PhD Midway Seminar

Simulation Tool and its application

Raju Rimal

03 March, 2017

Introduction

My PhD Plan

PhD Program	
Phase 1	Make a simulation Tool
Phase 2	Apply it for comparing different estimation Methods
Phase 3	Extend the simulation tool for model with background information
Phase 4	Apply it to test multi-matrix extension of PLS models such as LPLS and UPLS

- Make Simulation Tools for multi-response linear model data
- Using the tool, compare various **estimation techniques** and **understand** them
- **Extend** the simulation tool incorporating model with **background information**
- Apply this extended tool to test multi-matrix extension of partial least square (PLS) models such as LPLS and UPLS (both uses background information about X and Y for analysis)

What I learn

- Advanced Multivariate Model and technique to analyze it
- Programming concept for developing statistical packages and applications for various statistical methods
- Extending and improving existing methods in statistics
- And, obviously, to properly document what I have done

Today's Special

Today I will talk about:

- Simulation tool (simulatr) we are building
- A comparative study of various estimation techniques by simulating linear model data using simulatr

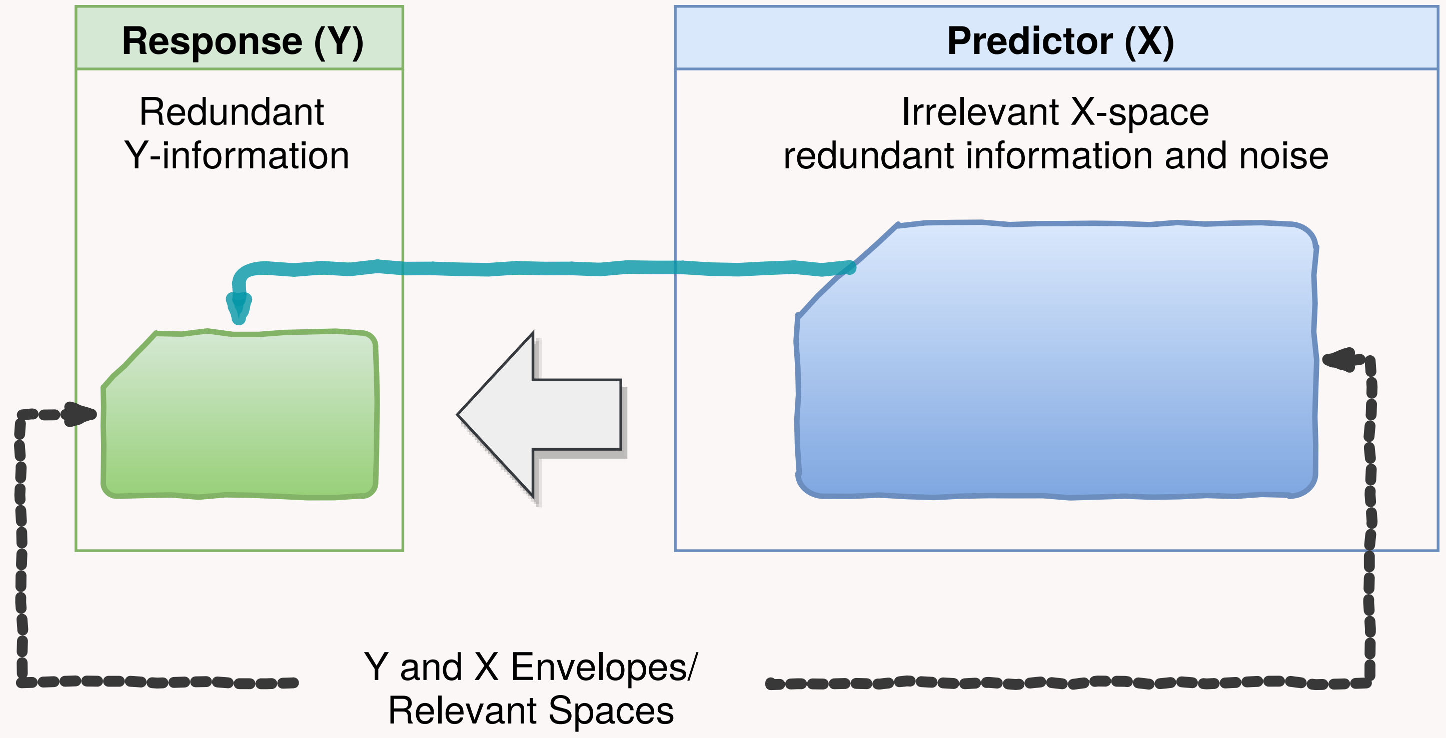
simrel-m: A versatile tool for simulating multi-response linear model data

Overview

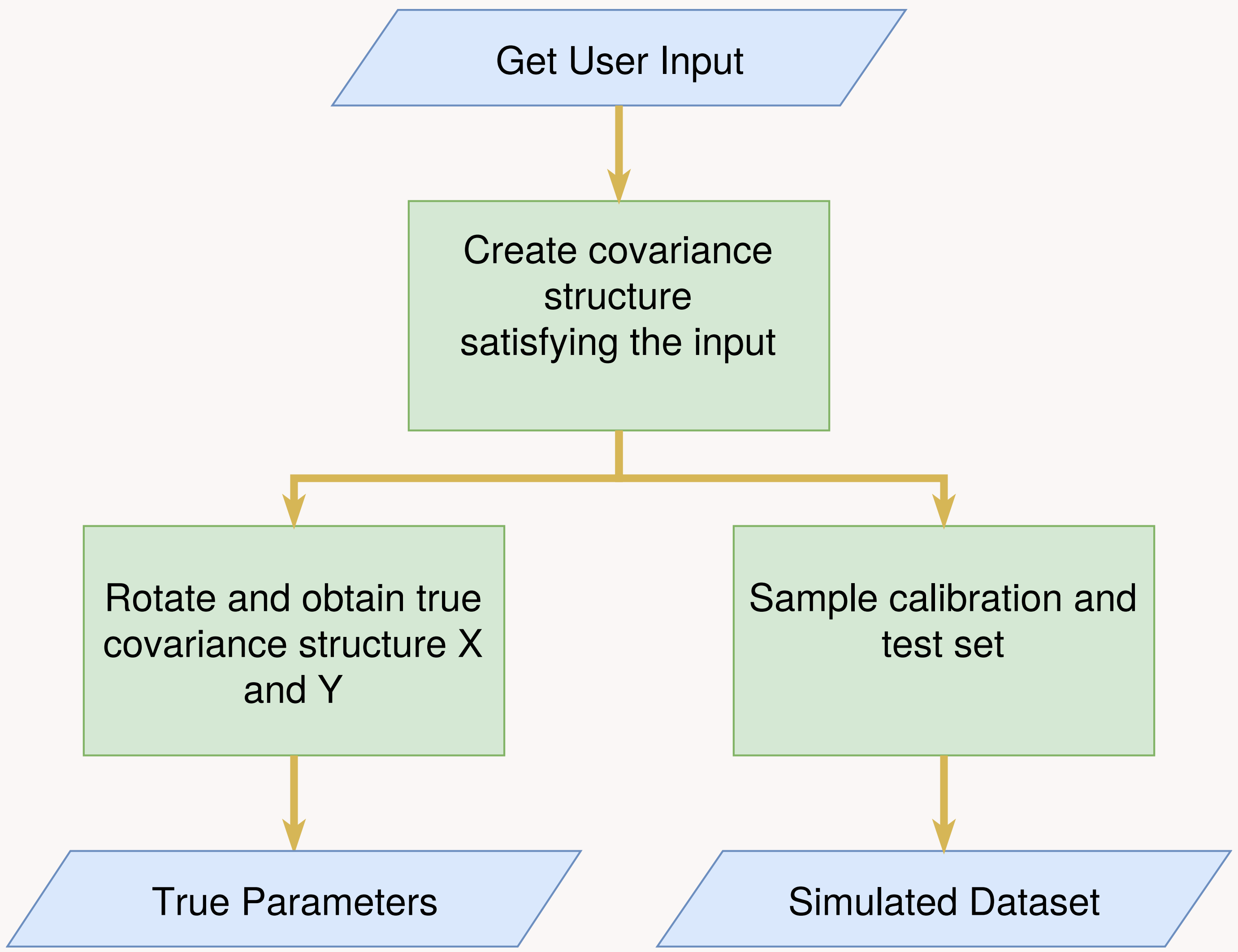
simrel-M is an extension of simrel (Sæbø, Almøy, & Helland, 2015) r-package for simulating multi-response data

- Uses the idea of reduction of random regression model by separating latent space of $\mathbf{X}$ into subspaces that is relevant and irrelevant for predicting each response
- The underlying concept is based on reparameterizing the population model,

$$\mathbf{Y} = \boldsymbol{\mu}_Y + \mathbf{B}^t(\mathbf{X} - \boldsymbol{\mu}_X) + \boldsymbol{\epsilon}, \text{ where } \boldsymbol{\epsilon} \sim N(0, \boldsymbol{\Sigma}_{Y|X})$$



Underlying procedure



- Collect population input parameter from users such as: number of variables, coefficient of determination and the position of relevant components
- Make a covariance matrix satisfying input parameters
- Rotate the covariance matrix orthogonally
- Sample calibration and validation sets

A comparative study of different estimation methods using simulated data

Overview

Four estimtion methods were considered

Ordinary Least Squares (OLS)

- Although unbiased, suffer highly from multicollinearity
- Widely used and can be used as reference for comparison

Partial Least Squares (PLS)

- Well established and widely used method
- Based on Latent Structure and free of multicollinearity problem

Envelope

- Relatively new method (Cook, Helland, & Su, 2013) and is also based on reduction of regression model
- Based on Maximum Likelihood but works better than OLS in p approaches n

Bayes PLS

- Bayesian Estimation of regression coefficient
- Promising performance was shown in previous studies (Helland, Sæbø, & Tjelmeland, 2012)

Simulation Design

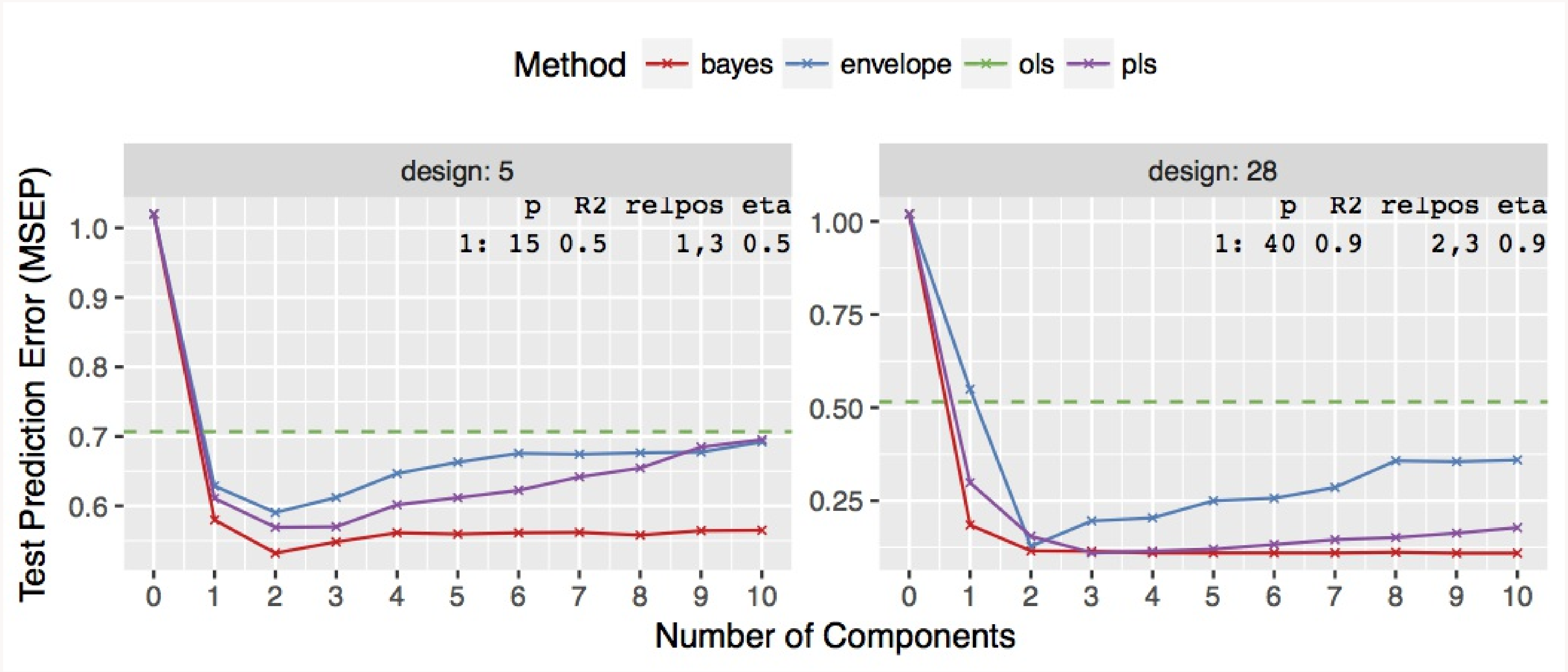
From the possible combination of following parameter combination, *32* single response calibration sets were simulated with *5* replication of each.

- **Number of sample observations**: *50*
- **Number of predictor variables**: *15* and *40*
- **Coefficient of determination** (R^2): *0.5* and *0.9*
- **Level of multicollinearity**: *0.5* and *0.9*
- **Position of relevant components**: *1* and *2*; *1* and *3*; *2* and *3*; *1*, *2* and *3*

A Systematic Comparison

- Bayes PLS has out-performed others methods in all kinds of data
- Envelope has performed better than OLS in all situations and PLS in some situations
- OLS prediction is very poor in noisy data with many predictor variables
- Position of relvant component and the decaying factor of eigenvalue has less impact on prediction in all the models

A Systematic Comparison



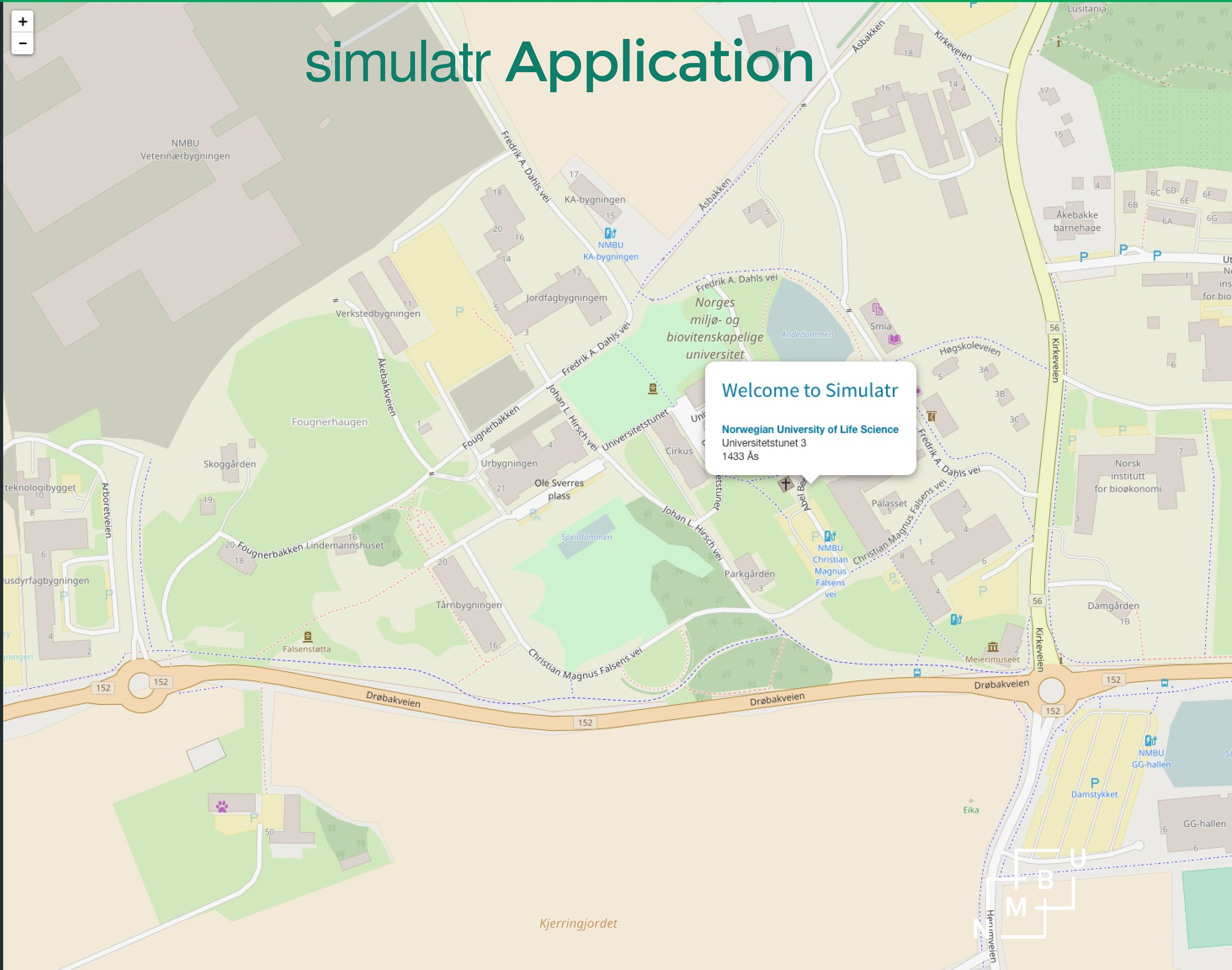
- Bayes PLS has approached to its minimum error with very few component and remained low for additional component
- PLS has moderate performance but better than envelope in many situations.
- OLS prediction is poor especially with large number of predictor
- Envelope method captured its minimum error and the error increased with additional components

Demonstration

NMBU

Seed

simulatr Application



162
818


salamat
Dakujem
teşekkür ederim
GRACiAS
SUWUN
धन्यवाद
TAK
TAKK
ASANTE
hvala
salamat
HVALA
mersi
شكرا جزيل
Ευχαριστώ
감사합니다
GRAZAS
kiitos
merci
GRAZZii
DANKE
salamat
TAKK
MAHALO
Рахмет
Thank You
kiitos
arigato
ARiGATO
takk
suwun
धन्यवाद
HVALA
GRAZiE
DAKUJEM
teşekkür ederim
MERCi
kiitos
hvala
mahalo
GRACiAS
TAKK
GRAZAS
ありがとう
DANKE
Благодарам
TAKK
ASANTE
多謝
salamat
SUWUN
SALAMAT
grazie
спасибо
gracias

References

References

Cook, R., Helland, I., & Su, Z. (2013). Envelopes and partial least squares regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *75*(5), 851–877.

Helland, I. S., Sæbø, S., & Tjelmeland. (2012). Near optimal prediction from relevant components. *Scandinavian Journal of Statistics*, *39*(4), 695–713.

Sæbø, S., Almøy, T., & Helland, I. S. (2015). Simrel—A versatile tool for linear model data simulation based on the concept of a relevant subspace and relevant predictors. *Chemometrics and Intelligent Laboratory Systems*, *146*, 128–135.