simrel-m

A simulation tool and its application

Raju Rimal

Supervisors:

Solve Sæbø and Trygve Almøy

http://mathatistics.github.io/nsm-17

11 June, 2017

Man is a tool-using animal. Without tools he is nothing, with tools he is all.

— Thomas Carlyle

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

Why `simrel-m`

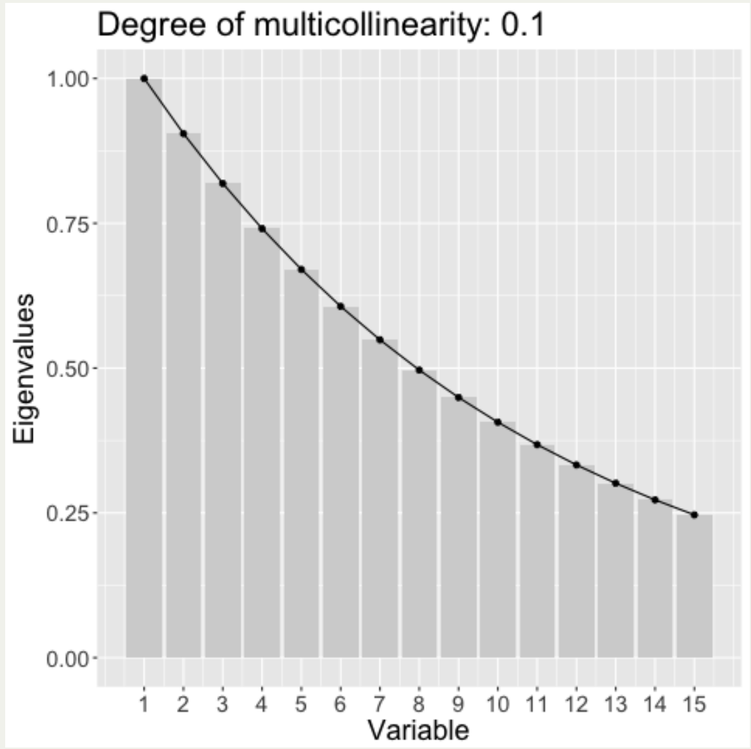
Why `simrel-m`

- By changing few parameters, we can simulate wide range of linear model data.
  For example,

Why `simrel-m`

- By changing few parameters, we can simulate wide range of linear model data. For example,

  1. Controlling degree of multicollinearity in the simulated data
  2. Specifying the relevant principle components for prediction

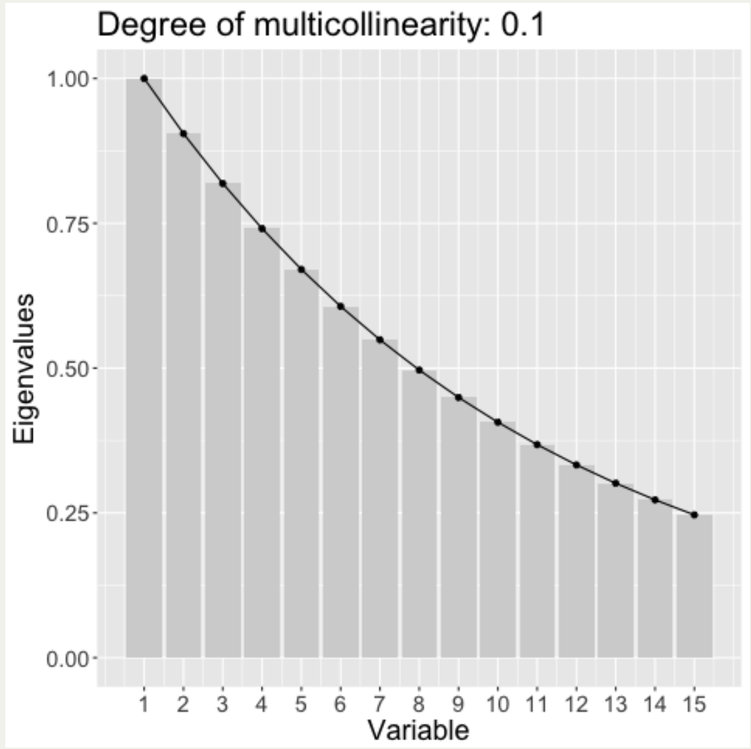


Why `simrel-m`

- By changing few parameters, we can simulate wide range of linear model data. For example,

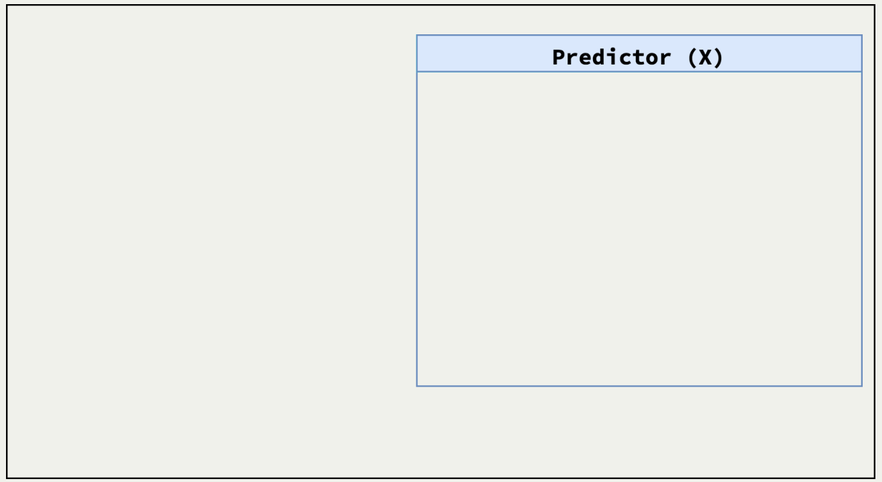
  1. Controlling degree of multicollinearity in the simulated data
  2. Specifying the relevant principle components for prediction

- It is easy to use and has wide application



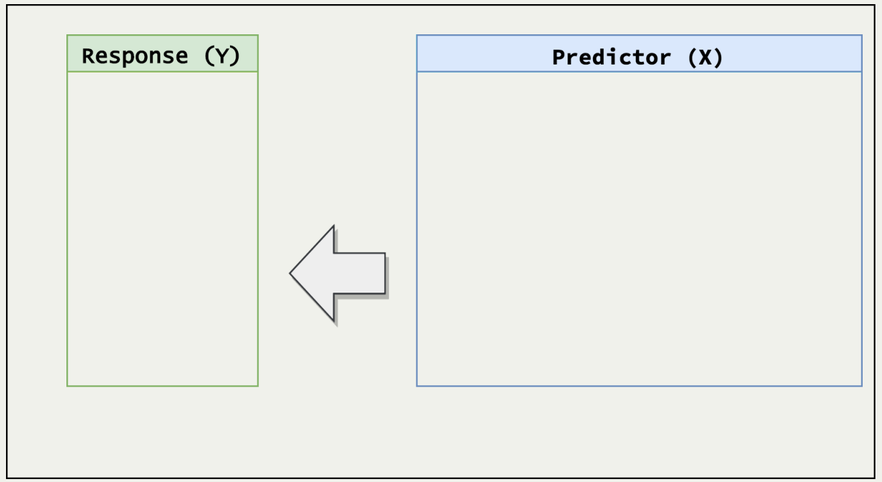
The idea behind

The idea behind



- Based on `simrel`[1] package
- Predictor Space (Blue Box)

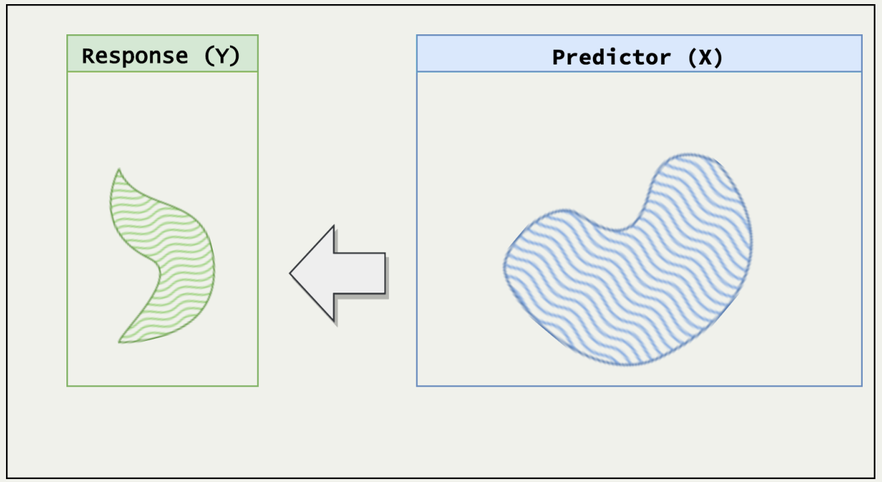
The idea behind



- Based on `simrel`[1] package
- Predictor Space (Blue Box)
- A model defines its relationship with

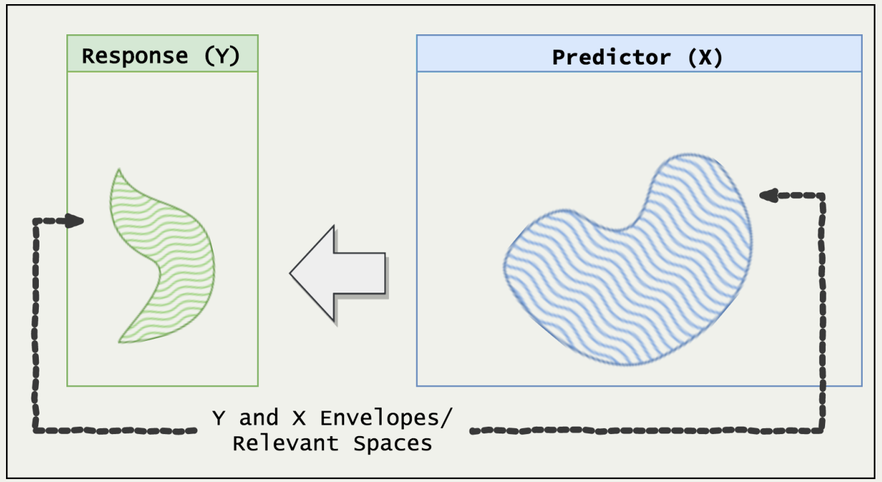
Response Space (Green Box)

The idea behind



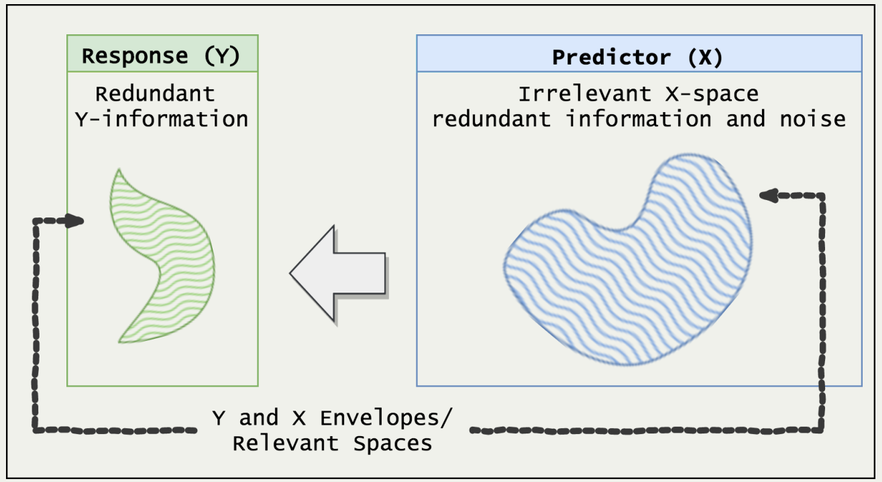
- Based on `simrel`[1] package
- Predictor Space (Blue Box)
- A model defines its relationship with Response Space (Green Box)
- Subspace within these spaces (a reduced regression model) contains information for this relationship

The idea behind



- Based on `simrel`[1] package
- Predictor Space (Blue Box)
- A model defines its relationship with Response Space (Green Box)
- Subspace within these spaces (a reduced regression model) contains information for this relationship

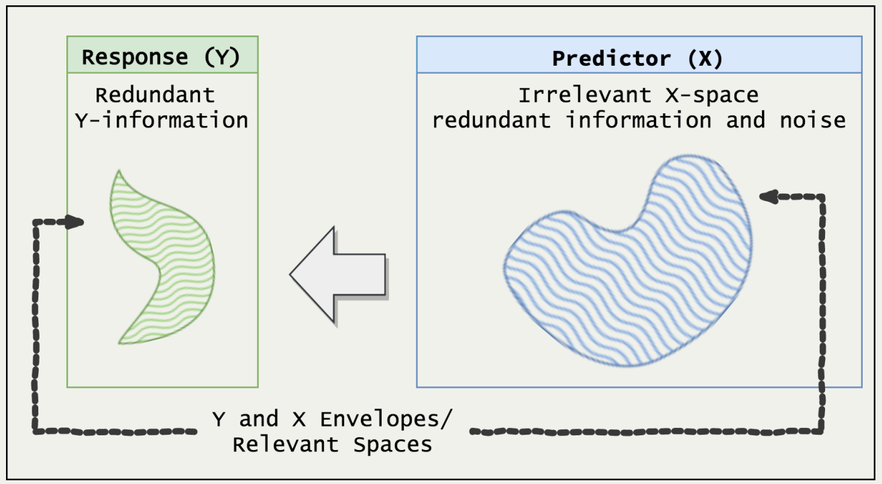
The idea behind



- Based on `simrel`[1] package
- Predictor Space (Blue Box)
- A model defines its relationship with Response Space (Green Box)
- Subspace within these spaces (a reduced regression model) contains information for this relationship

- Set of orthogonal variables (Z) span the relevant predictor subspace (predictor components)

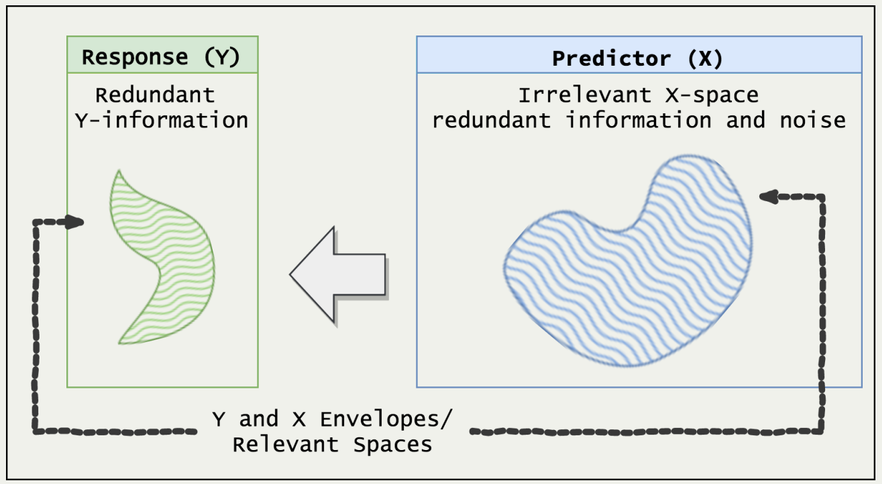
The idea behind



- Based on `simrel`[1] package
- Predictor Space (Blue Box)
- A model defines its relationship with Response Space (Green Box)
- Subspace within these spaces (a reduced regression model) contains information for this relationship

- Set of orthogonal variables (Z) span the relevant predictor subspace (predictor components)
- Set of orthogonal variables (W) span the response subspace (response components)

The idea behind

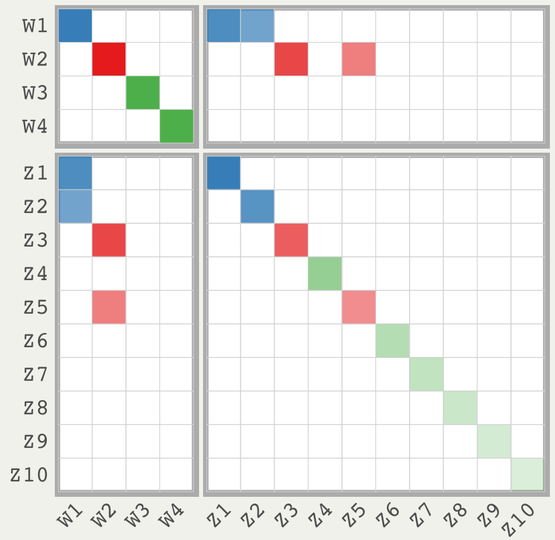


- Based on `simrel`[1] package
- Predictor Space (Blue Box)
- A model defines its relationship with Response Space (Green Box)
- Subspace within these spaces (a reduced regression model) contains information for this relationship

- Set of orthogonal variables (Z) span the relevant predictor subspace (predictor components)
- Set of orthogonal variables (W) span the response subspace (response components)
- Implement this idea to construct the relevant covariance matrix and make simulation with it

How it works

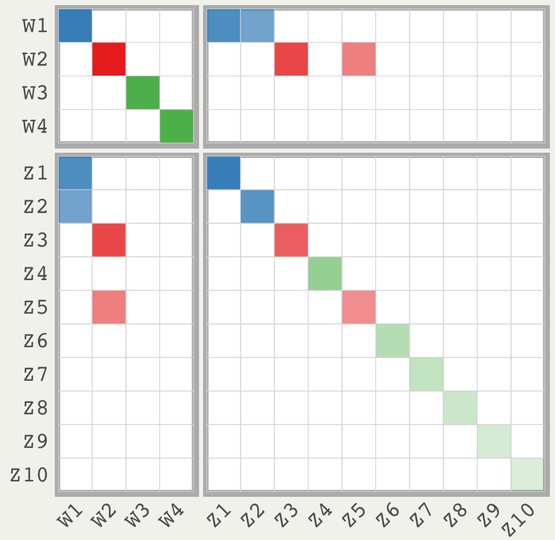
How it works



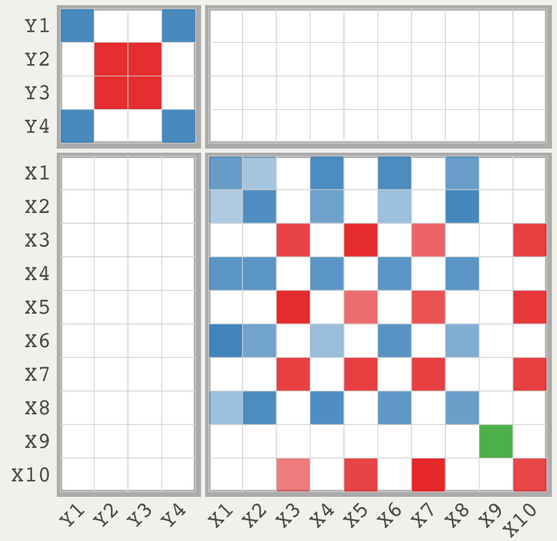
Relevant for: W2 W1 None

- Gets parameter setting from users
- Creates Covariance matrix

How it works



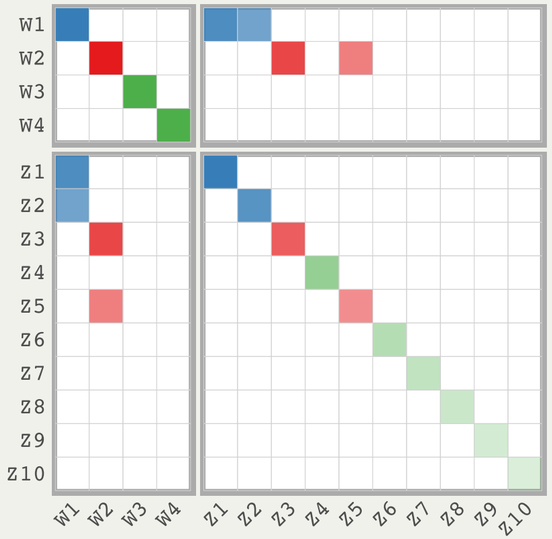
Relevant for: W2 W1 None



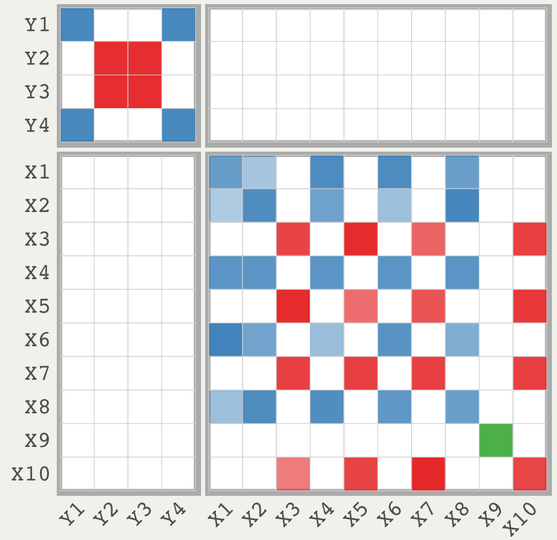
Relevant for: Y2, Y3 Y1, Y4 None

- Gets parameter setting from users
- Creates Covariance matrix
- Creates Rotation Matrix
- Rotates the sampled Latent variables

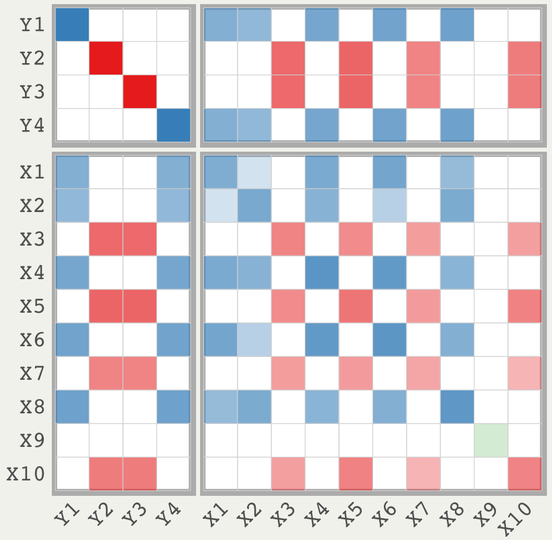
How it works



Relevant for: W2 W1 None



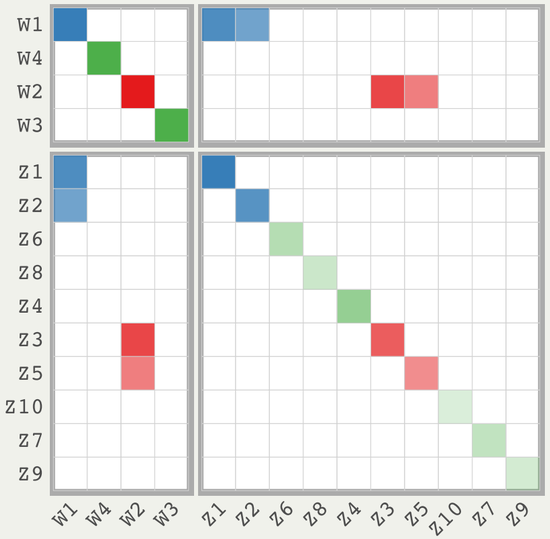
Relevant for: Y2, Y3 Y1, Y4 None



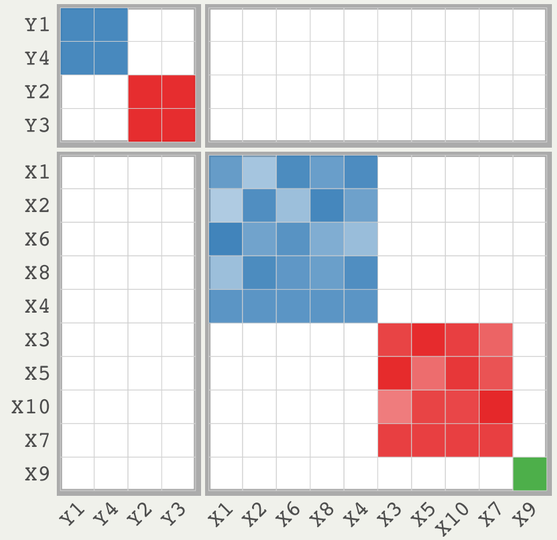
Relevant for: Y2, Y3 Y1, Y4 None

- Gets parameter setting from users
- Creates Covariance matrix
- Creates Rotation Matrix
- Rotates the sampled Latent variables

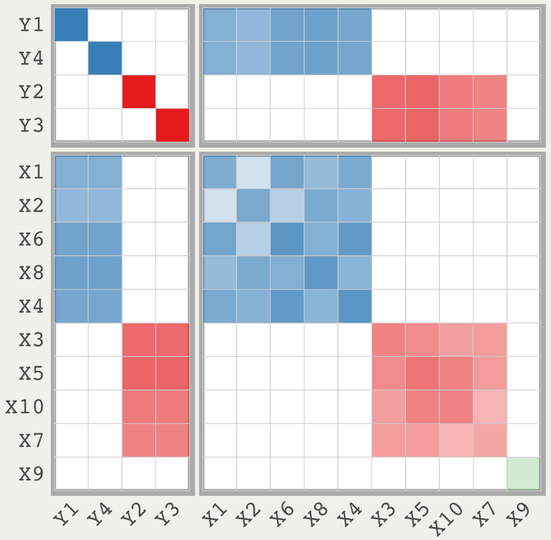
How it works



Relevant for: W2 W1 None



Relevant for: Y2, Y3 Y1, Y4 None



Relevant for: Y2, Y3 Y1, Y4 None

- Gets parameter setting from users
- Creates Covariance matrix
- Creates Rotation Matrix
- Rotates the sampled Latent variables

A web interface

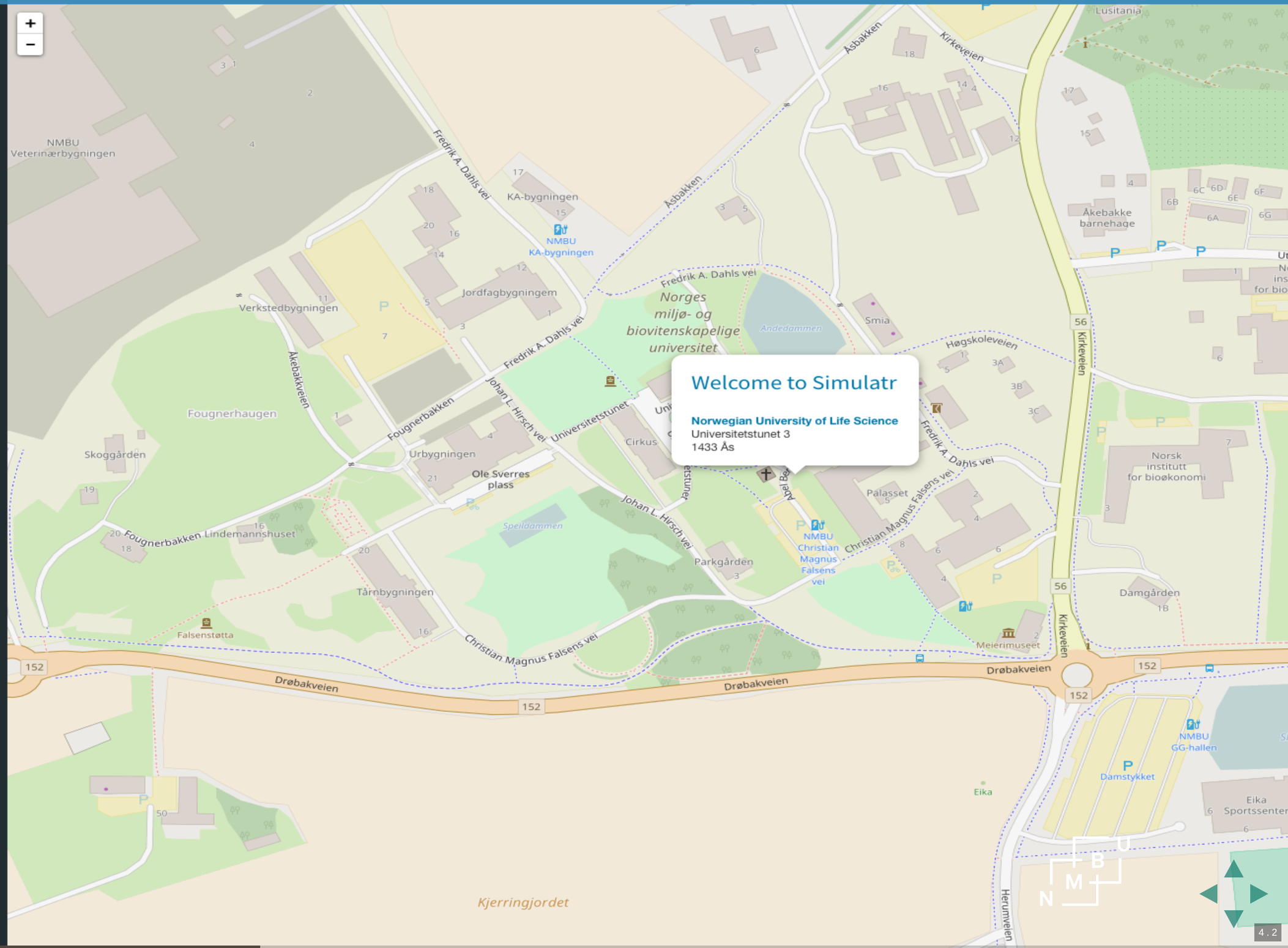
SimulatR NMBU

Simulate Now

123

New Seed

Parameter Settings



SimulatR NMBU

Simulate Now

123

New Seed

Parameter Settings

Type of simulation:

Multivariate Simulation

N: Train

200

N: Test

50

N: Predictors

15

Rel.Pred

5, 4

Coef. Determination

0.8, 0.7

RelPos.Comp

1, 2; 3, 4, 6

Gamma

0 0.6 1

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

N: Response

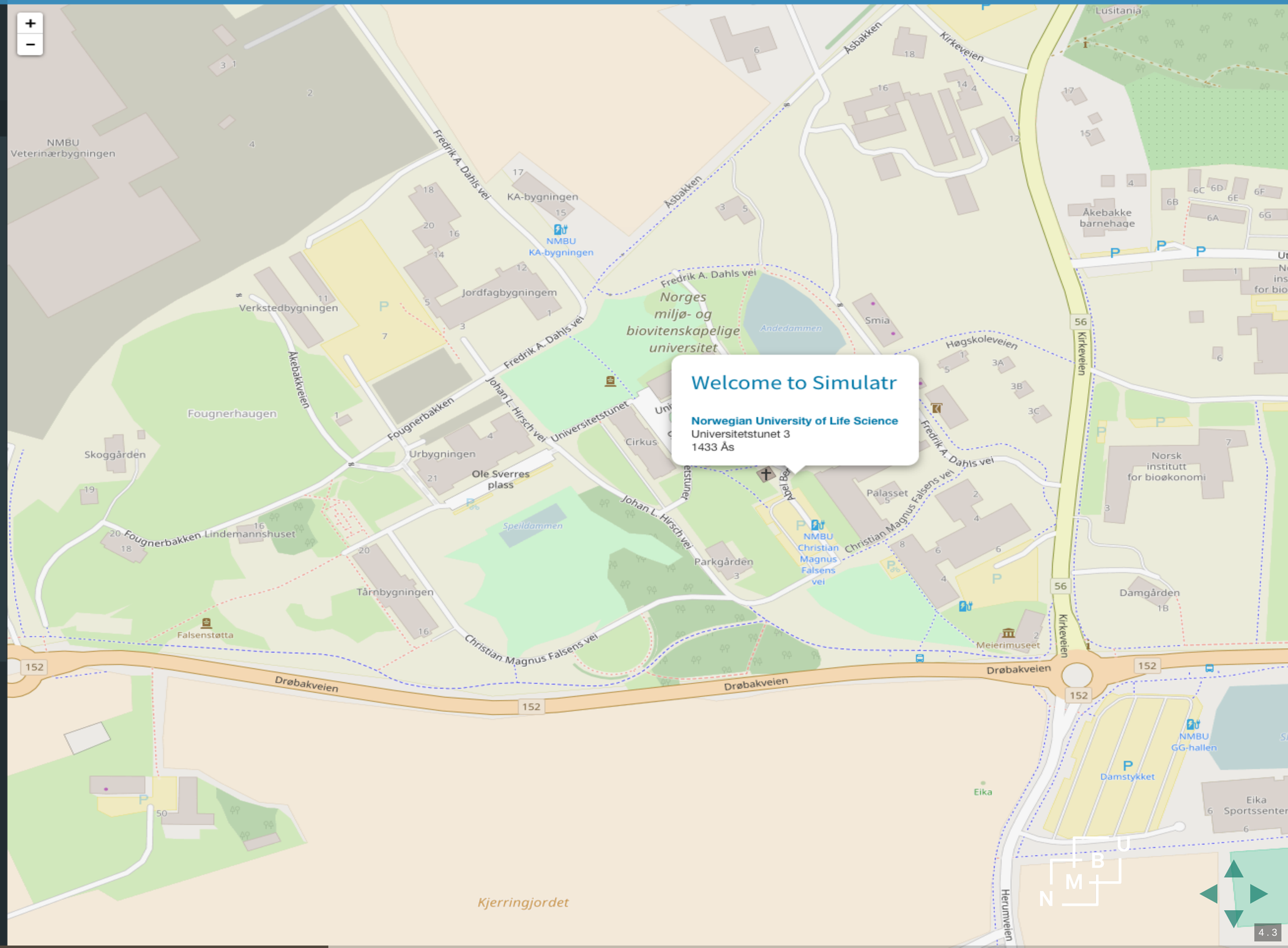
4

Response Mixup

1, 3; 2, 4

Display Extra Plots

Covariance Plot R-squared Plot



4.3

SimulatR NMBU

Simulate Now

123

New Seed

Parameter Settings

Type of simulation:

Multivariate Simulation

N: Train

200

N: Test

50

N: Predictors

15

Rel.Pred

5, 4

Coef. Determination

0.8, 0.7

RelPos.Comp

1, 2; 3, 4, 6

Gamma

0 0.6 1

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

N: Response

4

Response Mixup

1, 3; 2, 4

Display Extra Plots

Covariance Plot R-squared Plot

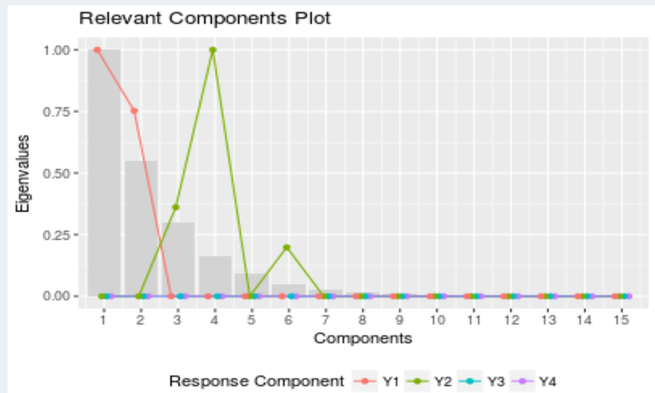
Simulation Overview

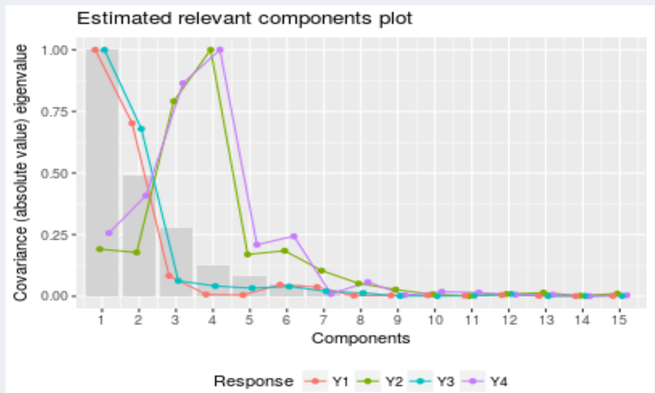
Estimation

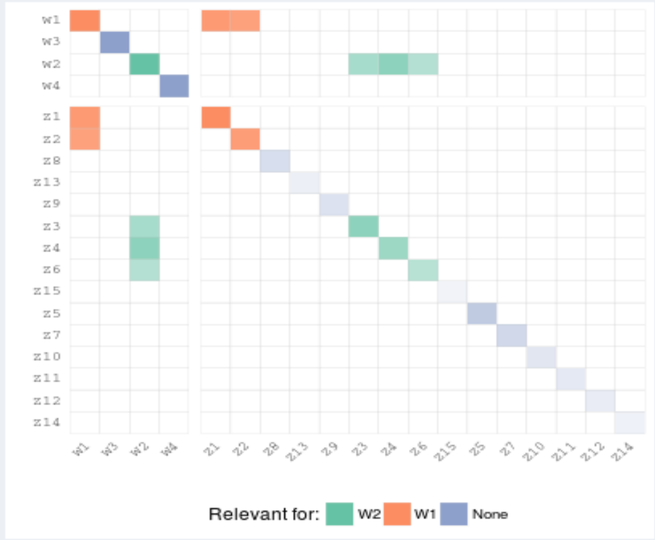
Method Comparison

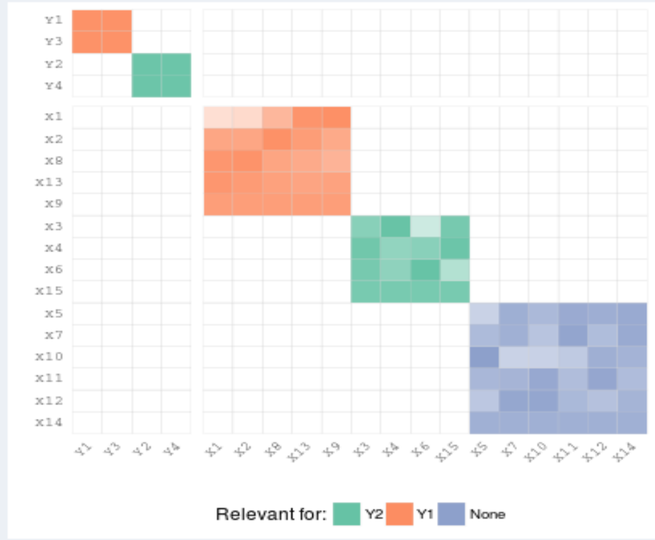
Flowchart

SimObj RData JSON









SimulatR NMBU

Simulate Now

123

New Seed

Parameter Settings

Type of simulation:

Multivariate Simulation

N: Train

200

N: Test

50

N: Predictors

15

Rel.Pred

5, 4

Coef. Determination

0.8, 0.7

RelPos.Comp

1, 2; 3, 4, 6

Gamma

0 0.2 1

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

N: Response

4

Response Mixup

1, 3; 2, 4

Display Extra Plots

Covariance Plot R-squared Plot

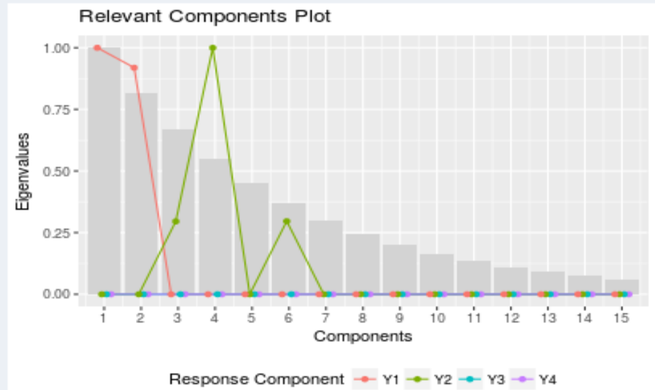
Simulation Overview

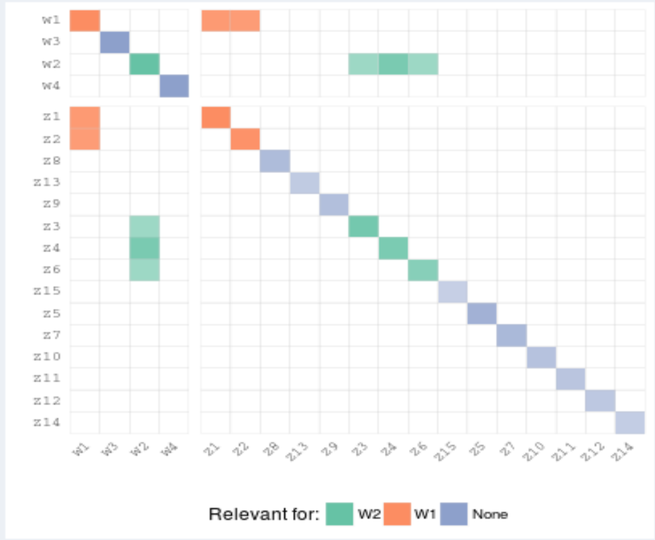
Estimation

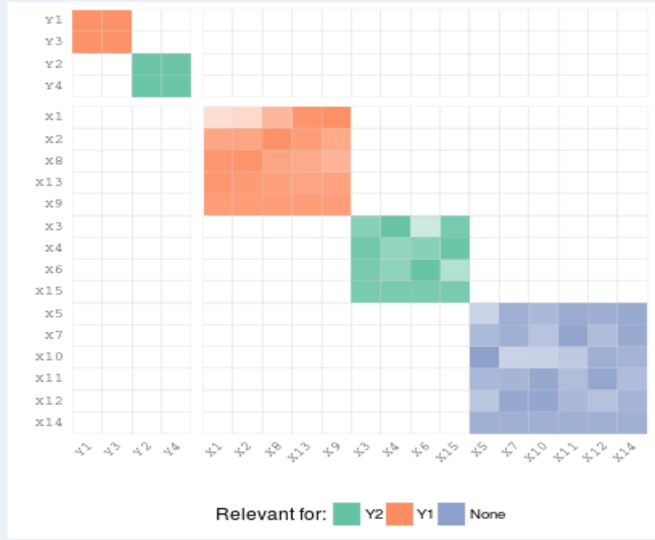
Method Comparison

Flowchart

SimObj RData JSON







SimulatR NMBU

Simulate Now

1963

New Seed

Parameter Settings

Type of simulation:

Multivariate Simulation

N: Train

200

N: Test

50

N: Predictors

15

Rel.Pred

5, 4

Coef. Determination

0.8, 0.7

RelPos.Comp

1, 2; 3, 4, 6

Gamma

0 0.2 1

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

N: Response

4

Response Mixup

1, 3; 2, 4

Display Extra Plots

Covariance Plot R-squared Plot

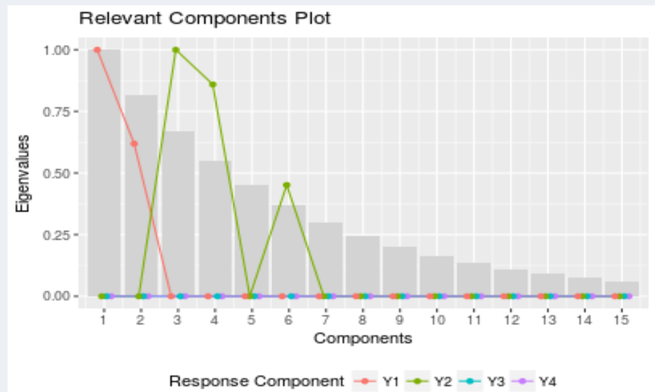
Simulation Overview

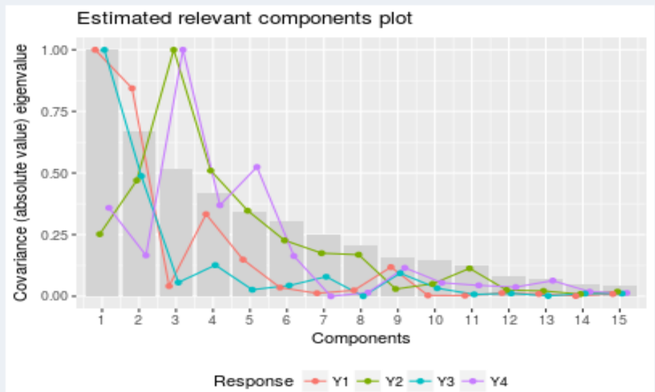
Estimation

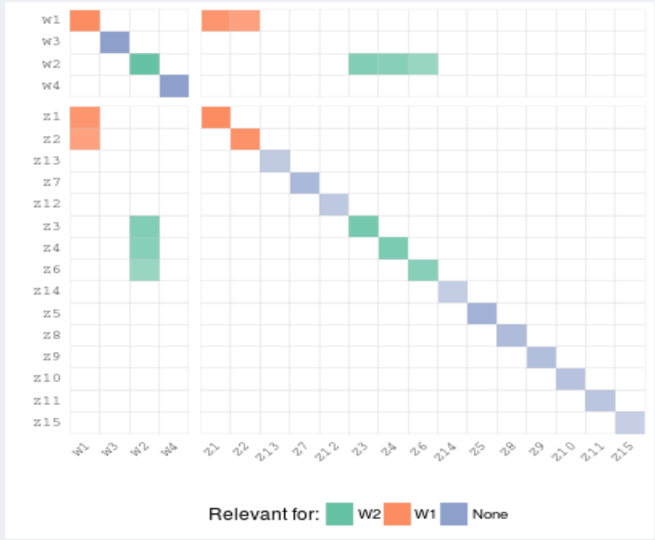
Method Comparison

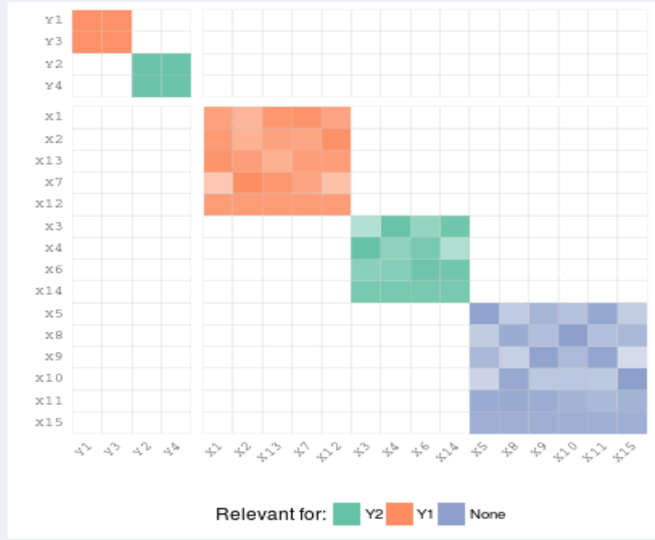
Flowchart

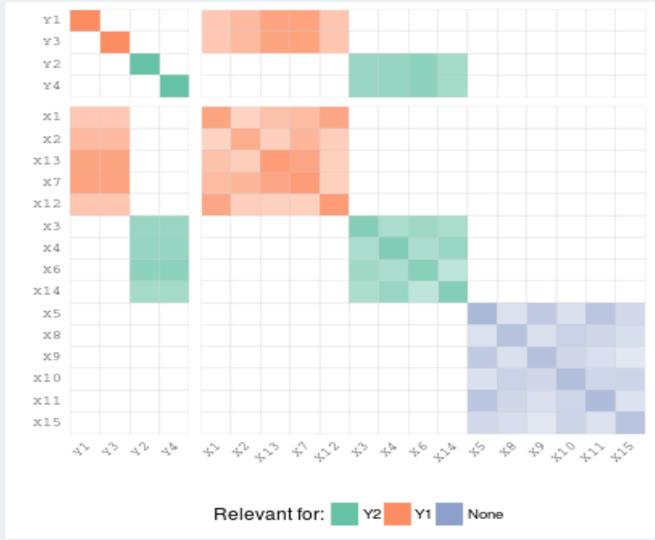
SimObj RData JSON











4.6

How to get it

Install simrel-m:

```
devtools::install_github(
  "therimalaya/simulatr",
  quiet = TRUE
)
```

Run the shiny app:

```
shiny::runGitHub(
  "AppSimulatr",
  "therimalaya"
)
```

Documentation:

https://therimalaya.github.io/simulatr/

An example of comparison of estimation methods

Design Properties

Consider two sets of data, both having following common properties,

Number of observation	100
Number of variables	16
Number of predictors relevant for each response components	5, 5, 5
Number of response variables	5
Relevant position of response component	1, 6; 2, 5; 3, 4
Position of Response components to rotate together	1, 4; 2, 5; 3

Design Properties

Consider two sets of data, both having following common properties,

Number of observation	100
Number of variables	16
Number of predictors relevant for each response components	5, 5, 5
Number of response variables	5
Relevant position of response component	1, 6; 2, 5; 3, 4
Position of Response components to rotate together	1, 4; 2, 5; 3

The difference between the two datasets are

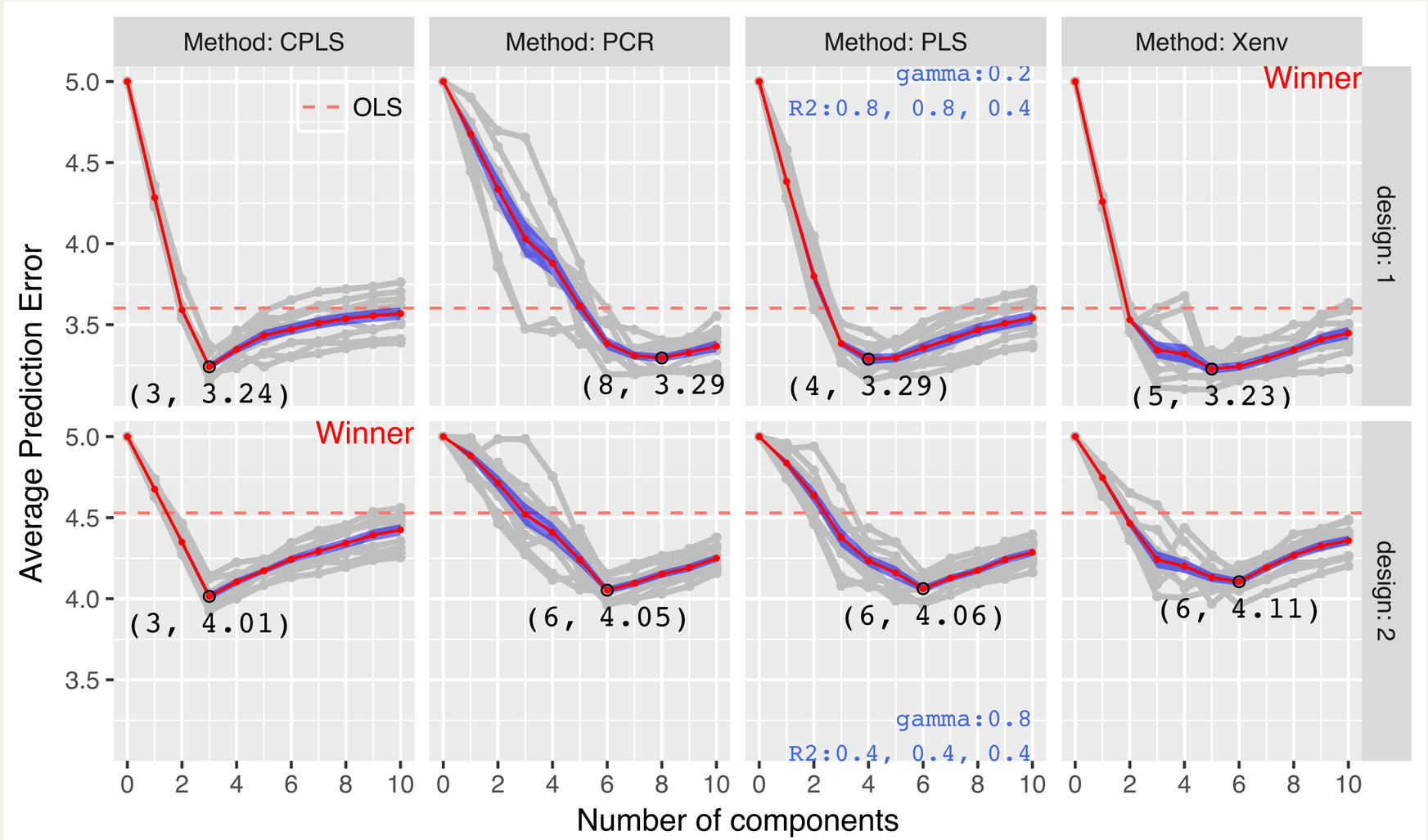
	Design1	**Design2**
Decay of eigenvalue (γ)	0.2	0.8
Coef. of Determination (ρ^2)	0.8, 0.8, 0.4	0.4, 0.4, 0.4

Estimation Methods

For comparison, let's consider the following estimation methods,

- Ordinary Least Squares (`ols`)
- Principle Component Regression (`pcr`)
- Partial Least Squares (`pls`) [2]
- Canonical Partial Least Squares (`cpls`) [3]
- Envelope Estimation of predictor space (`env`) [4]

A Comparison



Some Cases

Case I

- Testing new estimation Methods
- Studying its properties
- Studying its performance in data with various properties

Case I

- Testing new estimation Methods
- Studying its properties
- Studying its performance in data with various properties

Case II

- Educational use
- Students can learn how a method such as variable selection removes irrelevant variables
- Students can observe and study the loading weights on relevant and irrelevant principle components

Case I

- Testing new estimation Methods
- Studying its properties
- Studying its performance in data with various properties

Case II

- Educational use
- Students can learn how a method such as variable selection removes irrelevant variables
- Students can observe and study the loading weights on relevant and irrelevant principle components

Case III

- Comparing various methods (estimation methods, variable selection techniques)


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

References

References

[1] S. Sæbø, T. Almøy, I.S. Helland, 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 (2015) 128–35.

[2] H. Wold, Partial least squares, Encyclopedia of Statistical Sciences (1985).

[3] U.G. Indahl, K.H. Liland, T. Næs, Canonical partial least squaresa unified pls approach to classification and regression problems, Journal of Chemometrics 23(9) (2009) 495–504.

[4] R.D. Cook, B. Li, F. Chiaromonte, Envelope models for parsimonious and efficient multivariate linear regression, Statistica Sinica (2010) 927–60.

[5] I.S. Helland, Partial least squares regression and statistical models, Scandinavian Journal of Statistics (1990) 97–114.