Robust Multivariate Outlier Diagnostics in Chemometrics with Application to Spectrally Overlapping Drugs

Aylin ALIN [1] Claudio AGOSTINELLI [2] Georgi GERGOV [3]
Plamen KATSAROV [4] Yahya AL-DEGS [5]

[1]Department of Statistics, Dokuz Eylul University, Izmir, Turkey

[2]Dipartimento di Matematica, Università di Trento, Trento, Italy

[3]Faculty of Pharmacy, Medical University, Sofia, Bulgaria

[4]Department of Pharmaceutical Sciences, Faculty of Pharmacy, Medical University-Plovdiv, Bulgaria

[5]Department of Chemistry, Hashemite University , Jordan

- Introduction
- Robust SIMPLS Algorithm-RWSIMPLS [Alin and Agostinelli, 2017].
- Robust Diagnostics for Detecting Outliers and Influential Observations
- Numerical Example

$$\boldsymbol{Y} = E(\boldsymbol{Y}|\boldsymbol{x_1}, \ldots, \boldsymbol{x_k}) + \boldsymbol{\epsilon} = \boldsymbol{\beta X} + \boldsymbol{\epsilon} \quad (1)$$

For drug concentration problem ;

- $\boldsymbol{Y} \Rightarrow nxm$ matrix of m concentration values.
- $\boldsymbol{X} \Rightarrow nx(k+1)$ matrix of k absorbance values calculated as $A = -log_{10}R$ where R represents the reflectance.
- $\boldsymbol{\beta} = (\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \ldots, \boldsymbol{\beta_m}) \Rightarrow (k+1)xm$ matrix of unknown parameters with $\boldsymbol{\beta_s} = (\beta_{j0}, \ldots, \beta_{sk})$ for $s = 1, \ldots, m$.
- $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon_1}, \ldots, \boldsymbol{\epsilon_m})$ is a nxm matrix of errors.

- SIMPLS is one of the mostly used PLSR algorithm.
- It is extremely sensitive to outliers.
- For multivariate response variable $\Rightarrow$ RSIMPLS by [Hubert and Branden, 2003].
- For univariate response variable $\Rightarrow$ Partial version of M-regression estimator (PRM) by [Serrneels et al., 2005]
- For multivariate response variable $\Rightarrow$ Robustly Weighted SIMPLS algorithm (RWSIMPLS) by[Alin and Agostinelli, 2017]

- We propose new robust diagnostic tools based on the estimates from the RWSIMPLS to identify leverage points, regression outliers, influential observations and points causing non-normality in both concentration and absorbance values of overlapping binary drug systems.
- We compare the performance of the proposed diagnostics with the ones obtained from the classical SIMPLS on two data sets:
  - Biscuit dough data
  - Paracetamol (PAR) and Caffeine (CAF) data set

$$\boldsymbol{Y} = E(\boldsymbol{Y}|\boldsymbol{x_1}, \ldots, \boldsymbol{x_k}) + \boldsymbol{\epsilon} = \boldsymbol{\beta X} + \boldsymbol{\epsilon} \quad (2)$$

$$\boldsymbol{X} = \boldsymbol{TP^T} + \boldsymbol{E} \qquad \boldsymbol{Y} = \boldsymbol{UQ^T} + \boldsymbol{F} \quad (3)$$

- For convenience, we assume $\boldsymbol{X}$ and $\boldsymbol{Y}$ are mean-centered.
- $\boldsymbol{T} = \boldsymbol{XW} = (\boldsymbol{t_1}, \ldots, \boldsymbol{t_A}) \Rightarrow nxA$ component score matrix for $\boldsymbol{X}$
- $\boldsymbol{U} = \boldsymbol{YC} = (\boldsymbol{u_1}, \ldots, \boldsymbol{u_A}) \Rightarrow nxA$ component score matrix for $\boldsymbol{Y}$
- $A \Rightarrow$ the number of components $\leq min(k, m, n)$.
- $\boldsymbol{P} = (\boldsymbol{p_1}, \ldots, \boldsymbol{p_A}) \Rightarrow kxA$ loading matrix for $\boldsymbol{X}$.
- $\boldsymbol{Q} = (\boldsymbol{q_1}, \ldots, \boldsymbol{q_A}) \Rightarrow mxA$ loading matrix for $\boldsymbol{Y}$.
- $\boldsymbol{W} \Rightarrow kxA$ weight matrix for $\boldsymbol{X}$.
- $\boldsymbol{C} \Rightarrow mxA$ matrix are the weight matrix for $\boldsymbol{Y}$.

- The first score vectors $\boldsymbol{t_1}$ and $\boldsymbol{u_1}$ are the solutions to the following maximization problem.

$$\begin{aligned} cov(\boldsymbol{t}, \boldsymbol{u}) =& cov(\boldsymbol{Xw}, \boldsymbol{Yc}) = \boldsymbol{t}^T \boldsymbol{u} \\ =& (\boldsymbol{Xw})^T \boldsymbol{Yc} = \boldsymbol{w}^T \boldsymbol{X}^T \boldsymbol{Yc} \Rightarrow max \end{aligned} \tag{4}$$

  under the constraints of $\|\boldsymbol{t}\| = \|\boldsymbol{u}\| = 1$.
- The maximum of Equation (4) is attained for the vectors $\boldsymbol{w_1}$ and $\boldsymbol{c_1}$ corresponding to the largest singular values of $\boldsymbol{X}^T \boldsymbol{Y}$.
- Orthogonality of the previous score vectors; i.e. $\boldsymbol{t_a t_j} = 0$ and $\boldsymbol{u_a u_j} = 0$ for $1 \leq a < j \leq A$.
- In SIMPLS orthogonality is provided by finding the subsequent score vectors on deflated $\boldsymbol{X}^T \boldsymbol{Y}$.

Below are the steps of the SIMPLS algorithm.

Step 1: Centralize $\boldsymbol{X}$ and $\boldsymbol{Y}$ as in (5)

$$\begin{aligned} x_{ij} &= x_{ij} - mean(\boldsymbol{x_j}) \quad j = 1, \ldots, k \\ y_{is} &= y_{is} - mean(\boldsymbol{y_s}) \quad s = 1, \ldots, m \end{aligned} \tag{5}$$

and calculate $\boldsymbol{S_1} = \boldsymbol{X}^T \boldsymbol{Y}$

Step 2: $\boldsymbol{w_1}$ is obtained as the left singular vector of $\boldsymbol{S_1}$

Step 3: $\boldsymbol{w_1} = \frac{\boldsymbol{w_1}}{\|\boldsymbol{w_1}\|}$

Step 4: $\boldsymbol{t_1} = \boldsymbol{X}\boldsymbol{w_1}$

Step 5: $\boldsymbol{t_1} = \frac{\boldsymbol{t_1}}{\|\boldsymbol{t_1}\|}$

Step 6: $\boldsymbol{p_1} = \boldsymbol{X}^T \boldsymbol{t_1}$

Step 7: $\boldsymbol{q_1} = \boldsymbol{Y}^T \boldsymbol{t_1}$

Step 8: $\boldsymbol{S_{a+1} = S_a - v_a(v_a^T S_a)}$ where $\boldsymbol{v_1 = p_1}$ and $\boldsymbol{v_a = p_a - V_{a-1}(V_{a-1}^T p_a)}$ for $a > 1$ with $\boldsymbol{V_{a-1}} = (\boldsymbol{v_1, v_2, \ldots, v_{a-1}})$.

Repeat the steps 2-8 until all A components are extracted. Calculate the parameter estimates $\hat{\boldsymbol{\beta}}$ for the model in (2) as in Equation (6).

$$\hat{\boldsymbol{\beta}} = \boldsymbol{W(P^T W)^{-1} Q^T} \tag{6}$$

- $\delta(r_i(\hat{\beta})) = \frac{f^*(r_i|\hat{\beta})}{m^*(r_i|\hat{\sigma})} - 1 \Rightarrow$ The Pearson residual.
- $f^*(r_i|\hat{\beta}) = \int (k; t; h) d\hat{F}_n(t) \Rightarrow$ Kernel density estimator.
- $m^*(r_i|\hat{\sigma}) = \int (k; t; h) dM(t|\sigma) \Rightarrow$ The kernel smoothed model density estimator of $m(e_i|\hat{\sigma})$.
- In this study, we use Gaussian kernel density.
- The smoothing parameter $h = \sqrt{k\sigma^2}$ [Markatou et al., 1998])
- Given in Equation (7) is the weight for ith data point

$$\omega_{r_i} = \omega(r_i(\hat{\beta}); m(\cdot|\hat{\sigma}), \hat{F}_n) = \min\{1, \frac{\left[A(\delta(r_i(\hat{\beta}))) + 1\right]^+}{\delta(r_i(\hat{\beta})) + 1}\} \quad (7)$$

- The weight takes a value in $[0, 1]$.
- The function $A(.)$ ([Lindsay, 1994]) is a residual adjustment function.
- $A(\delta) = 2(\delta + 1)^{1/2} - 1$ is for the Hellinger distance residual function.

- The RWSIMPLS algorithm is based on deflating the weighted covariance matrix $\boldsymbol{S}^{\boldsymbol{\omega}} = \boldsymbol{X}^{\boldsymbol{\omega}\top}\boldsymbol{Y}^{\boldsymbol{\omega}}$. The weighted matrices $\boldsymbol{X}^{\boldsymbol{\omega}}$ and $\boldsymbol{Y}^{\boldsymbol{\omega}}$ are obtained by multiplying each row by the squared root of

$$\omega_{r_i} = \underset{s}{median}\{\omega_{r_{is}}\} \quad for \quad s = 1, \ldots, m; \quad i = 1, \ldots, n. \quad (8)$$

- The quantity $\omega_{r_{is}}$ is the weight of the ith residual corresponding to the sth response.

- Outliers:
  - (1) Regression (Vertical) outliers
    - The Classical Mahalonobis Distance for Residuals (9)

$$MD(\boldsymbol{r_i}) = \sqrt{\boldsymbol{r_i}^\top \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}^{-1} \boldsymbol{r_i}} \tag{9}$$

    - The cut-off value for $RMD(\boldsymbol{r_i^\omega}) \Rightarrow \sqrt{\chi^2_{m,(1-\alpha)}}$.
  - (2) Leverage points
    - Diagonal elements of the hat matrix $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$
    - Mahalonobis Distance for Explanatory Variables

$$MD_i = \sqrt{(\boldsymbol{x}_i - \boldsymbol{\mu_X})^\top \hat{\boldsymbol{\Sigma}}_{\boldsymbol{X}}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu_X})} \tag{10}$$

    - $h_{ii} = \frac{MD_i^2}{n-1} + \frac{1}{n}$.
    - The cut-off value for $RMD(\boldsymbol{r_i^\omega}) \Rightarrow \sqrt{\chi^2_{k,(1-\alpha)}}$.
    - Good Leverage and Bad Leverage
- Influential Points
  - Leave-one-out diagnostics (such as Cook's Distance [Cook, 1977]) are frequently used to flag the influential points.

- Robust Mahalonobis Distance for Scores

$$RMD(\boldsymbol{t}_i^{\boldsymbol{w}}) = \sqrt{(\boldsymbol{t}_i^{\boldsymbol{w}} - \boldsymbol{M}(\boldsymbol{T}^{\boldsymbol{w}}))^\top (\hat{\boldsymbol{\Sigma}}^{\boldsymbol{\omega}}_{\boldsymbol{T}^{\boldsymbol{w}}})^{-1} (\boldsymbol{t}_i^{\boldsymbol{w}} - \boldsymbol{M}(\boldsymbol{T}^{\boldsymbol{w}})} \tag{11}$$

$\boldsymbol{t}_i^{\boldsymbol{w}} = (t_{i1}^w, t_{i2}^w, \ldots, t_{iA}^w)^\top \Rightarrow$ The ith score from score matrix $\boldsymbol{T}^w$ calculated as

$$\boldsymbol{T}^{\boldsymbol{w}} = \boldsymbol{X}\boldsymbol{W}^{\boldsymbol{\omega}} \tag{12}$$

- $\boldsymbol{W}^{\boldsymbol{\omega}} \Rightarrow$ The robust weight matrix of $\boldsymbol{X}$ from the RWSIMPLS method
- $\boldsymbol{X} \Rightarrow$ The explanatory variable matrix unweighted but robustly centered as defined in [Alin and Agostinelli, 2017]
- $\boldsymbol{M}(\boldsymbol{T}^{\boldsymbol{w}}) \Rightarrow$ The $A \times 1$ L_1-median vector
-

$$\hat{\boldsymbol{\Sigma}}^{\boldsymbol{\omega}}_{\boldsymbol{T}^{\boldsymbol{w}}} = \frac{\sum_{i=1}^n \omega_{r_i} \boldsymbol{t}_i^{\boldsymbol{w}} \boldsymbol{t}_i^{\boldsymbol{w}\top}}{\sum_{i=1}^n \omega_{r_i} - 1}. \tag{13}$$

- The cut-off value for $RMD(\boldsymbol{t}_i^{\boldsymbol{w}}) \Rightarrow \sqrt{\chi^2_{A,(1-\alpha)}}$.

- Robust Mahalonobis Distance for Residuals (14)

$$RMD(\boldsymbol{r}_i^{\omega}) = \sqrt{\boldsymbol{r}_i^{\omega\top}(\hat{\boldsymbol{\Sigma}}_{\epsilon}^{\omega})^{-1}\boldsymbol{r}_i^{\omega}} \tag{14}$$

  - $\boldsymbol{r}_i^{\omega} = (r_{i1}^{\omega}, r_{i2}^{\omega}, \ldots, r_{im}^{\omega})^{\top} = \boldsymbol{r}_i(\hat{\boldsymbol{\beta}}^{\omega}) - \boldsymbol{M}(\boldsymbol{r}(\hat{\boldsymbol{\beta}}^{\omega}))$.
  - $\boldsymbol{r}_i(\hat{\boldsymbol{\beta}}^{\omega}) = \boldsymbol{y}_i - \boldsymbol{x}_i{}^{\top}\hat{\boldsymbol{\beta}}^{\omega}$.
  - $\boldsymbol{M}(\boldsymbol{r}(\hat{\boldsymbol{\beta}}^{\omega})) \Rightarrow$ the $m \times 1$ L_1-median vector computed from the collection of $(\boldsymbol{r}_1(\hat{\boldsymbol{\beta}}^{\omega}), \boldsymbol{r}_2(\hat{\boldsymbol{\beta}}^{\omega}), \ldots, \boldsymbol{r}_m(\hat{\boldsymbol{\beta}}^{\omega}))$.
  - $\boldsymbol{x}_i{}^{\top} = (x_{i1}, x_{i2}, \ldots, x_{ik})^{\top} \rightarrow i$th row of unweighted but robustly centered $\boldsymbol{X}$ matrix.
  - $\hat{\boldsymbol{\beta}} \Rightarrow m \times k$ RWSIMPLS coefficient estimate matrix.
  - The $m \times m$ robust covariance matrix for residuals is

$$\hat{\boldsymbol{\Sigma}}_{\epsilon}^{\omega} = \frac{\sum_{i=1}^{n} \omega_{r_i} \boldsymbol{r}_i^{\omega} \boldsymbol{r}_i^{\omega\top}}{\sum_{i=1}^{n} \omega_{r_i} - A}. \tag{15}$$

- The cut-off value for $RMD(\boldsymbol{r}_i^{\omega}) \Rightarrow \sqrt{\chi^2_{m,(1-\alpha)}}$.

- In classical ordinary least squares based regression, the Cook's Distance gives the squared distance from $\hat{\boldsymbol{\beta}}$ to $\hat{\boldsymbol{\beta}}_{(i)}$ relative to the fixed geometry of $\boldsymbol{X}^\top \boldsymbol{X}$ [Cook and Weisburg, 1982].
- For SIMPLS, the Cook's Distance can be calculated as to measure the squared distance from $\hat{\boldsymbol{\beta}}$ to $\hat{\boldsymbol{\beta}}_{(i)}$ relative to the fixed geometry of $\boldsymbol{T}^\top \boldsymbol{T}$.
- The Robust Cook's Distance (16)

$$RCD_{is}^{w} = \left(\frac{\tilde{r}_{is}(\boldsymbol{T}^{\boldsymbol{w}})}{\hat{\sigma}(\boldsymbol{T}^{\boldsymbol{w}})_s \sqrt{1 - h_{ii}^{w}}} \right)^2 \frac{h_{ii}^{w}}{A(1 - h_{ii}^{w})} \tag{16}$$

- $$\hat{\sigma}(\boldsymbol{T}^{\boldsymbol{w}})_s = \sqrt{\frac{\sum_{i=1}^{n} \omega_{r_i} \tilde{r}(t)_{is}^2}{\sum_{i=1}^{n} \omega_{r_i} - A}} \tag{17}$$

- $\tilde{r}_{is}(\boldsymbol{T}^{\boldsymbol{w}}) = r_{is}(\boldsymbol{T}^{\boldsymbol{w}}) - \boldsymbol{M}(\boldsymbol{r}(\boldsymbol{T}^{\boldsymbol{w}}))_s$

- $r_{is}(\boldsymbol{T}^w) = y_{is} - \hat{y}_{is}^w$.
- $\boldsymbol{M}(\boldsymbol{r}(\boldsymbol{T}^w))_s \Rightarrow$ the sth element of the $m \times 1$ L_1 median vector of $(\boldsymbol{r_1}(\boldsymbol{T}^w), \boldsymbol{r_2}(\boldsymbol{T}^w), \ldots, \boldsymbol{r_m}(\boldsymbol{T}^w))$.
- $\hat{y}_{is}^w \Rightarrow$ the fitted value at the ith row and sth column of the matrix in (18), and h_{ii}^w is the ith diagonal element of $\boldsymbol{H}^w$.

$$\hat{\boldsymbol{Y}}^{\boldsymbol{w}} = \boldsymbol{H}^{\boldsymbol{w}}\boldsymbol{Y} = \boldsymbol{T}^{\boldsymbol{w}}(\boldsymbol{T}^{\boldsymbol{w}\top}\boldsymbol{T}^{\boldsymbol{w}})^{-1}\boldsymbol{T}^{\boldsymbol{w}\top}\boldsymbol{Y} \tag{18}$$

- Cut-off value=1 [Cook and Weisburg, 1982] .

- Robust diagnostic to identify outliers with respect to

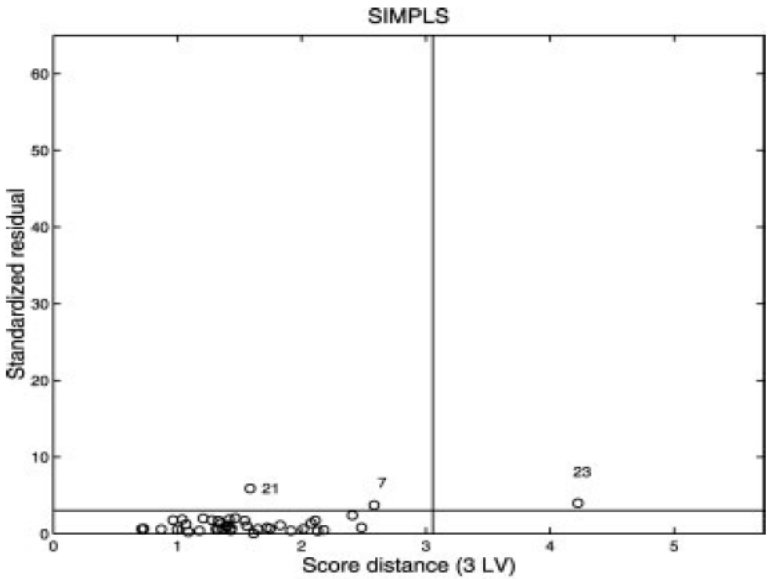
$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{P}^{\top} + \boldsymbol{E} \tag{19}$$

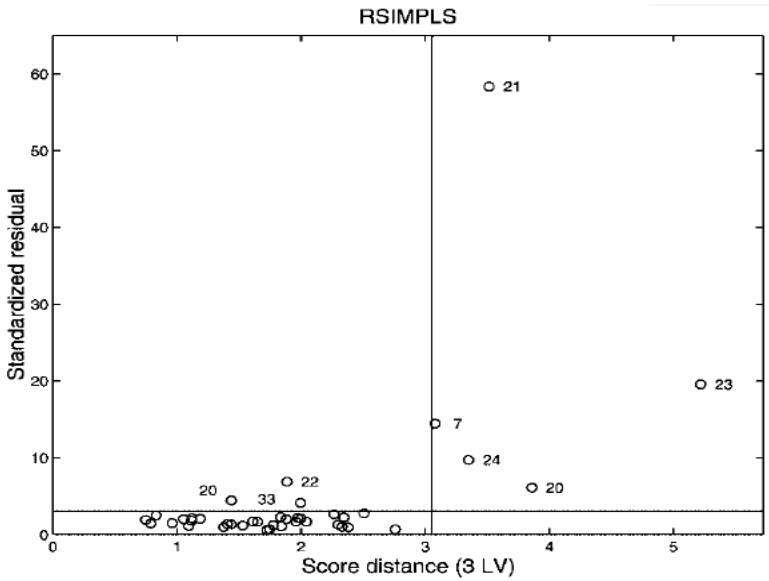
$$ROD(\boldsymbol{t}_i^{\boldsymbol{w}}) = \sqrt{\boldsymbol{x}_i - \boldsymbol{P}^{\boldsymbol{\omega}}\boldsymbol{t}_i^{\boldsymbol{w}}} \tag{20}$$

- $\boldsymbol{x_i} \Rightarrow k \times 1$ robustly centered explanatory variable vector for ith data.
- $\boldsymbol{P}^{\boldsymbol{\omega}} \Rightarrow$ Robust loading matrix for $\boldsymbol{X}$ from RWSIMPLS.
- Bootstrap method is used to obtain the cut-off value for $ROD(\boldsymbol{t}_i^{\boldsymbol{w}})$.
- Plot of Mahalonobis distance for scores $RMD(\boldsymbol{t}_i^{\boldsymbol{w}})$ against the orthogonal distance for scores $ROD(\boldsymbol{t}_i^{\boldsymbol{w}})$.

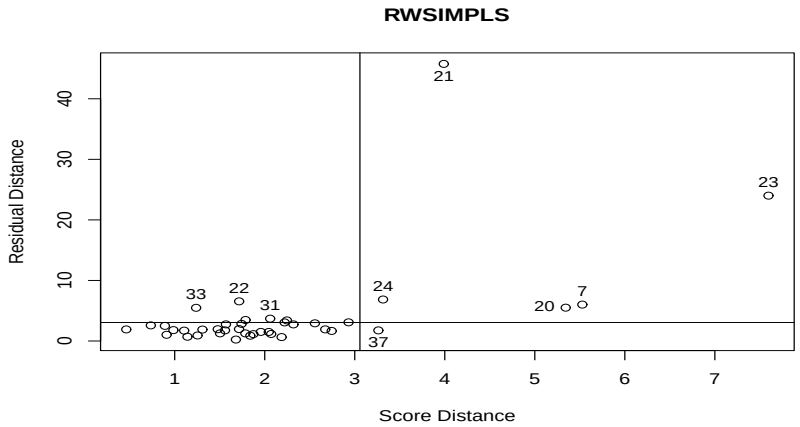
- The aim is to predict the concentrations of flour, sucrose and water of biscuit doughs based on 40 NIR spectra with measurements every 2nm from 1200 up to 2400 nm.
- After applying same preprocessing as in [Hubert and Branden, 2003], we end up with 600 measurements for 40 biscuit dough samples.
- Observation 23 is known to be an outlier.
- 3 components were used for diagnostics.

SIMPLS and RSIMPLS Diagnostics for Biscuit Dough Data-[Hubert and Branden, 2003]

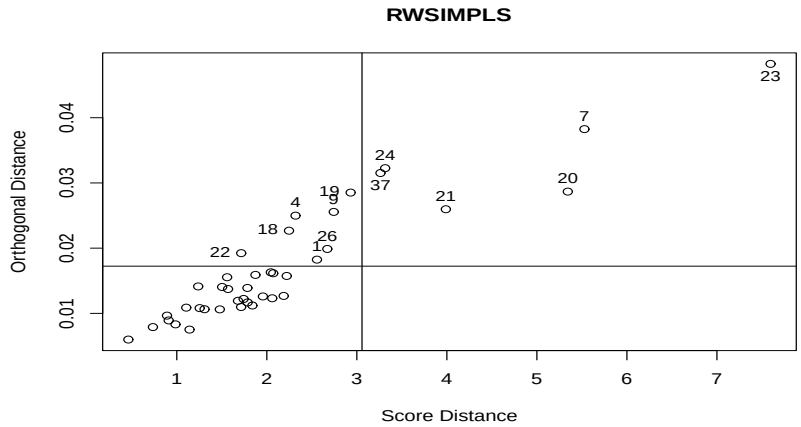




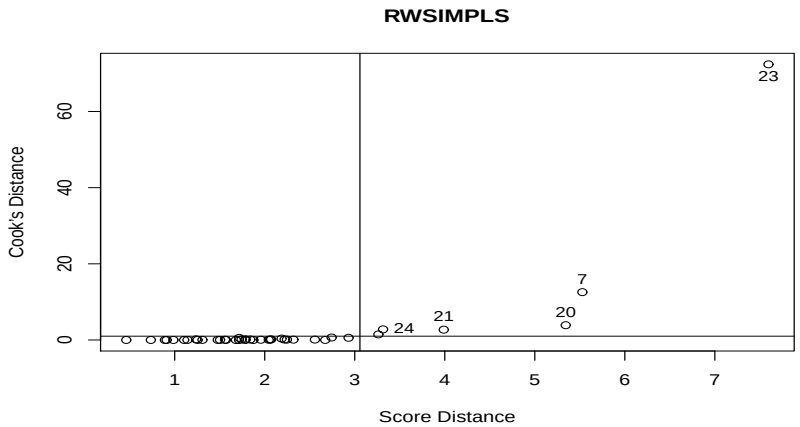
RWSIMPLS Diagnostics for Biscuit Dough Data



(a)



(b)



(c)

- Paracetamol which is one of the most popular and most commonly used analgesic and antipyretic drugs around the world, available without a prescription, both in mono and multi-component preparations.
- Coffee and caffeine-containing products affect the cardiovascular system, with their positive inotropic and chronotropic effects, and the central nervous system, with their locomotor activity stimulation and anxiogenic-like effects [Cappelletti et al., 2015].
- 100 mg/L standard solutions of both drugs were prepared from the corresponding pure materials.
- Both drugs were kindly donated from local pharmaceutical company (Hikma Pharmaceuticals, Amman, Jordan).

- For both drugs, initial pH was adjusted to 7.0 using diluted 0.01 M NaOH solution. Initial screening studies indicated insignificant effect of pH on spectral behavior of both drugs.
- For calibration sample, 25 mixtures of PAR and CAF were prepared.
- 15-mixture validation set with the new concentration levels outside the ranges of the calibration set were prepared in order to test the prediction performance of the models and to decide for the number of components.
- Absorbency of drugs were obtained using a double beam spectrophotometer (Thermo Electron Corporation Nicollet evolution 100).
- The spectra of drugs for both calibration and validation sets were recorded over the spectral range of 216 – 300 nm at 2 nm intervals.

Table: The Concentration Levels for Calibration and Validation Sets for PAR/CAF Drug System

	Calibration Set (mg/L)		Validation Set (mg/L)	
Number	Paracetamol	Caffeine	Paracetamol	Caffeine
1	10	10	2	4
2	10	1	6	8
3	1	1	8	3
4	1	20	11	2
5	20	5	17	8
6	5	20	8	14
7	20	10	13	16
8	10	5	4	19
9	5	5	19	7
10	5	15	3	1
11	15	20	6	9
12	20	15	9	16
13	15	10	11	11
14	10	20	0	14
15	20	20	14	0
16	20	1		
17	1	15		
18	15	1		
19	1	10		

$$RMSEP_j = \sqrt{\frac{\sum_{i=1}^{\nu} \left(\hat{y}_{i(valid)_j} - y_{i(valid)}\right)^2}{\nu - A}} \tag{21}$$

- $\nu \Rightarrow$ the number of samples in validation set.
- $\hat{y}_{i(valid)_j} = \boldsymbol{x}_{\boldsymbol{i(valid)}}^{\top} \hat{\boldsymbol{\beta}}_{\boldsymbol{j}}$ for $j =$ RWSIMPLS, SIMPLS, PRM with $\hat{\boldsymbol{\beta}}$ estimated from the calibration sample.

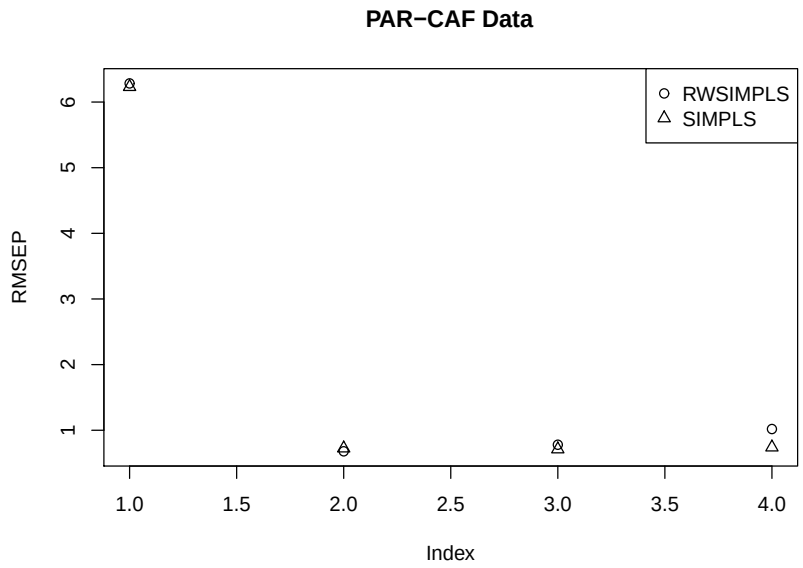


Figure: Root Mean Square Error for Prediction (RMSEP) vs The Number of Components for PAR/CAF data

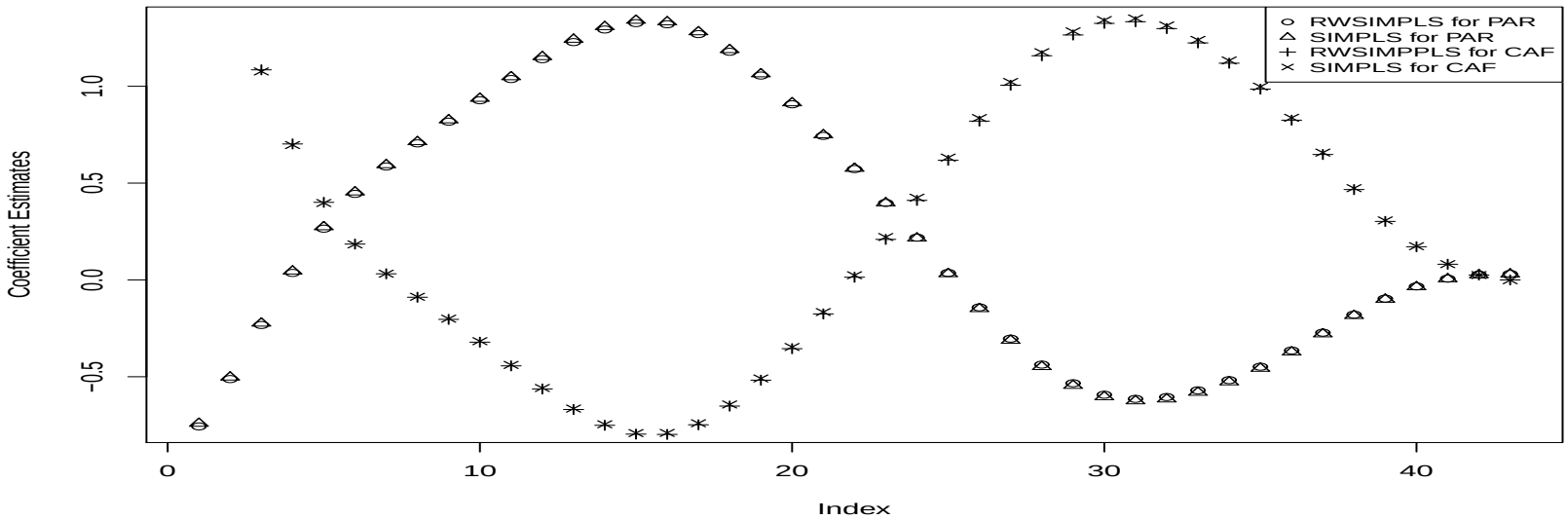
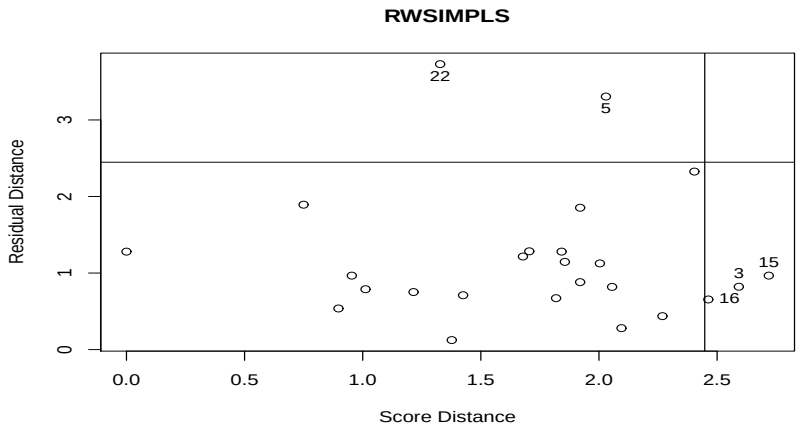


Figure: Coefficient Estimates for PAR/CAF Data with A=2

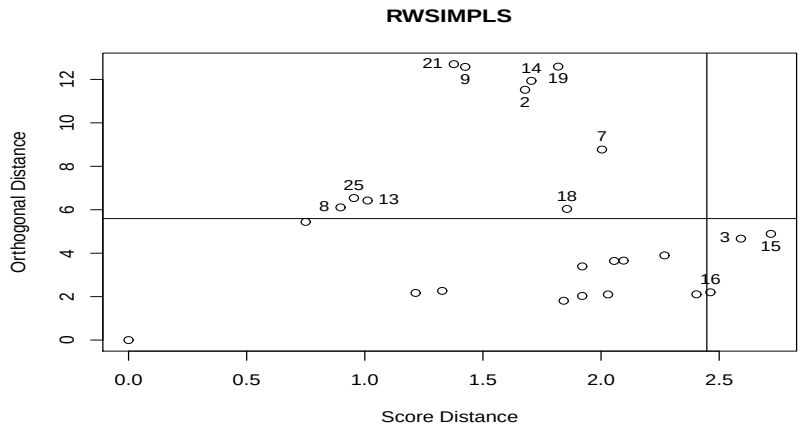
Table: The Scale Estimates For PAR/CAF Drug System with A=2

	RWSIMPLS	SIMPLS
PAR	0.2170177	0.2300721
CAF	0.2818257	0.2923287

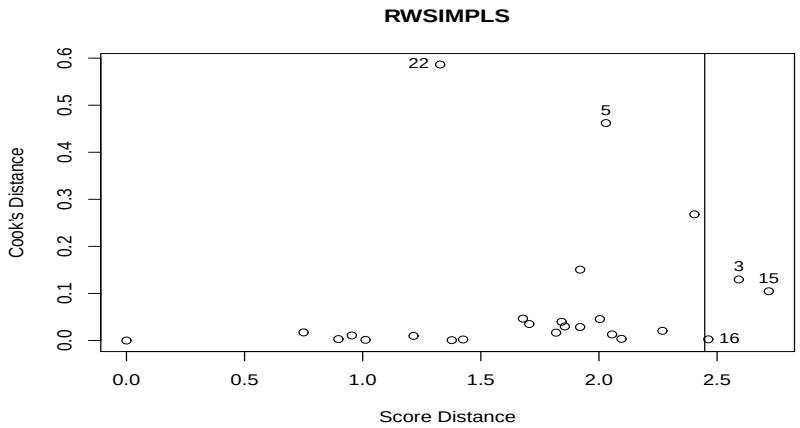
Binary Drug System: Paracetemol/Caffeine (PAR/CAF)



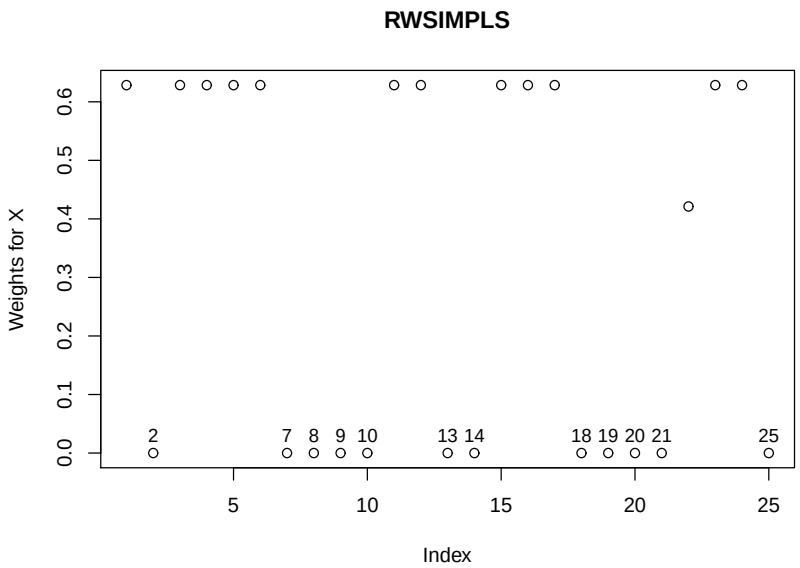
(a)



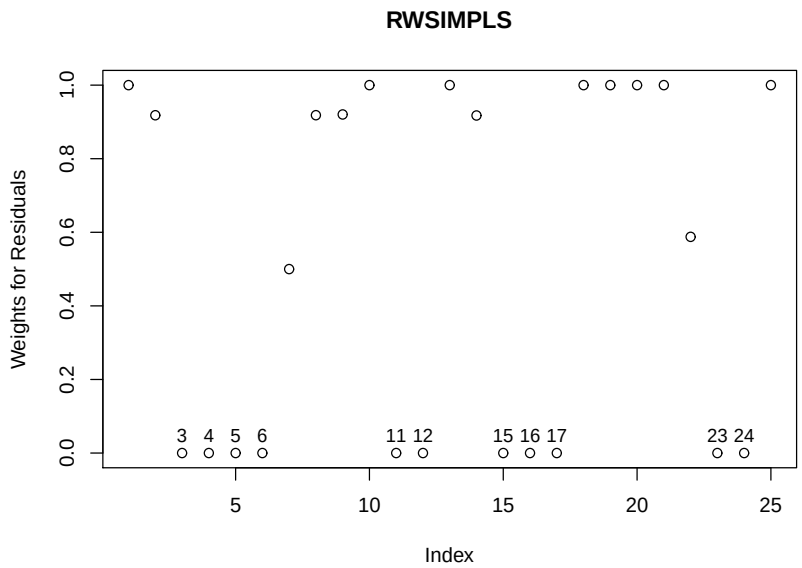
(b)



(c)



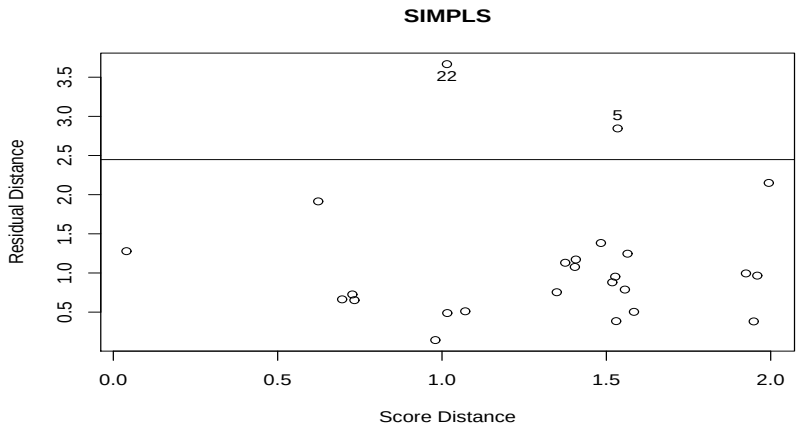
(a)



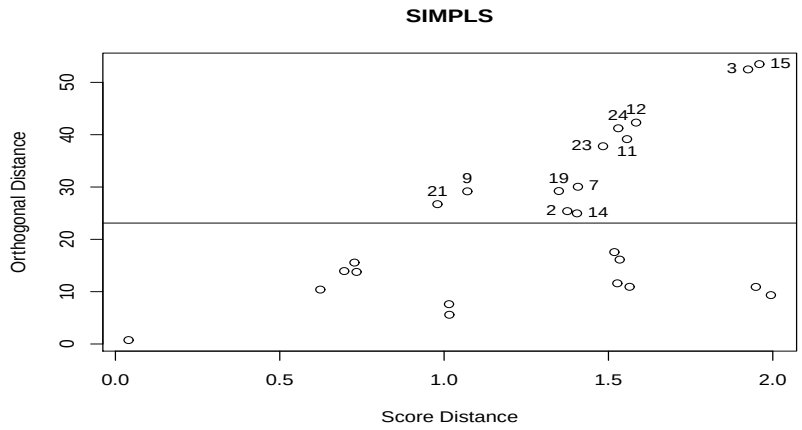
(b)

Figure: RWSIMPLS Weights for PAR/CAF Drug System

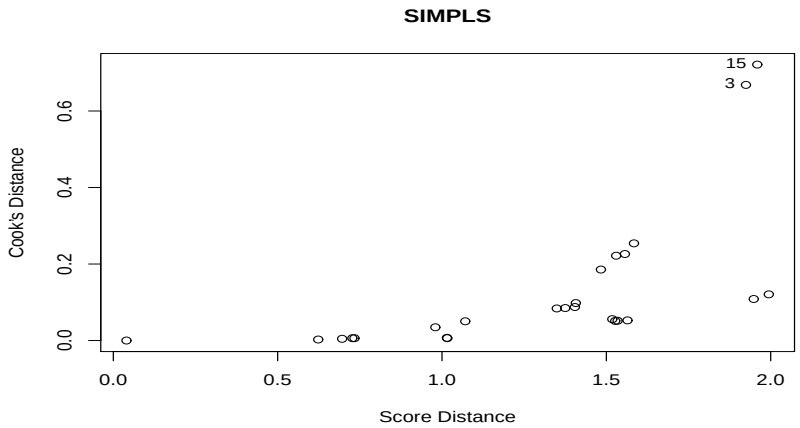
Binary Drug System: Paracetemol/Caffeine (PAR/CAF)



(a)



(b)



(c)

Thanks for your attention! Any questions? ...


Alin, A. and Agostinelli, C. (2017).
Robust iteratively reweighted SIMPLS.
Journal of Chemometrics, 31(3):e2881.


Cappelletti, S., Daria, P., Sani, G., and Aromatario, M. (2015).
Caffeine: Cognitive and physical performance enhancer or psychoactive drug?
Current Neuropharmacology, 13:71–88.


Cook, R. (1977).
Detection of influential observations in linear regression.
Technometrics, 19:15–18.


Cook, R. and Weisburg, S. (1982).
Residuals and Influence in Regression.
Chapman and Hall.


DeJong, S. (1993).
SIMPLS: An alternative approach tp partial least squares regression.
Chemometrics and Intelligent Laboratory Systems, 18:251–263.


Hubert, M. and Branden, K. (2003).
Robust methods for partial least squares regression.
Journal of Chemometrics, 17:537–549.


Lindsay, B. (1994).
Efficiency versus robustness: The case for minimum Hellinger distance and related methods.
Annals of Statistics, 22:1018–1114.


Markatou, M. (1996).
Robust statistical inference: Weighted likelihoods or usual M-estimation?
Communications in Statistics: Theory Methods, 25:2597–2613.


Markatou, M., Basu, A., and Lindsay, B. (1998).
Weighted likelihood estimating equations with a bootstrap root search.
Journal of American Statistical Associations, 93:740–750.


Serrneels, S., Croux, C., Filzmoser, P., and P.J., V. E. (2005).
Partial robust M-regression.
Chemometrics and Intelligent Laboratory Systems, 79:55–64.