

Accepted Manuscript

A tool for simulating multi-response linear model data

Raju Rimal, Trygve Almøy, Solve Sæbø

PII: S0169-7439(17)30482-3

DOI: [10.1016/j.chemolab.2018.02.009](https://doi.org/10.1016/j.chemolab.2018.02.009)

Reference: CHEMOM 3593

To appear in: *Chemometrics and Intelligent Laboratory Systems*

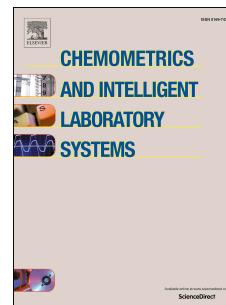
Received Date: 26 July 2017

Revised Date: 14 February 2018

Accepted Date: 19 February 2018

Please cite this article as: R. Rimal, T. Almøy, S. Sæbø, A tool for simulating multi-response linear model data, *Chemometrics and Intelligent Laboratory Systems* (2018), doi: 10.1016/j.chemolab.2018.02.009.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



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*Prorector, Norwegian University of Life Sciences, Ås, Norway*

Abstract

Data science is generating enormous amounts of data, and new and advanced analytical methods are constantly being developed to cope with the challenge of extracting information from such “big-data”. Researchers often use simulated data to assess and document the properties of these new methods, and in this paper we present an extension to the R-package `simrel`, which is a versatile and transparent tool for simulating linear model data with an extensive range of adjustable properties. The method is based on the concept of relevant components, and is equivalent to the newly developed envelope model. It is a multi-response extension of R-package `simrel` which is available in R-package repository CRAN, and as `simrel` the new approach is essentially based on random rotations of latent relevant components to obtain a predictor matrix \mathbf{X} , but in addition we introduce random rotations of latent components spanning a response space in order to obtain a multivariate response matrix \mathbf{Y} . The properties of the linear relation between \mathbf{X} and \mathbf{Y} are defined by a small set of input parameters which allow versatile and adjustable simulations. Sub-space rotations also allow for generating data suitable for testing variable selection methods in multi-response settings. The method is implemented as an update to the R-package `simrel`.

Keywords: `simrel` package in r, data simulation, linear model

*Corresponding Author

Email addresses: `raju.rimal@nmbu.no` (Raju Rimal), `trygve.almoy@nmbu.no` (Trygve Almøy), `solve.sabo@nmbu.no` (Solve Sæbø)

Download English Version:

<https://daneshyari.com/en/article/7561958>

Download Persian Version:

<https://daneshyari.com/article/7561958>

[Daneshyari.com](https://daneshyari.com)