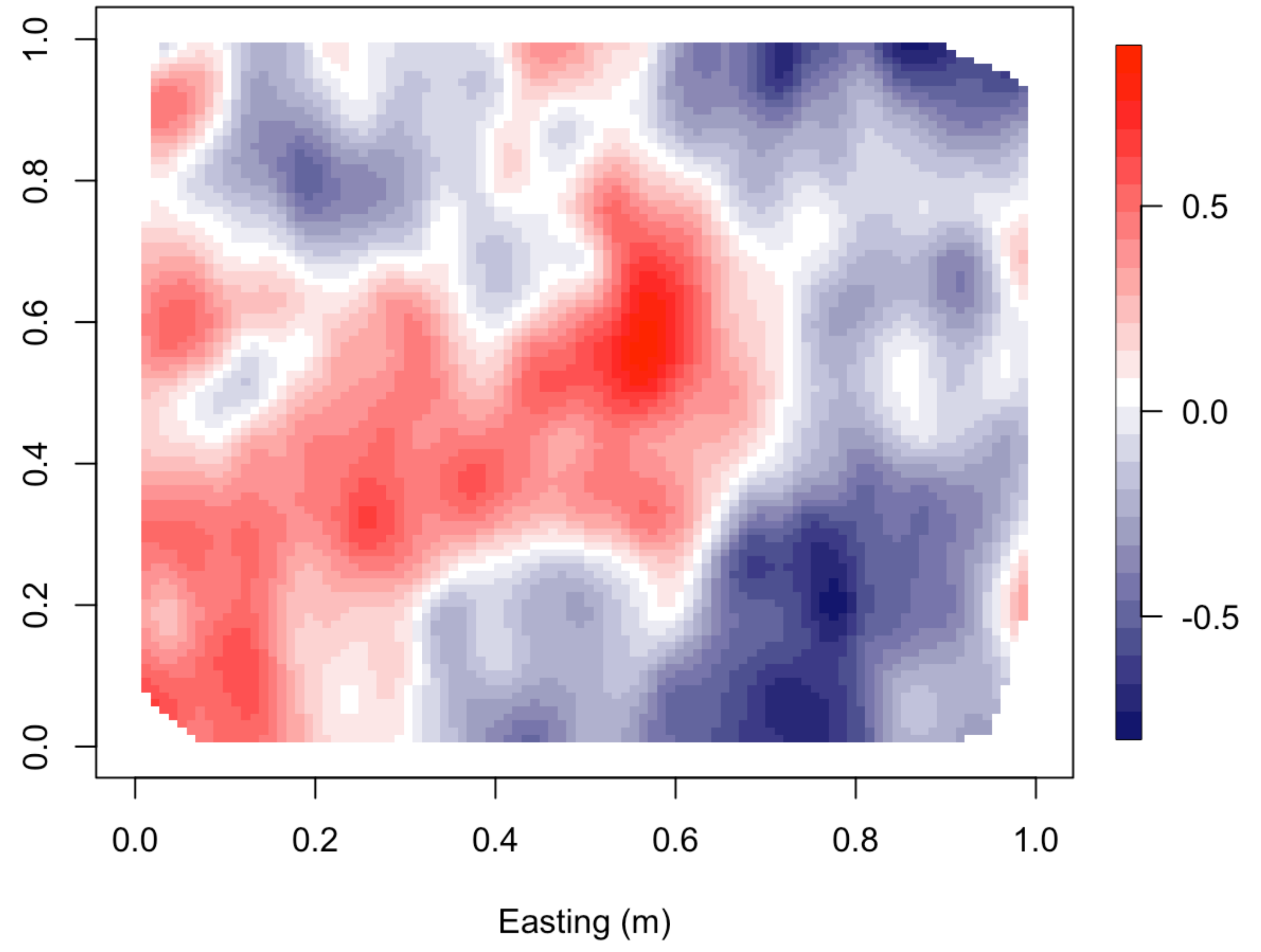
Dataset 2



$Y(s)$

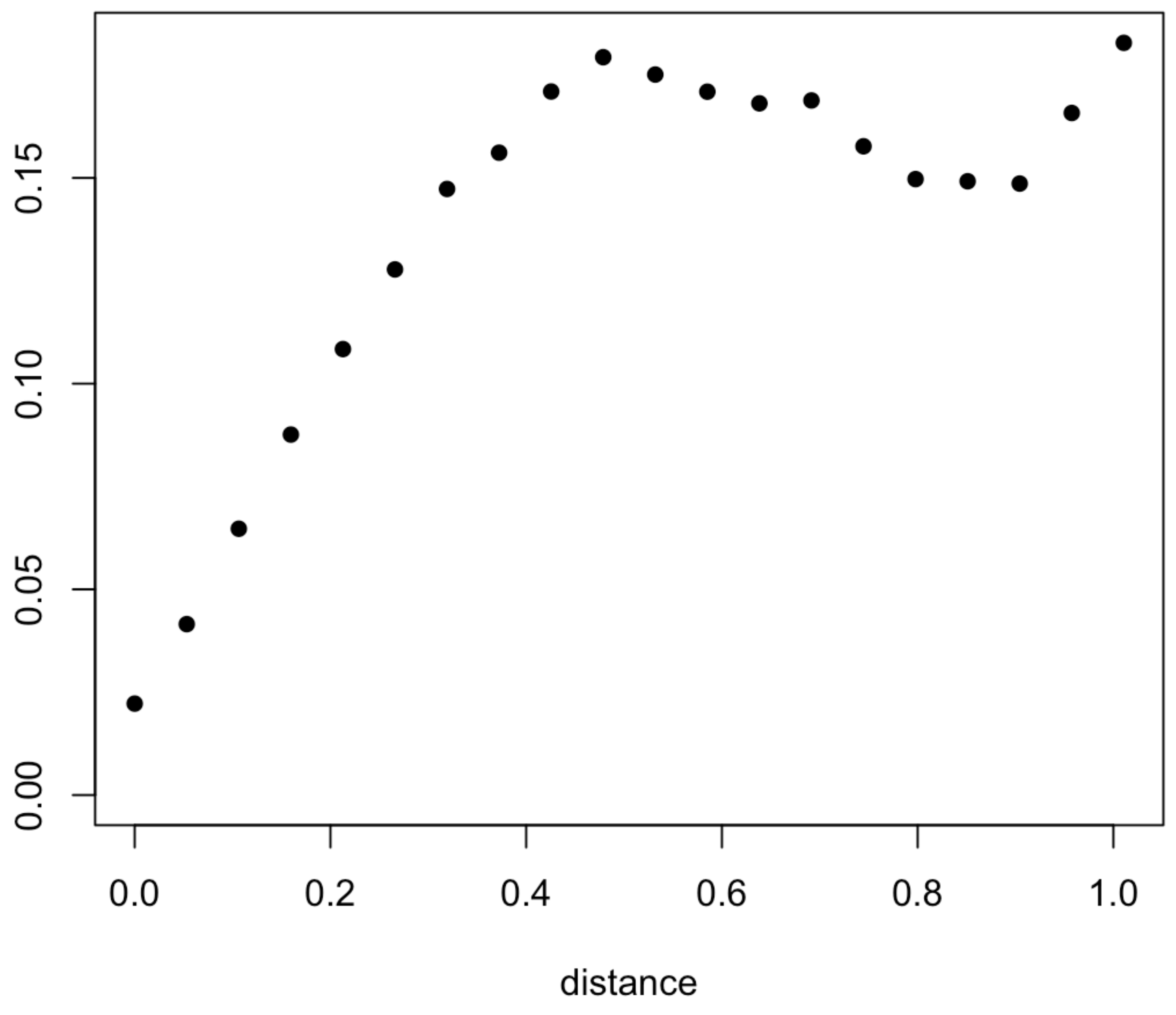


$X(s)$

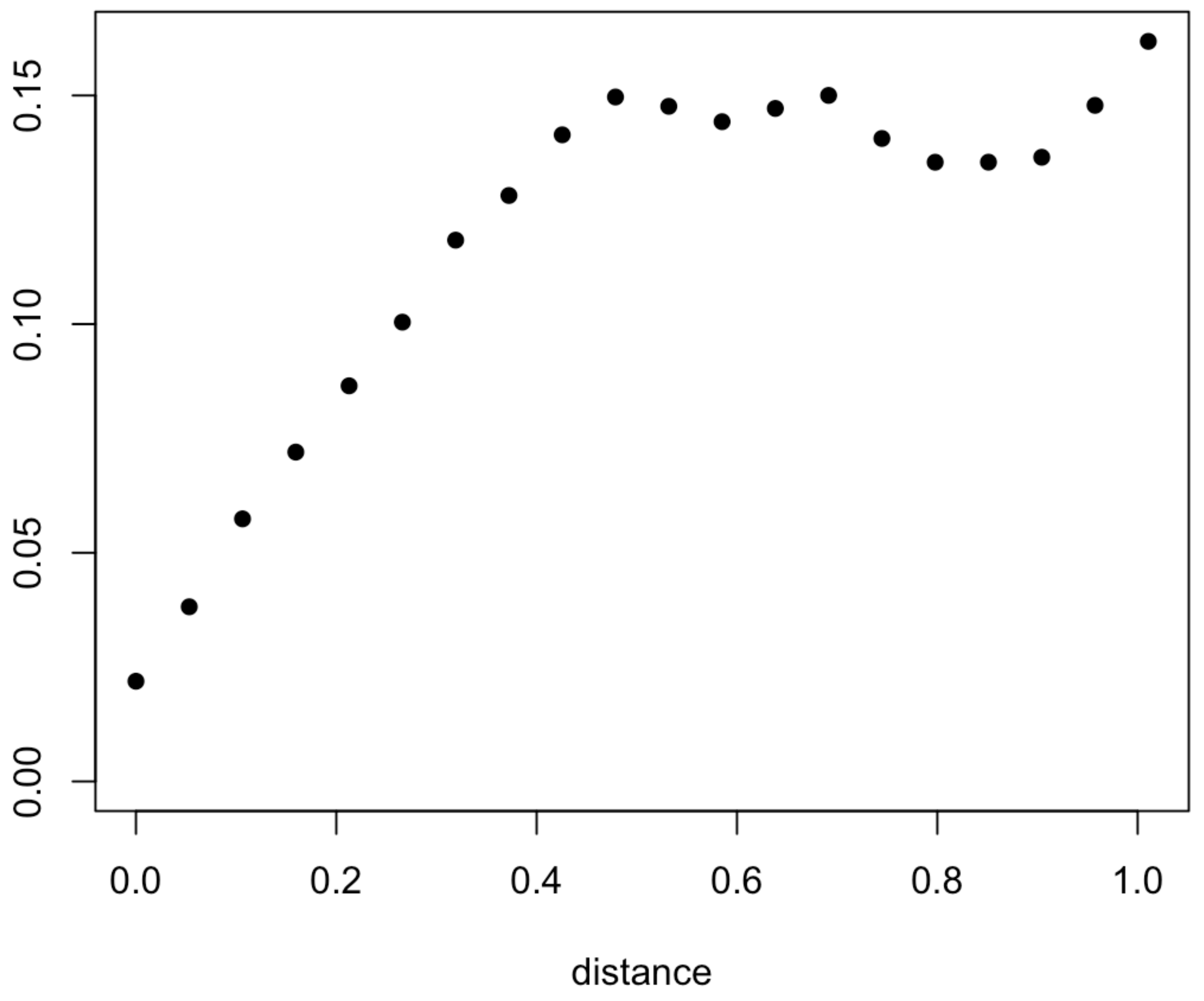


Residuals $Y(s) - \hat{\beta}X(s)$

Dataset 2



Variogram of $Y(s)$



Variogram of residuals
$Y(s) - \hat{\beta}X(s)$

Variogram of residuals suggests residual spatial variation

Spatial linear mixed effect models (SLMM)

When purely covariate based models does not suffice, one needs to leverage the information from locations

SLMM: $Y(s_i) = X(s_i)'\beta + w(s_i) + \epsilon(s_i)$

Linear fixed effect

Spatial random effect

iid random errors

$w(s_i)$ is introduced to model spatial patterns in $Y(s_i)$ that is not explained by $X(s_i)$

Process-level model

Usually goal is predicting $Y(s)$ at any location s in the domain D

E.g., Conceptually pollutant level exists at all possible sites

SLMM: $Y(s) = X(s)'\beta + w(s) + \epsilon(s)$ for any location $s \in D$

Need to model $w(s)$ as a smooth function or stochastic process over D

Many approaches to model and estimate $w(s)$: basis function expansions, penalized regression splines, Gaussian Processes

Gaussian processes

$w(s)$ is often modeled as a Gaussian Process (GP)

$$w(\cdot) \sim GP(0, C(\cdot,\cdot))$$

$$w = w(S) = (w(s_1), \ldots, w(s_n))' \sim N(0, C)$$

$$C_{ij} = Cov(w(s_i), w(s_j)) = C(s_i, s_j)$$

Gaussian processes

$$w(\cdot) \sim GP(0, C(\cdot, \cdot)),$$

$$w = w(S) = (w(s_1), \ldots, w(s_n))' \sim N(0, C)$$

The covariance function models the spatial dependence

Stationarity and Isotropy: $C(s_i, s_j) = C(||h||)$ where $h = s_i - s_j$

Gaussian processes

$$w(\cdot) \sim GP(0, C(\cdot, \cdot)),$$

$$w = w(S) = (w(s_1), \ldots, w(s_n))' \sim N(0, C)$$

Matérn covariance function: A common, flexible family of covariances C specified using a spatial variance σ^2, spatial decay ϕ, and smoothness ν

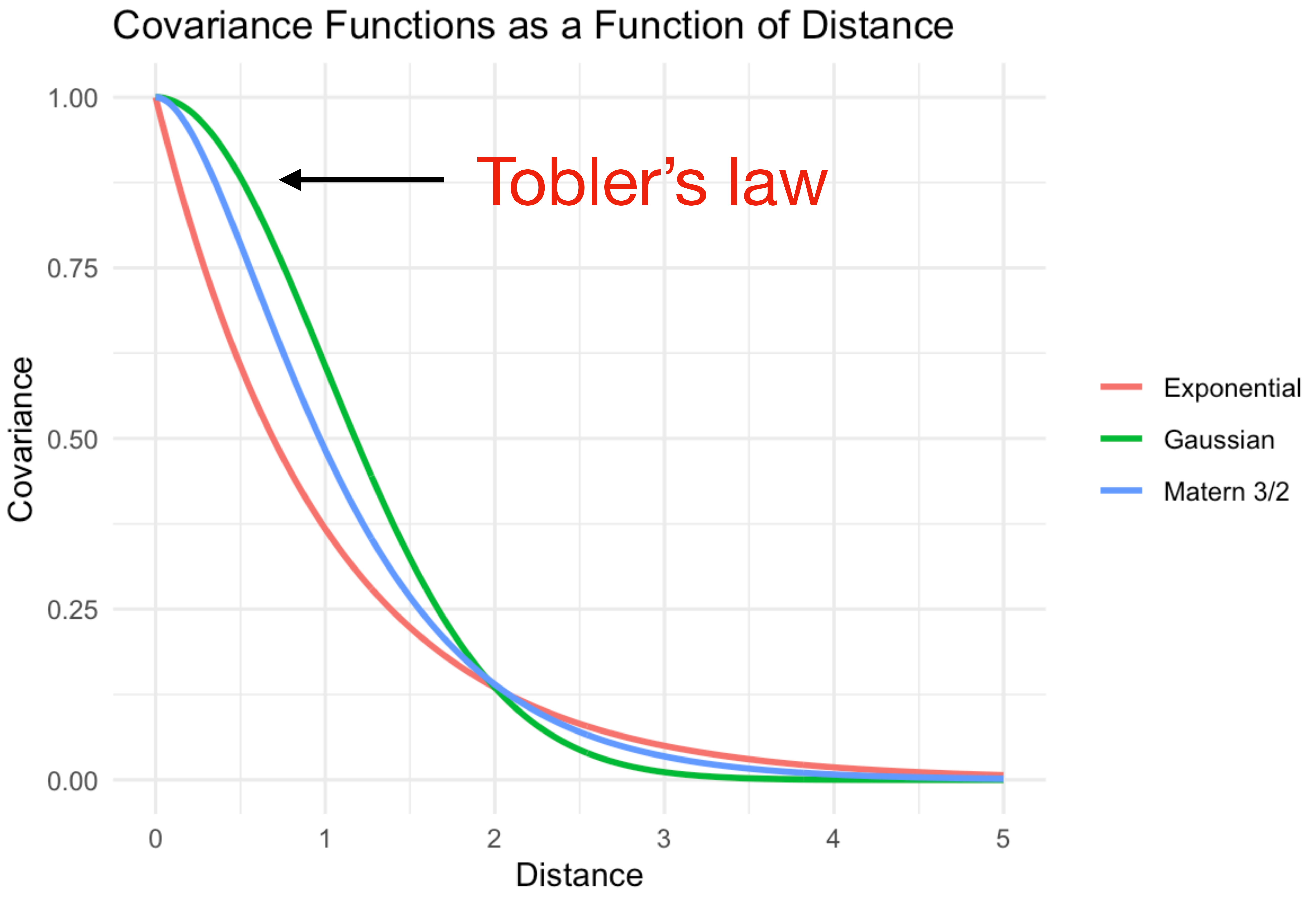
$\nu = 1/2$ (Exponential covariance): $C(\|h\|) = \sigma^2 \exp(-\phi\|h\|)$

$\nu = 3/2$: $C(\|h\|) = \sigma^2(1 + \phi\|h\|)\exp(-\phi\|h\|)$

$\nu = \infty$ (Gaussian covariance): $C(\|h\|) = \sigma^2 \exp(-\phi\|h\|^2)$

Gaussian processes

Matérn covariance function:



Gaussian processes

$$w(\cdot) \sim GP(0, C(\cdot, \cdot)),$$

$$w = w(S) = (w(s_1), \ldots, w(s_n))' \sim N(0, C)$$

Estimation of covariance parameters using Gaussian likelihood maximization

Process-level modeling: $w(s)$ is defined for any s in a region R

Allow predictions at any location s_0 via *kriging*

$$\begin{pmatrix} w(s_0) \\ w \end{pmatrix} \sim N\left(0, \begin{bmatrix} C(s_0, s_0) & C(s_0, S) \\ C(S, s_0) & C \end{bmatrix}\right)$$

$$w(s_0) \mid w \sim N(\mu(s_0), v(s_0))$$

Gaussian processes

$$w(\cdot) \sim GP(0, C(\cdot, \cdot)),$$

$$w = w(S) = (w(s_1), \ldots, w(s_n))' \sim N(0, C)$$

Flexibility and robustness: $w(s)$ for suitable covariance functions C can non-parametrically model any smooth fixed function $f(s)$ (van der Vaart 2008, 2011)

Spatial linear mixed effect models (SLMM)

SLMM: $Y(s_i) = X(s_i)'\beta + w(s_i) + \epsilon(s_i),\ i = 1,\ldots,n$

$w(\cdot) \sim GP(0, C(\cdot,\cdot))$, C is Matérn covariance with parameters σ^2, ϕ, ν

$w = w(S) = (w(s_1), \ldots, w(s_n))' \sim N(0, C)$

$\epsilon(s_i) \sim_{\text{iid}} N(0, \tau^2)$, τ^2 is often called the *nugget*

Spatial linear mixed effect models (SLMM)

Marginal model: $Y \sim N(X\beta, C + \tau^2 I)$

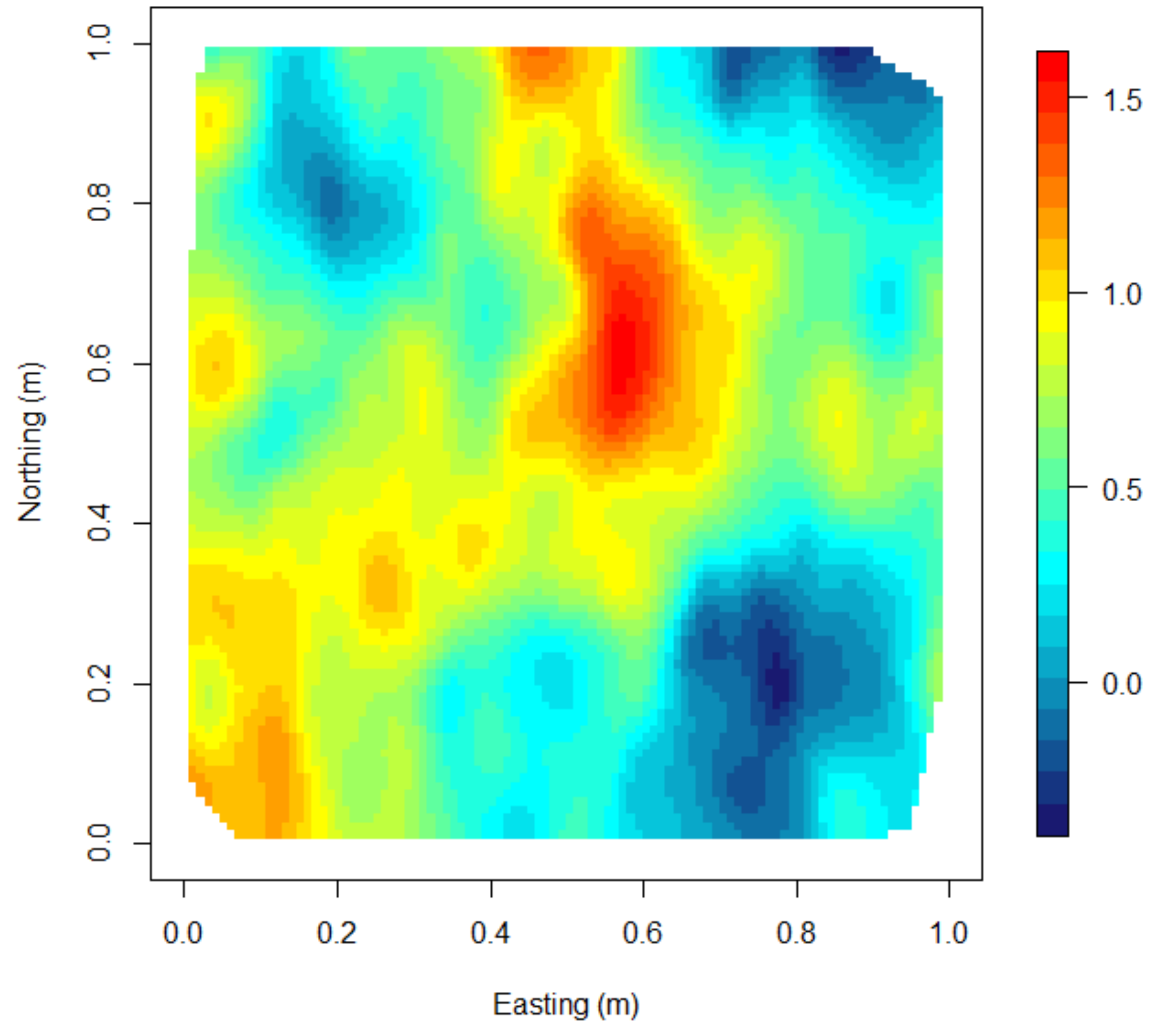
Parameters: Regression coefficient β and covariance parameters $\theta = (\tau^2, \sigma^2, \phi, \nu)$

- Estimates using maximum likelihood estimation (MLE)

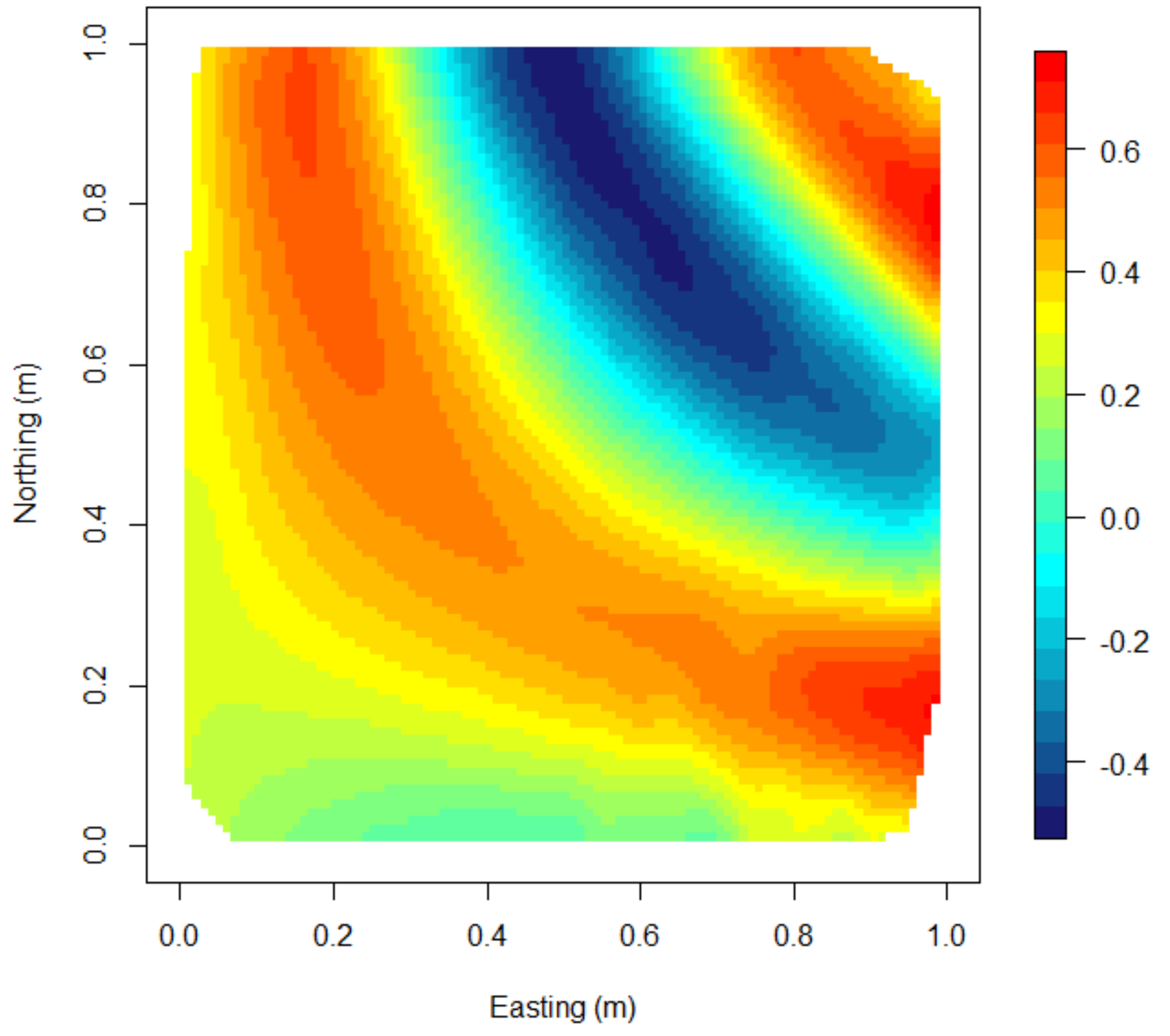
Predictions at a new location using kriging

- Based on the conditional normal distribution of $Y(s_0) \mid Y$

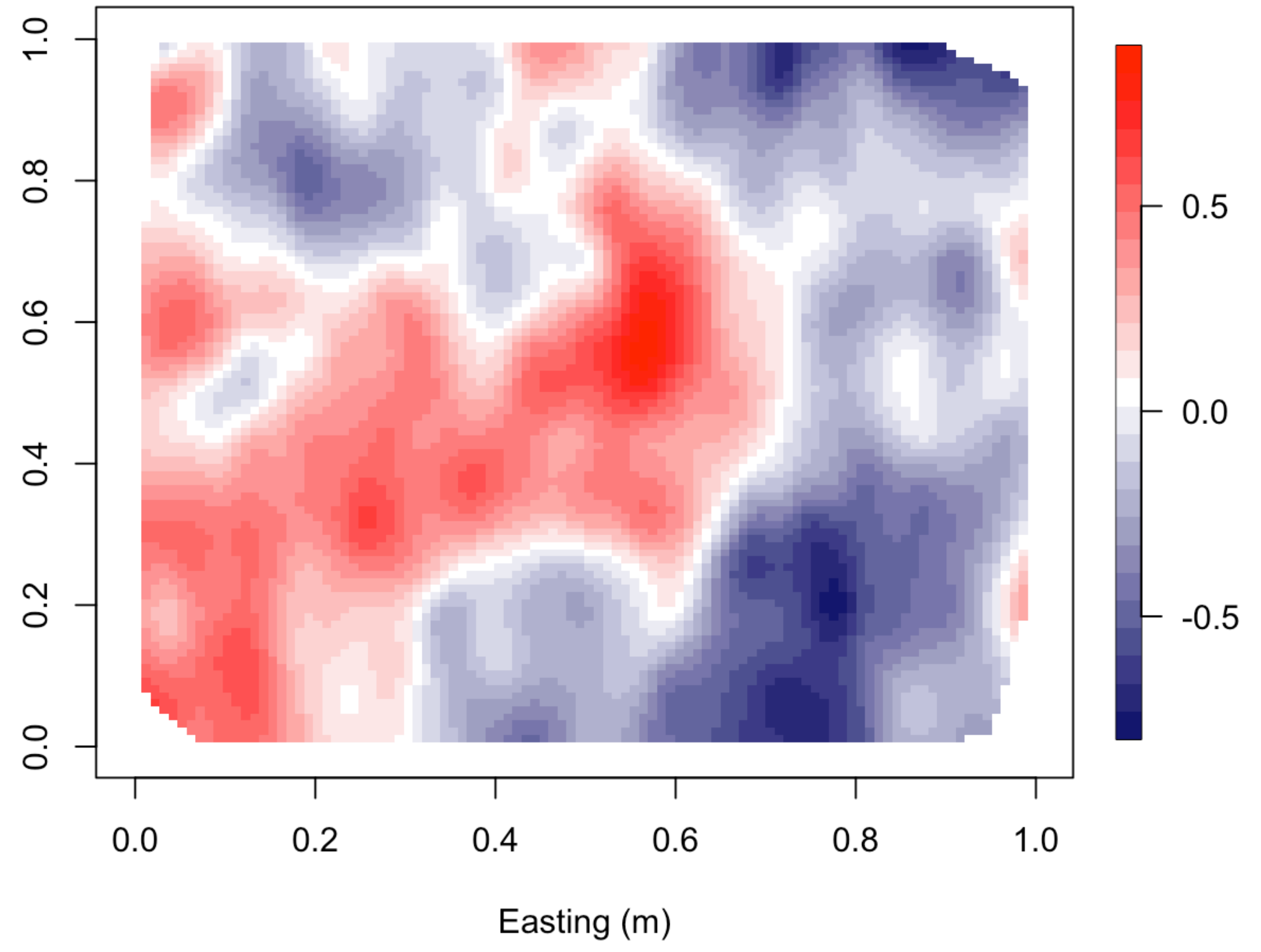
Dataset 2



$Y(s)$



$X(s)$



Residuals $Y(s) - \hat{\beta}X(s)$

Dataset 2

Model: $Y \sim N(X^*\beta^*, C + \tau^2 I), X^* = [1 : X], \beta^* = (\beta_0', \beta_1')'$

Parameters estimated using *likfit* function of geoR package

Parameter	**Value**
β_0	0.68
β_1	-0.50
τ^2	0.01
σ^2	0.14
ϕ [1]	0.26
ν [2]	0.5

[1] In geoR, ϕ is the inverse of our definition of ϕ

[2] Fixed at 0.5, i.e., using the exponential covariance

Dataset 2

Model comparison metrics: Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC)

- Lower values better
- AIC and BIC values are available from the output of likfit

	AIC	**BIC**
Model with Spatial	-146.1	-126.1
Model without Spatial	324.1	336.1

Spatial model is clearly favored

Dataset 2

Prediction: Available from krige.conv function of geoR

Data split:

80% for estimation of parameters (train),

20% for validation of predictions (test)

Dataset 2

Prediction metrics:

Root mean square prediction error (RMSPE) $= \sqrt{\frac{1}{n_{test}} \sum_{i=1}^{n_{test}} (y_i - \hat{y}_i)^2}$

- Compares the point predictions $\hat{y}_i$
- Lower values is better

Dataset 2

Prediction metrics:

Mean coverage probability (CP) of 95% prediction intervals

$$= \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} I(y_i \in (\hat{y}_{i,0.025}, \hat{y}_{i,0.975}))$$

Evaluates the coverage of the interval predictions $(\hat{y}_{i,0.025}, \hat{y}_{i,0.975})$

Ideally should be close to 95%

Otherwise we will have under or over coverage

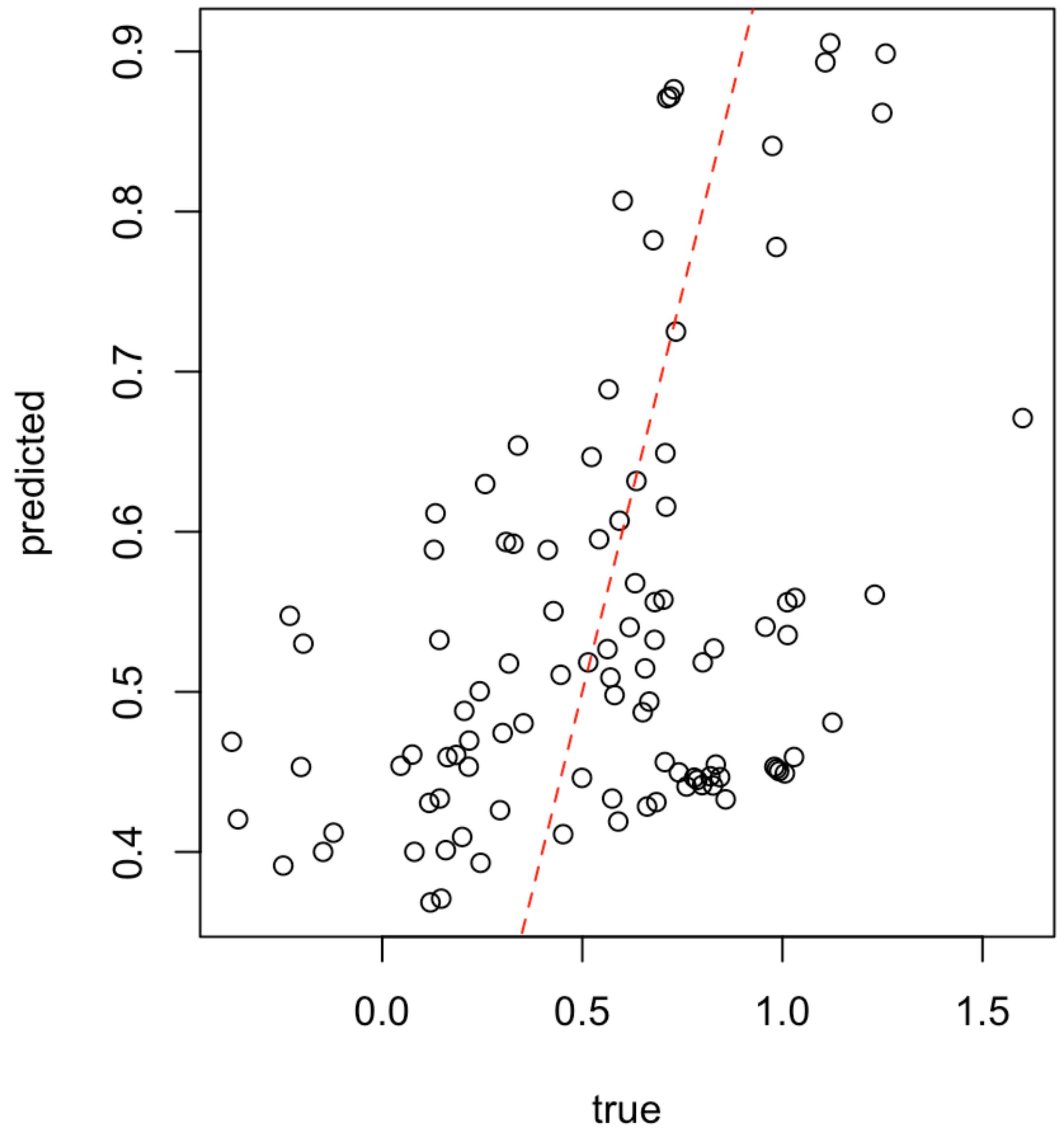
Mean prediction interval width (PIW) $= \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} (\hat{y}_{i,0.975} - \hat{y}_{i,0.025})$

If CP ≈ 0.95, then smaller PIW is better

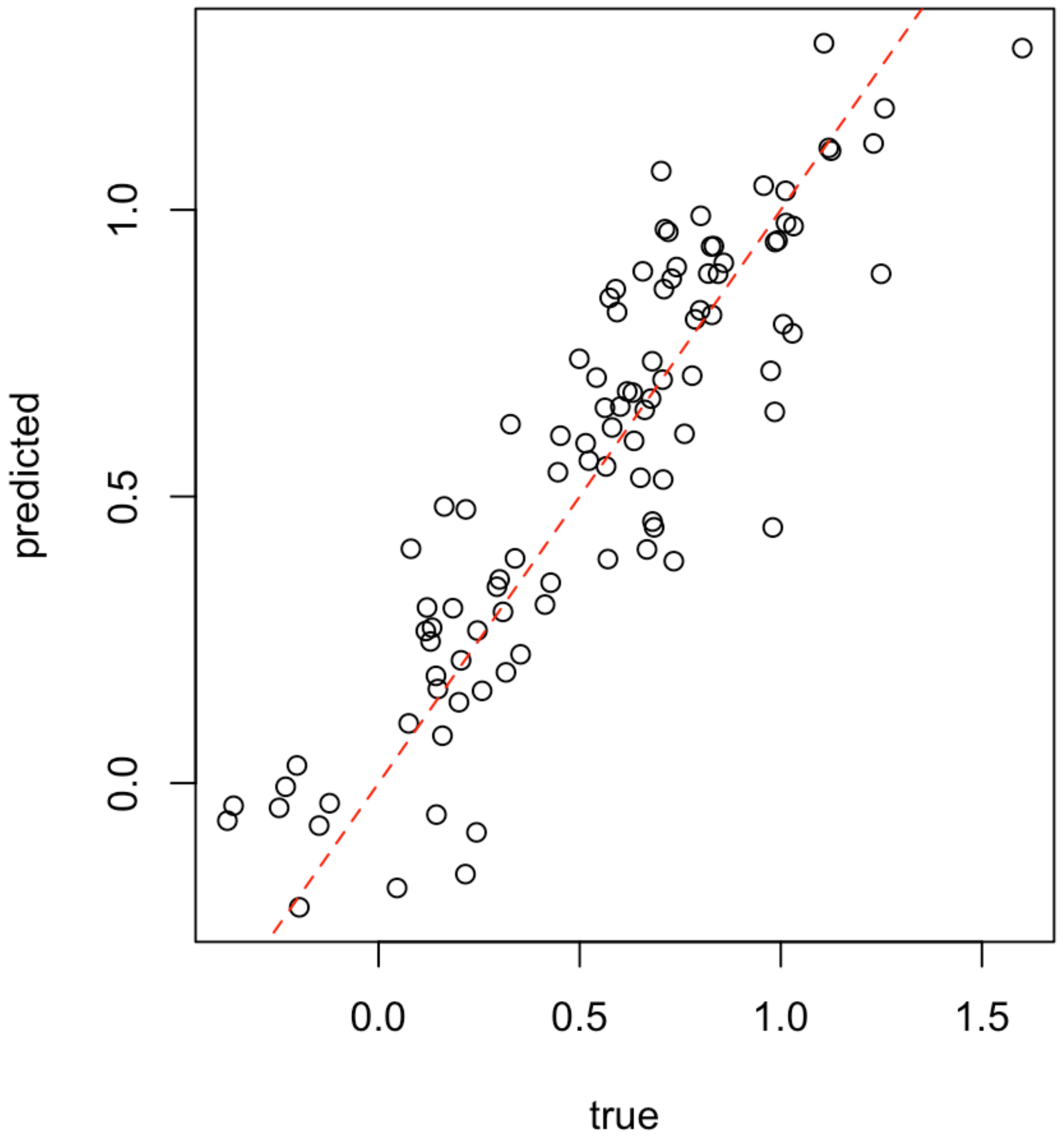
Dataset 2

Prediction:

Metric	Spatial	Non-Spatial
RMSPE	0.18	0.36
CP	0.95	0.95
PIW	0.68	1.42

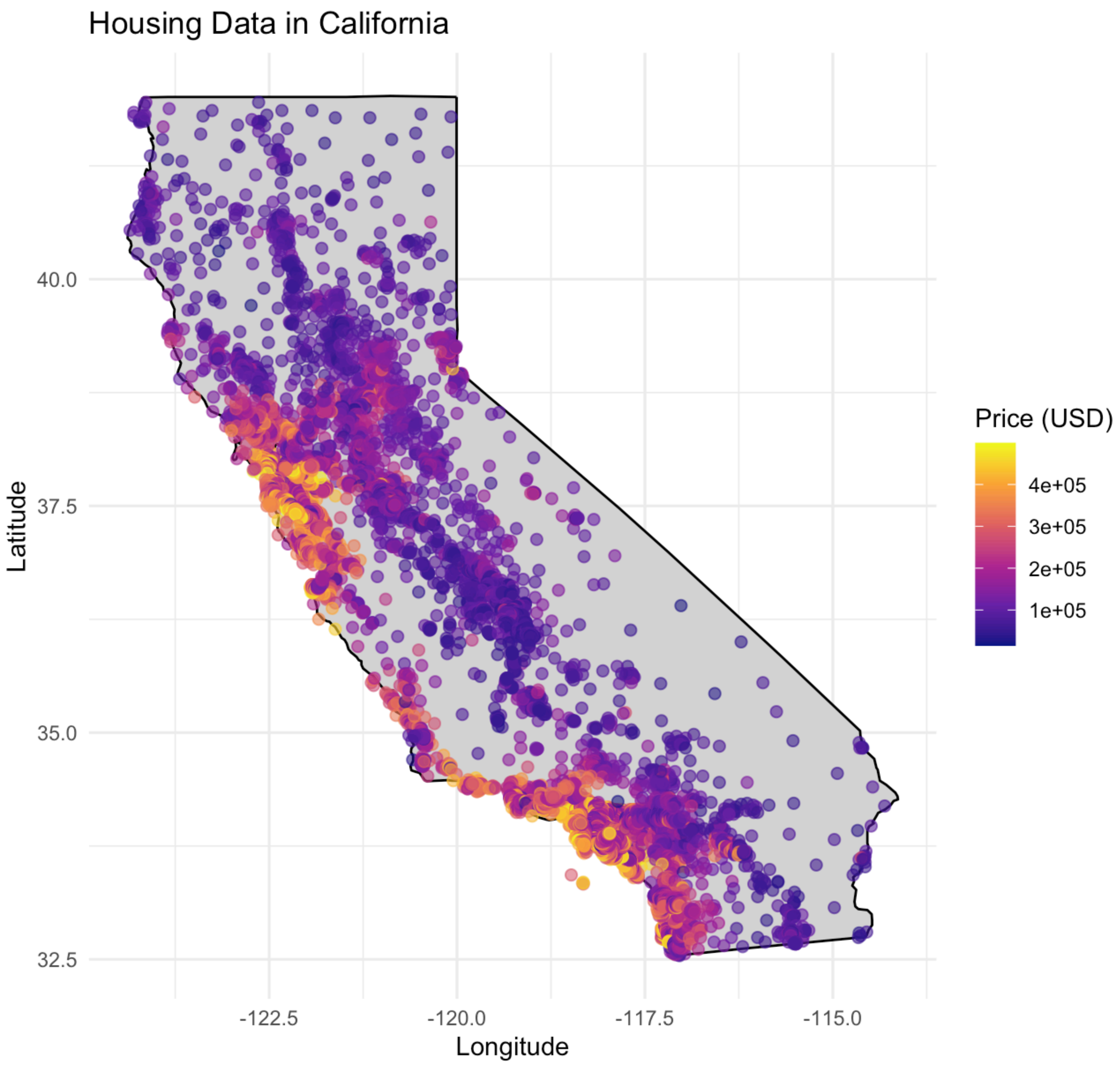


Non-spatial model fit on test data



Spatial model fit on test data

House prices in California



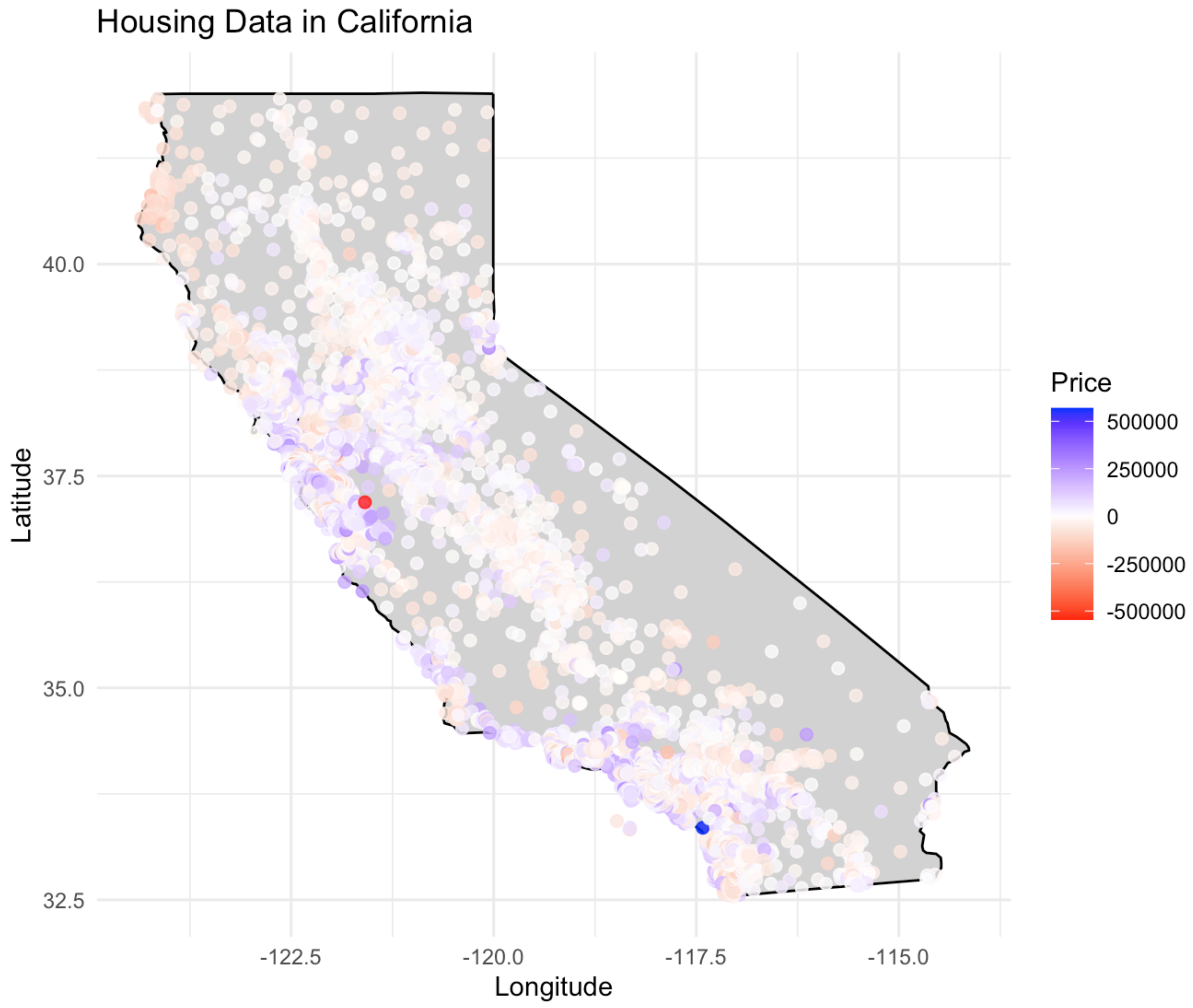
Covariates:

- Median income
- Median house age
- Total rooms
- Total bedrooms
- Population
- Number of Households
- Ocean proximity

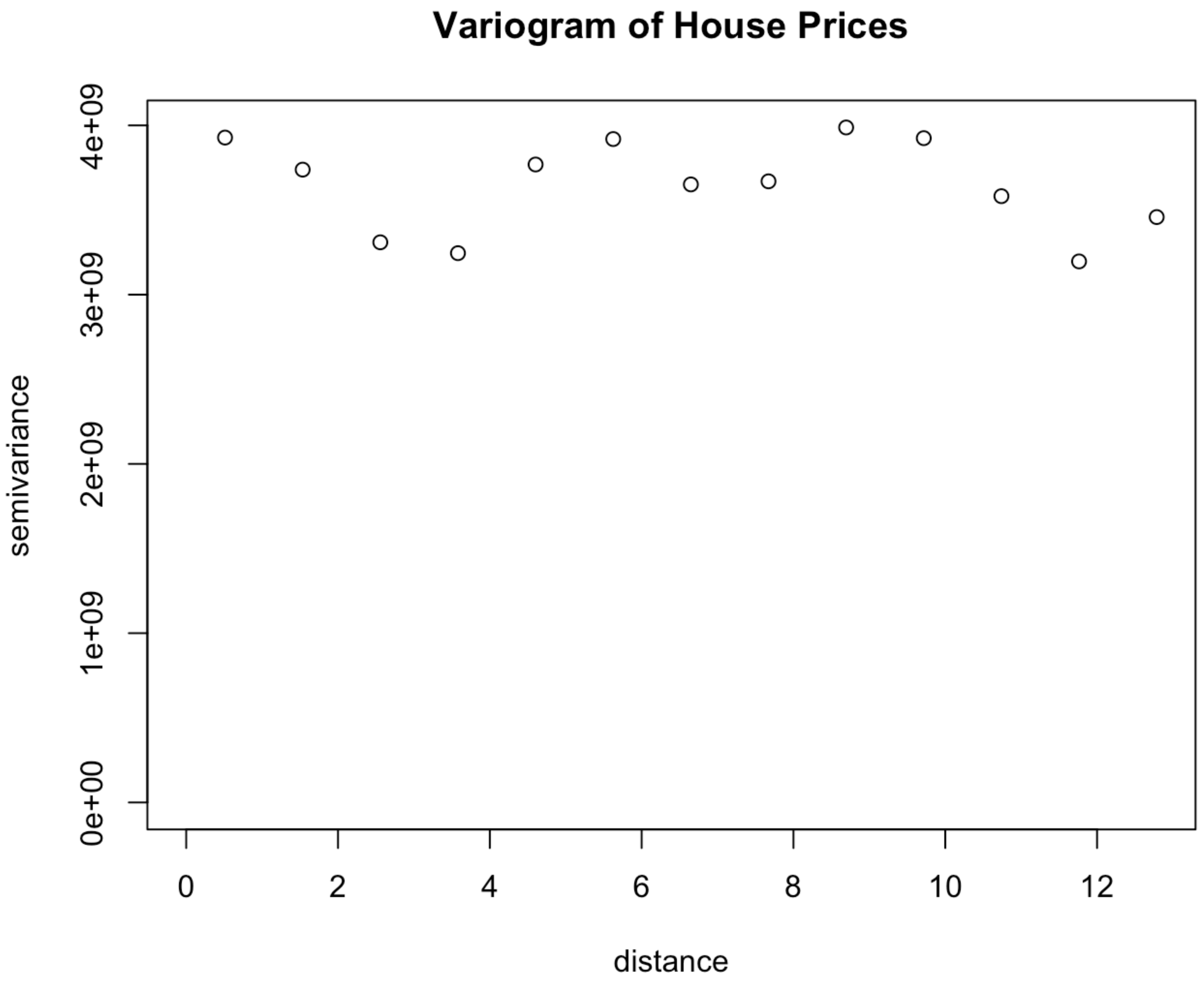
Data available on Kaggle.com

House prices in California

Linear model analysis



Map of residuals



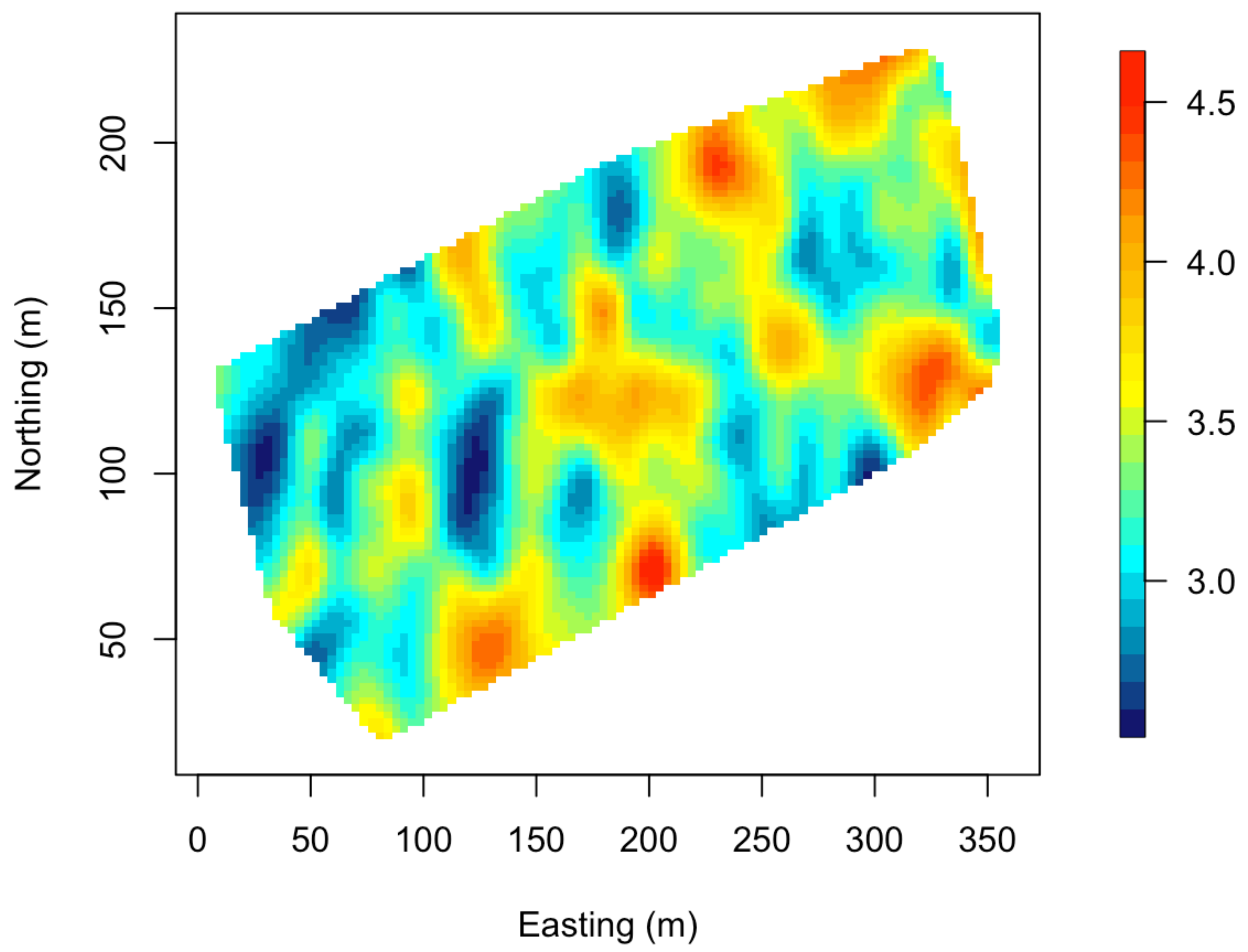
Variogram of residuals

Western Experimental Forestry (WEF) data

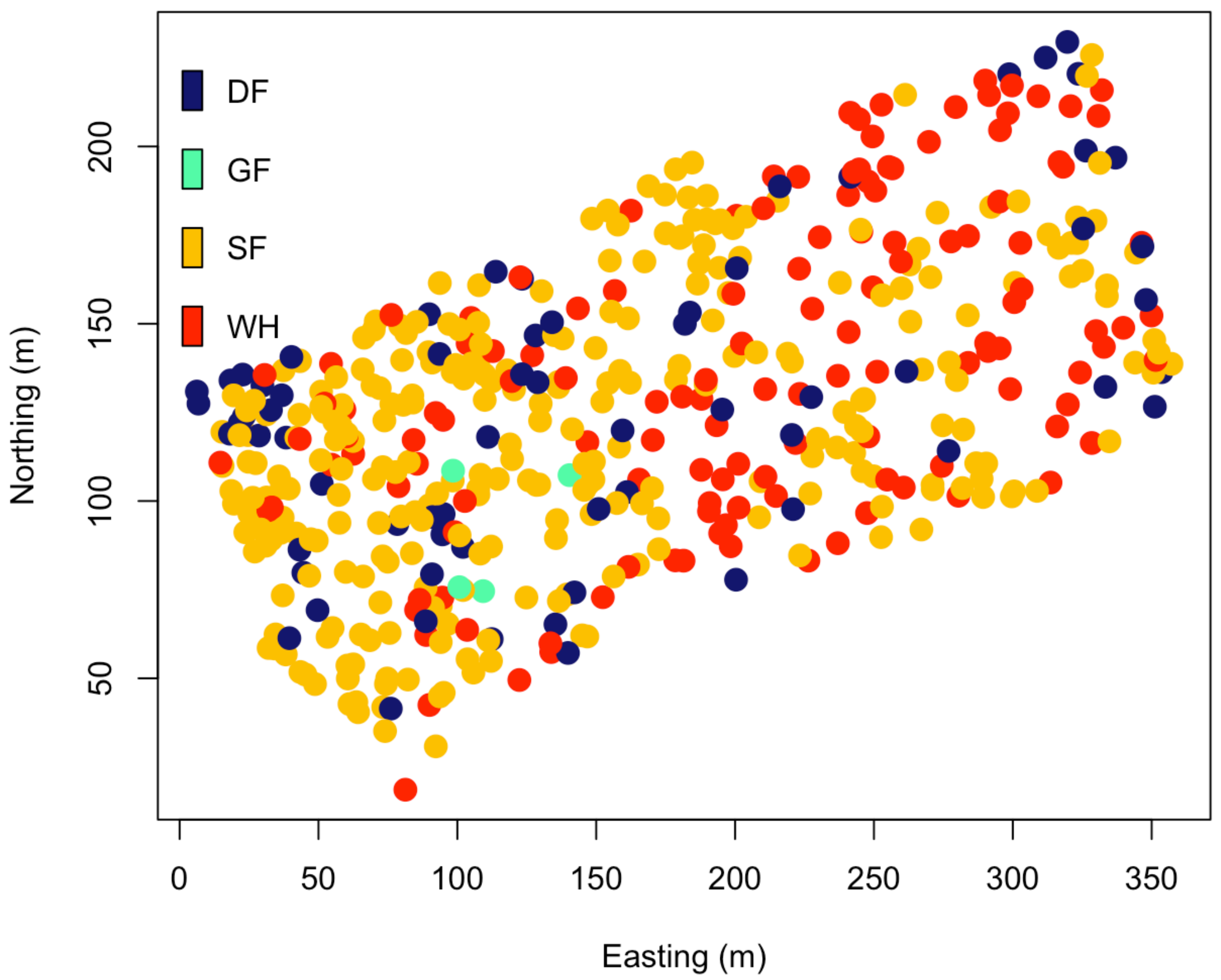
Data from *spBayes* package on census of all trees in a 10 ha. stand in Oregon

Response of interest: log(Diameter at breast height), i.e., log(DBH)

Covariate: Tree species (Categorical variable based on 4 species)



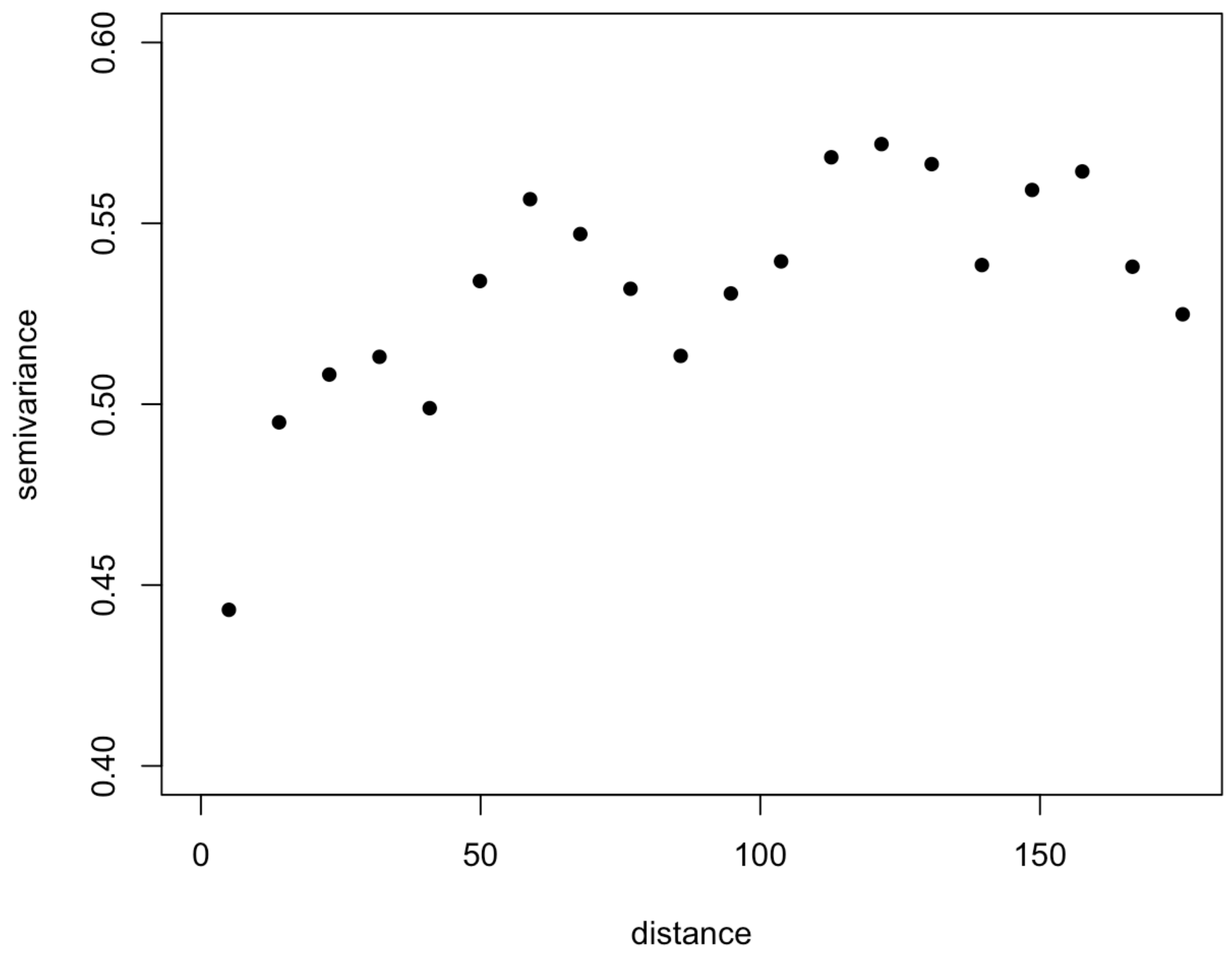
log(DBH)



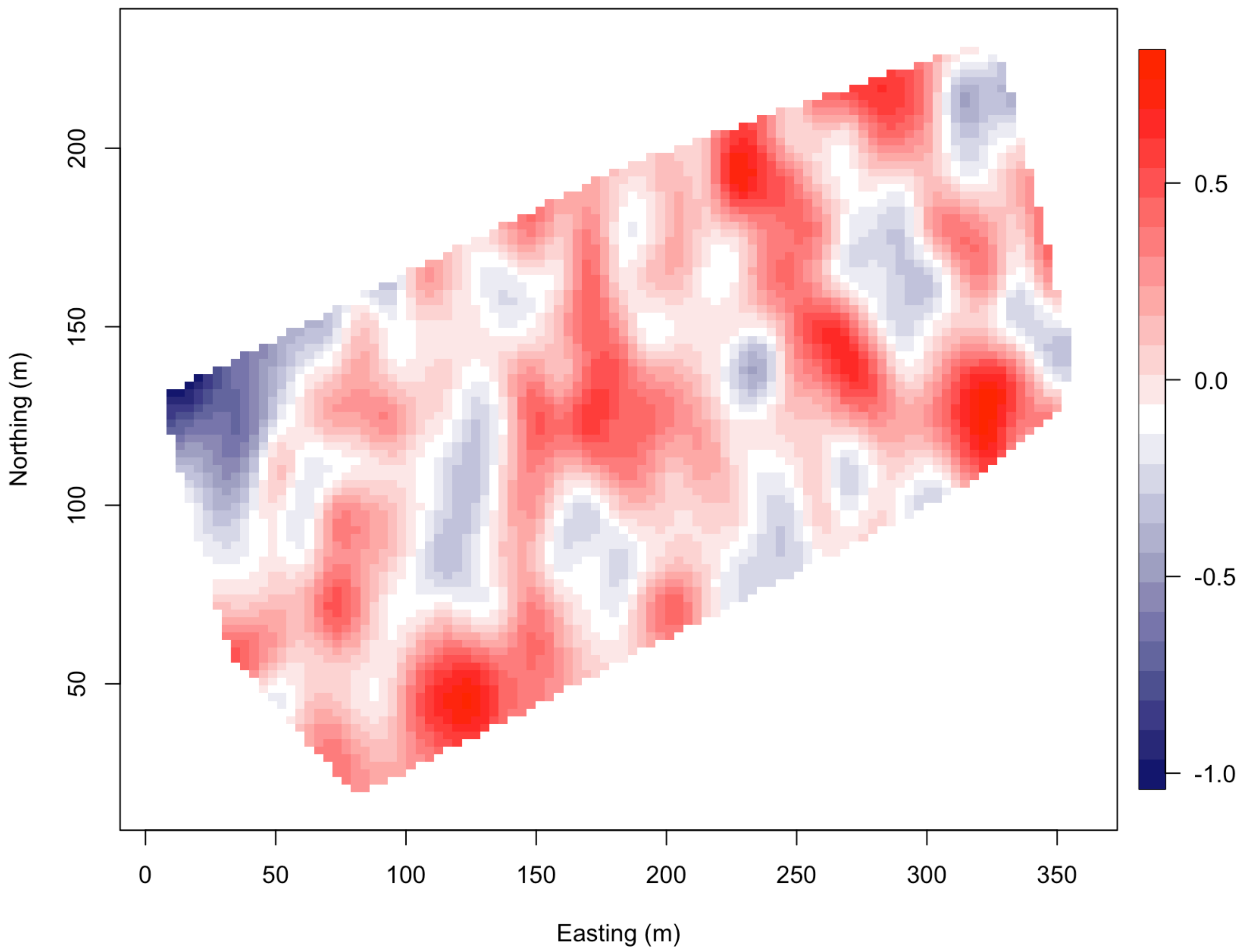
Species

Western Experimental Forestry (WEF) data

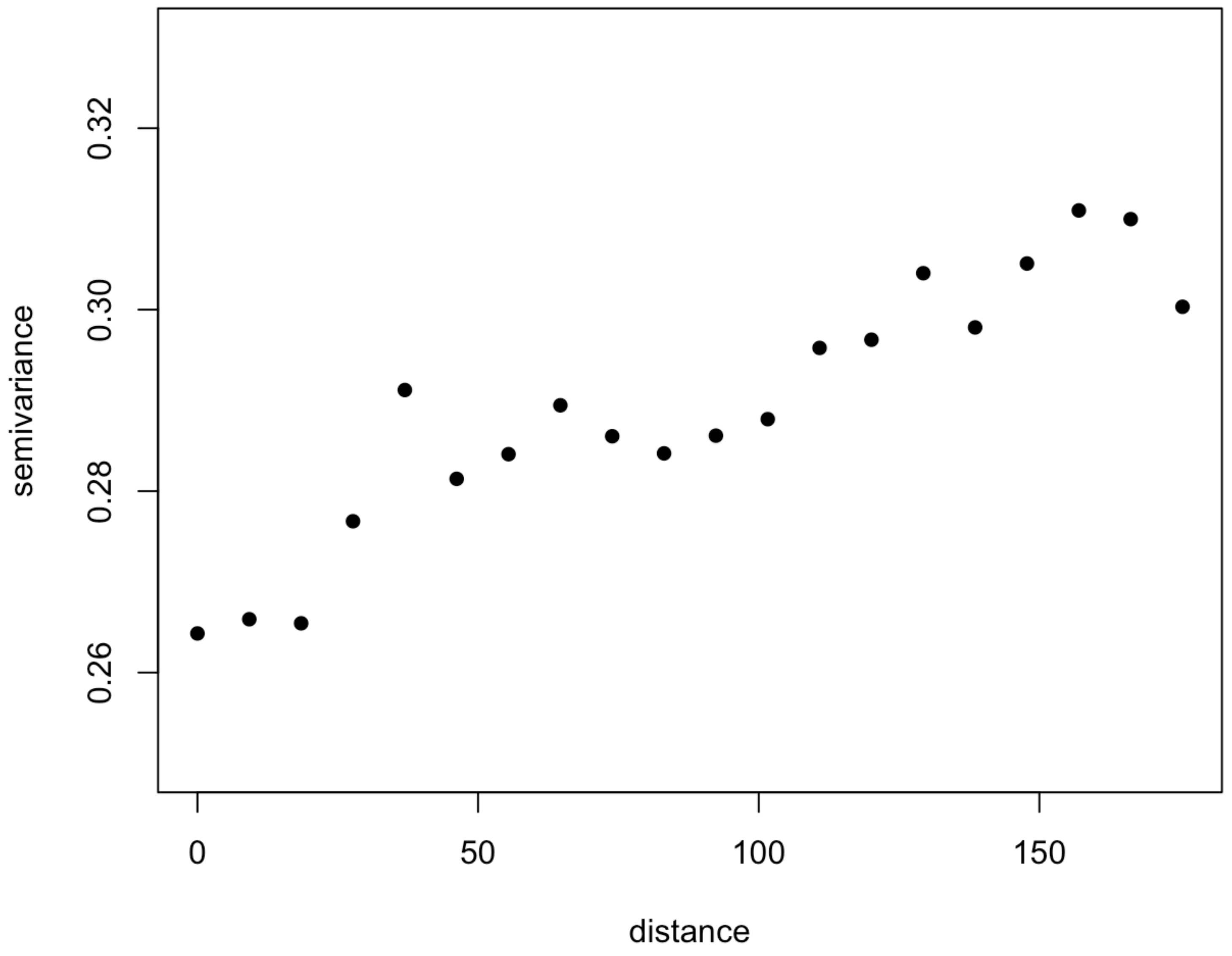
Linear model analysis:



Variogram of log(DBH)



Map of residuals

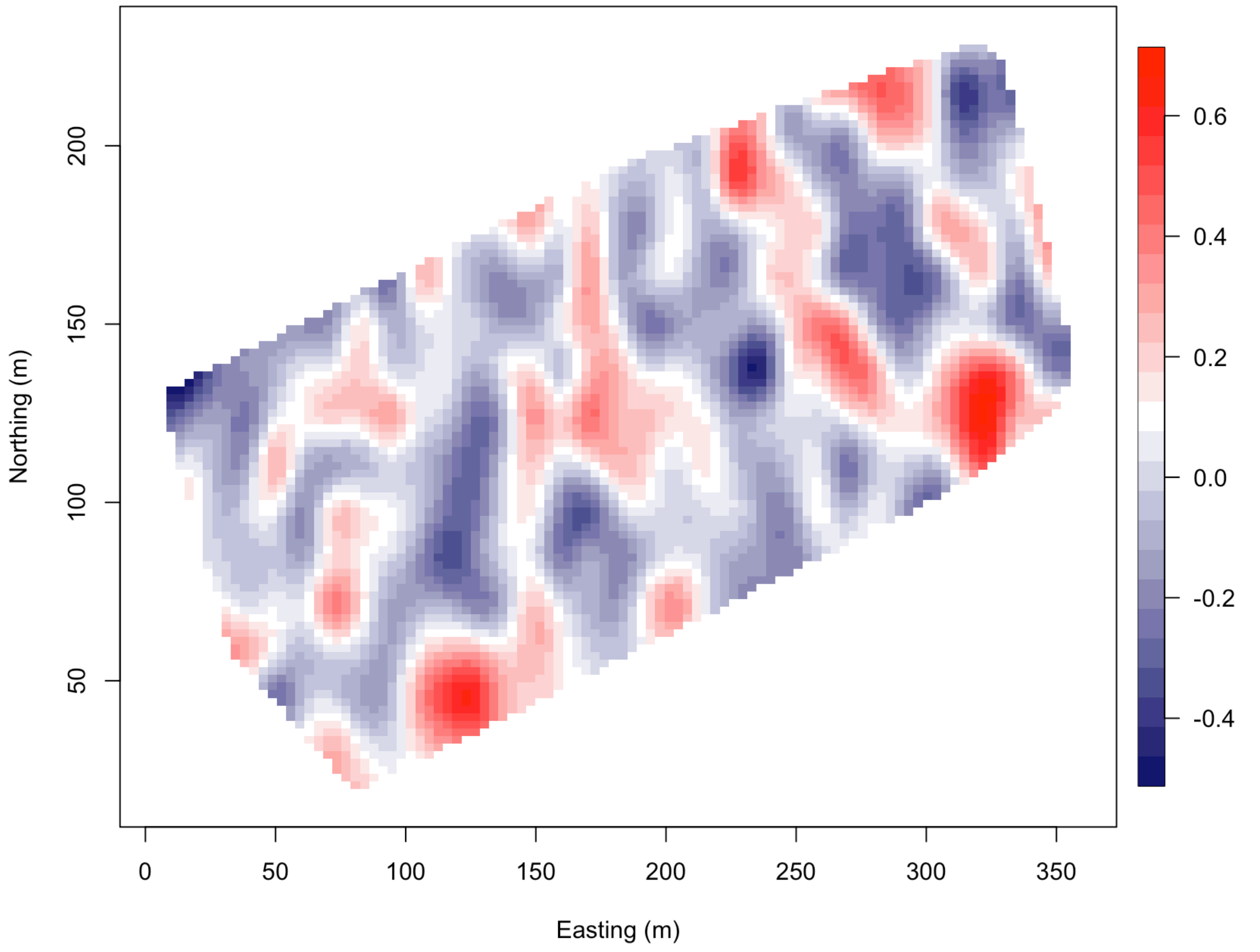


Variogram of residuals

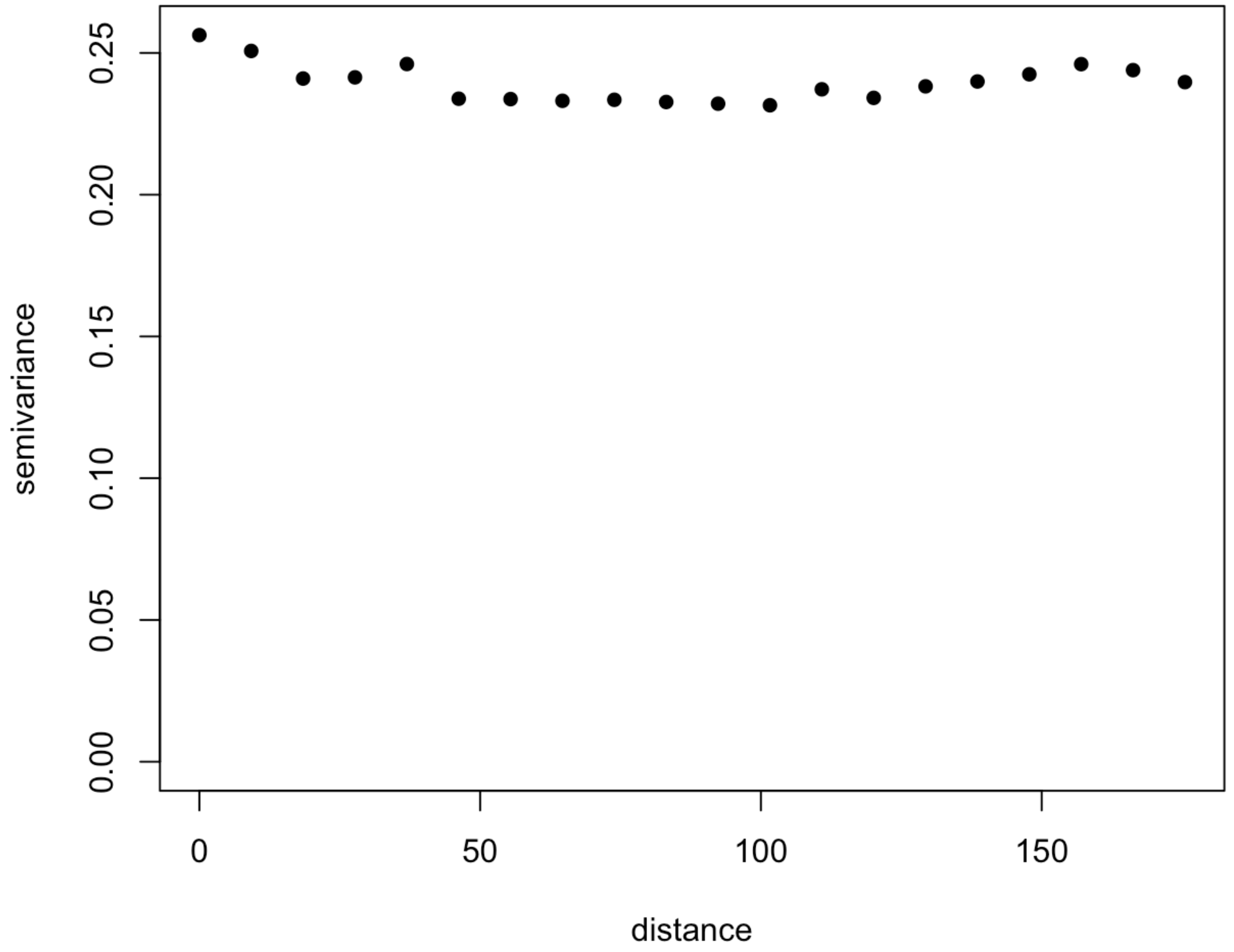
Western Experimental Forestry (WEF) data

Model comparisons:

Metric	Spatial	Non-Spatial
AIC	797	825
BIC	826	846



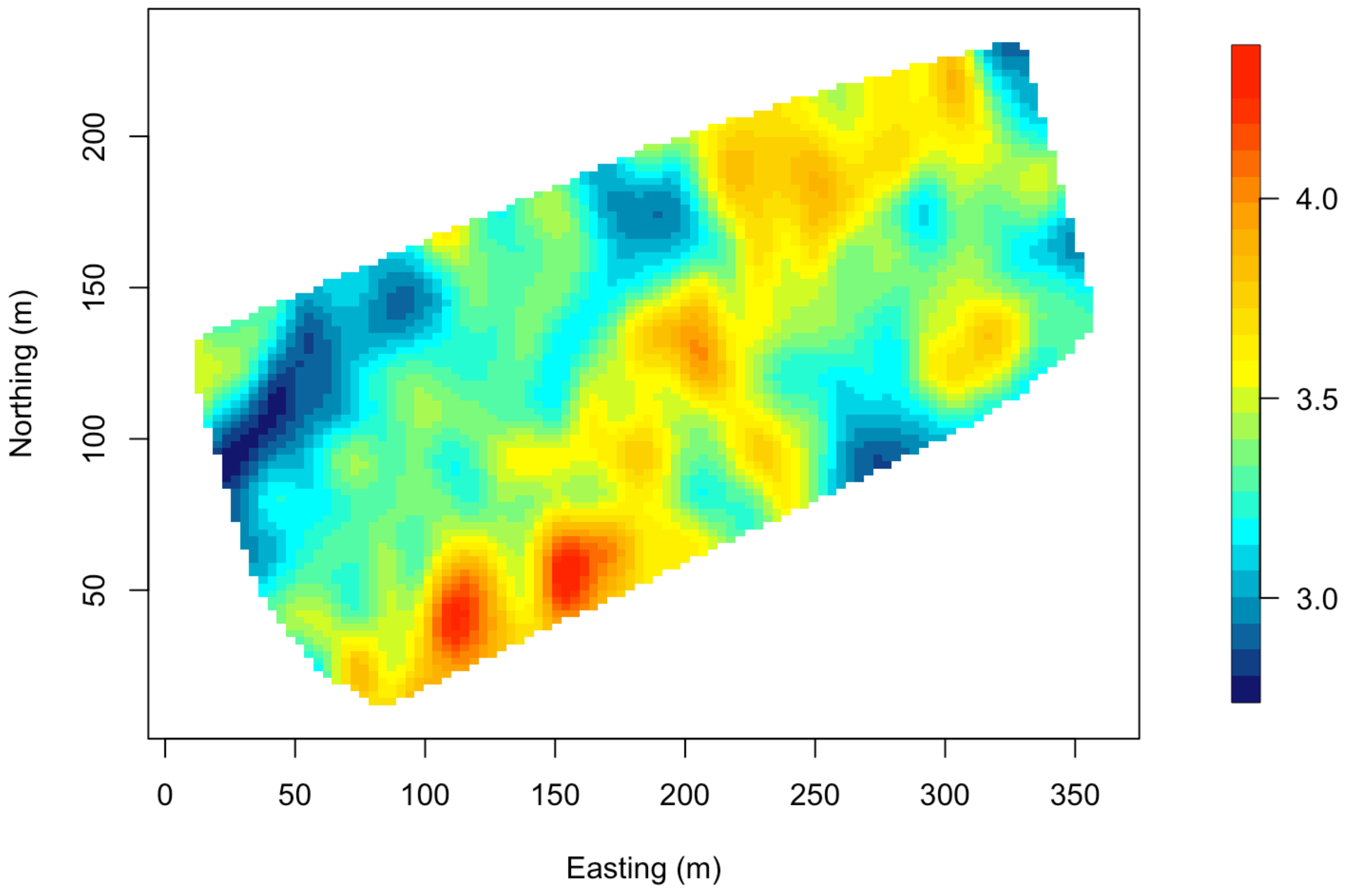
Map of residuals
from the spatial model



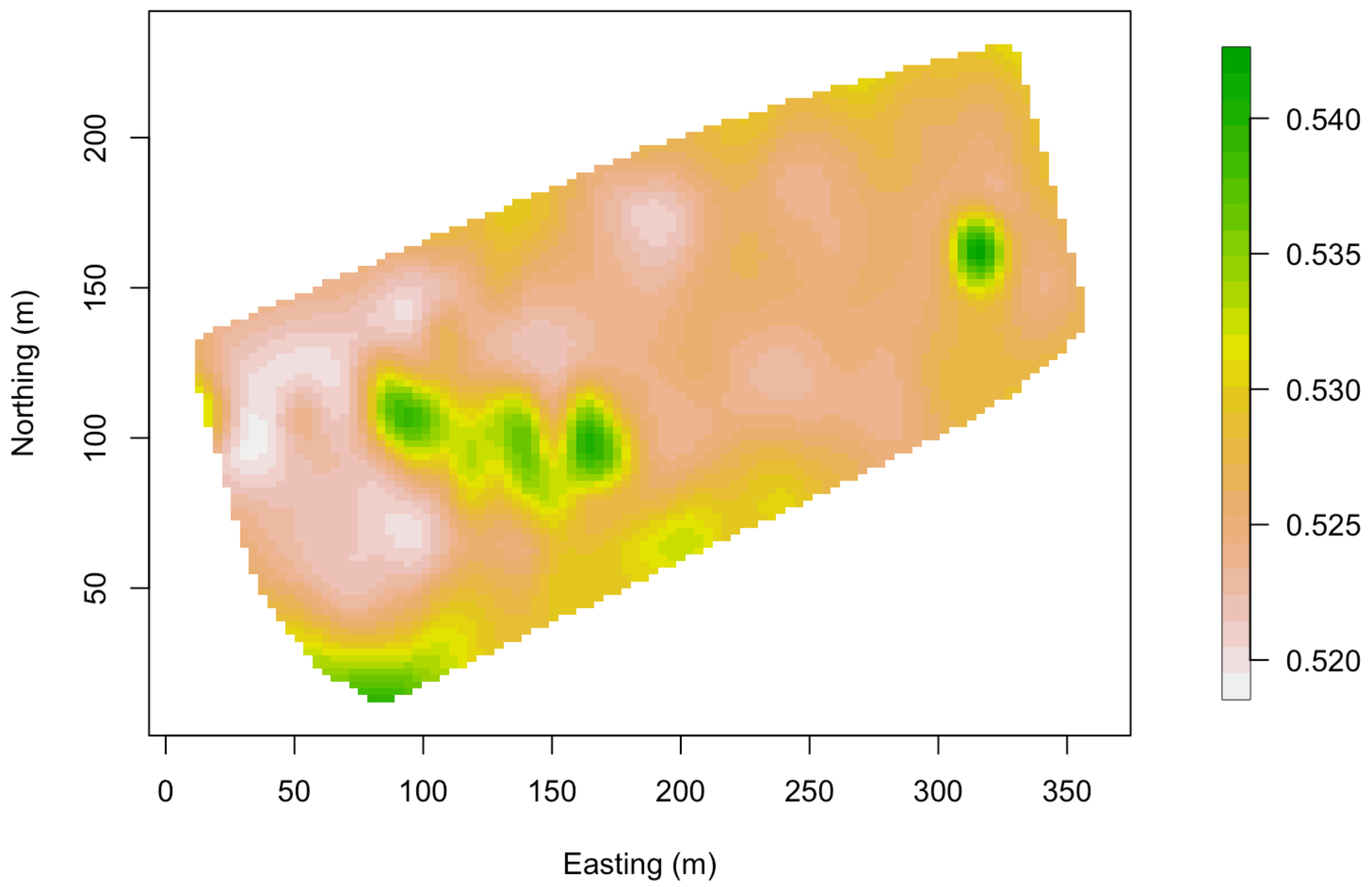
Variogram of residuals
from the spatial model

Western Experimental Forestry (WEF) data

Predictions:



Map of predicted DBH

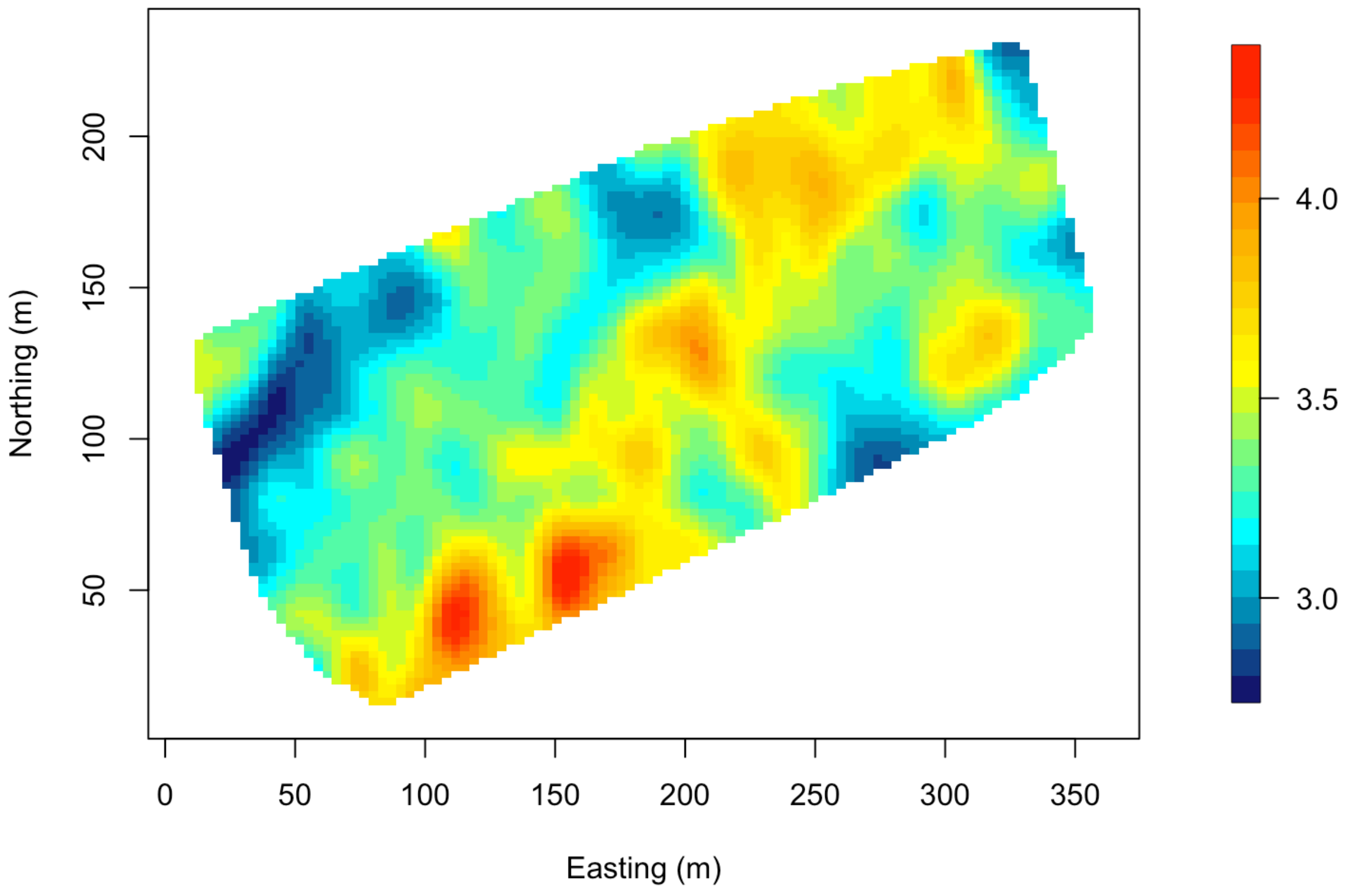


Map of standard deviation of predicted DBH

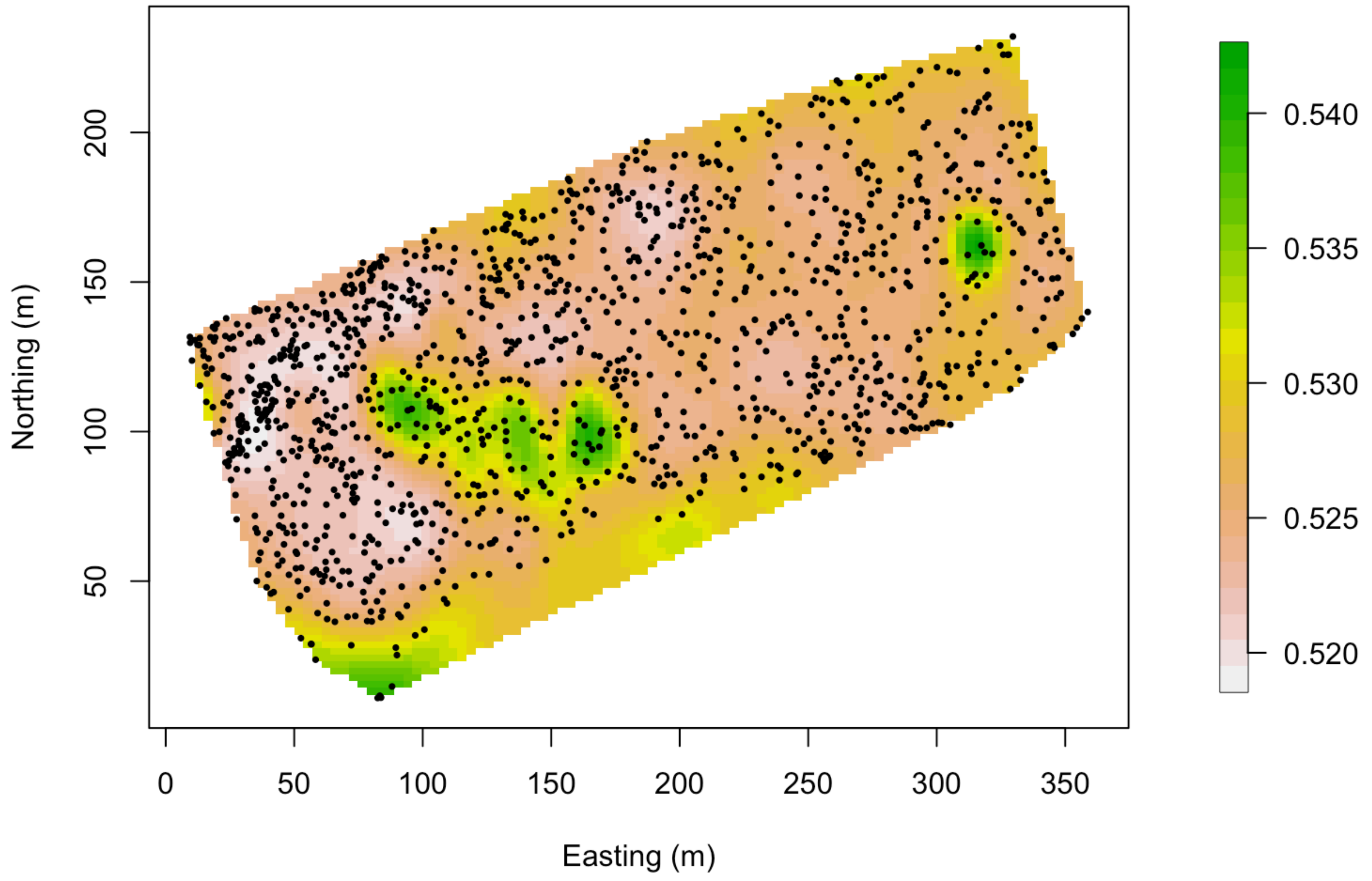
Metric	Spatial	Non-spatial
RMSPE	0.54	0.56
CP	0.96	0.98
PIW	2.1	2.2

Western Experimental Forestry (WEF) data

Predictions:



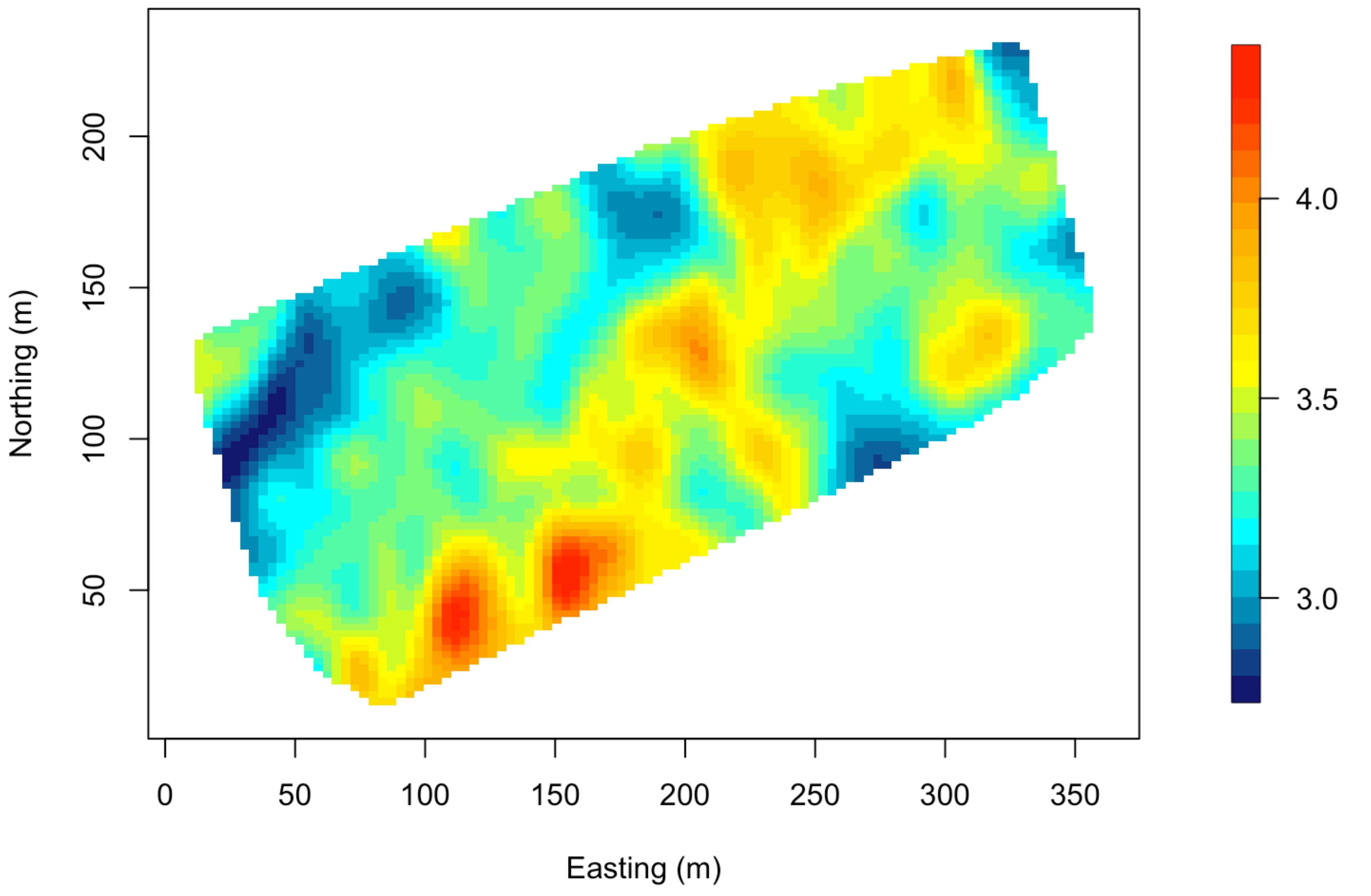
Map of predicted DBH



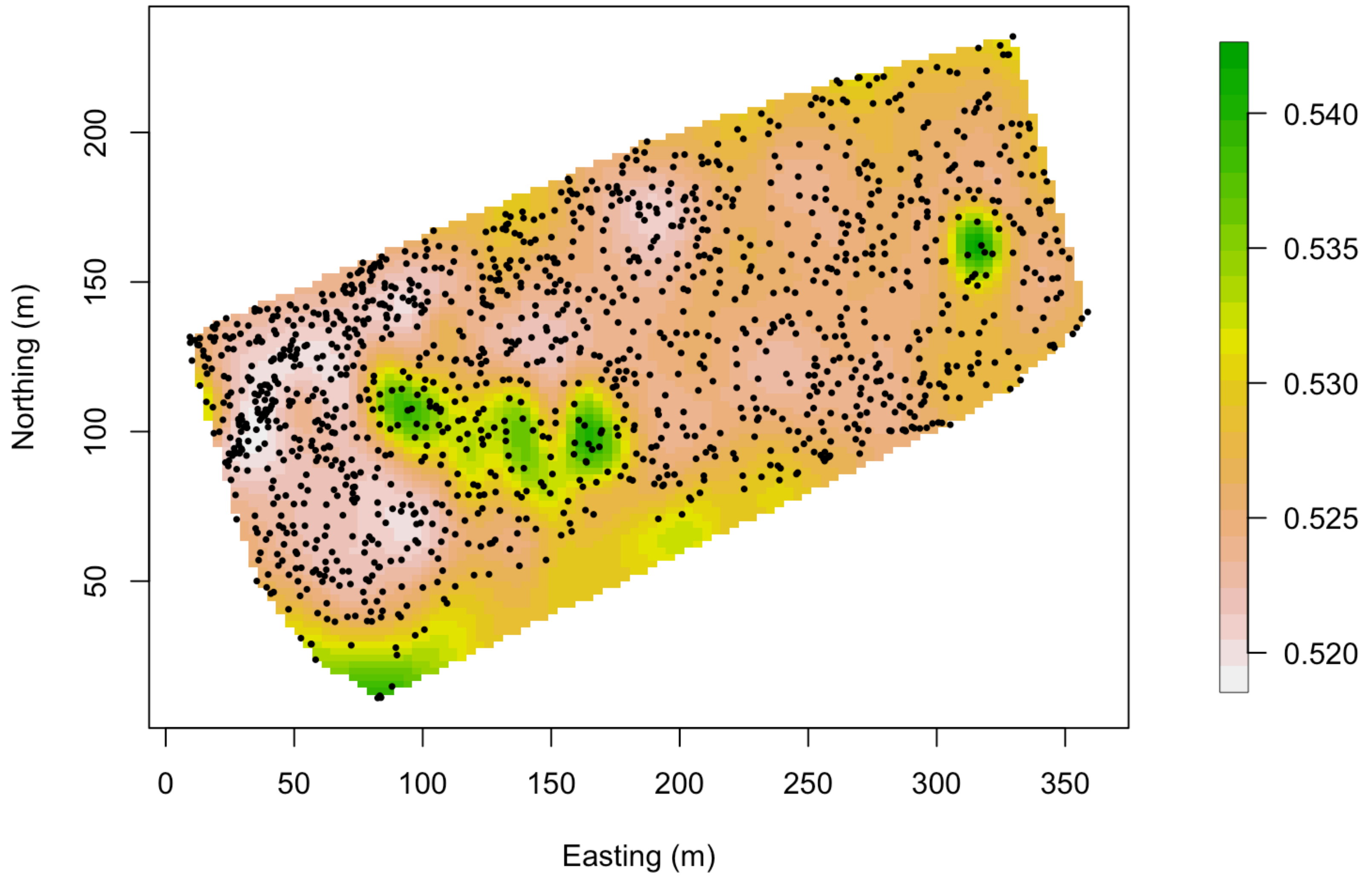
Map of standard deviation of predicted DBH

Western Experimental Forestry (WEF) data

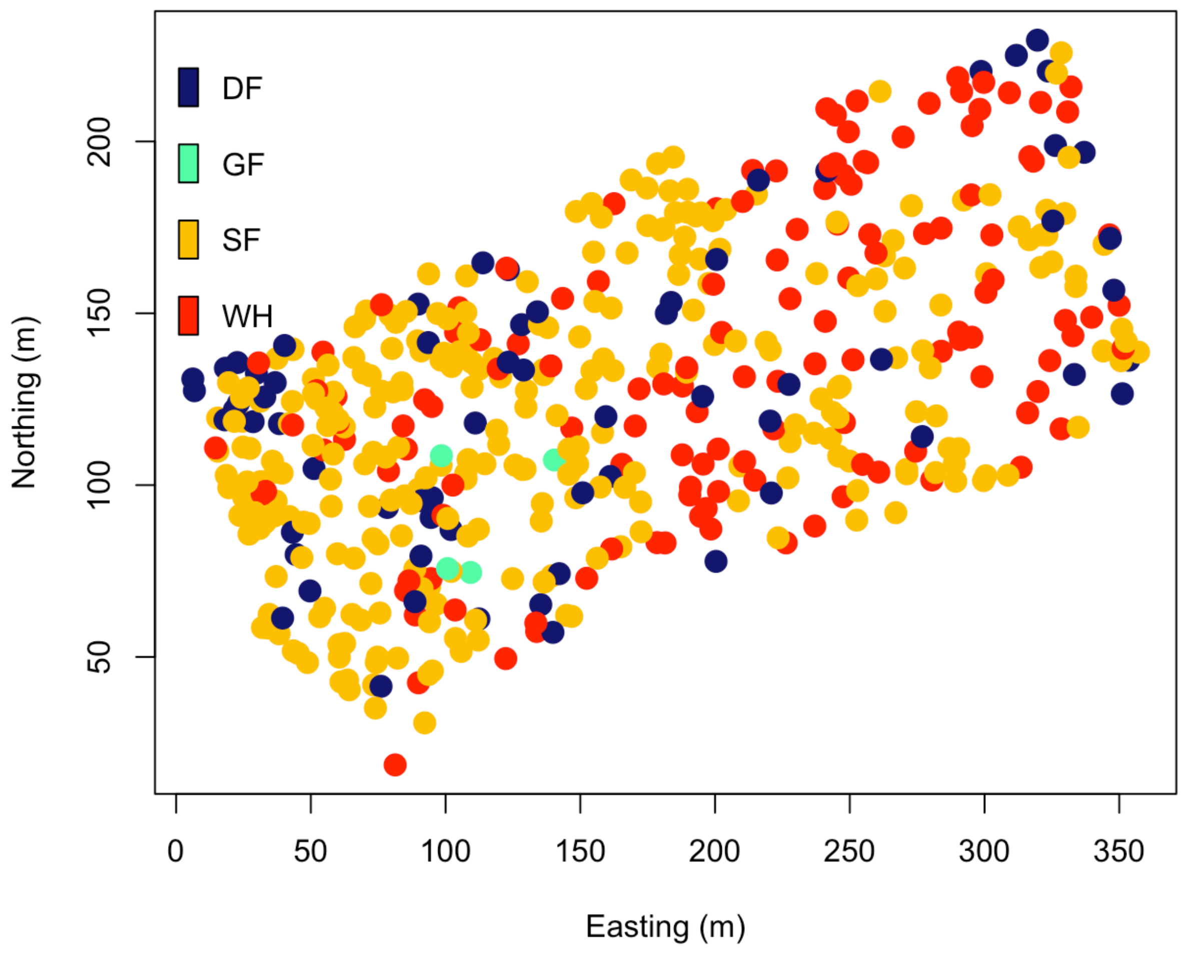
Predictions:



Map of predicted DBH



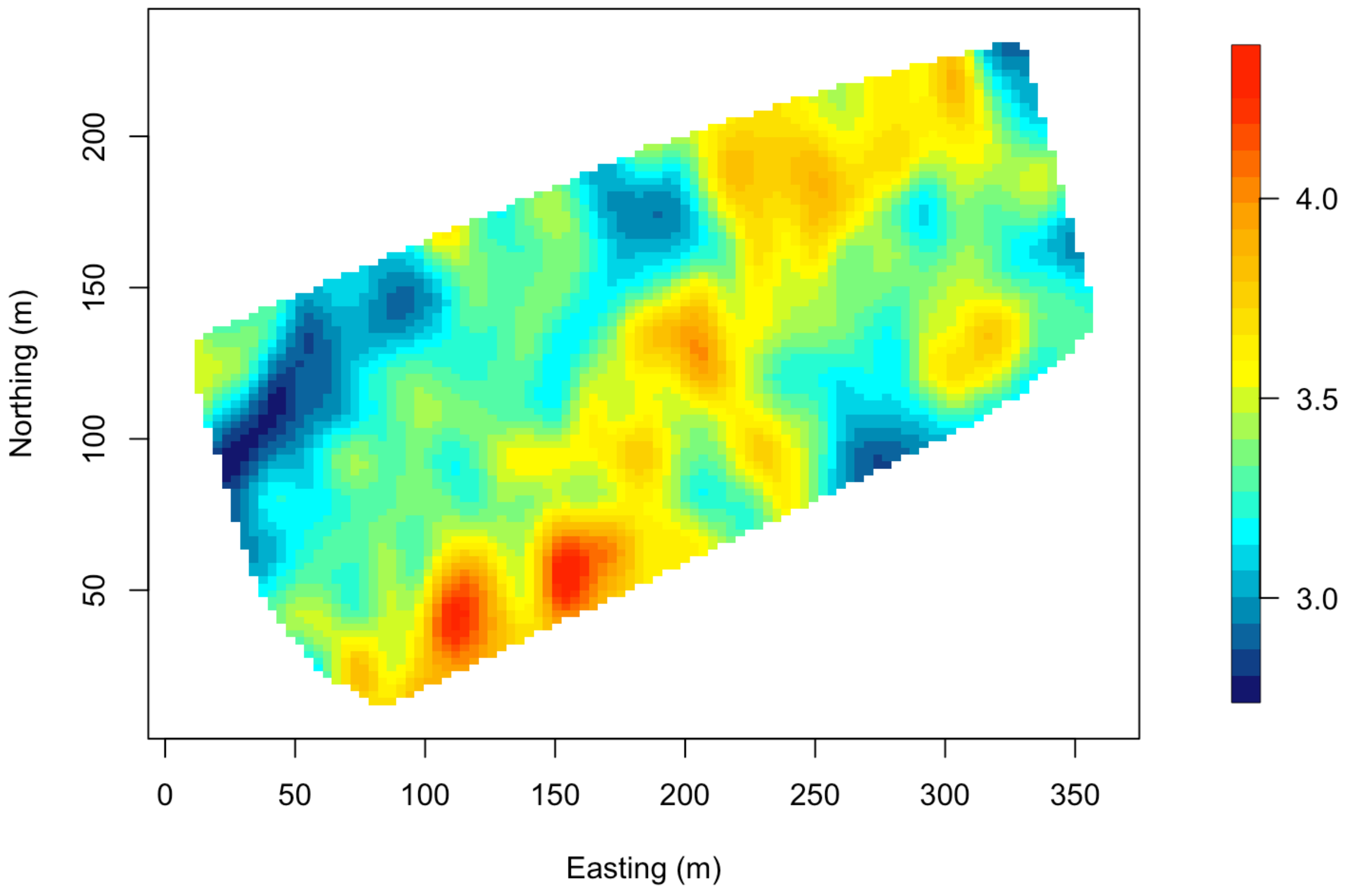
Map of standard deviation of predicted DBH with data locations



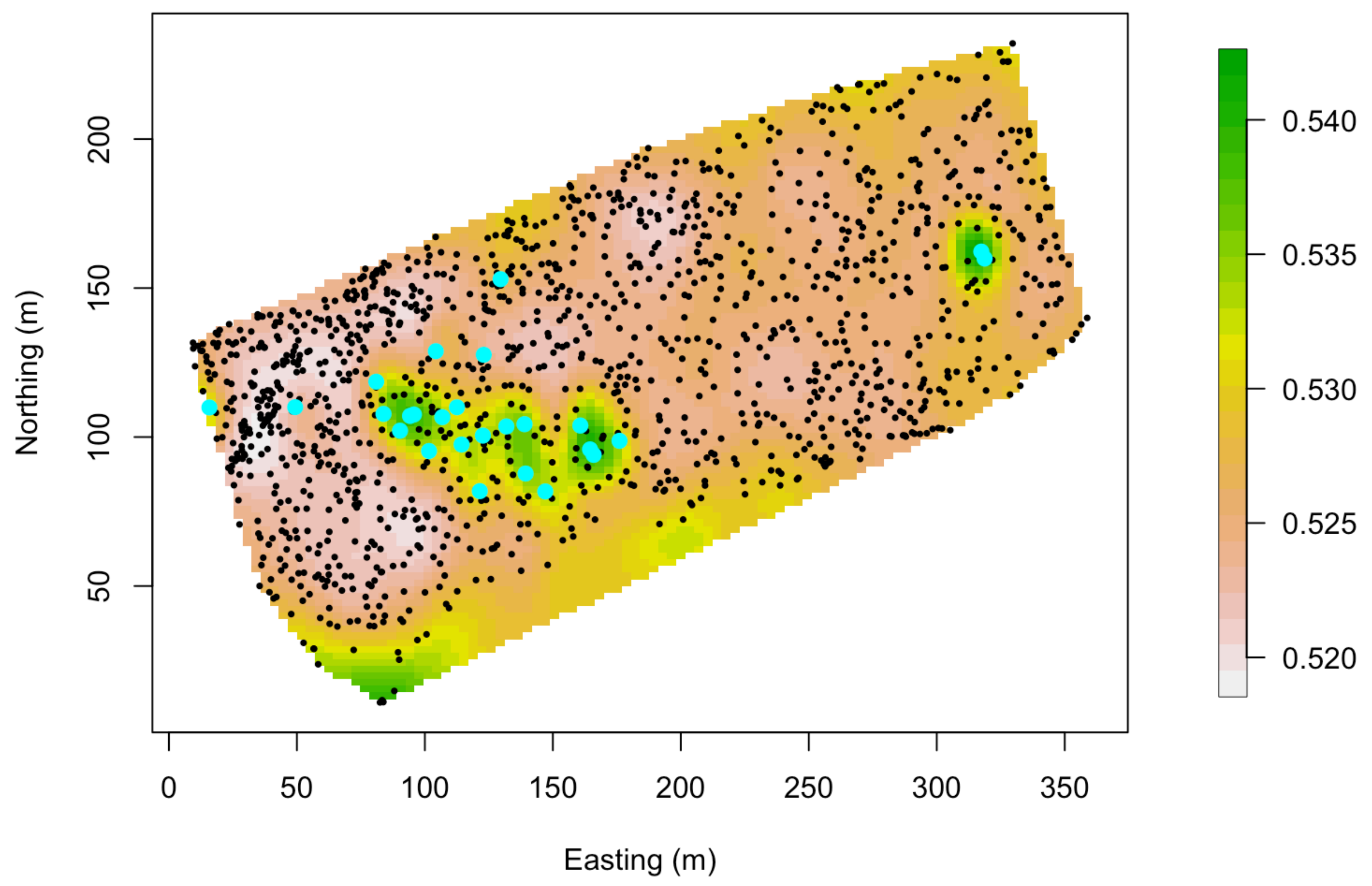
Map of species type for training data

Western Experimental Forestry (WEF) data

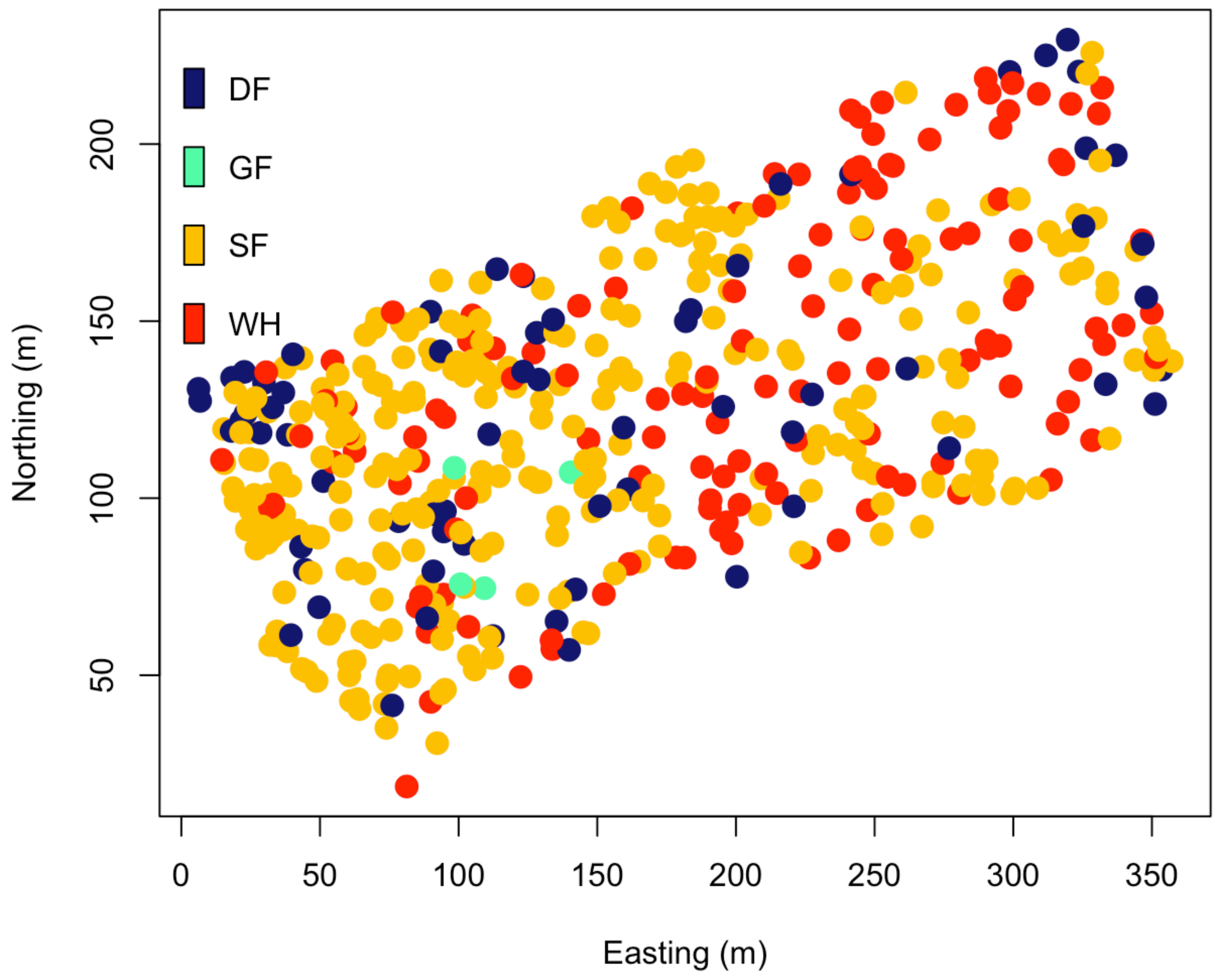
Predictions:



Map of predicted DBH



Map of standard deviation of predicted DBH (highlighting locations of Species GF)



Map of species type for training data

Big spatial data

Data at n locations: $S = \{s_1, \ldots, s_n\}$

Marginal model: $Y \sim N(X\beta, \Sigma(\theta))$ where $\Sigma(\theta) = C + \tau^2 I$

Parameter estimation using MLE

Log-likelihood: $l(\beta, \theta \mid Y) = -\frac{1}{2}\log\det(\Sigma(\theta)) - \frac{1}{2}(Y - X\beta)^\top \Sigma(\theta)^{-1}(Y - X\beta)$

Needs evaluation of $\det(\Sigma(\theta))$ and quadratic forms of $\Sigma(\theta)^{-1}$

Big spatial data

Prediction at a new location s_0: $Y(s_0) \mid Y, \theta, \beta = N\left(\mu(s_0), \sigma^2(s_0)\right)$

Conditional mean: $\mu(s_0) = X'(s_0)\beta + C(s_0, S)\Sigma^{-1}(Y - X\beta)$

Conditional variance: $\sigma^2(s_0) = C(s_0, s_0) + \tau^2 - C(s_0, S)\Sigma^{-1}C(S, s_0)$

Again needs evaluation of quadratic forms of Σ^{-1}

Computational details

Σ is a dense $n \times n$ matrix

Both $\det(\Sigma)$ and Σ^{-1} are best computed via the Cholesky decomposition

Cholesky decomposition: Any symmetric matrix A can be factorized as $A = LDL'$ where L is lower triangular and D is diagonal

Cholesky decomposition requires $O(n^2)$ storage and $O(n^3)$ memory

Not feasible for large n

Methods for spatial big data

Low-rank models

Spectral approximations

Lattice-based

Multi-resolution approaches

Covariance tapering

Stochastic Partial Differential Equations

Nearest-neighbor models

See Heaton et al. (2019) for a review

Methods for spatial big data

Low-rank models

Spectral approximations

Lattice-based

Multi-resolution approaches

Covariance tapering

Stochastic Partial Differential Equations

Nearest-neighbor models

See Heaton et al. (2019) for a review

Nearest Neighbor Gaussian Processes

GP regression model: $Y \sim N(X\beta, \Sigma(\theta))$

Likelihood factorization: $p(Y) = p(Y_1) \times \prod_{i=2}^{n} p(Y_i \,|\, Y_1, \ldots, Y_{i-1})$

Vecchia's GP likelihood approximation (Vecchia, 1988, JRSSB):

$$p(Y) \approx p(Y_1) \times \prod_{i=2}^{n} p(Y_i \,|\, Y_{N(i)})$$

$N(i) =$ set of m nearest neighbors of location s_i among $s_1, \ldots, s_{i-1}$

Reduces computation time from $O(n^3)$ to $O(nm^3)$

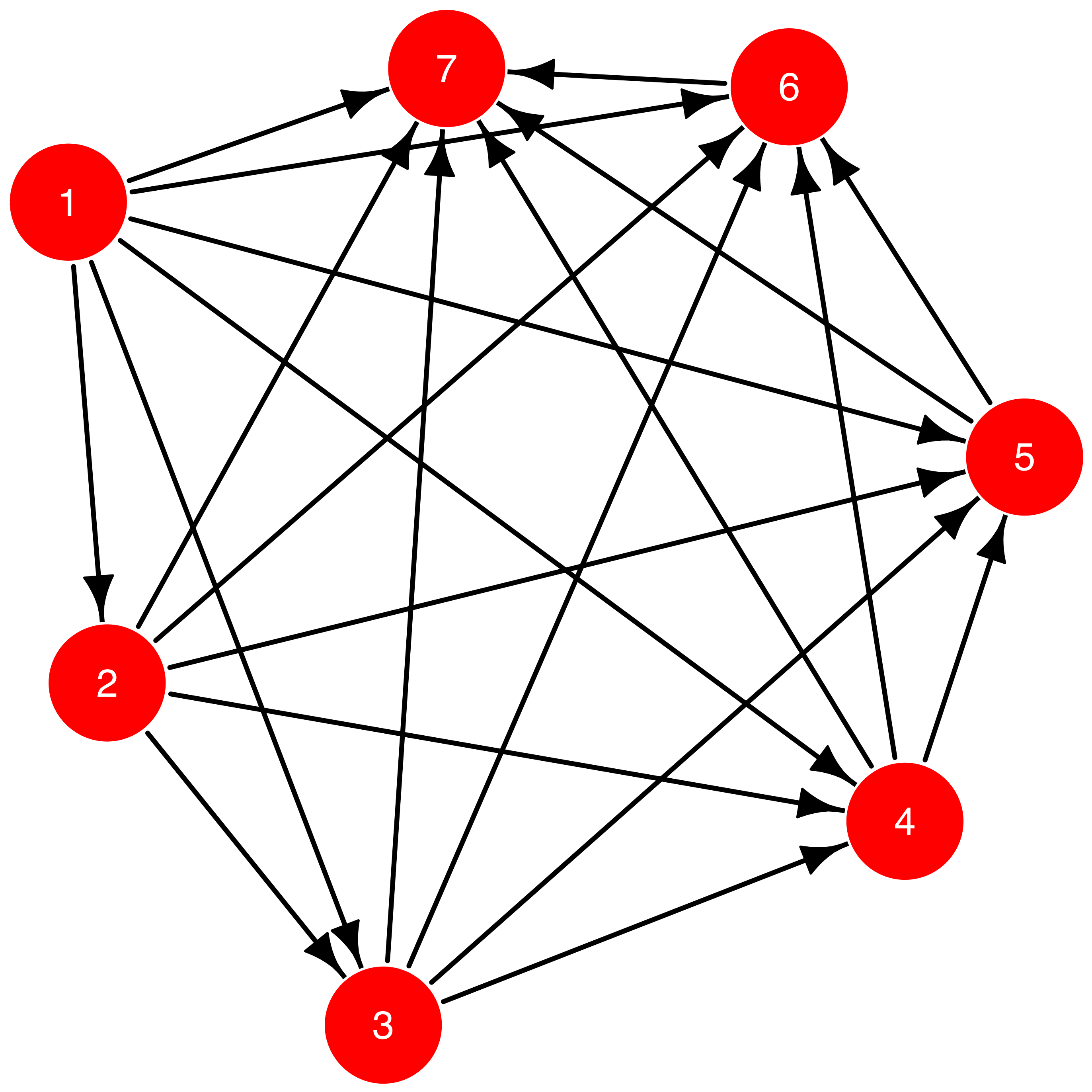
Nearest Neighbor Gaussian Processes

NNGP (Datta et al, 2016, JASA): Vecchia's approximation corresponds to a distribution $N(0, \tilde{\Sigma})$ and can be extended to a valid Gaussian process (NNGP)

Nearest Neighbor Gaussian Processes

NNGP (Datta et al, 2016, JASA): Vecchia's approximation is the likelihood of a distribution $N(0, \tilde{\Sigma})$ and can be extended to a valid Gaussian process (NNGP)

NNGP likelihood factorizes on a sparse directed acyclic graph (DAG)



Complete DAG

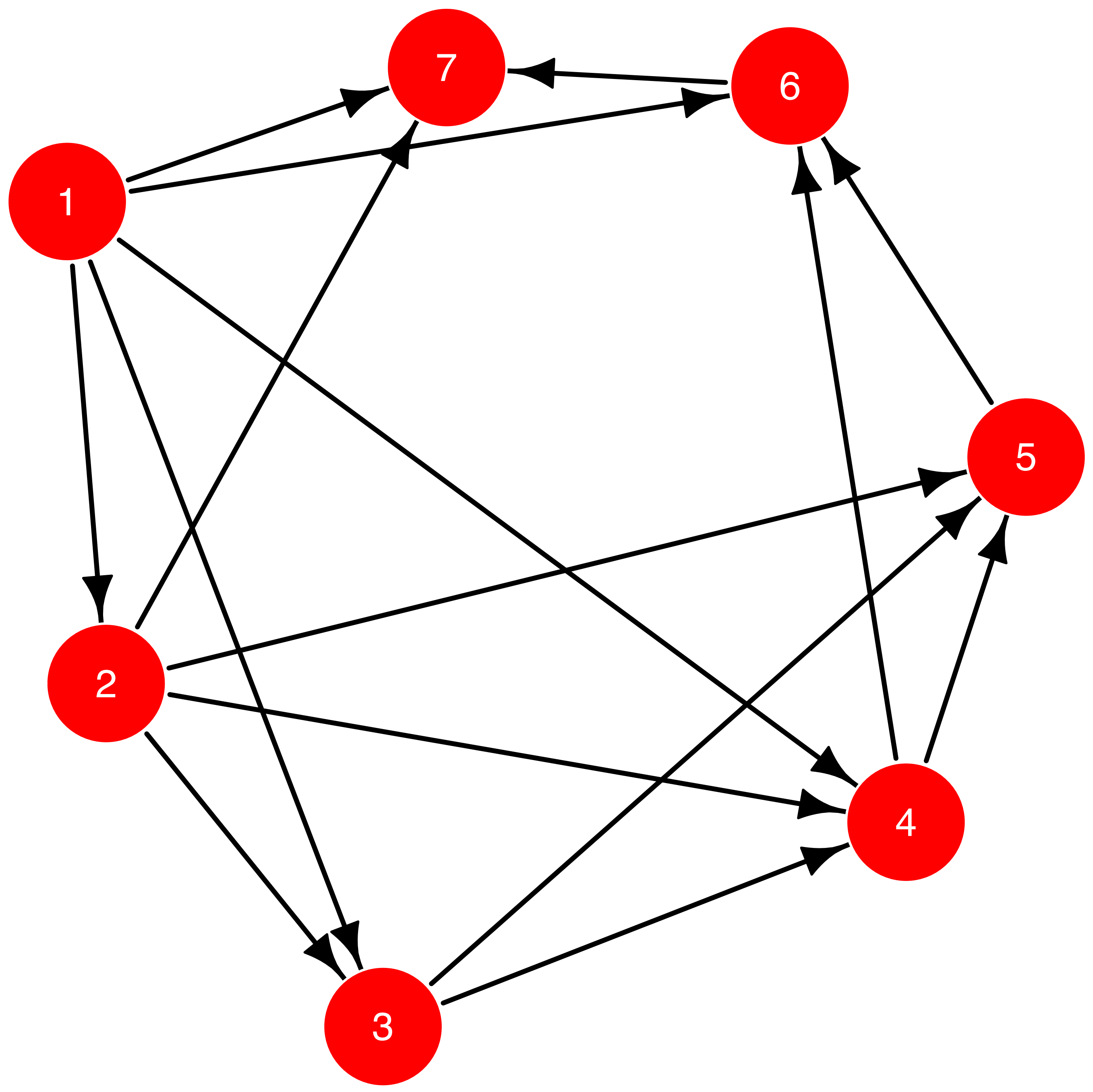
Full GP likelihood

$$p(y) = p(y_1)p(y_2 \mid y_1)$$
$$\times p(y_3 \mid y_1, y_2)p(y_4 \mid y_1, y_2, y_3)p(y_5 \mid y_1, y_2, y_3, y_4)$$
$$\times p(y_6 \mid y_1, y_2, y_3, y_4, y_5)p(y_7 \mid y_1, y_2, y_3, y_4, y_5, y_6)$$

Nearest Neighbor Gaussian Processes

NNGP (Datta et al, 2016, JASA): Vecchia's approximation is the likelihood of a distribution $N(0, \tilde{\Sigma})$ and can be extended to a valid Gaussian process (NNGP)

NNGP likelihood factorizes on a sparse directed acyclic graph (DAG)



3-NN DAG

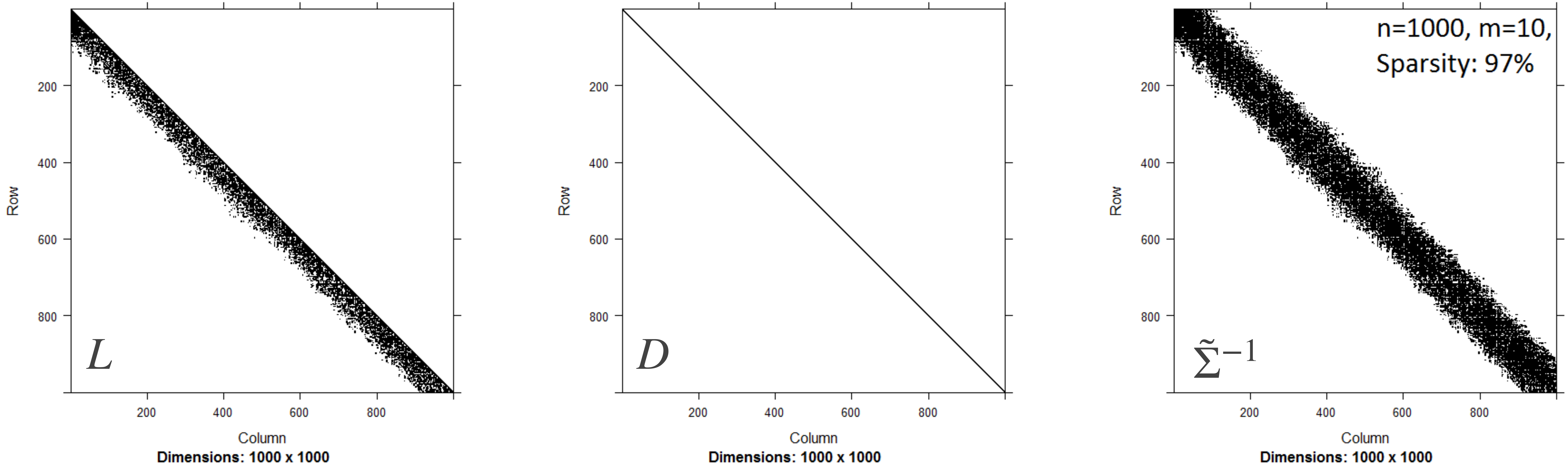
NNGP likelihood

$$p(y) = p(y_1)p(y_2 \mid y_1)$$
$$\times\, p(y_3 \mid y_1, y_2)p(y_4 \mid y_1, y_2, y_3)p(y_5 \mid y_1, y_2, y_3, \cancel{y_4})$$
$$\times\, p(y_6 \mid \cancel{y_1}, \cancel{y_2}, y_3, y_4, y_5)p(y_7 \mid y_1, y_2, \cancel{y_3}, \cancel{y_4}, \cancel{y_5}, y_6)$$

Nearest Neighbor Gaussian Processes

The NNGP precision matrix admits the factorization $\tilde{\Sigma}^{-1} = L'DL$

- D is diagonal with entries d_i
- L is lower triangular and row sparse
  - Sparsity determined by the nearest-neighbor DAG
- $\tilde{\Sigma}^{-1}$ is also sparse



Nearest Neighbor Gaussian Processes

Estimation:

The NNGP precision matrix $\tilde{\Sigma}^{-1} = L'DL$

D is diagonal with entries d_i

L is lower triangular and row sparse

L and D can be computed in $O(nm^3)$ time

$$\det(\tilde{\Sigma}) = \frac{1}{\prod_i d_i}$$

$$x'\tilde{\Sigma}^{-1}x = (Lx)'D(Lx) = \sum_i v_i^2 d_i \text{ where } v = Lx$$

Total time to evaluate NNGP likelihood is $O(nm^3)$

Nearest Neighbor Gaussian Processes

Predictions:

NNGP prediction at a new location s_0:

$$Y(s_0) \mid Y, \theta, \beta = Y(s_0) \mid Y_{N_0}, \theta, \beta = = N\left(\tilde{\mu}(s_0), \tilde{\sigma}^2(s_0)\right)$$

$N_0 = m$ nearest neighbors of s_0 among $s_1, \ldots, s_n$

Conditional mean: $\tilde{\mu}(s_0) = X'(s_0)\beta + C(s_0, N_0)\Sigma^{-1}_{N_0,N_0}(Y_{N_0} - X_{N_0}\beta)$

Conditional variance: $\tilde{\sigma}^2(s_0) = C(s_0, s_0) + \tau^2 - C(s_0, N_0)\Sigma^{-1}_{N_0,N_0}C(N_0, s_0)$

Nearest Neighbor Gaussian Processes

Software (R package):

BRISC (Saha and Datta)

Frequentist implementation

Estimation with bootstrapped uncertainty

Prediction with uncertainty

Simulation of large spatial data

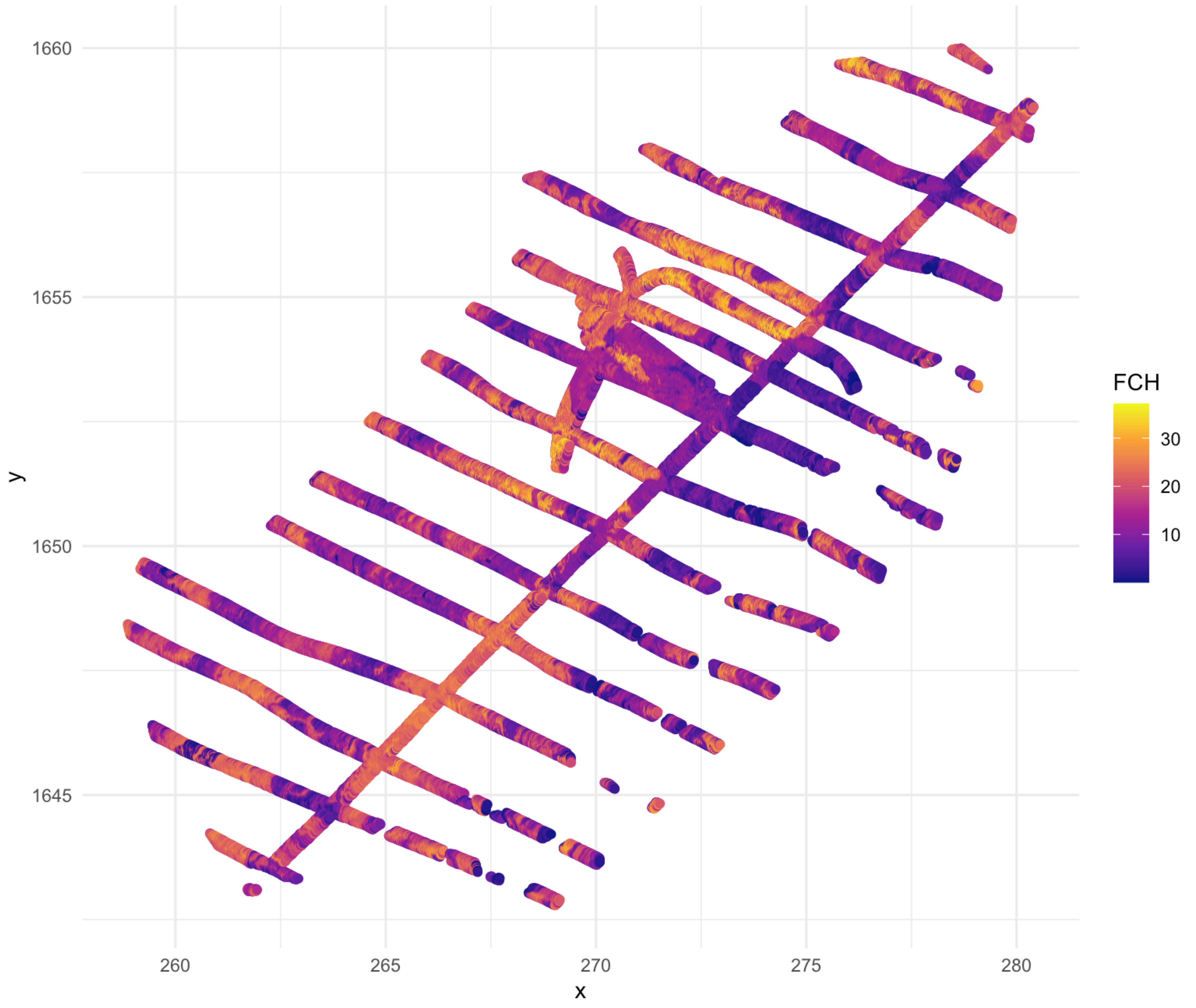
spNNGP (Finley, Datta, and Banerjee)

Bayesian implementation

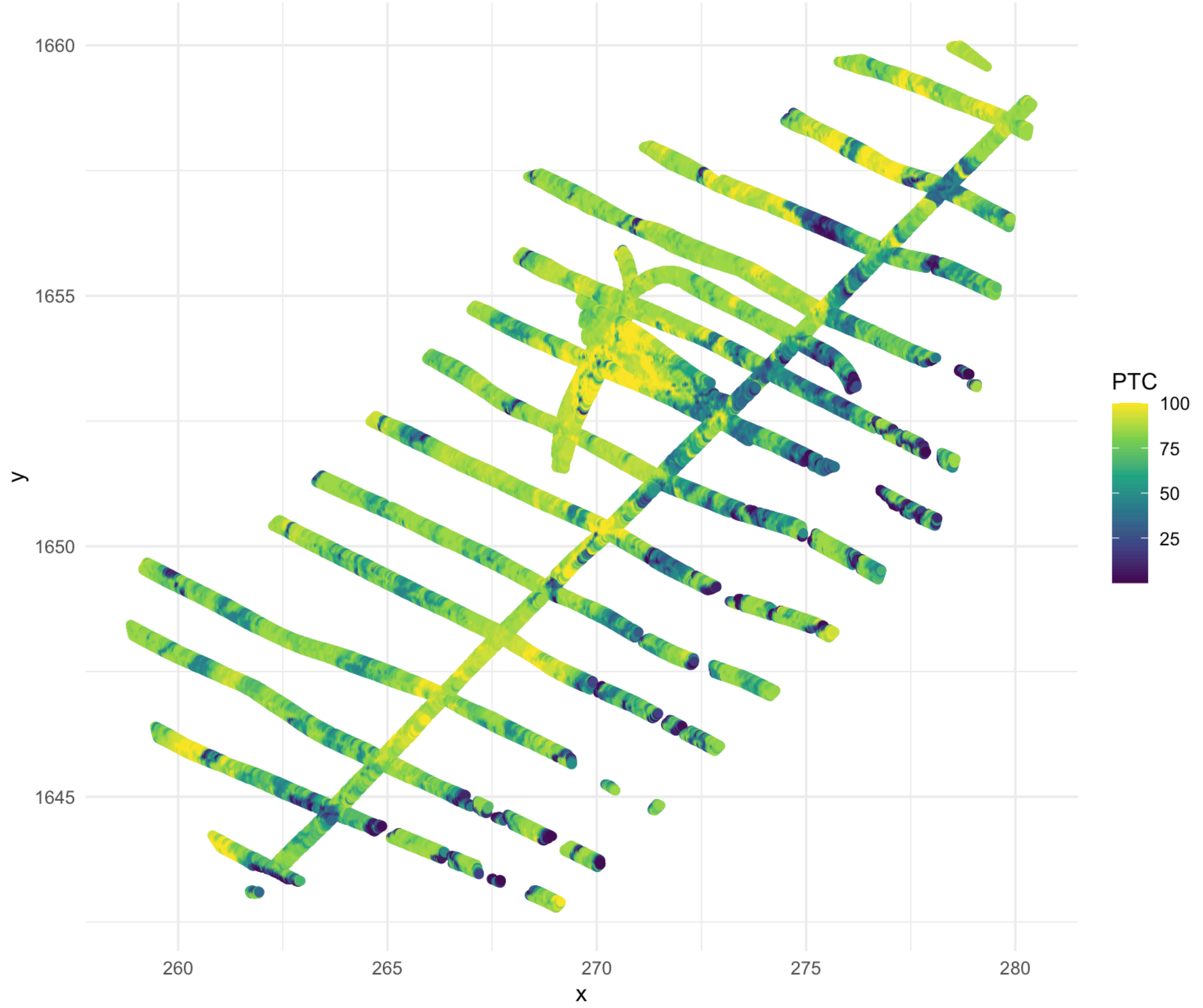
Full posterior distributions using MCMC

Bonanza Creek Experimental Forest Data

Forest canopy height (FCH) estimates at 180,000 locations NASA Goddard's LiDAR in Bonanza Creek Experimental Forest, Alaska



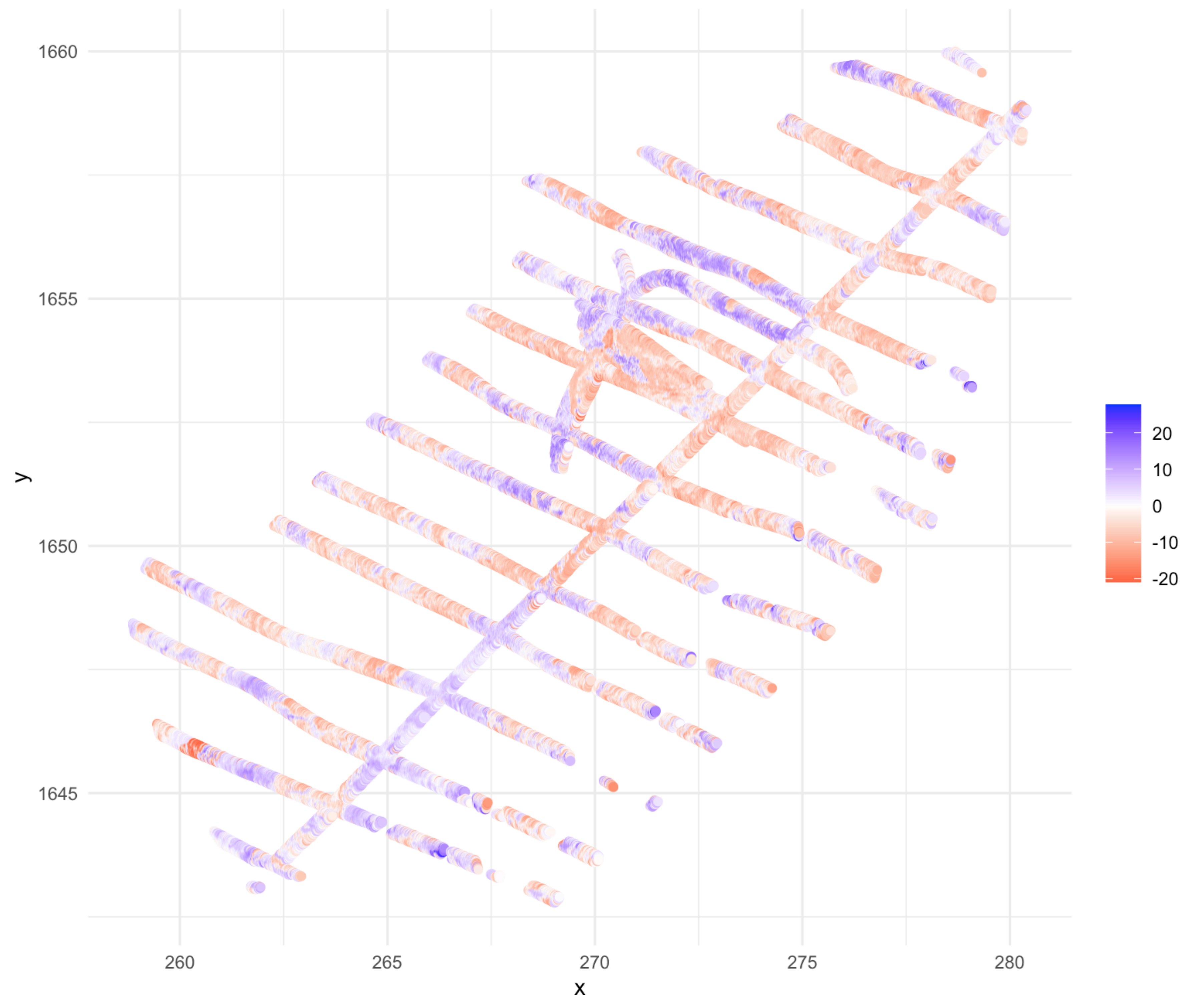
Forest canopy height (FCH)



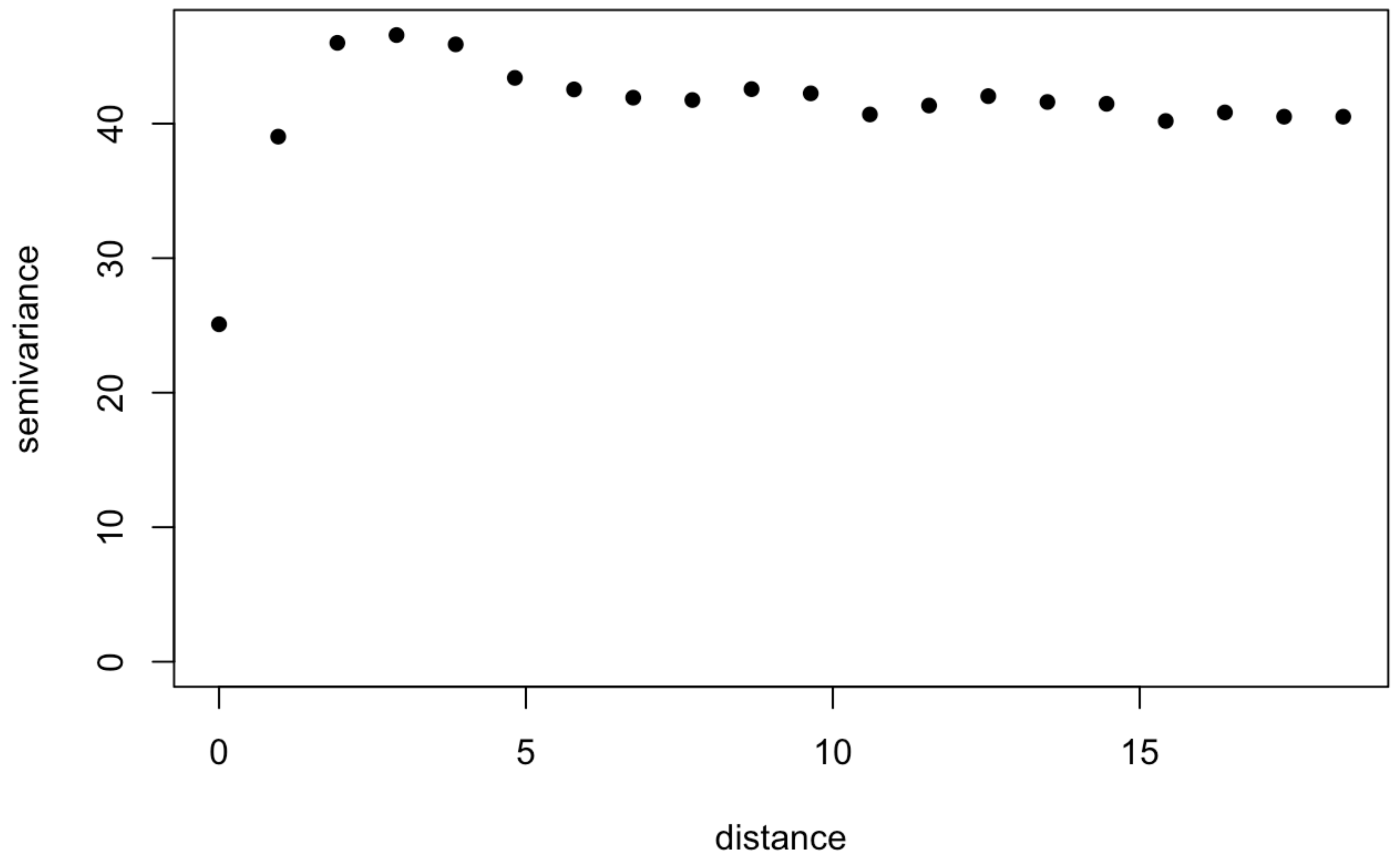
Covariate — Percent tree cover (PTC)

Bonanza Creek Experimental Forest Data

Linear model:



Residuals

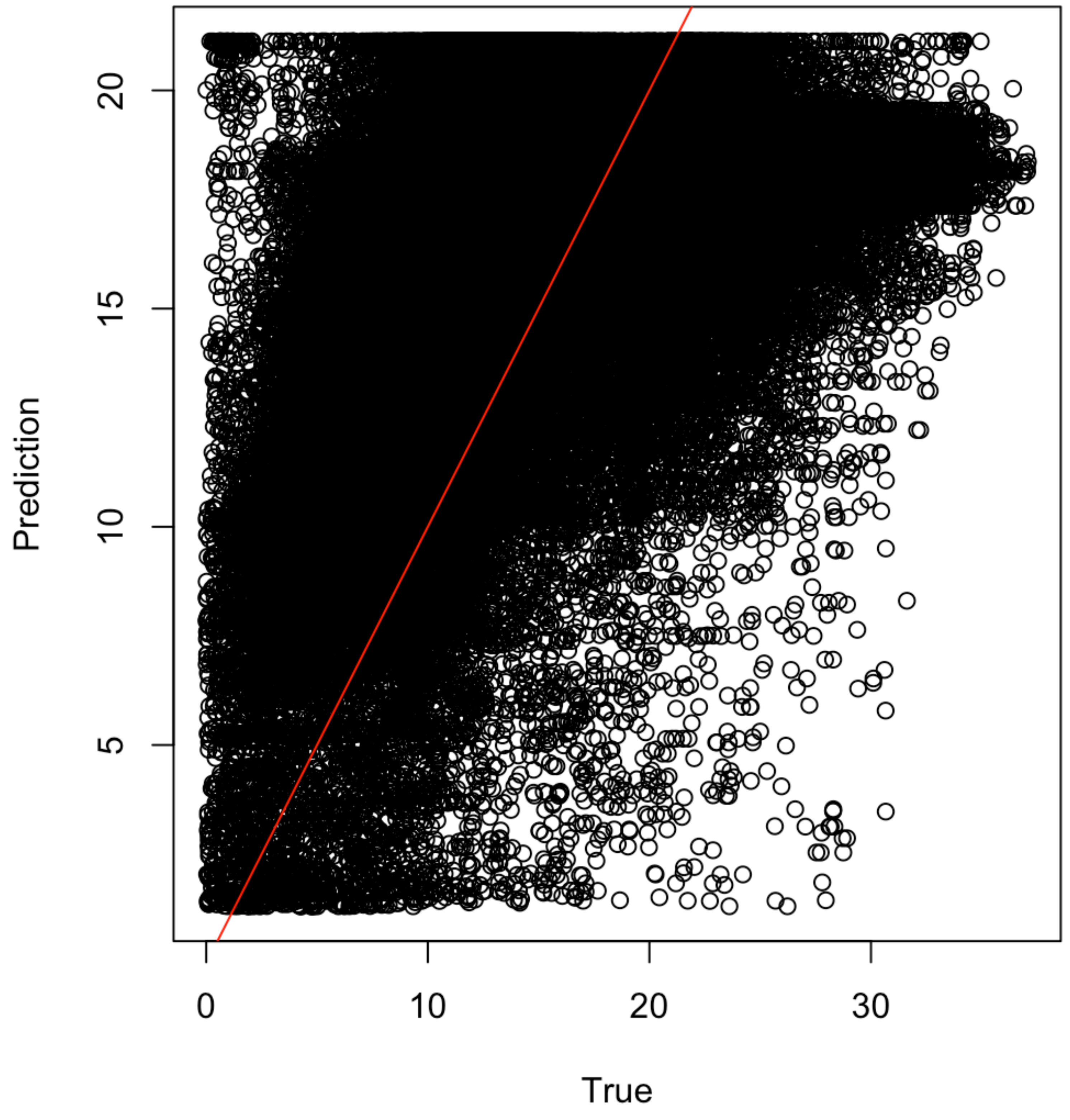


Variogram of a subset of residuals

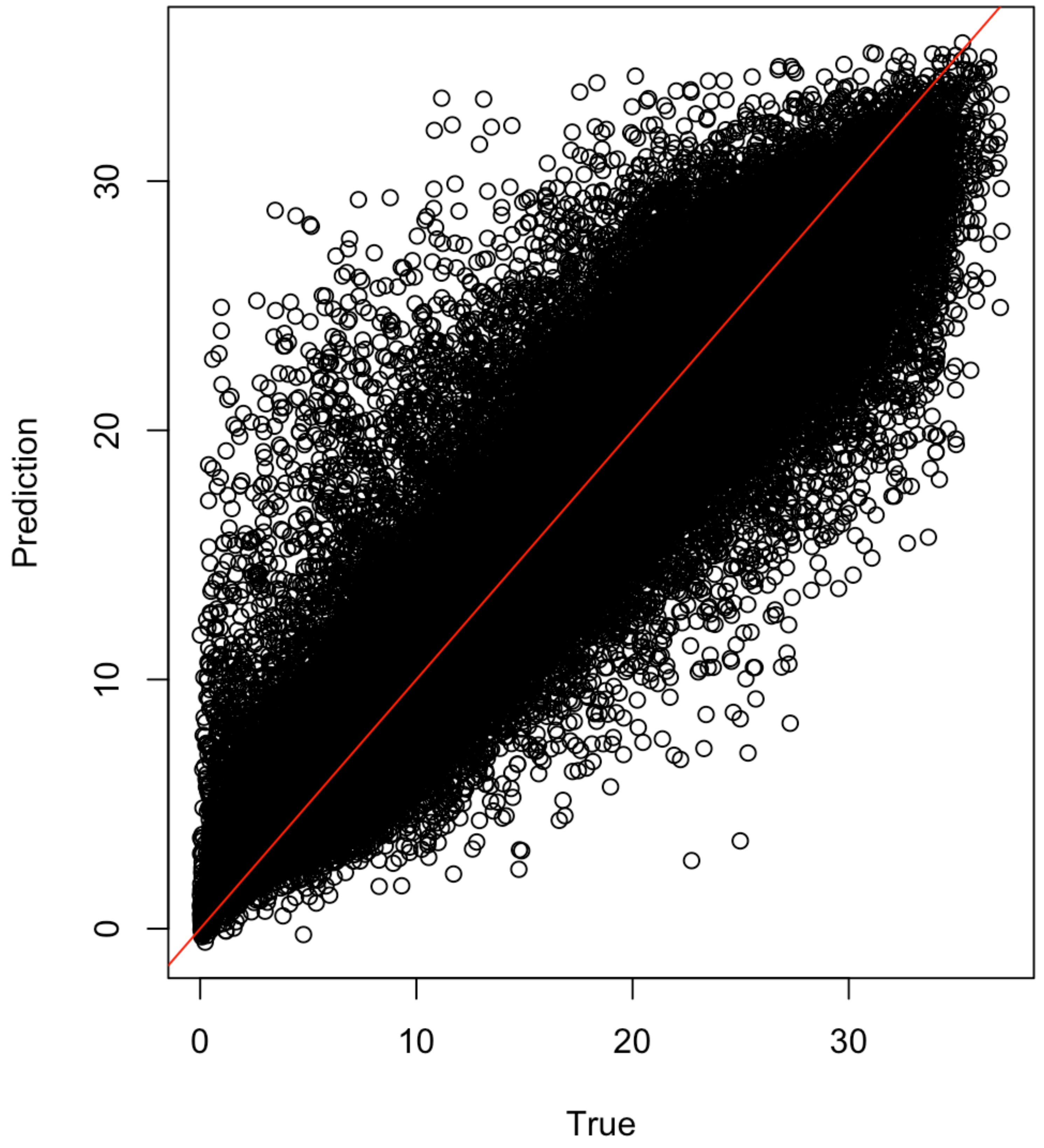
Bonanza Creek Experimental Forest Data

Spatial model: Fitted using *BRISC_estimation,* predictions using *BRISC_prediction*

Metric	Spatial	Non-Spatial
RMSPE	2.92	6.59
CP	0.94	0.96
CIW	11.15	25.84



Non-spatial model

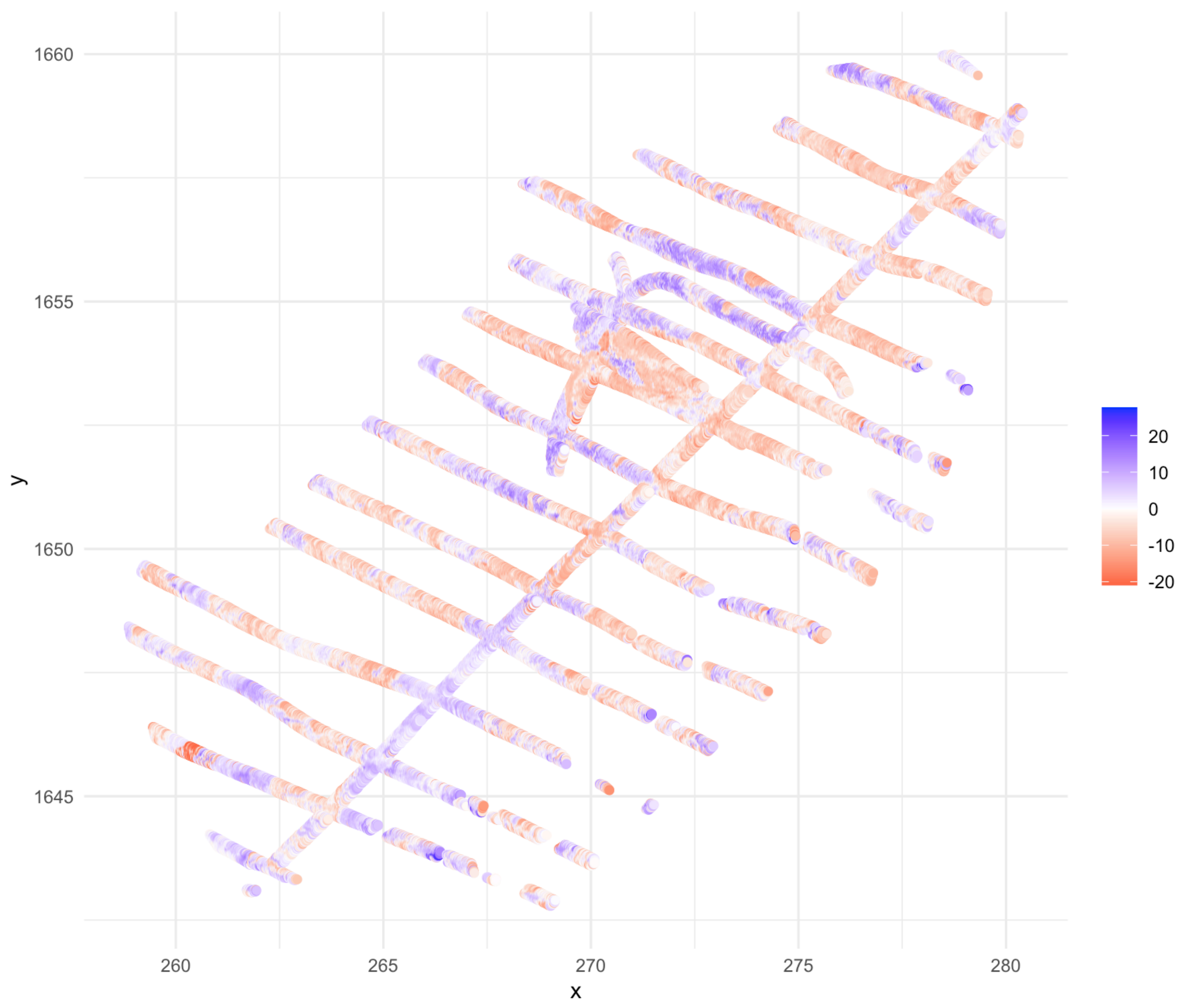


Spatial model

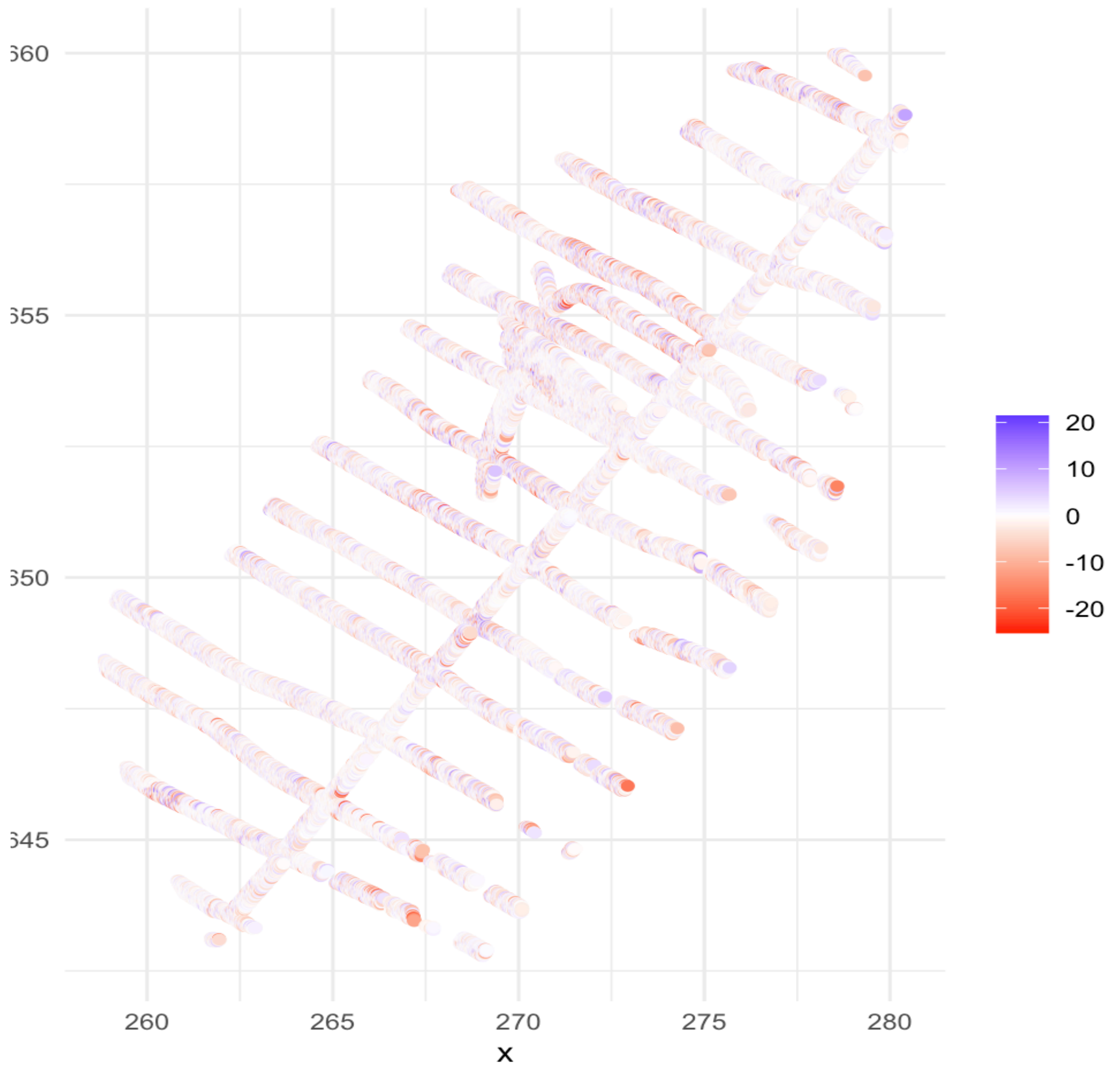
Bonanza Creek Experimental Forest Data

Spatial model: Fitted using *BRISC_estimation,* predictions using *BRISC_prediction*

Metric	Spatial	Non-Spatial
RMSPE	2.92	6.59
CP	0.94	0.96
CIW	11.15	25.84



Spatial
model residuals



Non-spatial
model residuals

Summary

Introduction to geostatistics
- Data setup and analysis objectives

Exploratory data analysis to understand need for spatial modeling
- Maps and variograms of data and linear model residuals

Spatial linear mixed effect models
- Process level modeling and Gaussian processes
- Parameter estimation
- Prediction (kriging) with uncertainty quantification

Summary

Model comparison

- Estimation: AIC, BIC
- Prediction: RMSPE, coverage probability and width of prediction intervals,

Big spatial data

- Computing challenges
- Fast alternatives (Nearest Neighbor Gaussian Process)

Spatial analysis using geoR and BRISC R-packages

References

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2003). *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC.

Cressie, N., & Wikle, C. K. (2011). Statistics for spatio-temporal data. John Wiley & Sons.

Stein, M. L. (2012). Interpolation of spatial data: some theory for kriging. Springer Science & Business Media.

Van Der Vaart, A. W., & Van Zanten, J. H. (2008). Rates of contraction of posterior distributions based on Gaussian process priors.

Van Der Vaart, A., & Van Zanten, H. (2011). Information Rates of Nonparametric Gaussian Process Methods. Journal of Machine Learning Research, 12(6).

Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guinness, J., Guhaniyogi, R., ... & Zammit-Mangion, A. (2019). A case study competition among methods for analyzing large spatial data. Journal of Agricultural, Biological and Environmental Statistics, 24, 398-425.

Vecchia, A. V. (1988). Estimation and model identification for continuous spatial processes. Journal of the Royal Statistical Society Series B: Statistical Methodology , 50 (2), 297-312.

Datta, A., Banerjee, S., Finley, A. O., & Gelfand, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. Journal of the American Statistical Association, 111(514), 800-812.