University of Zurich[UZH]

Department of Mathematics

Lucerne University of Applied Sciences and Arts
HOCHSCHULE LUZERN
FH Zentralschweiz

varycoef: An R Package to Model Spatially Varying Coefficients

Jakob A. Dambon[1,2], Fabio Sigrist[2], Reinhard Furrer[1,3]

[1]Department of Mathematics, University of Zurich
[2]Institute of Financial Services Zug, Lucerne University of Applied Sciences and Arts
[3]Department of Computational Science, University of Zurich

Contact

@JakobDambon
jakob.dambon@math.uzh.ch
http://user.math.uzh.ch/dambon

1. Introduction

In regression models for spatial data, it is often assumed that the marginal effects on the response are constant over space. In practice, this assumption might often be questionable. Spatially varying coefficient (SVC) models extend the linear regression model, by allowing the coefficients to vary spatially. The underlying model equation is

$$y(s) = \beta_1(s)x^{(1)}(s) + \beta_2(s)x^{(2)}(s) + \cdots + \beta_p(s)x^{(p)}(s) + \epsilon(s).$$

where s is a location in a domain $D \subseteq R^2$. This model offers a higher degree of flexibility while being easy to interpret.

2. Data and Motivation

Our novel methodology is motivated by a data set of apartment transactions in Switzerland containing the selling price, six covariates and approximated coordinates for each transaction. Real estate mass appraisal is a classical application for SVC models. We define our model as

$$\begin{aligned}\log price = &\ \beta_1(s) + \beta_2(s)\log HNF03 + \beta_3(s)Z.AGE\\ &+\beta_4(s)Z.AGE.sq + \beta_5(s)MICRO + \beta_6(s)STAND\\ &+\beta_7(s)RENOV + \beta_8(s)DAQREQ + \epsilon,\end{aligned}$$

training it on data from 6 consecutive quarters and giving predictions for a following 7th quarter. Tab.1 gives an overview of the data and the SVCs. To the best of our knowledge, there exists no geostatistical methodology for our application as there are too many observations (ca. 1.5×10^4) and too many SVCs.

3. Methodology

Similarly to [Gelfand et al., 2003], we assume that the SVCs follow zero mean Gaussian processes with a stationary covariance function. Each SVC is defined by

$$\eta_j(\cdot) \sim N\left(\mathbf{0}, \mathbf{\Sigma}^{(j)}\right)$$

with $\mathbf{\Sigma}^{(j)}$ depending on some range ρ_j and variance σ_j^2 parameters. Using two data matrices $\boldsymbol{X}$ and $\boldsymbol{W}$, the general SVC model is given by

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

where $\boldsymbol{\mu}$ are the fixed effects, $\boldsymbol{\eta}$ are the joint SVCs modelling the random effects (RE), and $\boldsymbol{\epsilon}$ is the error term with no spatial structure. Further, we assume mutual prior independence of the Gaussian processes.
The response $\boldsymbol{y}$ is normally distributed and we can state its likelihood function given some observed data $\boldsymbol{X}$ and $\boldsymbol{W}$. We use maximum likelihood estimation (MLE) to estimate the mean and covariance parameters. We apply covariance tapering as well as likelihood regularization suggested by [Furrer et al., 2006] and [Fuglstad et al., 2018], respectively, to deal with the large data set. Predictions for SVCs as well as the predictive variance can be computed using the empirical best linear unbiased predictor (EBLUP).

4. Availability

The new methodology has been implemented in the R package `varycoef` which is available on CRAN [Dambon et al., 2019].

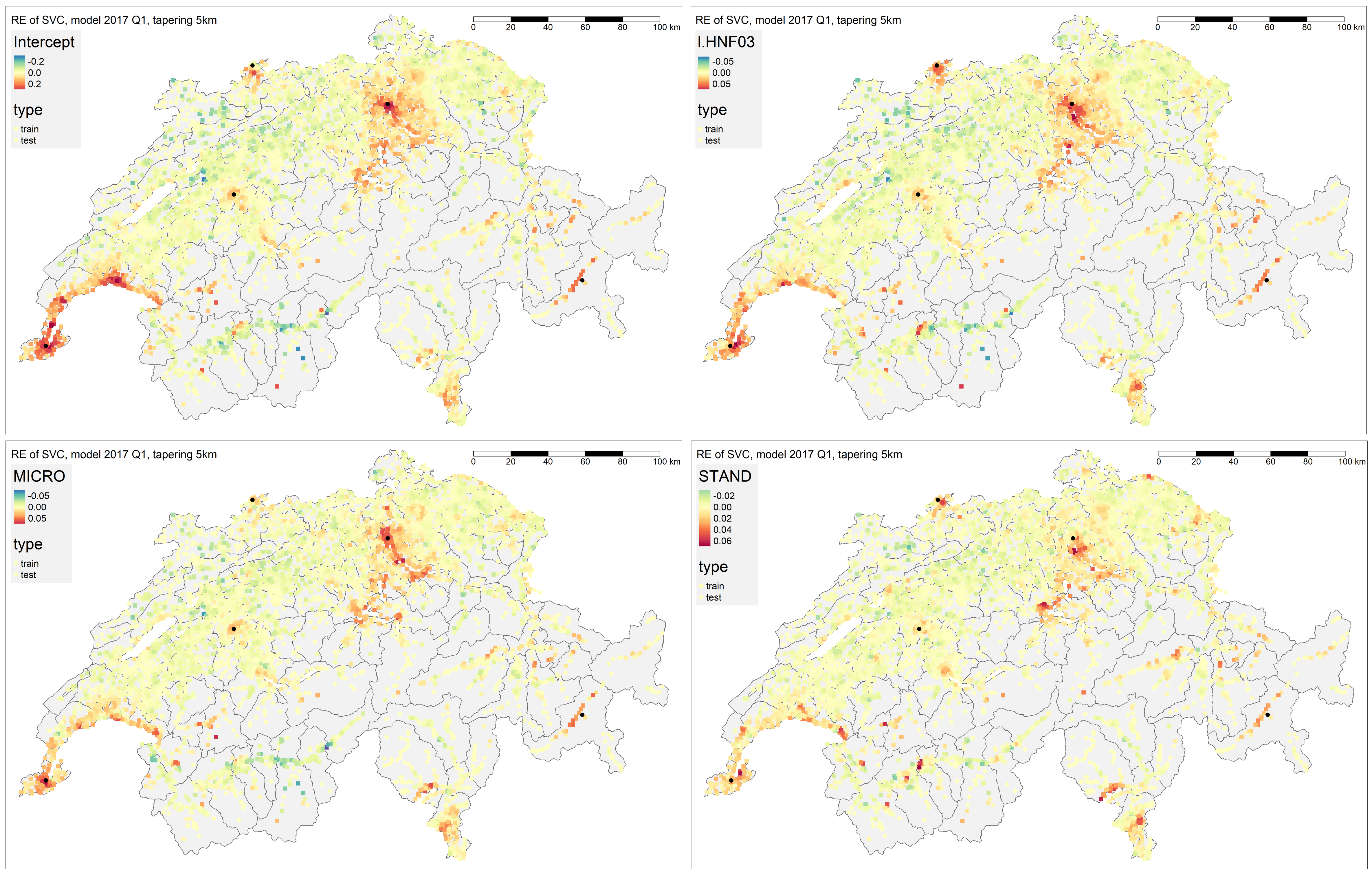

Fig.1: Selection of 4 SVC defined as zero mean Gaussian processes and modeled using proposed MLE method to predict for Q1 2017. Black dots indicate the cities of Geneva, Bern, Basel, Zurich, and St. Moritz (West to East). Covariance tapering range set to 5 km.

5. Results

In an extensive simulation study we were able to show that our proposed novel methodology is one of few, if not the only one, to accurately estimate the covariance parameters of the Gaussian processes.
We analyzed the predictive performance in the before mentioned real estate application using a temporal cross validation (CV) with 4 folds to compare our approach with other SVC methods such as GWR [Fotheringham et al., 2002] or ESF [Murakami and Griffith, 2015]. Fig.1 and Tab.1 show the first fold results for our proposed MLE method. Fig.2 shows the computed root-mean-squared errors (RMSE) over all CV folds.

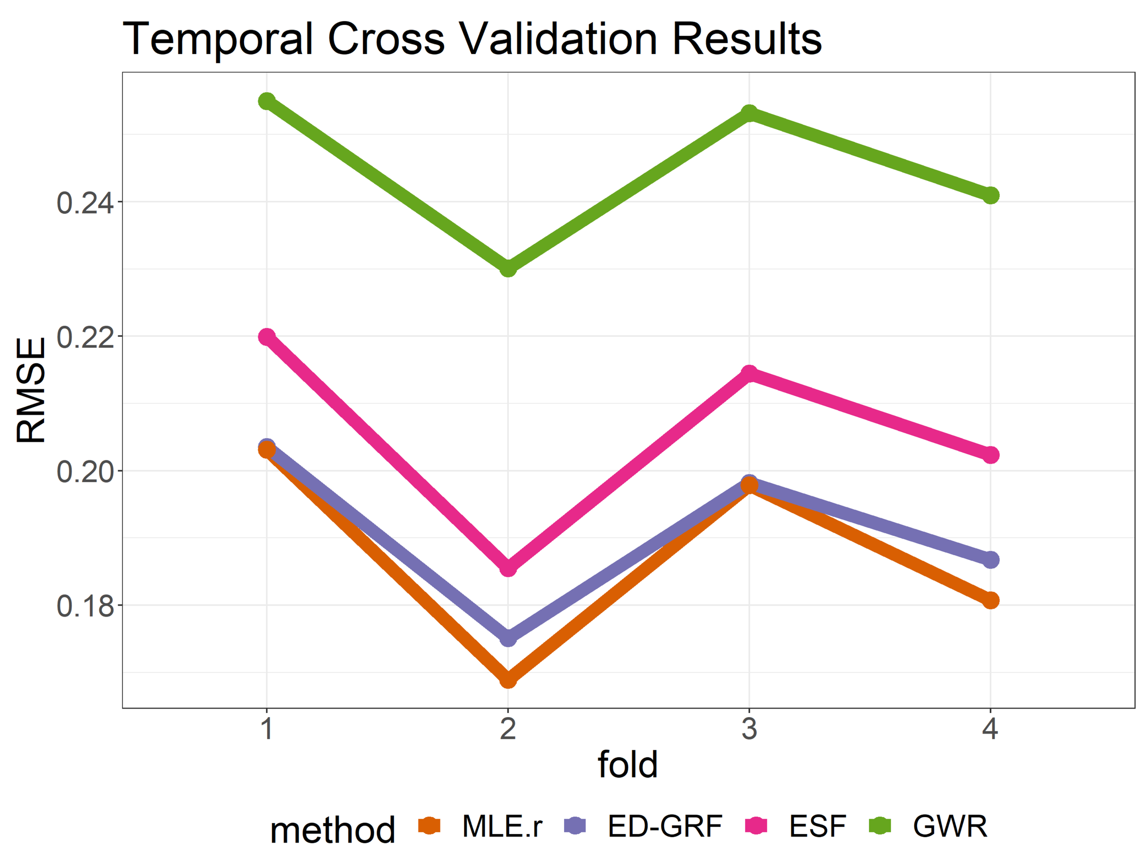

Fig.2: RMSE of the temporal CV over all folds. There are 3 SVC models where the following methods have been applied: our novel regularized MLE (MLE.r), GWR, and ESF. A classical geostatistical model with external drift using MLE (ED-GRF) was added to the CV. Validation in fold 1 was done on Q1 2017, fold 2 on Q2, etc.

Conclusion and Future Work

Our new methodology offers a high degree of flexibility to model SVCs based on geostatistics. It scales to large data both in the number of observations and SVCs. The prediction accuracy outperforms most other available SVC methods.

Future work includes runtime optimization, SVC selection, and the extension on spatio-temporal varying coefficients.

Tab.1: Description of all SVC covariates in the model with corresponding covariate ranges. The last column are ML-estimated fixed effects of a SVC model to predict for Q1 2017.

SVC	Covariate description	Range	$\hat{\mu}_j$
Intercept		1	8.64
HNF03	Area in square meters	3 – 5.7	0.88
Z.AGE	Standardized Age	-0.85 – 4	-0.17
Z.AGE.sq	Squared Z.AGE	0 – 16	0.03
MICRO	Rating of location	1 – 5	0.08
STAND	Rating of standard	1 – 5	0.08
RENOV	Need for renovation	0 – 4	0.04
DAQREQ	Dummy variable indicating last quarter of training data	0 or 1	0.02

References

1. [Dambon et al., 2019] `varycoef`: Varying Coefficients, R package version 0.2.9. https://CRAN.R-project.org/package=varycoef
2. [Fotheringham et al., 2002] Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. *Chichester: Wiley.*
3. [Fuglstad et al., 2018] Constructing Priors that Penalize the Complexity of Gaussian Random Fields. *JASA 114(525)*
4. [Furrer et al., 2006] Covariance Tapering for Interpolation of Large Spatial Datasets. *Journal of Computational and Graphical Statistics 15(3)*
5. [Gelfand et al., 2003] Spatial Modeling with Spatially Varying Coefficient Processes. *JASA 98(462)*
6. [Murakami and Griffith, 2015] Random Effects Specifications in Eigenvector Spatial Filtering: A Simulation Study. *Journal of Geographical Systems 17(4)*