

A versatile tool for simulating linear model data

Raju Rimal, Trygve Almøy and Solve Sæbø

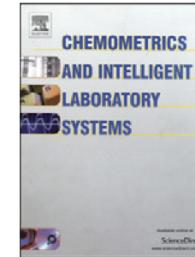
Wednesday, 04 July 2018 | <https://goo.gl/9xDZqk>



Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab



Software Description

simrel – A versatile tool for linear model data simulation based on the concept of a relevant subspace and relevant predictors



Solve Sæbø ^{a,*}, Trygve Almøy ^a, Inge S. Helland ^b

^a Dep. Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, P.O. Box 5003, NO-1432 Ås, Norway

^b Department of Mathematics, University of Oslo, P.O. Box 1053, NO-0316 Oslo, Norway

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

— Thomas Carlyle, Scottish Philosopher



Contents lists available at [ScienceDirect](#)

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemometrics



A tool for simulating multi-response linear model data

Raju Rimal ^{a,*}, Trygve Almøy ^a, Solve Sæbø ^b



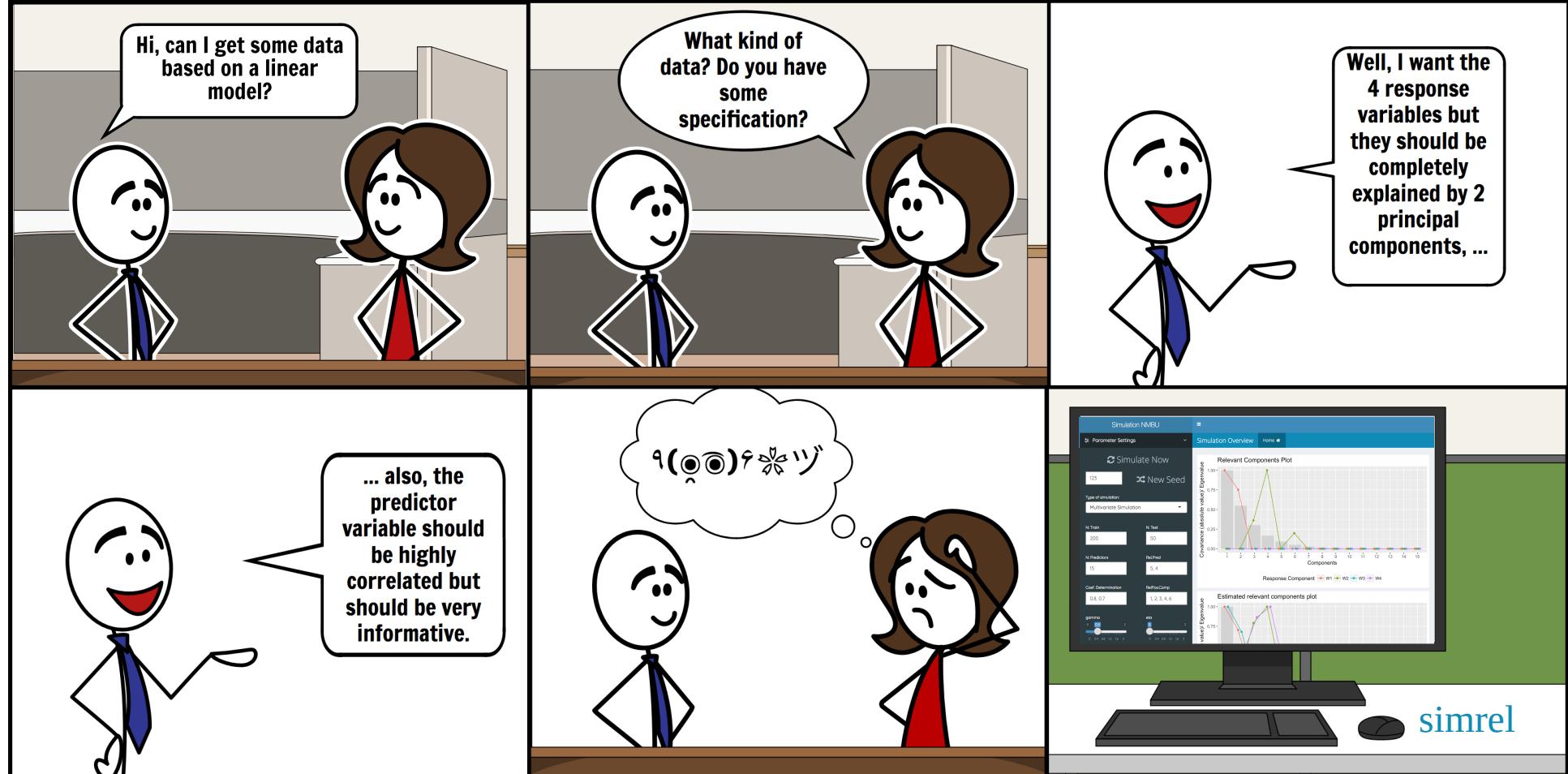
^a Faculty of Chemistry and Bioinformatics, Norwegian University of Life Sciences, Ås, Norway

^b Prorektor, Norwegian University of Life Sciences, Ås, Norway

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

— Thomas Carlyle, Scottish Philosopher

Why simrel



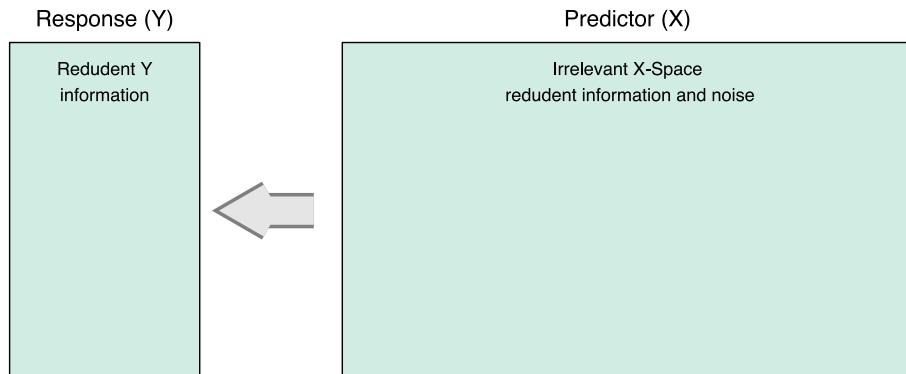
The Model

$$\begin{bmatrix} y \\ x \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mu_y \\ \mu_x \end{bmatrix}, \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix}_{(m+p) \times (m+p)} \right)$$

There are $\frac{1}{2}(p+m)(p+m+1)$ unknowns to identify this model.

Relevant space within a model

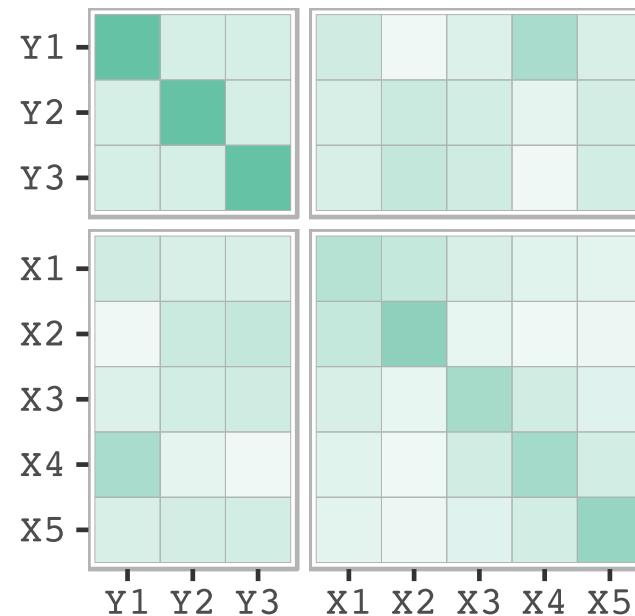
A concept behind reduction of regression model



equivalently,

$$\mathbf{y} = \beta_0 + \boldsymbol{\beta}^t \mathbf{x} + \boldsymbol{\epsilon}$$

also, we can express, $\boldsymbol{\beta} = \Sigma_{xx}^{-1} \Sigma_{xy}$



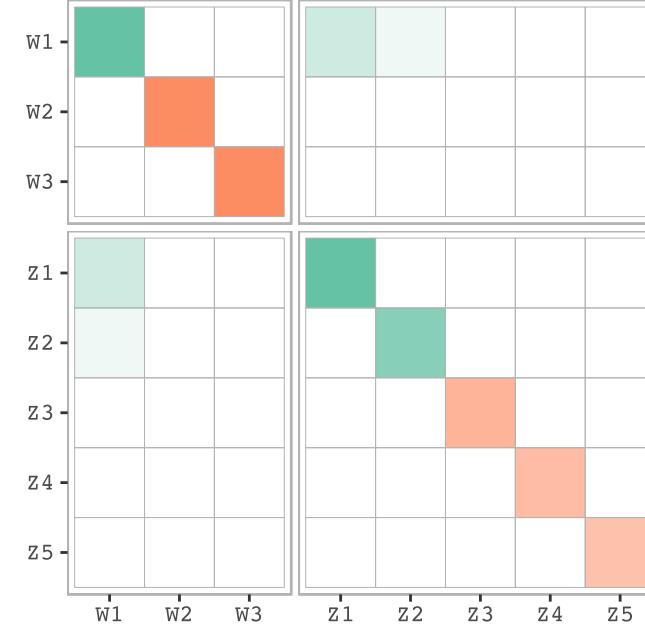
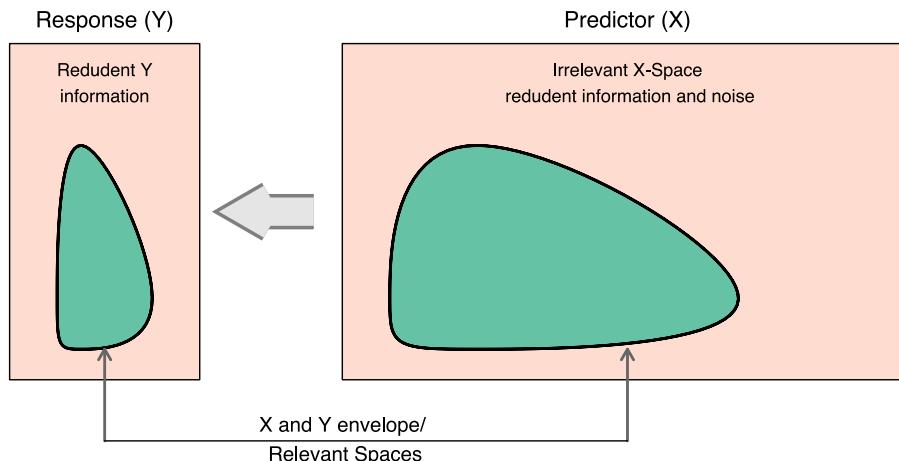
Reduction of Regression Model

$$\begin{bmatrix} w \\ z \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mu_w \\ \mu_z \end{bmatrix}, \begin{bmatrix} \Sigma_{ww} & \Sigma_{wz} \\ \Sigma_{zw} & \Sigma_{zz} \end{bmatrix}_{(m+p) \times (m+p)} \right)$$

$\frac{1}{2}(p + m)(p + m + 1)$ unknowns can be reduced and parameterized.

Relevant space within a model

A concept behind reduction of regression model



A subspace in predictor space spanned by subset of *predictor components* is relevant for the response. Also, only a subspace of response space spanned by subset of *response components* is informative.

Shiny App

Application of simrel



Research

Most of the research papers use simulated data.

Here are just few mentions:

- Theoretical evaluation of prediction error in linear regression with a bivariate response variable containing missing data¹
- A note on fast envelope estimation²
- Near optimal prediction from relevant components³
- A simulation study on comparison of prediction methods when only a few components are relevant⁴

Education

- Teaching Statistics
- Creating Examples

Let's Get
Started

An Example (comparison of two estimators)



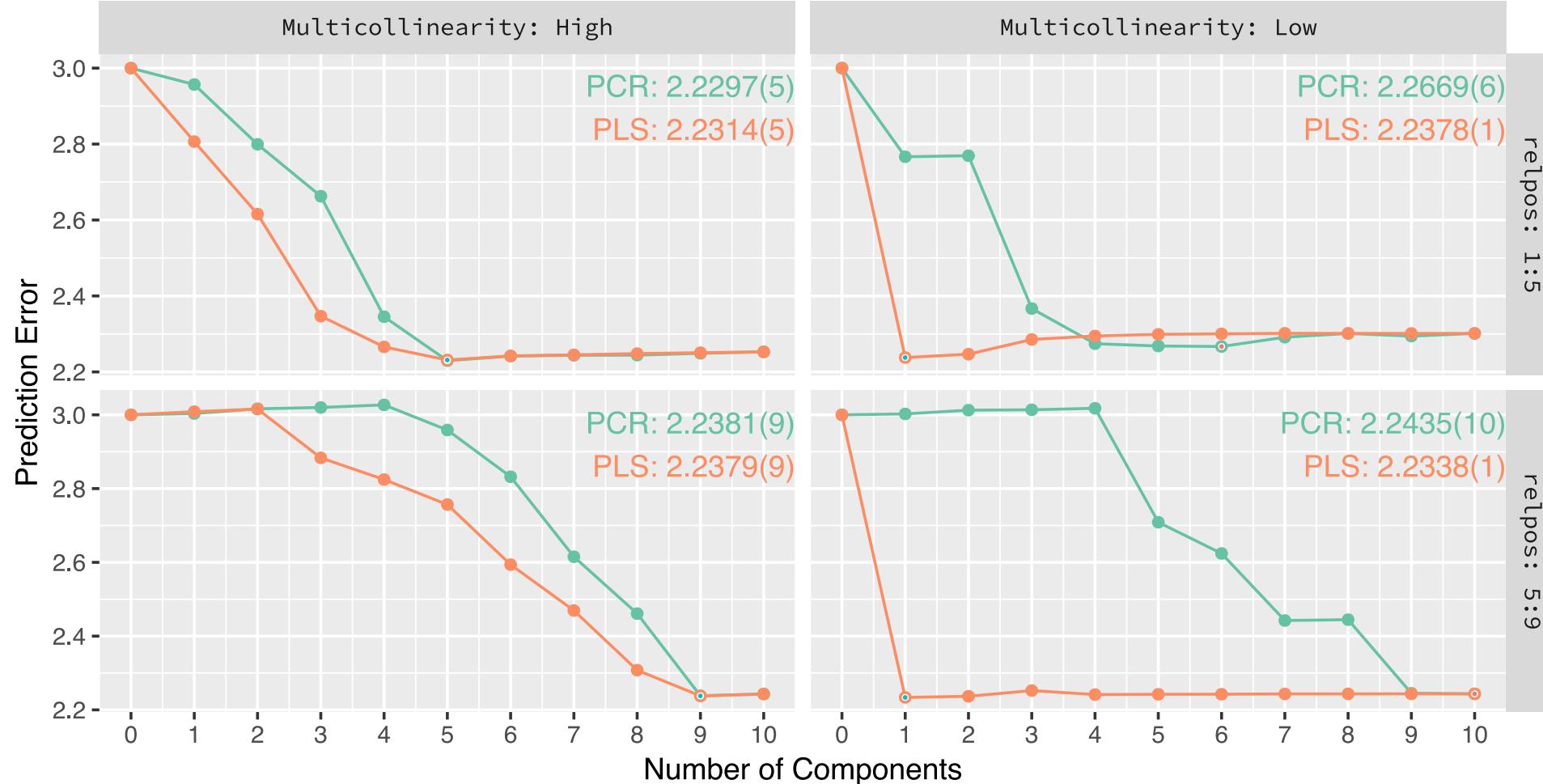
```
design <- crossing(  
  gamma = c(0.1, 1.2),  
  relpos = c("1:5", "5:9")) %>%  
  mutate(relpos = map(  
    relpos, ~eval(parse(text = .x))),  
    sim_obj = map2(gamma, relpos,  
      ~simrel(  
        n      = 500,  
        p      = 10, m = 3,  
        q      = 10,  
        relpos = list(.y),  
        ypos   = list(1:3),  
        gamma  = .x,  
        eta    = 0, R2 = 0.8,  
        type   = "multivariate"  
      ))  
    ))
```

```
# A tibble: 4 x 3  
  gamma relpos sim_obj  
  <dbl> <chr>  <list>  
1 0.1  1:5    <S3: simrel>  
2 0.1  5:9    <S3: simrel>  
3 1.2  1:5    <S3: simrel>  
4 1.2  5:9    <S3: simrel>
```

Estimator under comparison

- Principal Component Regression (PCR)
- Partial Least Squares (PLS)

An Example



Shiny Application and Installation

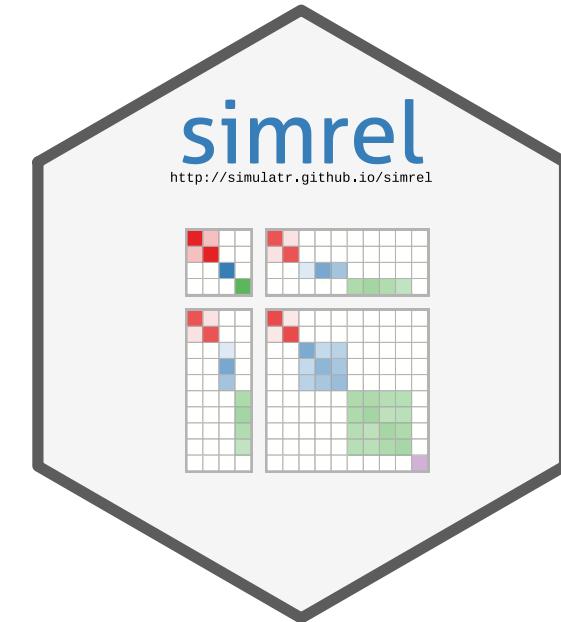


Install R-package:

```
if (!require(devtools)) install.packages("devtools")
devtools::install_github("simulatr/simrel")
```

Run Shiny Application

```
if (!require(simrel)) install.packages("simrel")
shiny::runGitHub("simulatr/AppSimulatr")
```



Acknowledgement



Solve Sæbø
NMBU



Trygve Almøy
BioStatistics, NMBU

Thank You



For the opportunity

The image features a large, bold "Thank You" in the center. Surrounding it are numerous words from various languages, each with its name and a phonetic transcription below it. The languages include: Bulgarian (Саламат, salamat), Czech (Dakujem), Turkish (teşekkür ederim), Korean (수운, SUWUN), Russian (спасибо, spasibo), Spanish (GRACIAS, GRACIAS), French (MERCI, MERCI), German (DANK, DANKE), Italian (grazie), Polish (dziękuje), Hungarian (szégyel), and others like HVALA, TAKK, MERSI, GRAZAS, etc.

Below the central "Thank You" are additional words in various languages:

- Arabic: لیزج ارکش (Lizj Arkesh)
- French: MERCI (MERCI)
- Spanish: GRACIAS (GRACIAS)
- Italian: grazie
- Polish: dziękuje
- Hungarian: szégyel
- Russian: спасибо (spasibo)
- Chinese: 多謝 (Duoxie)
- Japanese: ありがとう (Arigatou)
- Armenian: ԾՈՒՅՈ (Sziuyu)
- Georgian: მარათ (Marat)
- Ukrainian: Спасибо (Spasibo)
- Swedish: Tack (TACK)
- Dutch: dank je (dank je)
- Portuguese: Obrigado (OBIGADO)
- Chinese: 謝謝 (Xie Xie)
- Armenian: ՇՈՒՅՈ (Shuyu)

References



1. Gangsei LE, Almøy T, Sæbø S. Theoretical evaluation of prediction error in linear regression with a bivariate response variable containing missing data. *Communications in Statistics - Theory and Methods*. 2016;0(0):1-9.
doi:[10.1080/03610926.2016.1222434](https://doi.org/10.1080/03610926.2016.1222434).
2. Cook RD, Forzani L, Su Z. A note on fast envelope estimation. *Journal of Multivariate Analysis*. 2016;150:42-54.
3. Helland IS, Saebø S, Tjelmeland H, others. Near optimal prediction from relevant components. *Scandinavian Journal of Statistics*. 2012;39(4):695-713.
4. Almøy T. A simulation study on comparison of prediction methods when only a few components are relevant. *Computational Statistics & Data Analysis*. 1996;21(1):87-107. doi:[10.1016/0167-9473\(95\)00006-2](https://doi.org/10.1016/0167-9473(95)00006-2).
5. Rimal R, Almøy T, ø SS. A tool for simulating multi-response linear model data. *Chemometrics and Intelligent Laboratory Systems*. 2018;176:1-10. doi:[10.1016/j.chemolab.2018.02.009](https://doi.org/10.1016/j.chemolab.2018.02.009).
6. Sæbø S, Almøy T, Helland IS. 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*. 2015.
7. Helland IS, Sæbø S, Almøy T, Rimal R. Model and estimators for partial least squares regression. *Journal of Chemometrics*. e3044. doi:[10.1002/cem.3044](https://doi.org/10.1002/cem.3044).