SHORT COMMUNICATION

A brief note on a new faster covariate’s selection (fCovSel) algorithm

Puneet Mishra

Correspondence
Puneet Mishra, Wageningen Food and Biobased Research, Bornse Weilanden 9, P.O. Box 17, 6700AA, Wageningen, The Netherlands.
Email: puneet.mishra@wur.nl

Abstract
Covariate’s selection (CovSel) is a variable selection technique in the domain of chemometrics and is commonly used for extracting variables carrying high covariance with response variables. CovSel is a special case of partial least square analysis, where at each weight estimation step, the weight vector is of the binary form to select the variables. In the earlier algorithm of the CovSel which is identical to the nonlinear iterative partial least square (NIPALS), there is a key step of predictor matrix deflation which makes it a time-consuming approach leading to longer time consumption during tasks such as cross-validation or in multiblock multiway CovSel scenarios where a wide number of variables combinations are usually explored for model optimization.

We present a new CovSel algorithm called faster CovSel (fCovSel) which drops the need for predictor matrix deflation. By dropping the predictor matrix deflation step, the method naturally becomes faster than the CovSel based on NIPALS. Mathematical and analytical comparisons of the CovSel and the fCovSel in terms of achieving the same solution and time requirements are presented.

KEYWORDS
data compression, feature selection, latent space modeling, multivariate

1 | INTRODUCTION

In the domain of chemometrics and analytical chemistry, multivariate signals are widely met [1,2]. Furthermore, multivariate signals generated from several analytical techniques such as optical spectroscopy are highly collinear [3]. Often, such multivariate multicollinear signals cannot be directly interpreted but demand chemometric approaches which can handle the multicollinearity and simultaneously allow exploring meaningful information [3]. For example, to explore the data, variance-based subspace modeling techniques such as principal component analysis (PCA) are widely used [4]. In the presence of response variables, the common approach to subspace modeling is the partial least square (PLS) which models the subspaces as the function of covariance between the predictor and the response variables [5,6]. Note that both the PCA and PLS are bilinear modeling approaches where the analysis results in a bilinear decomposition of predictor matrix into a set of scores and loading vectors, where the loadings are the key directions maximizing the covariance (in the case of PLS) or variance (in the case of PCA) and scores are the data projected to those directions.
Note that apart from subspace modeling, one of the key tasks while exploring the multivariate multicollinear data sets is to extract information about the key variables of interest [7, 8]. This task is often also referred to as variables selection [9, 10] in the domain of chemometrics or feature selection [11] in the domain of machine learning. The key variables of interest are information-rich subset of variables that can be used for wide ranges of tasks such as the development of predictive models, data pattern interpretation, development of low-cost multispectral sensors in the case of optical spectroscopy, development of selective chemical sensing applications, and selection of key markers in omics data analysis [8–10, 12, 13]. In the domain of chemometrics, there are three distinct types of variable selection approaches existing, that is, filter, wrapper, and embedded methods [7]. The main difference between the methods is the amount of interaction the variables selection has with the underlying model for which the variables need to be extracted. For example, in the case of the filter-based method, there is almost no direct interaction of the variable selection strategy with the underlying model [11]. An example of the filter-based method can be the selection of variables based on their correlation with the response. The second case of the wrapper method involves the exploration of all possible individual variables and their combinations in an exhaustive search to find the best possible subset of variables. Unlike the filter-based methods, in the wrapper methods, the variable selection is performed by evaluating the model performance; however, due to the random nature of such exhaustive searches, for example, genetic algorithms [14, 15], the results are not repeatable. The third case is the embedded methods involving variable selection as a joint step of model calibration by optimizing training criteria such as successive addition, elimination, and a joint exploration of both [7, 16]. More details on variables selection in the chemometric domain can be found in some recent review articles [9, 10, 12, 13, 17].

Of all the three classes of methods, the embedded methods are of high interest as they undertake the step of variable selection and modeling as a joint step to find the information-rich subset of variables being highly predictive at the same time. CovSel is one such powerful method allowing the extraction of informative variables by repeated steps of covariance maximization and orthogonalization [7]. CovSel is similar to the nonlinear iterative partial least square (NIPALS) algorithm which involves successive steps of covariance maximization and repeated orthogonalization of predictor matrices [5,18]. In the later part of the manuscript, the similarity between the CovSel and NIPALS will be highlighted in detail. A wide range of applications of the CovSel can be found in scientific literature [19–21]. Furthermore, wide extensions of the CovSel for multiblock [8,20] and multiway multiblock [22] cases can also be found in the scientific literature. The main reason behind the CovSel being so popular and extendable to frameworks such as multiblock and multiway is because the CovSel is a special case of PLS and almost any extensions that are possible for PLS are also directly applicable to the CovSel. However, just like the NIPALS algorithm is a slow algorithm in the chemometric domain due to the requirement of deflation of the predictor matrix [23,24], the deflation of the predictor matrix also makes the current CovSel [7] inherently a slower algorithm. Due to the presence of the predictor matrix deflation step, the extension of the CovSel such as for multiblock and multiway can even be slower than the CovSel on a single data block. In the domain of chemometrics, recently, a major focus has been on having fast and stable PLS algorithms, and one out of many is the nondeflation PLS approach where the predictor matrix deflation step is replaced by the Gram–Schmidt (GS) orthogonalization step of the scores of the predictor matrix [23–25]. Such an algorithm is naturally faster as there is no predictor matrix deflation step [23–25]. Since the CovSel is a special case of PLS, therefore, using the concept of nondeflating PLS [23–25], a nondeflating version of the CovSel can be achieved which will naturally be faster due to no predictor matrix deflation step. Furthermore, the extension of such nondeflating CovSel approaches to multiblock [8] and multiway [22] scenarios will also be naturally faster, as those methods just involve the deflation of multiple predictor matrices or deflation of wide matrices of unfolded multiway arrays, respectively. To the best awareness of the author, a nondeflating version of the CovSel is never developed and the one developed in this study can have wide practical implications for single block [7], multiblock [8], and multiway [22] CovSel modeling.

The aim of this study was to present a new CovSel algorithm called faster CovSel (fCovSel) which drops the need for the predictor matrix deflation. By dropping the predictor matrix deflation, the method naturally becomes faster than the CovSel. Mathematical and analytical comparisons of the CovSel and the fCovSel in terms of achieving the exact same solution and time requirements are presented.

2 THEORY

This section will supply the theoretical basis for the fCovSel algorithm without deflation of predictor matrix. All matrices are presented in bold upper case. All vectors are in bold lower case, and constants are in lower case. Let \( X \) be the
predictor matrix of size \( n \times p \), where \( n \) is the number of samples and \( p \) be the number of variables. Let \( Y \) be the multiresponse matrix of size \( n \times k \), where \( n \) is the number of samples and \( k \) be the number of responses. The CovSel algorithm proposed in the earlier studies \([7,8,22]\) involves two main steps. The first step is to find the variables that carry the maximum covariance with the response variable, and the second step is to use the column vector for the selected variable to deflate both the predictor matrix and response variables. The process is repeated until the desired number of variables is extracted being \( n < p \); if \( n < p \), then the total number of variables can be extracted \( = n \). The deflation of the predictor matrix is a computationally expensive step in the traditional CovSel algorithm. Furthermore, the CovSel is almost identical to the NIPALS algorithm as shown in Table 1, except for the step where the scores are the column vector of predictor matrix, while for NIPALS, they are estimated as the product of predictor matrix and the normalized loading weight. One can assume that the CovSel has the loading weights of the form \([0,0,0,…,1,…,0,0,0]\), where the multiplication of the loading weight with the predictor matrix leads to the selection of the column vector as the score vector. The CovSel algorithm presented in the earlier study \([7]\) uses the squared covariance; however, in the algorithm presented in Table 1, that step is replaced with the determination of the absolute value (mod) to show the similarity between the CovSel and NIPALS. Note that selecting variable using the absolute value (mod) of the covariances leads to the exact same variable selection as that based on squared covariances as the square operation is just making all the covariances positive to identify the maximum covariance. For the multiresponse case of CovSel, the sum of the absolute (mode) covariances can be used to select the variable, and later, the column vector corresponding to the selected variable can be used for the deflating the predictor matrix. The algorithm for the fCovSel is presented in Table 2. The main key difference in the fCovSel and the CovSel algorithm can be noted as the GS orthogonalization \([26]\) of the selected column vector of the predictor matrix concerning all the earlier selected variables to remove any

### Table 1: Algorithms of CovSel and NIPALS presented for comparative purpose

| CovSel for single (y) or multiresponse (Y) | NIPALS for single response (y) |
|-------------------------------------------|--------------------------------|
| \( i = 1 : n \)                           | \( i = 1 : n \)               |
| Step 1: \( V_i = X_i' Y \)                | Step 1: \( v_i = X_i' y \)    |
| Step 2: \( k = \text{argmax}(\text{sum}(|V_i|, 2)) \) | Step 2: \( v_i = \frac{v_i}{|v_i|} \) |
| Step 3: \( \tau = X_k \)                 | Step 3: \( \tau = X_k v_i \)  |
| Step 4: \( t_i = \frac{\tau}{||\tau||} \) | Step 4: \( t_i = \frac{\tau}{||\tau||} \) |
| Step 5: \( X_{i+1} = X_i - t_i t_i' X_i \) | Step 5: \( X_{i+1} = X_i - t_i t_i' X_i \) |

**Note:** All matrices are presented in bold upper case. All vectors are in bold lower case and constants are in lower case.

### Table 2: The fCovSel algorithm without deflation of predictor matrix

| fCovSel algorithm for single response (y) or multi response (Y) |
|---------------------------------------------------------------|
| for \( i = 1 : \text{number of variables} \)                  |
| Step 1: \( V_i = X_i' Y_i \)                                 |
| Step 2: \( k = \text{argmax}(\text{sum}(|V_i|, 2)), \text{note that } k \text{ is the selected variable for step } i \) |
| Step 3: \( \tau = X_k \)                                     |
| Step 4: if \( i > 1 \), \( \tau = \tau - T_{i-1} T_{i-1}' \tau \), end (GS reorthogonalization) |
| Step 5: \( t_i = \frac{\tau}{||\tau||} \)                    |
| Step 6: \( Y_{i+1} = Y_i - t_i t_i' Y_i \)                   |
| Step 7: \( T(i) = t_i \)                                     |

**Note:** All matrices are presented in bold upper case. All vectors are in bold lower case and constants are in lower case.
recently proposed response-oriented sequential alternation modeling [23] of the response variable. This idea is inspired from the nondeflating PLS [24, 25] and the recently proposed response-oriented sequential alternation modeling [23]. The process of the fCovSel can then be repeated to the desired number of variables. In the later part of the study, it has been analytically shown that the fCovSel and the CovSel will lead to the same solution in terms of the selected variables. Furthermore, fCovSel will also lead to the same solution as CovSel for the multiresponse case.

The key difference between the fCovSel and the traditional CovSel is that the computationally expensive predictor matrix deflation step is replaced by two main steps, that is, the deflation of the response and the reorthogonalization of the scores as proposed in previous studies [23, 25]. It can be mathematically shown that the predictor matrix deflation can be skipped with such two steps. The main idea behind the CovSel is to use the covariance information between predictor \(X(n \times p)\) and the response \(Y(n \times k)\) to select the variable carrying maximum covariance. In successive steps, the predictor matrix is deflated to capture covariances. For a particular step, the covariance \(V_{i+1}(p \times k)\) estimation can be expressed as Equation (1):

\[
V_{i+1} = X_i^T Y
\]

Furthermore, the \(X_i(n \times p)\) at a particular step can be expressed as \((I - T_{1i}T_{1i}^T)X, a(n \times p)\) size matrix, where the \(T_{1i}(n \times l)\) is the matrix of orthonormal scores extracted up to the \(i^{th}\) step. Note that \(I\) is the identity matrix of size \((n \times n)\). Resubstituting Equation (1) with \(X_i = (I - T_{1i}T_{1i}^T)X\), Equation (2) is achieved:

\[
V_{i+1} = (I - T_{1i}T_{1i}^T)X_i^T Y
\]

Distributing the transpose in \((I - T_{1i}T_{1i}^T)X_i^T = X_i^T(I - T_{1i}T_{1i}^T)^T\) leads to Equation (3):

\[
V_{i+1} = X_i^T(I - T_{1i}T_{1i}^T)^T Y = X_i^T Y_i
\]

where, \(Y_i = (I - T_{1i}T_{1i}^T)^T Y\)

From Equation (3), it can be noted that the covariance estimation \(V_{i+1}\) can be done without the deflation of predictor matrix \(X\) while needing the deflation of the responses \(Y\). This is slightly different from the current CovSel algorithm (also NIPALS) where the deflation of predictor matrix is needed while the deflation of response does not contribute anything extra. Since responses \(Y\) are already known, the only extra thing needed to estimate the covariance according to Equation (3) is the orthonormal scores matrix \(T\). Such orthonormal scores can be obtained with the scores reorthogonalization step proposed in Indahl [25] as explained in Equation (4)

\[
\tau_i = (I - T_{1i-1}T_{1i-1}^T) \tau_i
\]

In Equation (4), the term \((I - T_{1i-1}T_{1i-1}^T)\) of size \((n \times n)\) is the orthogonal projection operator spanning the variability orthogonal to the \(T_{1i-1}(n \times i - 1)\) scores. The multiplication of the orthogonal projection operator with the non-orthogonalized scores removes all the variability explained by the \(T_{1i-1}\) scores.

The nonnormalized score vector \(\tau(n \times 1)\) extracted with Equation (4) is the same as selecting the columns as the score vector from deflated \(X\) matrix. The difference in the former approach is that in Equation (4), the column is selected before deflation and only the single nonnormalized score vector is deflated at a particular time, while in later, the whole \(X\) matrix is first deflated and then the column vector is selected as the nonnormalized score for that step.

Please note that the \(\tau(n \times 1)\) in Equation (4) is the \(i^{th}\) column vector of the predictor matrix carrying the maximum covariance with the response variables. Furthermore, Equation (4) followed by the normalization in Equation (5) defines a recurrence equation for computing the orthonormal scores \(T(n \times 1)\). The associated nested sequence of vector equations can be solved from the starting score vector \(t_1(n \times 1)\), and the starting score vector \(t_1\) can be estimated without the need of any deflation operation. All scores can be accumulated as column vectors in a score matrix (Equation 6) which can later be used for postprocessing operations such as estimation of regression coefficients.
\[ t_i = \frac{\tau_i}{\|\tau_i\|} \]  
\[ T = [T \ t_i] \]

Since fCovSel and CovSel are cases of NIPALS, it is possible to extract the regression vectors using the traditional regression coefficient \( b \) (total variables \( \times \) 1) estimation equation for NIPALS approach (Equation 7) as has been reported in several earlier works [23,27].

\[ b = W(PW)^{-1}q \]

Note that the \( W \) (weight matrix) in the case of fCovSel or CovSel is the identity matrix of size same as the total number of selected variables as the loading weights are of the form \([0,0,0 \ldots 1 \ldots 0,0,0]\). \( P \) is the loading matrix (total selected variables \( \times \) total selected variables) which can be obtained by postprocessing as \( P = X' T \), and the \( q \) (1 \( \times \) total selected variables) can be obtained by postprocessing \( q = T' y \) or \( Q = T' Y \), for multiresp case. Since the postprocessing operation is the same for fCovSel and CovSel, therefore, this study will only compare the time requirements for variable extraction as that is the main step where CovSel and fCovSel are different.

Note that since the fCovSel algorithm leads to the same solution as the CovSel algorithm [7] presented in the earlier study, this means that all the operations that were possible with the CovSel algorithm such as the identification of key variables, building regression or classification predictive models, are also naturally possible with fCovSel. Furthermore, all extensions such as multiblock [8] and multiway [22] can directly incorporate the fCovSel algorithm without compromising any results but achieving results in less time due to no predictor matrix deflation step in the fCovSel algorithm. The fCovSel algorithm will be made available at: https://github.com/puneetmishra2

3 | DATASETS FOR COMPARISON OF COVSEL AND FCOSSEL

The apricot data set was used to show the similarity of the fCovSel concerning the CovSel analysis and the benefits of the fCovSel such as speed gain in optimization of multiblock models such as SO-CovSel [8]. The apricot data set was a two-block data, where NIR (near infrared 800–2770 nm) and MIR (mid-infrared 4000–650 cm\(^{-1}\)) spectra were acquired on 750 apricot puree samples. As a reference property, SSC (soluble solids content %) was used. More details on the data set can be found in earlier study related to the measurement of samples and reference analysis [28]. The NIR data block had 769 variables, while the MIR data block had 579 variables. The CovSel and the fCovSel analysis aimed to select the variables that carry high covariance with the SSC.

4 | RESULTS AND DISCUSSION

The proposed fCovSel algorithm allows to achieve the same solution as the CovSel approach but in a much faster way due to no deflation of the predictor matrix. The similarity in the results can be noted with the CovSel and the fCovSel analysis performed on the MIR data to explain the SSC. To verify if the order of extracted variables was the same, the variable indexes extracted by the CovSel and the fCovSel are plotted as correlation plots. A correlation of 1 suggested that the order of extracted variables was the same for the CovSel and the fCovSel verifying the identical variable selection (Figure 1).

To have a comparison of execution speeds, the CovSel and the fCovSel analyses were performed as a function of the increasing number of variables and the results are shown in Figure 2. It can be noted that the fCovSel always performed faster than the CovSel algorithm requiring predictor matrix deflation. The effect of sample size was also explored, and the results are shown in Figure 3. The reduction of sample size from 750 to 579 had minimal effect on the execution speed of both the CovSel and the fCovSel algorithm. The size was reduced from 750 to 579 as there were only 579 variables in the data set. The analysis presented in Figures 2 and 3 shows that the fCovSel performed faster than the CovSel; however, the time gain was a couple of seconds. This may appear as a slight improvement, but the real benefit of the
speed will be when the exhaustive operations are performed. For example, one of the main steps in the CovSel based variable selection is to incorporate the cross-validation procedure to select the optimal number of variables. The second even more exhaustive task is to perform cross-validation in the multiblock variable selection scenario such as for the SO-CovSel [8], requiring all combinations of selected variables to be explored. In such a multiblock cross-validation scenario, there will be repeated variable selection analysis involved; hence, a faster algorithm such as the fCovSel can show its real benefit in the time gain. As an example, the SO-CovSel and the SO-fCovSel analyses were performed on two blocks of apricot data set (NIR and MIR) to explain the variance in SSC. The analysis involved exploration of variable combinations in the range from (1, 1) to (20, 20). In Figure 4, the x-axis is mentioned as 1 to 20 but indicates (1, 1) to (20, 20). The result of the analysis clearly shows the potential benefit of the fCovSel over the CovSel in terms of time saving for exhaustive modeling tasks such as multiblock modeling. In the same way, the algorithm can also speed up the multiway multiblock CovSel approach [22] as the most time-consuming step in the multiway CovSel is the deflation of a wide matrix achieved by unfolding the multiway array.
We presented a newer faster algorithm for covariate selection called fCovSel which is naturally faster than the CovSel algorithm due to no predictor matrix deflation step involved. The comparison of execution speed based on real data set showed that for the same number of variables extracted, the execution time for the fCovSel was lower than the CovSel. Furthermore, the exploration of the fCovSel in the exhaustive multiblock variable combinations exploration showed a huge time reduction with the use of the fCovSel compared with the CovSel. The fCovSel can become the preferred choice for covariates selection over the CovSel when a faster execution speed is needed and particularly serves as the backbone to the multiblock and multiway extension of the CovSel as their optimization is usually exhaustive and time-consuming.

**CONCLUSIONS**

We presented a newer faster algorithm for covariate selection called fCovSel which is naturally faster than the CovSel algorithm due to no predictor matrix deflation step involved. The comparison of execution speed based on real data set showed that for the same number of variables extracted, the execution time for the fCovSel was lower than the CovSel. Furthermore, the exploration of the fCovSel in the exhaustive multiblock variable combinations exploration showed a huge time reduction with the use of the fCovSel compared with the CovSel. The fCovSel can become the preferred choice for covariates selection over the CovSel when a faster execution speed is needed and particularly serves as the backbone to the multiblock and multiway extension of the CovSel as their optimization is usually exhaustive and time-consuming.

**PEER REVIEW**

The peer review history for this article is available at https://publons.com/publon/10.1002/cem.3397.

**ORCID**

Puneet Mishra https://orcid.org/0000-0001-8895-798X
REFERENCES

1. Wold S, Albano C, Dunn WJ, et al. Multivariate data analysis in chemistry. In: Kowalski BR, ed. Chemometrics: Mathematics and Statistics in Chemistry. Springer Netherlands: Dordrecht; 1984:17-95.
2. Andersson M. A comparison of nine PLS1 algorithms. J Chemometr. 2009;23(10):518-529.
3. Saeyes W, Nguyen Do Trong N, Van Beers R, Nicoli BM. Multivariate calibration of spectroscopic sensors for postharvest quality evaluation: A review. Postharvest Biology and Technology. 2019;158:110981. doi:10.1016/j.postharvbio.2019.110981
4. Bro R, Smilde AK. Principal component analysis. Anal Methods. 2014;6(9):2812-2831.
5. Wold, S., PLS Modeling With Latent Variables in Two or More Dimensions. 1987.
6. Wold S, Sjostrom M, Eriksson L. PLS-regression: A basic tool of chemometrics. Chemom Intel Lab Syst. 2001;58(2):109-130.
7. Roger JM, Palagos B, Bertrand D, Fernandez-Ahumada E. CovSel: Variable selection for highly multivariate and multi-response calibration: Application to IR spectroscopy. Chemom Intel Lab Syst. 2011;106(2):216-223.
8. Biancolillo A, Marini F, Roger J-M. SO-CovSel: A novel method for variable selection in a multiblock framework. J Chemometr. 2020;34(2):e3120. doi:10.1002/cem.3120
9. Höskuldsson A. Variable and subset selection in PLS regression. Chemom Intel Lab Syst. 2001;55(1):23-38.
10. Zou XB, Xiaobo Z, Jiewen Z, Povey MJW, Holmes M, Hanpin M. Variables selection methods in near-infrared spectroscopy. Anal Chim Acta. 2010;667(1-2):14-32.
11. Chandrashekar G, Sahin F. A survey on feature selection methods. Computers & Electrical Engineering. 2014;40(1):16-28.
12. Mehmoord T, Liland KH, Snipen L, Sebo S. A review of variable selection methods in partial least squares regression. Chemom Intel Lab Syst. 2012;118:62-69.
13. Biancolillo A, Liland KH, Måge I, Nørgaard L, Saudland A, Wagner J, Thiis T, Bro R. Variable selection in multi-block regression. Chemom Intel Lab Syst. 2016;156:89-101.
14. Leardi R, Boggia R, Terrile M. Genetic algorithms as a strategy for feature selection. J Chemometr. 1992;6(5):267-281.
15. Stefansson P, Liland KH, Thiis T, Burud I. Fast method for GA-PLS with simultaneous feature selection and identification of optimal preprocessing technique for datasets with many observations. J Chemometr. 2020;34:e3195. doi:10.1002/cem.3195
16. Nørgaard L, Saudland A, Wagner J, Nielsen JP, Munck L, Engelsen SB. Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy. Appl Spectrosc. 2000;54(3):413-419.
17. Mehmoord T, Sebo S, Liland K. Comparison of variable selection methods in Partial Least Squares Regression. J Chemometr. 2020;34:e3226. doi:10.1002/cem.3226
18. Wold S, Ruhe A, Wold H, Dunn IIIWJ. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. SIAM Journal on Scientific and Statistical Computing. 1984;5(3):735-743.
19. Mishra P, Woltering EJ. Identifying key wavenumbers that improve prediction of amylose in rice samples utilising advanced wavenumber selection techniques. Talanta. 2021;224:121908. doi:10.1016/j.talanta.2020.121908
20. Mishra P, Marini F, Brouwer B, et al. Sequential fusion of information from two portable spectrometers for improved prediction of moisture and soluble solids content in pear fruit. Talanta. 2021;223(Pt 2):121733. doi:10.1016/j.talanta.2020.121733
21. Mishra P, Woltering E, Brouwer B, Hoveveen-van Echtelt E. Improving moisture and soluble solids content prediction in pear fruit using near-infrared spectroscopy with variable selection and model updating approach. Postharvest Biology and Technology. 2021;171:111348. doi:10.1016/j.postharvbio.2020.111348
22. Biancolillo AM, Marini F, Roger J-M. N-CovSel, a new strategy for feature selection in N-way data. Conference Proceeding of 17th Scandinavian Symposium on Chemometrics: Denmark. 2021.
23. Liland KH, Næs T, Indahl UG. ROSA—A fast extension of partial least squares regression for multiblock data analysis. J Chemometr. 2016;30(11):651-662.
24. Björck Å, Indahl UG. Fast and stable partial least squares modelling: A benchmark study with theoretical comments. J Chemometr. 2017;31(8):e2898.
25. Indahl UG. The geometry of PLS1 explained properly: 10 key notes on mathematical properties of and some alternative algorithmic approaches to PLS1 modelling. J Chemometr. 2014;28(3):168-180.
26. Draper NR, Smith H. Applied Regression Analysis. John Wiley & Sons; 1998.
27. Liland KH, Stefansson P, Indahl UG. Much faster cross-validation in PLSR-modelling by avoiding redundant calculations. J Chemometr. 2020;34(3):e3201.
28. Bureau S, Ruiz D, Reich M, et al. Application of ATR-FTIR for a rapid and simultaneous determination of sugars and organic acids in apricot fruit. Food Chem. 2009;115(3):1133-1140.

How to cite this article: Mishra P. A brief note on a new faster covariate’s selection (fCovSel) algorithm. Journal of Chemometrics. 2022;36(5):e3397. doi:10.1002/cem.3397