CT-GUI: A graphical user interface to perform calibration transfer for multivariate calibrations

Puneet Mishra

Food and Biobased Research, Wageningen University and Research, Bornse Weilanden 9, 6708, WG, Wageningen, the Netherlands

ARTICLE INFO

Keywords:
Orthogonal Projection
Instrument transfer
Model maintenance

ABSTRACT

A new MATLAB based graphical user interface integrating several multivariate calibration transfer approaches is presented. Several methods for standard-based and standard-free calibration transfer are integrated. A demonstration of the toolbox and functionalities of different approaches was judged using a calibration transfer case of the corn dataset. All methods were tested on the corn dataset and the results showed that the toolbox is ready to be used. The best feature of the toolbox is that it is a push-button toolbox that even non-expert can use to achieve complex calibration transfer tasks. The CT-GUI toolbox is available at https://tinyurl.com/PMishra.

1. Introduction

Multivariate chemometric calibrations are highly specific to the primary instrument (spectrometer) on which the data was first collected and modelled [1–3]. Hence, simply reusing a model made for an instrument on a new instrument is not possible as different instruments have intrinsic differences such as different instrument components such as the detector, illumination, difference in spectra range or resolution, or differences in the surrounding environment in which the instrument is operated [2]. To deal with this challenge, extensive development in the chemometrics domain have taken place and several calibration transfer (CT) strategies are now available to transfer models from one instrument to another [2,3].

In the domain of the chemometrics, two types of CT methods are available i.e., standard-based [2] and standard-free [3]. The standard-based methods are fairly developed and widely used in practice, while the standard-free methods are recent. The standard-based methods such as direct standardization (DS) [4] and piece direct standardization (PDS) [5], requires standard samples to be measured on both instruments to model instrumental differences and/or compensate for them [2]. By measuring the same standard sample on different instruments, it is assumed that the differences in the signals are due to the intrinsic differences in the instruments. The instrument differences can be modelled as the transfer function. The transfer function can be used to transform the data before any model application [4,5]. However, in many scenarios, the measurement of standard samples on both instruments is not possible [6–8], for example, when instruments are in remote locations and the primary instrument is damaged. Hence, there is an increasing trend towards the development of standard-free CT methods to support model sharing across different instruments. Recently, several approaches to perform standard-free CT have been proposed in the literature [3]. Approaches range from techniques that require non-standard samples such as the dynamic orthogonal projections (DOP) [9,10] where the need for standard samples is replaced by the measurement of new samples just on the slave instrument, domain adaption (DA) techniques such as transfer component analysis (TCA) [11,12], domain-invariant PLS regression [23], and parameter-free calibration enhancement (PFCE) framework which allows updating of model based on some new measurements from a different instrument [13].

Although there have been significant developments in the chemometrics domain related to new CT methods [2,3], the use of such methods is limited to highly skilled chemometricians with competence in programming. Wider usage of such methods by non-experts requires methods to be available in easy-to-use interfaces, where the CT can be performed with basic or even no skills in programming. The best solution for non-expert users is the graphical user interface (GUI), as it requires minimal programming skills as well reduces the time required to adapt or standardize the codes from scratch. GUIs are gaining popularity and recently several chemometric toolboxes have become available, such as the batch process modelling and monitoring toolbox (MV BATCH) [14], the data pre-processing and multivariate statistical process control toolbox (PreScreen) [15], the trilinear data discriminant analysis toolbox (TTWD-DA) [16], the multivariate calibration toolbox (TOMCAT) [17], the hyperspectral image analysis toolbox (HyperTools) [18], the multi-block analysis toolbox (MBA-GUI) [19], the batch effect correction toolbox (FRUITNIR-GUI) [12], the multivariate exploratory data analysis
toolbox (MEDA) [20] and the N-Way toolbox [21]. Until now, there is no dedicated GUI that integrates several CT techniques. A command-line toolbox called SAISIR [22] provides a subset of CT techniques but they are for highly skilled programmers and a non-expert lacking programming skill may never implement them.

The objective of this work is to provide a new MATLAB based GUI integrating several multivariate CT approaches. Several methods for standard-based and standard-free CT are integrated. A demonstration of the toolbox and functionalities of different approaches was judged using a CT case of the corn dataset. All methods were tested on the corn dataset and the results showed that the toolbox is ready to be used.

2. Software description

The CT-GUI was built utilizing the application builder in MATLAB version 2018b (MathWorks, Natick, MA, USA). The application can be downloaded and installed in MATLAB (preferred versions 2018b or higher) or can be used as a stand-alone executable or can be run through the ‘.mlapp’ files in the MATLAB command line. If the user does not have MATLAB version 2018b or greater then it is recommended to install the free MATLAB runtime tool and run the app as a standalone. All the MATLAB functions can be downloaded from: https://tinyurl.com/PMishra2. The standalone application can be downloaded from: https://tinyurl.com/PMishra. The free MATLAB runtime can be downloaded from: https://nl.mathworks.com/products/compiler/matlab-runtime.html. The compatible runtime for this toolbox is version 9.5 and 2018b. Before installing the standalone application, the MATLAB runtime environment should be installed.

To run the toolbox from the command line, the user should use the toolbox folder as the current folder and type ‘T1’ on the command line to start the main GUI interface. The GUI supports the input data formats.csv, ‘.xlsx’ and ‘.mat’. The input to the toolbox is the matrix of spectra of size $n \times 1 \times p$ where $n$ is the number of samples and $p$ is the number for variables. The extra 1 is the first row of the matrix which are the variables (for examples wavelengths in the case of optical spectroscopy). The reference property matrix must be of size $n \times 1$. The toolbox has three types of data pre-processing’s, i.e., smoothing, scatter correction and normalization, and derivation, which it inherits from the wide collection available in MBA-GUI [19].

3. Brief description of methods integrated

3.1. Direct standardization

DS [4] relates the response of a standard sample measured with one instrument ($R_1$) to its response obtained on another instrument ($R_2$) using a linear transformation matrix (F) as $R_1 = R_2 F$, where F is a squared matrix with dimension equal to the number of wavelengths. The F matrix is obtained as $F = R_2 R_1^T$. For any new data acquisition, the spectra are transformed with the F matrix such that the model made on the primary instrument can be directly used on the secondary instrument. In the GUI, DS is implemented by means of an in-house MATLAB code.

3.2. Piece wise direct standardization

In the DS method, the whole spectrum of the secondary instrument is used to fit each point of the spectrum of the primary instrument. However, in most cases, the spectral variation is present in localized regions. Hence, the new method PDS [5] was designed to map the local variation in the secondary instrument with each point on the primary instrument. The PDS [5] transformation matrix is used to transform any new spectrum measured on a secondary instrument, thus allowing calibration model transfer. A key point to note is that unlike DS, which is almost parameter-free, the PDS approach requires to decide the optimal window size to map the local variation in the secondary instrument data. In the GUI, PDS is implemented by means of an in-house MATLAB code.

3.3. Dynamic orthogonal projection

Dynamic orthogonal projection (DOP) is a model maintenance method developed to deal with physical, chemical, and environmental effects in spectroscopic modelling [9]. The approach is based on the correction of the calibration dataset based on the new reference measurements performed in different physical, chemical, and environmental conditions. The correction is performed using orthogonal projections based on the subspace defined by the difference of the calibration spectra and the new condition spectra. The DOP method estimates virtual standards, i.e., the spectra that should have been measured if the calibration conditions had not varied. This is accomplished by means of linear combinations of the original calibration data matrix, whose coefficients are calculated using kernels centred on reference values. Once the virtual standards are prepared then the difference spectra are calculated. The orthogonal basis for the difference spectra is estimated by singular value decomposition (SVD) and finally, the original spectra are projected orthogonally to that basis. This removes the external influences from the spectra and then the model recalibrated on these data becomes insensitive to the differences (physical, chemical, and environmental conditions). In the GUI, DOP is implemented by means of an in-house function [10]; once the data are corrected, then a PLS regression model is built as described previously. The codes for DOP were inherited from the previously published external effect removal FRUIT-NIR GUI toolbox [12].

3.4. Parameter free calibration enhancement

The parameter-free calibration enhancement framework (PFCE) is a recently proposed method to update existing calibrations based on the measurements from new instruments or different measurement conditions [13]. The PFCE can be used in three scenarios i.e., non-supervised PFCE (NS-PFCE), semi-supervised PFCE (SS-PFCE), and full-supervised PFCE (FS-PFCE). The NS-PFCE requires standard samples spectra from primary and secondary instrument to update the primary model by implementing a correlation constraint on the regression coefficients. The SS-PFCE and FS-PFCE methods integrate the spectra from the second instrument and corresponding reference values to update the models [13]. The NS-PFCE is like the DS approach to calibration transfer, as they both require the spectra from primary and secondary instruments to model the difference. The SS-PFCE and FS-PFCE can be considered like the DOP approach as they both require some new spectral and reference measurements to update the models. The PFCE was integrated with the open codes provided along with the manuscript [13]. While using the PFCE integrated into this toolbox, please also cite the original manuscript i.e. [13].

4. Data used for demonstration

To demonstrate the functionalities of the toolbox the publicly available corn data set was used (www.eigenvector.com/data/Corn). From the corn data set, data corresponding to two spectrometers i.e., M5 and MP6 was used. The aim was to transfer the model from M5 to MP6 spectrometer. The corn data set has 80 spectra of corn samples and corresponding reference properties. To demonstrate the functionalities of the toolbox, the protein content was used as the reference property. The 80 spectra were divided into calibration (60%) and independent test set (40%) using the Duplex algorithm. The calibration set was used for primary model development and calibration transfer, while the independent test set was used for evaluating the performance of transferred calibration. The mean spectra from the M5 and MP6 spectrometers are shown in Fig. 1. The differences in the mean spectra can be noted along with the whole spectral range, where the MP6 spectrometer has a globally lower signal intensity compared to the M5 spectrometer. Such a difference in mean spectral response for the same samples indicates the intrinsic difference in the spectrometers, which should be carefully dealt with while transferring the model from one spectrometer to another. A summary of
5. Demonstration

The starting interface of the toolbox is shown in Fig. 2. The interface provides options to load data from two instruments. Further, several pre-processing approaches can be explored to enhance the data. Later, from the drop-down menu, a relevant CT method can be selected and used.

An example of running the DS calibration transfer and general outputs from the toolbox are shown in Figs. 3–5. The DS-CT provides a plot for optimal latent variable selection where the PLS models with (bottom row Fig. 3) and without CT are optimized (top row Fig. 3). For example, the PLS model without CT suggests 9 LVs (Latent Variables) for model calibration, while the PLS after DS-CT suggest only 8 LVs to optimize the model sufficient to be used on a new instrument. Hence, 8 LVs can be used to calibrate the final model. The outputs of the final model calibrated based on 8 LVs are shown in Fig. 4A and B. After the model calibration, the test set can be loaded, and the model can be deployed. A summary of the results of the test set for corn data is shown in Fig. 4C. It can be noticed that after the DS-CT the RMSE (Root Mean Squared Error) for the calibration (Fig. 4B) and test set (Fig. 4C) were similar ~0.14%.

The CT methods also provide an extra plot presenting the regression vector of the final model (Fig. 5A) and the difference between the regression vectors before and after CT (Fig. 5B). The regression vector of the final models can be used for interpretation of the model, while the difference vector can be used to understand the main spectral regions carrying the differences in the spectral response of two instruments.

To demonstrate that all integrated methods are fully functional, the corn data was analyzed with each technique. A summary of the results for the calibration enhancement framework which will benefit users of spectroscopy in multiple situations. From my side, I am confident that this toolbox will serve an important purpose for transferring and updating spectral models.”

6. Conclusion

A new GUI for calibration transfer (CT) was presented. The toolbox functionalities were demonstrated on the corn data set. The result showed that the toolbox is fully functional, and all methods are ready to be used. On the corn data set, all integrated CT techniques performed similarly. The user may explore the toolbox and find the best-suited technique for their CT challenge. An exemplary tutorial video to use the GUI functionalities can be accessed at: CT-GUI quick test.

In the case of difficulty in software usage, the user can always post the comment in the YouTube video related to CT-GUI and the developer of toolbox will support the user in resolving the issue.

7. Independent testing

Name: Dr. Lalit Mohan Kandpal.
Job: Research scientist.
Affiliation: Ghent University, Belgium Email: LalitMohan.Kandpal@UGent.be Phone: +32465759700.

“Dr. Lalit Mohan Kandpal in Precision soil & crop engineering group, Department of Environment, Ghent University, B-9000 Gent, Belgium. I have tested the “CTGUI” graphical user-friendly interface based on MATLAB for the analysis of soil chemical data, and it appears to function as the authors described. When analysing the soil data with spectral sensor, it is important to standardize tailored procedure and retrieve results that can answer the analytical problem. Further, when the model needs to be deployed in the field, then very often the models made on laboratory instruments are required to be transferred to portable handheld field instruments. This is precisely what CTGUI does.

The toolbox according to the link provided has two options i.e., command line tools as well as the stand-alone tool. The main benefit of the standalone tool is that it does not require MATLAB software to run on computer. I have tested the standalone tool with the MATLAB 2018b runtime tool as well as command line on the MATLAB 2020b. The GUI is working well in both the MATLAB versions. Further, I have also performed the analysis according to the tutorial video provided with the toolbox at: https://www.youtube.com/watch?v=fl1EnUwV4. I was able to replicate all the analysis as demonstrated in the tutorial video.

The main benefit of the toolbox is that it provides most used CT techniques such as direct standardization, piecewise direct standardization to be easily used by even non-experts. Furthermore, the toolbox also integrates recent model updating approaches such as parameter free calibration enhancement framework which will benefit users of spectroscopy in multiple situations. From my side, I am confident that this toolbox will serve an important purpose for transferring and updating spectral models.”
Fig. 2. The starting interface of the toolbox. This interface allows user to upload data and pre-process it and select the relevant calibration transfer or model updating approach.

Fig. 3. A summary of latent variables optimization routine for direct standardization calibration transfer. PLS model (A) Explained variance in response variable as a function of latent variables, and (B) mean square error as a function of latent variables. DS corrected PLS model (C) Explained variance in response variable as a function of latent variables, and (D) mean square error as a function of latent variables. The user can use the (D) to select the optimal number of latent variables for final model.
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R.N. Feudale, N.A. Woody, H. Tan, A.J. Myles, S.D. Brown, J. Ferré, Transfer of multivariate calibration models: a review, Chemom. Intell. Lab. Syst. 64 (2002) 181–192.
[2] J.J. Workman, A review of calibration transfer practices and instrument differences in spectroscopy, Appl. Spectrosc. 72 (2017) 340–365.
[3] B. Malli, A. Birlutiu, T. Natschlager, Standard-free calibration transfer - an evaluation of different techniques, Chemom. Intell. Lab. Syst. 161 (2017) 49–60.
[4] Y. Wang, D.J. Veltkamp, B.R. Kowalski, Multivariate instrument standardization, Anal. Chem. 63 (1991) 2750–2756.
[5] Y. Wang, B.R. Kowalski, Calibration transfer and measurement stability of near-infrared spectrometers, Appl. Spectrosc. 46 (1992) 764–771.
[6] P. Mishra, R. Nikzad-Langerodi, Partial Least Square Regression versus Domain Invariant Partial Least Square Regression with Application to Near-Infrared Spectroscopy of Fresh Fruit, Infrared Physics & Technology, 2020, 103547.
[7] R. Nikzad-Langerodi, F. Sobieczky, Graph-based Calibration Transfer, 2020 arXiv preprint arXiv:2006.00089.
[8] R. Nikzad-Langerodi, W. Zellinger, S. Saminger-Platz, B.A. Moser, Domain adaptation for regression under Beer–Lambert’s law, Knowl. Base Syst. 210 (2020), 106447.
[9] M. Zealiter, J.M. Roger, V. Bellon-Maurel, Dynamic orthogonal projection. A new method to maintain the on-line robustness of multivariate calibrations. Application to NIR-based monitoring of wine fermentations, Chemom. Intell. Lab. Syst. 80 (2006) 227–235.
[10] P. Mishra, J.M. Roger, D.N. Rutledge, E. Woltering, Two standard-free approaches to correct for external influences on near-infrared spectra to make models widely applicable, Postharvest Biol. Technol. 170 (2020), 111326.
[11] S.J. Pan, I.W. Tsang, J.T. Kwok, Q. Yang, Domain adaptation via transfer component analysis, IEEE Trans. Neural Network. 22 (2011) 199–210.
[12] P. Mishra, J.M. Roger, F. Marini, A. Biancolillo, D.N. Rutledge, Fruitmir-Gui, A graphical user interface for correcting external influences in multi-batch near infrared experiments related to fruit quality prediction, Postharvest Biol. Technol. 22 (2) (2020) 199–210, 111414.
[13] J. Zhang, B. Li, Y. Hu, L. Zhou, G. Wang, G. Guo, Q. Zhang, S. Lei, A. Zhang, A parameter-free framework for calibration enhancement of near-infrared spectroscopy based on correlation constraint, Anal. Chim. Acta 1142 (2021) 169–178.
[14] J.M. Gonzalez-Martinez, J. Camacho, A. Ferrer, MVBatch, A matlab toolbox for batch process modeling and monitoring, Chemom. Intell. Lab. Syst. 183 (2020) 122–133.
[15] G. Yi, C. Herdsman, J. Morris, A MATLAB toolbox for data pre-processing and multivariate statistical process control, Chemom. Intell. Lab. Syst. 194 (2019), 103863.
[16] C.J.M. Morais, K.M.G. Lima, F.L. Martin, Twed-DA, A MATLAB toolbox for discriminant analysis based on trilinear three-way data, Chemom. Intell. Lab. Syst. 188 (2019) 46–53.
[17] M. Dasykowski, S. Sernee, K. Kaczmarek, P. Van Espen, C. Croux, B. Walczak, Tomcat, A MATLAB toolbox for multivariate calibration techniques, Chemom. Intell. Lab. Syst. 85 (2007) 269–277.
[18] N. Mobaraki, J.M. Amigo, HYPER-Tools. A graphical user-friendly interface for hyperspectral image analysis, Chemom. Intell. Lab. Syst. 172 (2018) 174–187.
[19] P. Mishra, J.M. Roger, D.N. Rutledge, A. Biancolillo, F. Martini, A. Nordon, D. Jouan-Rimbaud-Bouveresse, Mbai-Gui, A Chemometric Graphical User Interface for Multi-Block Data Visualisation, Regression, Classification, Variable Selection and Automated Pre-processing, Chemometrics and Intelligent Laboratory Systems, 2020, 104139.
[20] J. Camacho, A. Pérez-Villegas, R.A. Rodríguez-Gómez, E. Jiménez-Mazzas, Multivariate exploratory data analysis (MEDA) toolbox for matlab, Chemom. Intell. Lab. Syst. 143 (2015) 49–57.
[21] C.A. Andersson, R. Bro, The N-way toolbox for MATLAB, Chemom. Intell. Lab. Syst. 52 (2000) 1–4.
[22] C.B.Y. Cordella, D. Bertrand, Saisir, A new general chemometric toolbox, Trac. Trends Anal. Chem. 54 (2014) 75–82.
[23] Ramin Nikzad-Langerodi, et al., Domain-invariant partial-least-squares regression, Analytical chemistry 90 (11) (2018) 6693–6701, https://doi.org/10.1021/acs.analchem.8b04948.