University of Zurich[UZH]
Institute of Mathematics
Lucerne University of Applied Sciences and Arts
HOCHSCHULE LUZERN
FH Zentralschweiz

# Spatially Varying Coefficients Models: How MLE Stacks up Against Other Methods

*JSM 2019, Denver*

**Jakob Dambon**[1, 2], Fabio Sigrist[2], Reinhard Furrer[1]

[1] University of Zurich

[2] Lucerne University of Applied Sciences and Arts

# Introduction

– Start with *linear regression* model with $p$ covariates:

$$y_i = \beta^{(1)} x_i^{(1)} + \ \ldots \ + \beta^{(p)} x_i^{(p)} + \epsilon_i, \qquad \epsilon_i \sim N(0, \tau^2) \text{ i. i. d.}$$

– Assigning locations $\boldsymbol{s}_i$ to observation $i$, one can generalize it to:

$$y_i = \beta^{(1)}(\boldsymbol{s}_i) x_i^{(1)} + \ \ldots \ + \beta^{(p)}(\boldsymbol{s}_i) x_i^{(p)} + \epsilon_i, \qquad \epsilon_i \sim N(0, \tau^2) \text{ i. i. d.}$$

– *Spatially varying coefficients (SVC)*

# Different Methods and Assumptions on SVC

- Geographically Weighted Regression[(1)]

- Eigenvector Spatial Filtering[(2)]

- Bayesian SVC Processes[(3)]
  - Gaussian Predictive Processes[(4)]
  - Gaussian Markov Random Field Approximation[(5)]

$\boldsymbol{\beta}^{(j)}(\cdot) \sim N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$
a *Gaussian random field*

where $\boldsymbol{\Sigma}_j$ are of some specific *Matérn* class

# Motivation

- Use Case: *Real Estate Appraisal*
  - Strong believe in SVC[6, 7]
    - Area, age
  - **LARGE** data
- Methods having a Gaussian random field assumption do not scale.

# Goal

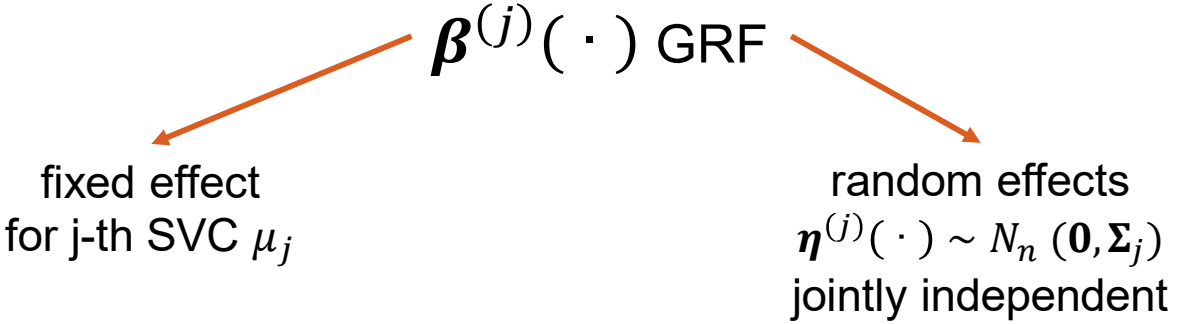
Establish a method that scales

in $n$ observations to the ten thousands

in $p$ SVC to the tens

using Gaussian random fields to model SVC.

# Model Assumption for MLE




– Covariance matrix $\boldsymbol{\Sigma}_j$ parametrized by a range $\rho_j$ and marginal variance $\sigma_j^2$

– Model in matrix notation

$$\boldsymbol{y} = \boldsymbol{X\mu} + \boldsymbol{W\eta} + \boldsymbol{\epsilon}$$

– Joint GRF $\boldsymbol{\eta}(\cdot) \sim N_{np}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}} = \mathrm{diag}(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_\mathrm{p}))$ with according parametrization for data matrix $\boldsymbol{W}$

# Maximum Likelihood Estimation

- Log likelihood: $\log L(\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{\mu}) \propto \log\det \boldsymbol{\Sigma}_{\mathbf{y}} + (\boldsymbol{y} - \boldsymbol{X\mu})^{\top} \boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1} (\boldsymbol{y} - \boldsymbol{X\mu})$

  where $\boldsymbol{\theta} = \left(\rho_1, \sigma_1^2, \ldots, \rho_p, \sigma_p^2, \tau^2\right)$ and $\boldsymbol{\Sigma}_{\mathbf{y}} = \boldsymbol{\Sigma}_{\boldsymbol{y}}(\boldsymbol{\theta}) = \boldsymbol{W\Sigma_{\eta} W}^{\top} + \tau^2 I_{n\times n}$

- *Numeric optimization required*, challenges:
  - High-dimensional parameter space $(\boldsymbol{\theta}, \boldsymbol{\mu}) \in \boldsymbol{\Omega} = (0, \infty)^{2p+1} \times \mathrm{R}^p$
  - Computational intensive tasks

Ad-hoc MLE does **not** scale in $p$ nor $n$

# Proposed Optimization

– Use *profile likelihood* and *generalized least square* estimate for $\boldsymbol{\mu}$

Dimension reduction of parameter space

$$\hat{\boldsymbol{\mu}} = \left(\boldsymbol{X}^\top \boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1}(\hat{\boldsymbol{\theta}})\, \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1}(\hat{\boldsymbol{\theta}})\, \boldsymbol{y}$$

– Independence assumption on Gaussian random field

Efficient updating of $\boldsymbol{\Sigma}_{\boldsymbol{y}}(\boldsymbol{\theta})$

$$\boldsymbol{\Sigma}^{taper} \odot \boldsymbol{\Sigma}_{\boldsymbol{y}}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}^{taper} \odot \left( \sum_{j=1}^{p} \boldsymbol{\Sigma}_j \odot \left( \boldsymbol{x}^{(j)} (\boldsymbol{x}^{(j)})^\top \right) \right) + \tau^2 I_{n \times n}$$

– Covariance matrix tapering[8]

Introducing sparse matrices to ease Cholesky decomposition

– Further regularize likelihood using priors[9, 10]

# Empirical Validation

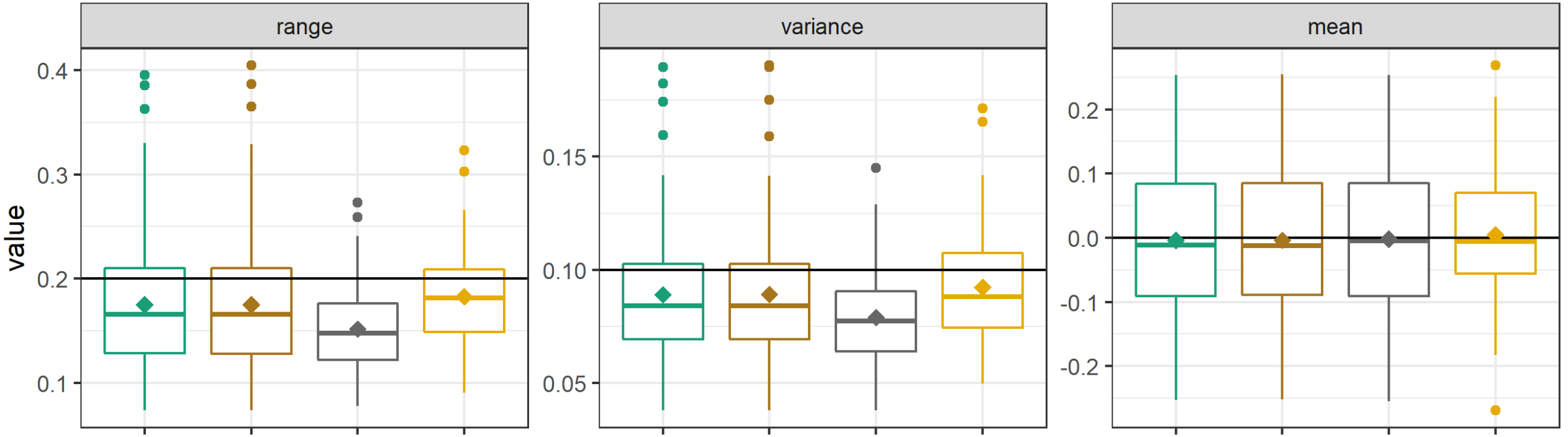
- Simulation Study
  - Estimate parameters
  - Give predictions
  - Repeat $N = 100$
- Use Case
  - Cross Validation

# Parameter Estimation for SVC $(\rho_j,\ \sigma_j^2,\ \mu_j)$

| Name | Parameter Optimization | Prior Regularization |
|---|---|---|
| MLE | $\theta, \mu$ | no |
| MLE with profile | $\theta$ | no |
| MLE with profile | $\theta$ | PC priors[9, 10] |
| GMRF-Approx. | - | PC priors[9, 10] |

Estimated parameters

simulations: 100, black line is the true value

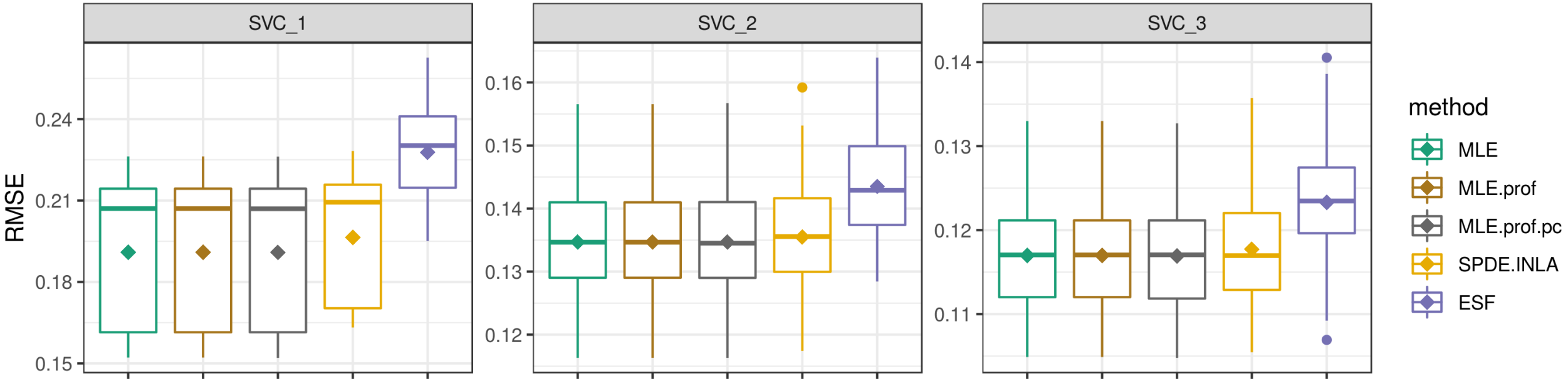


# Predictive Performance

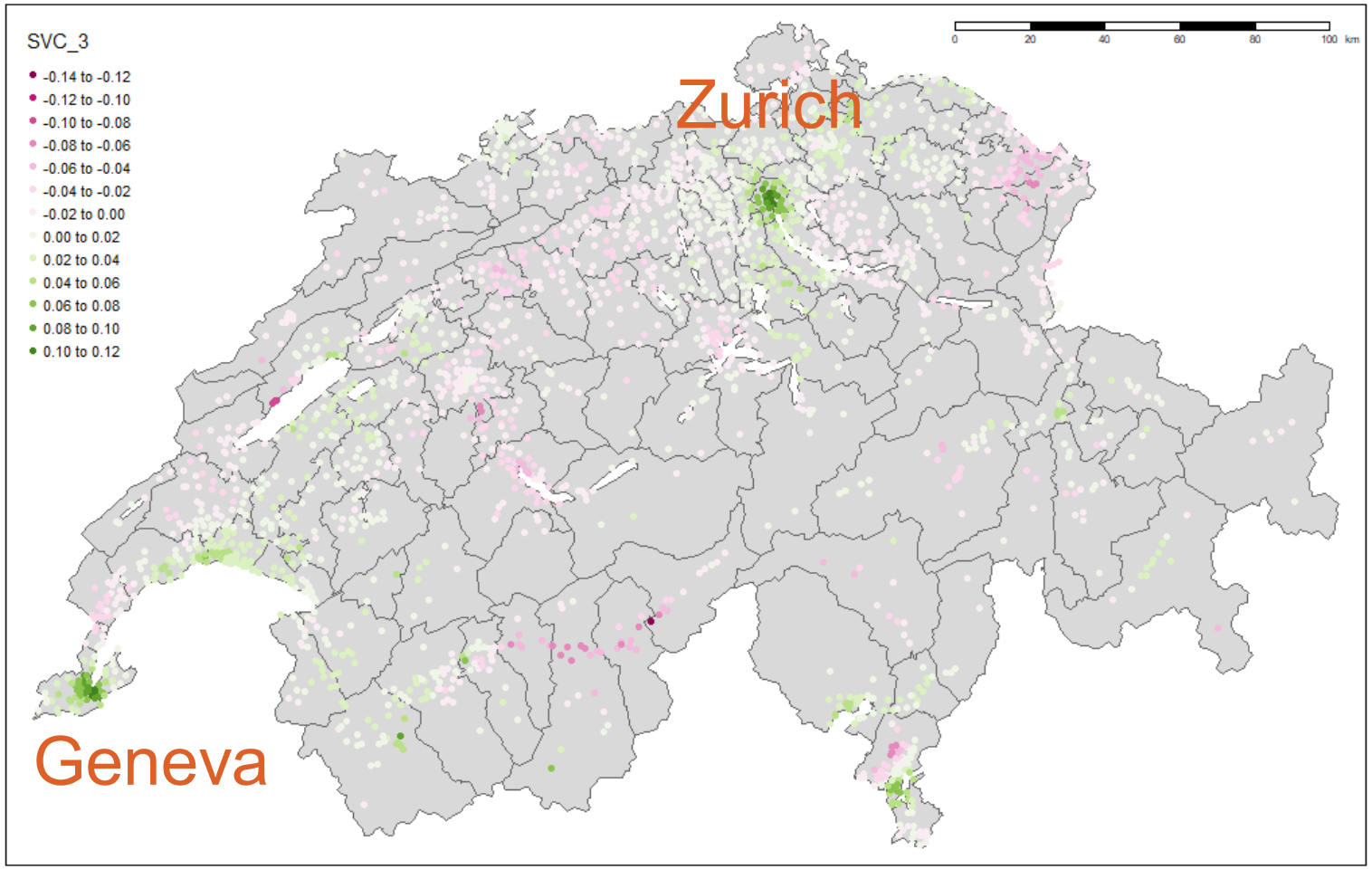
$$\mathrm{RMSE}_{test}^{(j)} = \sqrt{\sum_{i=1}^{n_{test}} \left(\beta^{(j)}(s_i) - \hat{\beta}^{(j)}(s_i)\right)^2 \Big/ n_{test}}$$

using the *empirical best linear unbiased predictor*

simulations: 100

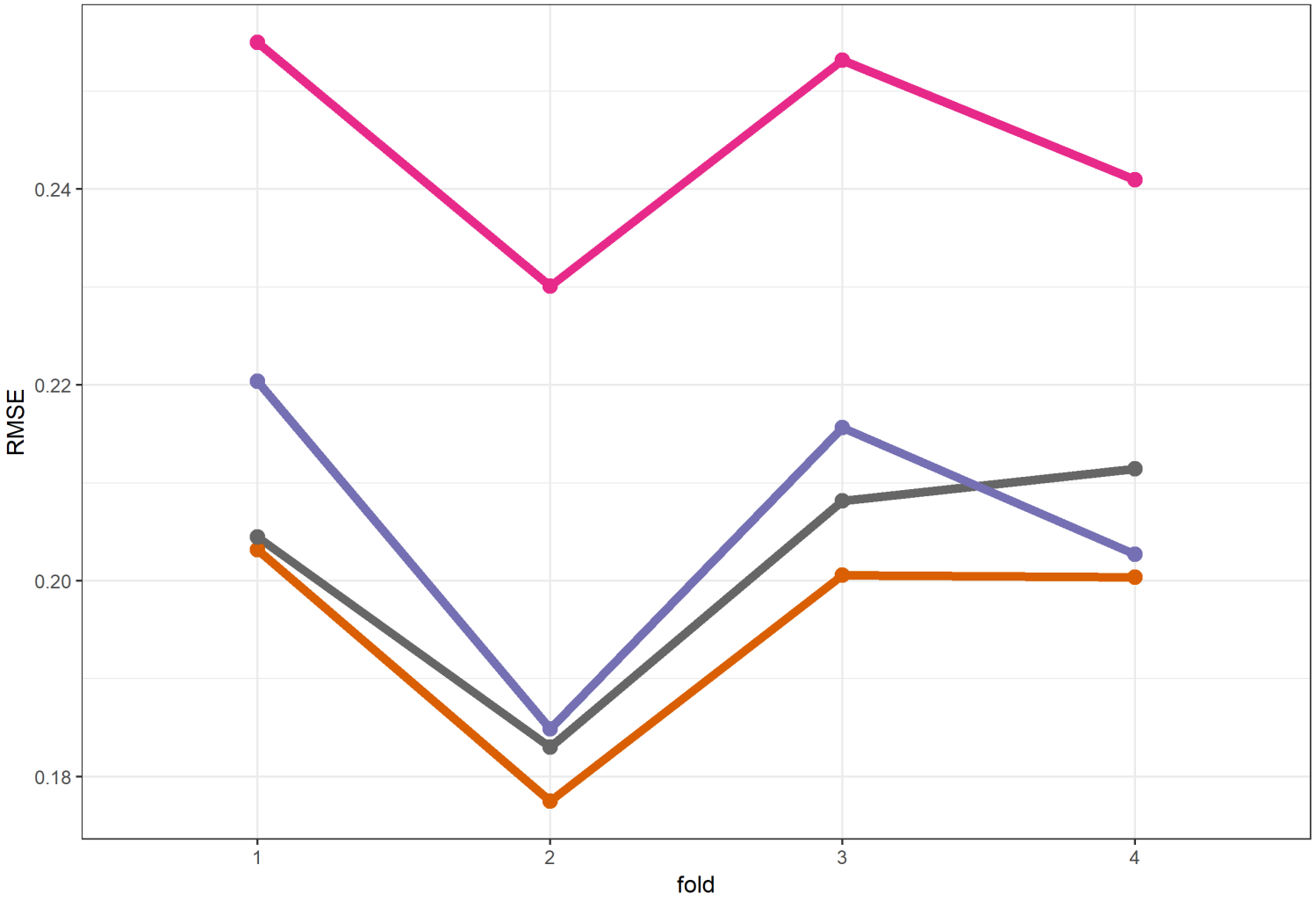


# Estimated SVC for Age Effect




$p = 8$
$n \approx 4'000$

# Application: Out-of-sample RMSE in a Cross Validation




**Thank you for your attention!**

@JakobDambon
jakob.dambon@math.uzh.ch

R-package **`varycoef`**

https://github.com/jakobdambon/varycoef

method

- MLE with profile likelihood & PC priors
- Classical geo-statistical model
- Eigenvector spatial filtering
- Geographically weighted regression

# Bibliography

(1) Fotheringham *et al.* (2002). Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. *Wiley.*

(2) Murakami and Griffith (2015). Random Effects Specifications in Eigenvector Spatial Filtering: A Simulation Study. *J. Geogr. Syst.* **17**, 311 - 331.

(3) Gelfand *et al.* (2003). Spatial Modeling with Spatially Varying Coefficient Processes. *JASA* **98**, 387-396.

(4) Banerjee *et al.* (2008). Gaussian Predictive Process Models for Large Spatial Data Sets. *JRSS: Series B* **70**, 825 - 848.

(5) Lindgren *et al.* (2011). An Explicit Link between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach. *JRSS: Series B* **73**, 423 - 498.

(6) Cao *et al.* (2019). A Big Data Based Geographically Weighted Regression Model for Public Housing Prices: A Case Study in Singapore. *Ann. Am. Assoc. Geogr.* **109**, 173 - 186.

(7) Geng *et al.* (2011). Geographically Weighted Regression Model (GWR) based Spatial Analysis of House Price in Shenzhen. In *2011 19th International Conference on Geoinformatics*, pp. 1 - 5.

(8) Furrer *et al.* (2006). Covariance Tapering for Interpolation of Large Spatial Datasets. *J. Comput. Graph. Stat.* **15**, 502 - 523.

(9) Simpson *et al.* (2017). Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Statist. Sci.* **32**, 1 - 28.

(10) Fuglstad *et al.* (2018). Constructing Priors that Penalize the Complexity of Gaussian Random Fields. *JASA* **114**, 445 - 452.

# Parametrization of $\boldsymbol{W}$

– Ordering of random effects

$$\boldsymbol{\eta} = \begin{pmatrix} \boldsymbol{\eta}_1 \\ \vdots \\ \boldsymbol{\eta}_p \end{pmatrix}$$

– Therefore,

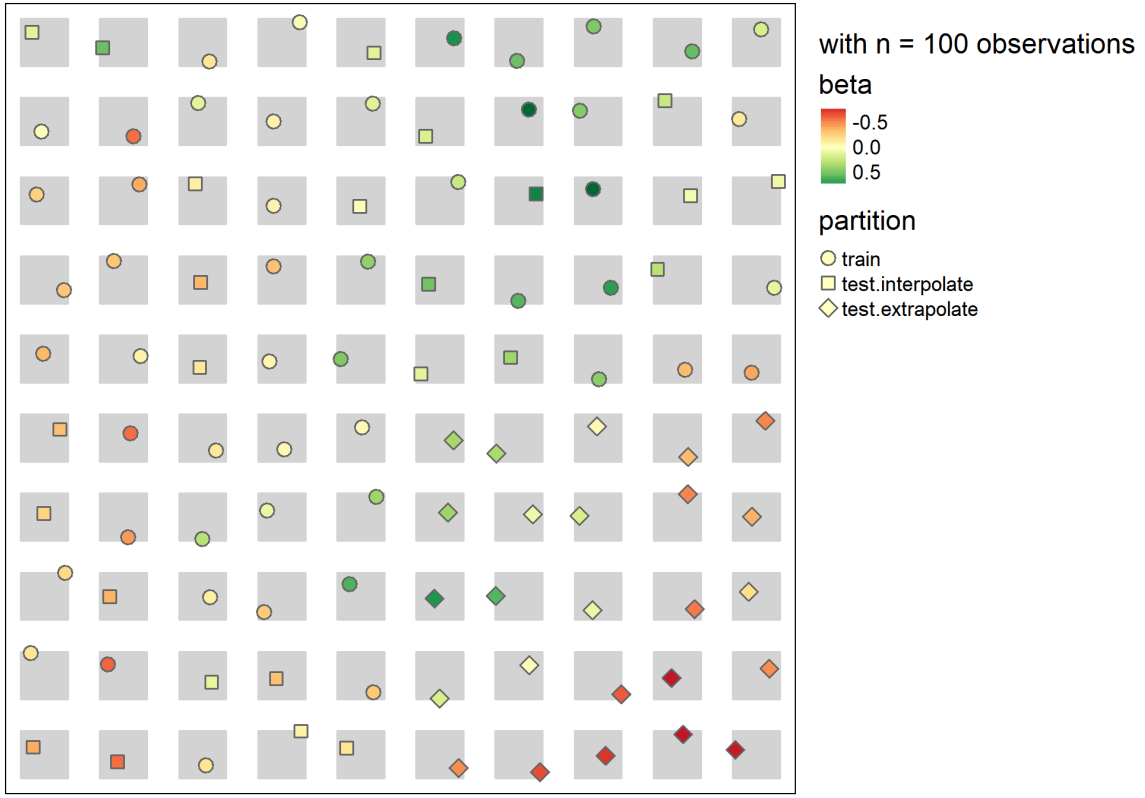
$$\boldsymbol{W} = \begin{pmatrix} \mathrm{diag}\left(\boldsymbol{x}^{(1)}\right) & \ldots & \mathrm{diag}\left(\boldsymbol{x}^{(p)}\right) \end{pmatrix}$$

# Simulation Set-Up

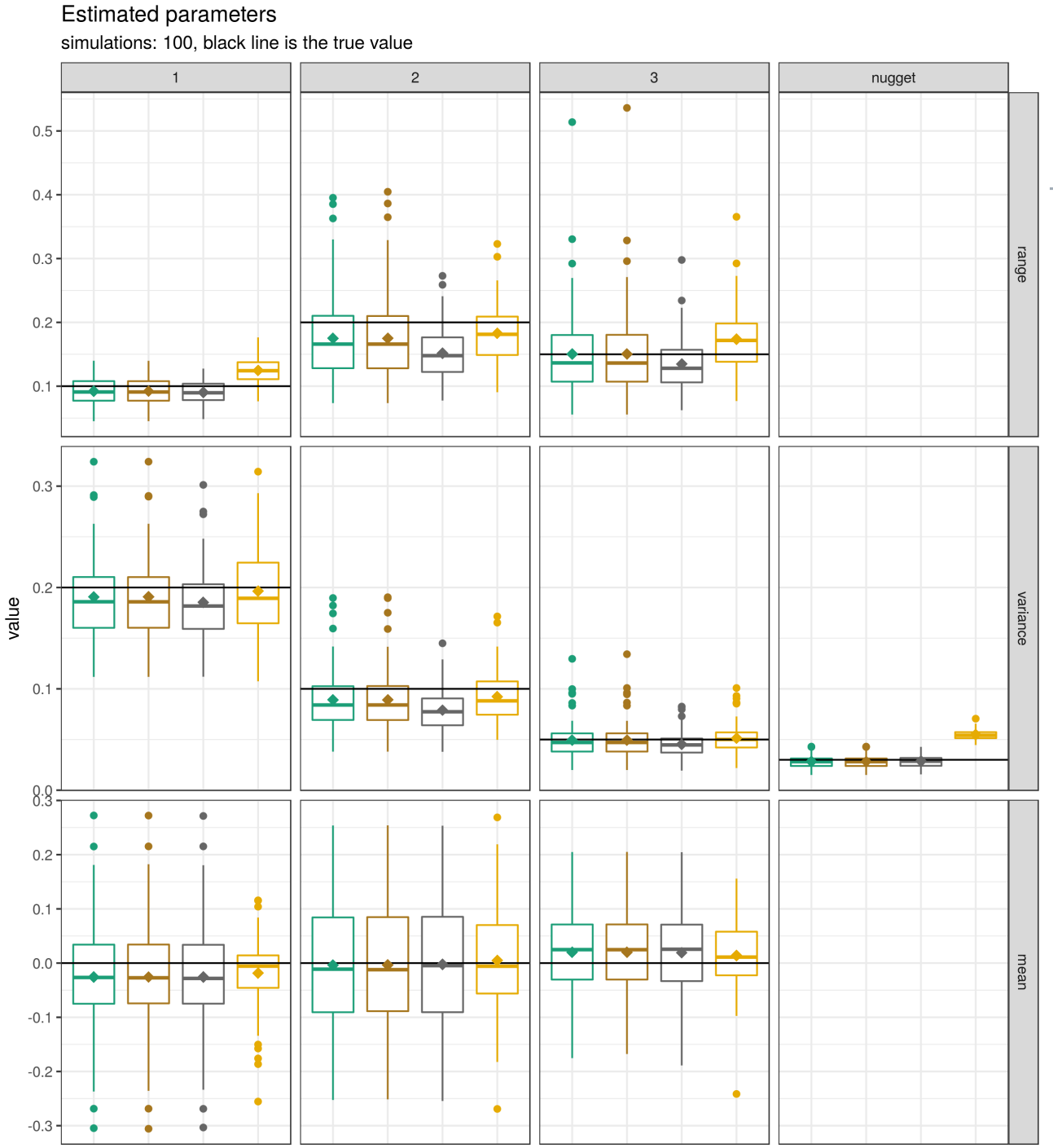
$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\mu} + \boldsymbol{W}\boldsymbol{\eta} + \boldsymbol{\epsilon}$$

- $\boldsymbol{X} = (\mathbf{1}_n \; \boldsymbol{x}^{(2)} \quad \ldots \quad \boldsymbol{x}^{(p)})$
  - Includes intercept
  - All other covariates $\boldsymbol{x}^{(j)} \sim Z$ i. i. d.
- $\boldsymbol{\eta}^{(j)}(\cdot) \sim N_n(\mathbf{0}, \boldsymbol{\Sigma}_j)$
  - fixed range and variance
  - $\boldsymbol{\mu} = \mathbf{0}$

Example of SVC sampled from Perturbed Grid

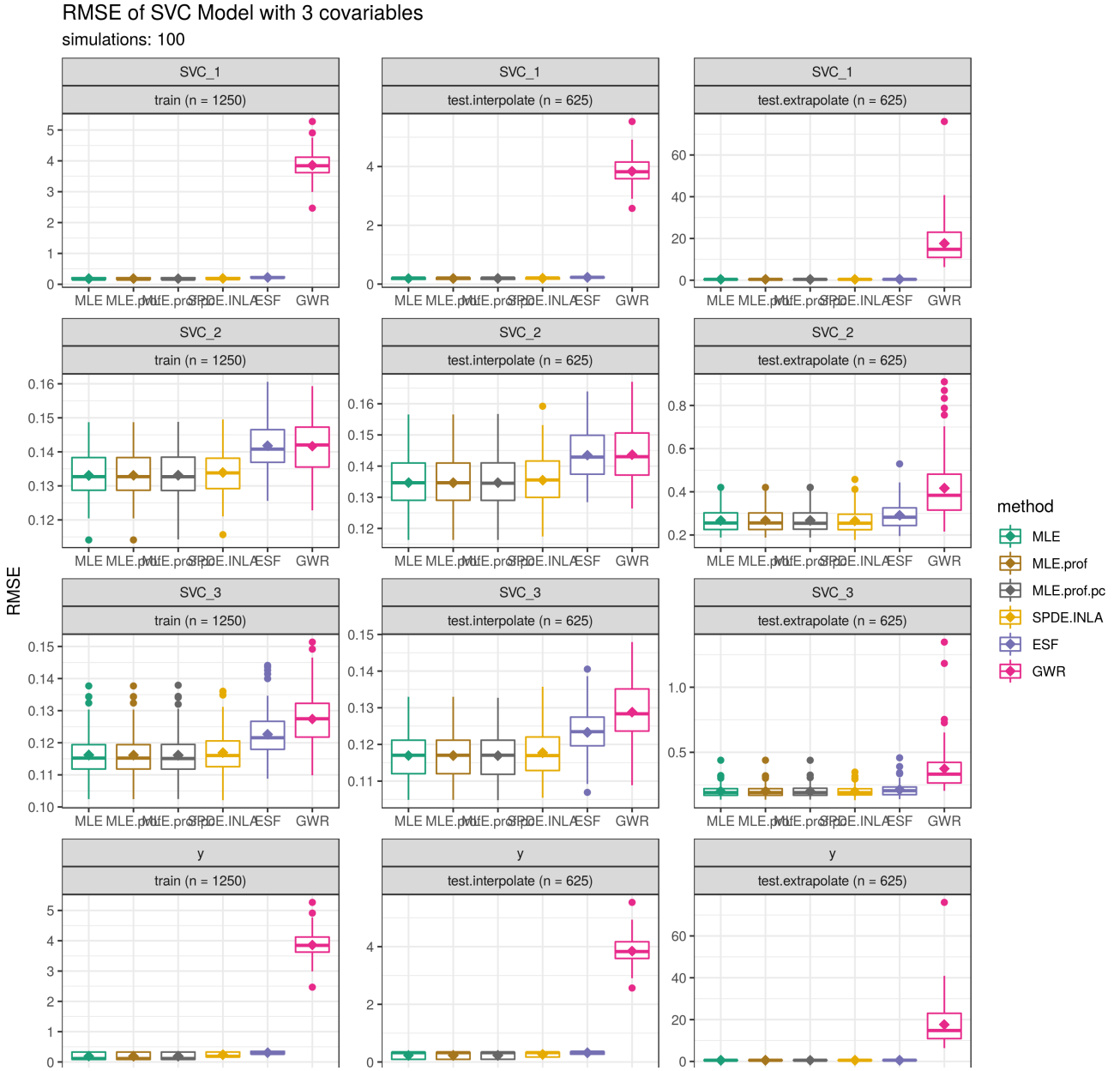
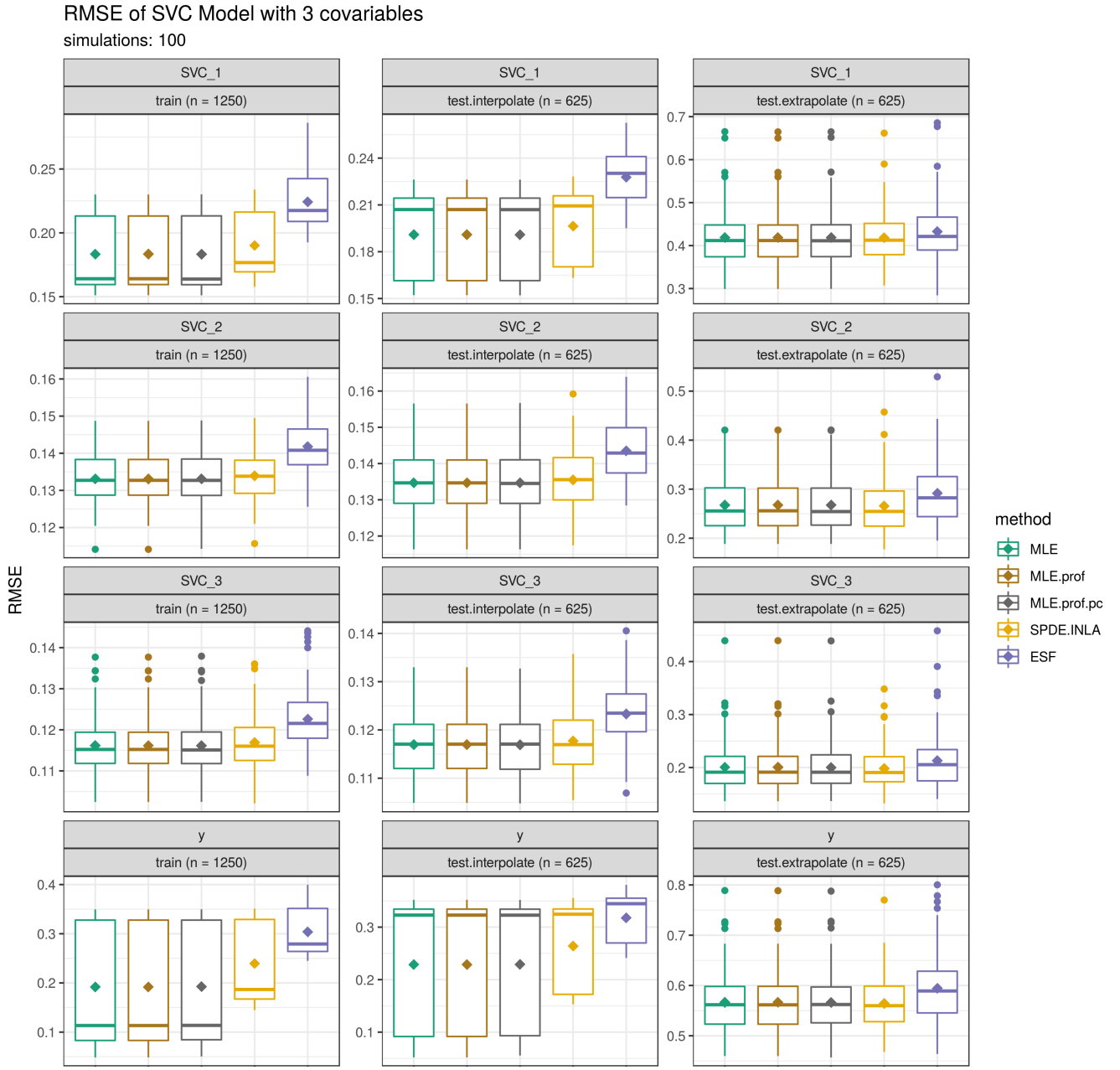


# All Parameter Estimates



# GWR Performance

- Intercept poorly modelled and predicted
- Performance on extrapolation very bad



# Overall Performance