Regularized spatial Bayesian principal component regression for forest inventory

Virpi Junttila[1], **Marko Laine**[2] **and Tuomo Kauranne**[1]
[1] Lappeenranta University of Technology, Finland
[2] Finnish Meteorological Institute, Finland

These days, model based forest inventory often relies on remotely sensed data covering the whole geographical area of the inventory site combined with measurements of the forest characteristics in given field sample plots. All the data can be spatially located, and thus a regression model with spatial residual correlation can be utilized.

Remotely sended data obtained by airborne Light Detection And Ranging (LiDAR) correlates well with many forest inventory characteristics, such as tree height, timber volume and biomass. To construct an accurate and precise model to predict the characteristics over thousands of hectares, LiDAR data is generally supplemented with several hundred field plot measurements. This can be costly and time consuming, especially in difficult terrain.

Different LiDAR and spatial data based sampling designs can reduce the number of field sample plots needed. However, problems arising from features of the LiDAR data, such as a large number of regressors compared to sample size (over-fitting), or strong correlation among the regressors (multicollinearity), may cause additional uncertainty to the model parameter estimates. To overcome these problems, a regularized spatial Bayesian linear model utilizing principal components of the regressors is proposed.

The proposed method is based on a hierarchical Bayesian model for predicting the inventory characteristics by using the LiDAR data. Regularization is performed using a penalty term as a hyper-parameter with shrinkage toward zero in the principal component space of the regressors, effectively picking a low dimensional subspace of the regressors. The spatial dependence is accounted by using a Gaussian process model. The estimation for the shrinkage and the spatial correlation parameters is done via Markov chain Monte Carlo (MCMC) simulation.

With a small number of training plots, this method automatically limits the number of principal components used according to the data, and prevents the use of components less explaining the variability of the original regressors. The model also utilizes the spatial structure of the model residuals if detected within the training set. This presentation describes the method and illustrates its performance with case studies of Finnish forest inventory sites.

Keywords: Spatial Bayesian regression, Regularization, Remotely sensed data based forest inventory, Principal components, Dimension reduction.