

## Abstract

*J. Chemom.*, **38**, e3588, 1-10 (2024)

Varmuza K., Filzmoser P.:

### **Adjusted Pareto scaling for multivariate calibration models.**

<https://doi.org/10.1002/cem.3588> [open access]

<https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/cem.3588>

The performance of multivariate calibration models  $\hat{y} = f(x)$  for the prediction of a numerical property  $y$  from a set of  $x$ -variables depends on the type of scaling of the  $x$ -variables. Common scaling methods are autoscaling (dividing the centered  $x$  by its standard deviation  $s$ ) and Pareto scaling (dividing the centered  $x$  by  $s^P$  with  $P = 0.5$ ).

The adjusted Pareto scaling presented here varies the exponent  $P$  between 0 (no scaling) and 1 (autoscaling) with the aim of obtaining an optimum prediction performance for  $\hat{y}$ .

Related scaling methods based on the variable spread are range scaling and vast scaling; while level scaling is based on the location (central value) of the variable.

These scaling methods and robust versions are compared for models created by partial least-squares (PLS) regression. The applied strategy repeated double cross validation (rdCV) evaluates the model performance for test set objects and considers its variability.

Results with three data sets from chemistry show: (a) the efficacy of the different scaling methods depends on the data structure; (b) optimization of the Pareto exponent  $P$  is recommended; (c) range scaling or vast scaling may be better than adjusted Pareto scaling; (d) in general a heuristic search for the best scaling method is advisable.

Overall, the consideration of different variants of scaling allow for a flexible adjustment of the variable contributions to the calibration model.