Food Chemistry 419 (2023) 136055

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

A methodological approach to preprocessing FTIR spectra of adulterated sesame oil

Rasool Khodabakhshian [a,*], Hajarsadat Seyedalibeyk Lavasani [a], Philipp Weller [b]

[a] Department of Biosystems Engineering, Ferdowsi University of Mashhad, 9177948978 Mashhad, Iran
[b] Institute for Instrumental Analytics and Bioanalytics, Mannheim University of Applied Sciences, 68163 Mannheim, Germany

ARTICLE INFO

Keywords:
Infrared spectroscopy
Food quality
PLSR
OSC
SNV
EMSC

ABSTRACT

Fourier transform infrared (FTIR) spectroscopy is established as an effective and fast method for the confirmation of the authenticity of food and among other, edible oils. However, no standard procedure is available for applying preprocessing as a vital step in obtaining accurate results from spectra. This study proposes a methodological approach to preprocessing FTIR spectra of sesame oil adulterated with vegetable oils (canola oil, corn oil, and sunflower oil). The primary preprocessing methods investigated are orthogonal signal correction (OSC), standard normal variate transformation (SNV), and extended multiplicative scatter correction (EMSC). Other preprocessing methods are used both as standalone methods and in combination with the primary preprocessing methods. The preprocessing results are compared using partial least squares regression (PLSR). OSC alone or with detrending were the most accurate in predicting the adulteration level of sesame oil, with a maximum coefficient of prediction (R_p^2) range of 0.910 to 0.971 for different adulterants.

1. Introduction

Infrared spectroscopy is an efficient and cost-effective analytical tool for analyzing food components (Shi et al., 2019) and for food authentication (Horn et al., 2018) and, in many cases, does not require sample preparation. As strongly demonstrated in previous studies in recent years, Fourier transform infrared (FTIR) spectroscopy has been used as an effective and non-destructive method for adulteration detection in food (Aboulwafa et al., 2019; Jamwal et al., 2021; Yang et al., 2020). For example, adulteration in turmeric powder adulterated with starch powder, sudan red, and metanil yellow was detected with over 90 percent precision using partial least squares-discriminant analysis (PLS-DA) (Khodabakhshian et al., 2021). Outside of food analysis, autonomous strategies have also been developed incorporating FTIR spectroscopy as a means for parameter-free spectral monitoring chemical processes (Fath et al., 2021). However, it should be noted that typically, IR spectroscopy is not suitable for samples such as blood (Wang et al., 2021) that contain water because water molecules imply high noise levels and strong absorption (Yao et al., 2016), which means that the results are susceptible to water absorption. Meanwhile, mid-infrared (MIR) approaches based on ultra-thin flow cells for aqueous samples are commercially available (Hofmann et al. 2017).

FTIR spectroscopy is used for monitoring and detecting the authenticity and adulteration of edible oils (Mousa et al., 2022). The infrared region used in FTIR extends from 4000 to 400 cm^{-1} (2500 to 15000 nm) (Manley, 2014). Due to health benefits and being more expensive than other vegetable oils, cold-pressed sesame oil is a good candidate and might be adulterated with cheaper oils (Dogruer et al., 2021). Edible oils contain 95–99% triglycerides (Bunaciu et al., 2020) and the FTIR spectra of different vegetable oils, including sesame oil, canola oil, corn oil, and sunflower oil, are particularly hard to decipher as their spectra are quite similar (Fadzlillah et al., 2014; Ozulku et al., 2017). Preprocessing therefore plays a significant role in the accurate detection of pure and adulterated samples.

As preprocessing is a vital step for obtaining reasonable results from spectral data (Rinnan et al., 2009), the impact of the methodological choice of preprocessing method on the results must be investigated (Nørgaard et al., 2020). The choice of a suitable preprocessing method is challenging because such methods may introduce unwanted variation or even remove valuable information (Biney et al., 2021). Studies have been conducted on common preprocessing methods and their effects on classification accuracy in the last decade (Khodabakhshian et al., 2022; Khodabakhshian et al., 2017; Moghimi et al., 2010; Rinnan et al., 2009; Shi et al., 2019). Furthermore, the effect of changing the order of

* Corresponding author.
E-mail address: khodabakhshian@um.ac.ir (R. Khodabakhshian).

https://doi.org/10.1016/j.foodchem.2023.136055
Received 19 July 2022; Received in revised form 3 March 2023; Accepted 26 March 2023
Available online 5 April 2023
0308-8146/© 2023 Elsevier Ltd. All rights reserved.

applying preprocessing methods and the possible effect on the model's accuracy have been studied (Horn et al., 2018). However, it is speculated that preprocessing spectra can hinder the performance of partial least squares (PLS) models (Mishra et al., 2021).

This study aims to investigate and demonstrate the effect of various preprocessing methods on the classification accuracy of sesame oil mixed with three vegetable oils (canola oil, corn oil, and sunflower oil) and to determine whether preprocessing can significantly affect the accuracy of classification. Different combinations of preprocessing methods are applied to the FTIR spectra of samples, and the accuracy of the results will be analyzed using PLSR. The strengths and limitations of every preprocessing method and its accuracy in detecting the adulteration in sesame oil will be discussed. Additionally, suggestions will be given on applying preprocessing methods to FTIR spectra of edible vegetable oils.

2. Materials and methods

2.1. Sample collection and preparation

Cold-pressed and filtered sesame oil was provided by Parand Oil in 2021. Canola, corn, and sunflower oils were purchased from supermarket chain stores in Mashhad, Razavi Khorasan Province, Iran. Each adulterant vegetable oil was added to sesame oil at seven levels: 5%, 10%, 15%, 25%, 35%, 50%, and 75% (v/v). In total, 25 samples, including 21 adulterated sesame oil samples and four pure oil samples (sesame, canola, corn, and sunflower oils), were prepared. The samples were stored in a cool, dry, and dark place at all times before the analysis. Each sample was vortexed in the laboratory before spectroscopy to ensure the accuracy of the results.

2.2. FTIR spectra acquisition

Transmittance spectra were obtained with an Avatar 370 FTIR spectrometer (Thermo Nicolet Corporation, Madison, WI, USA), and OMNIC spectroscopy software (Thermo Nicolet Corporation, Madison, WI, USA) was used to collect spectral data from the spectrometer. The transmittance spectra were converted to absorbance for analysis. For each measurement, the background spectra of the potassium bromide (KBr) disk were obtained first, and then a drop (0.1 mL) of the sample was dropped on the disk for spectra acquisition. Five measurements were obtained from each sample. All five replicates were measured in one day. The measured spectral range was 4000 cm^{-1} to 400 cm^{-1} with a resolution of 4 cm^{-1} and scanning interval of 1.9285 cm^{-1} which resulted in 1868 data points for each spectrum. As the beam passes through the sample, sample path lengths affect the results, therefore the data were normalized before the spectra can be further analyzed and used for classification. In total, 125 spectra from 25 samples with five replicates were obtained.

2.3. Spectral preprocessing

FTIR Spectra are preprocessed to remove any noise and unwanted information and to improve the accuracy of the resulting model (Yang et al., 2018). Preprocessing has been shown to improve the stability of the model (Li et al., 2020). However, it is important to take into account that preprocessing does not always improve the results, and an unsuitable method can even have a negative effect on a model's accuracy (Mishra et al., 2021). Often, studies utilize more than one preprocessing method, and the performance of different methods is compared (Brendel et al., 2020; Gerhardt et al., 2019). However, in most cases, only one preprocessing method, which can be a single method or a combination of methods, is reported and used for processing, modeling, and classification purposes (Jamwal et al., 2021; Obeidat et al., 2018; Santos et al., 2019). In the absence of a standard selection methodology, preprocessing techniques are chosen based on the researchers' experience or the previous studies done in their field.

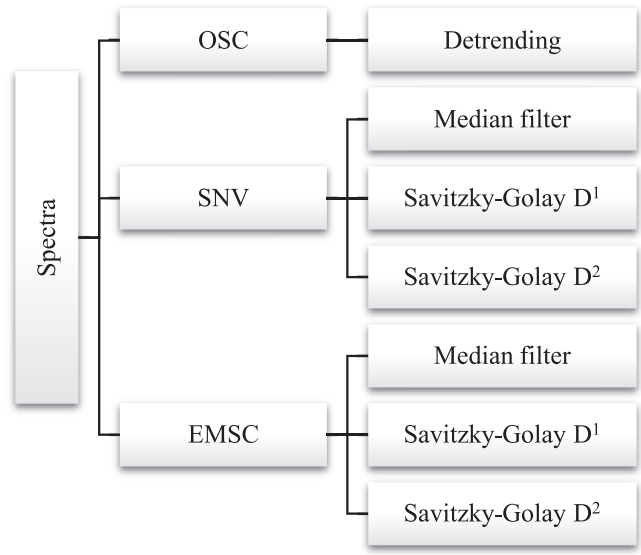

Fig. 1. Preprocessing methods and their combinations.

The primary preprocessing methods investigated in this study are orthogonal signal correction (OSC), standard normal variate transformation (SNV), and extended multiplicative scatter correction (EMSC). Other preprocessing methods, namely median filter smoothing, Savitzky-Golay first (D^1) and second (D^2) order derivatives and detrending are used both as standalone methods and in combination with the primary preprocessing methods as shown in Fig. 1. Each preprocessing has been developed for a specific purpose and affects the spectra differently. Before the preprocessing stage, the original spectra were pretreated with baseline correction and normalization. In this article, these pretreated spectra were modeled without any additional preprocessing for evaluating different preprocessing methods. Unscrambler X V10.4 (CAMO Software AS, Bedford, Massachusetts, United States) software was used for preprocessing the data.

The three primary preprocessing methods have been chosen based on their established history and their use in recent literature. Unlike other preprocessing methods, OSC aims to avoid removing important information (Parsaei-Khomami et al., 2022) and to only remove unwanted variation and is a popular choice for preprocessing in various fields (Biney et al., 2021) as it has shown to improve prediction accuracy (Biney et al., 2020; Munnaf & Mouazen, 2022). A modified version of this method, direct orthogonal signal correction (DOSC), can also be used. This algorithm is based on least square steps and was developed and modified by Westerhuits (Xie et al., 2022). Standard normal variate transformation (SNV) can be used to reduce the light scattering effect and also to correct baseline offsets (Upadhyay et al., 2022). Multiplicative scatter correction (MSC) was shown to be superior in correcting the effect of light scattering (Guo et al., 2021). MSC has been modified to the extended multiplicative scatter correction (EMSC) for better improving results.

2.4. Partial least squares regression (PLSR)

Partial least squares regression (PLSR) is widely used for developing regression models based on spectral data (Wang et al., 2021). PLS models are usually fitted using the nonlinear iterative partial least squares (NIPALS) algorithm (Kumar, 2021). In this study, NIPALS and kernel PLS had the same results up to four significant digits, except in some cases when Savitzky-Golay's first and second derivatives were present. In these cases, kernel PLS could not compute a model and the NIPALS algorithm was used instead. The category variable method is

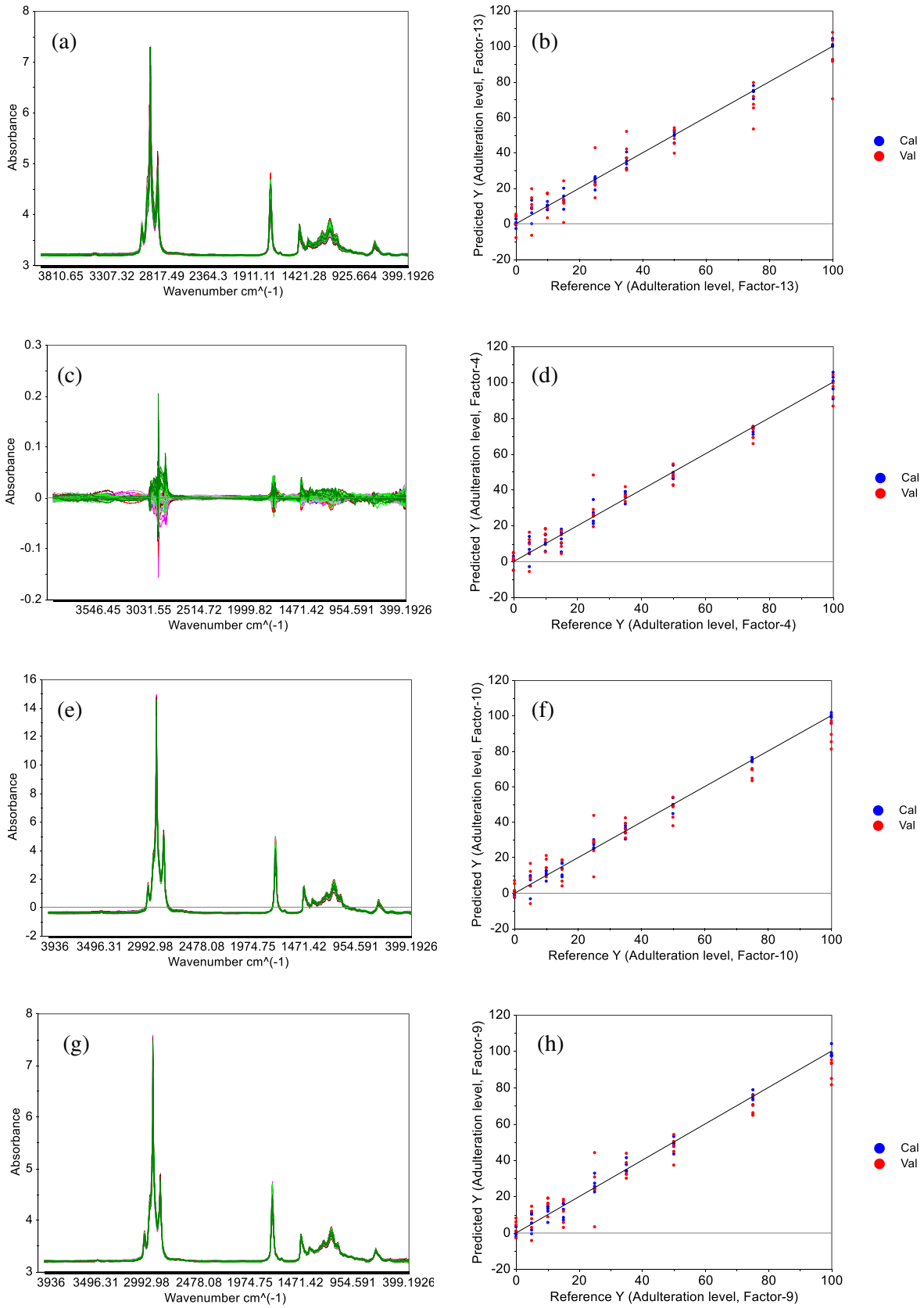

Fig. 2. FTIR spectra of sesame oil (pure and adulterated with canola oil) and quantification models developed by PLSR with no preprocessing (a, b), OSC (c, d), SNV (e, f), and EMSC (g, h).

Table 1
The results of PLSR for different preprocessing techniques (sesame oil adulterated with canola oil).

NO.	Preprocessing	Factor	R_c^2	RMSEC (%)	R_p^2	RMSEP (%)
1	None	13	0.990	3.193	0.935	9.247
2	OSC	4	0.983	4.175	0.967	6.613
3	SNV	10	0.991	2.995	0.952	7.929
4	EMSC	9	0.988	3.545	0.948	8.247
5	Detrending	11	0.986	3.741	0.928	9.676
6	Median filter	11	0.990	3.159	0.943	8.641
7	Savitzky-Golay D^1	12	0.994	2.465	0.953	7.832
8	Savitzky-Golay D^2	11	0.990	3.200	0.939	8.906
9	OSC, Detrending	4	0.985	3.882	0.971	6.173
10	SNV, Median filter	9	0.991	3.113	0.951	7.975
11	SNV, Savitzky-Golay D^1	10	0.993	2.751	0.960	7.199
12	SNV, Savitzky-Golay D^2	16	0.999	0.699	0.932	9.407
13	EMSC, Median filter	9	0.992	2.929	0.962	7.089
14	EMSC, Savitzky-Golay D^1	10	0.993	2.723	0.959	7.295
15	EMSC, Savitzky-Golay D^2	15	0.999	0.879	0.922	10.073

Table 2
The results of PLSR for different preprocessing techniques (sesame oil adulterated with corn oil).

NO.	Preprocessing	Factor	R_c^2	RMSEC (%)	R_p^2	RMSEP (%)
1	None	16	0.992	2.834	0.854	13.81
2	OSC	6	0.994	2.448	0.963	6.942
3	SNV	17	0.997	1.654	0.833	14.786
4	EMSC	15	0.996	2.057	0.869	13.075
5	Detrending	15	0.995	2.216	0.869	13.110
6	Median filter	16	0.996	2.073	0.826	15.089
7	Savitzky-Golay D^1	17	0.999	0.828	0.741	18.410
8	Savitzky-Golay D^2	16	0.998	1.582	0.559	24.004
9	OSC, Detrending	6	0.994	2.579	0.962	7.049
10	SNV, Median filter	17	0.997	1.681	0.831	14.847
11	SNV, Savitzky-Golay D^1	15	0.998	1.504	0.774	17.207
12	SNV, Savitzky-Golay D^2	16	0.997	1.628	0.566	23.823
13	EMSC, Median filter	17	0.998	1.437	0.888	12.106
14	EMSC, Savitzky-Golay D^1	15	0.998	1.649	0.767	17.440
15	EMSC, Savitzky-Golay D^2	16	0.997	1.650	0.565	23.849

chosen for cross validation, which is used when replicates of the physical samples are present. This method resulted in nine segments which corresponds to varying levels of adulteration and each segment includes the five replicates of one level of adulteration in the PLSR model.

Some methods, such as spectral variable selection, have been developed to increase the accuracy of PLSR models. In spectral variable selection, valuable spectral variables that affect the results the most are selected, and the rest of the spectra is discarded. Selecting spectral variables reduces the number of variables that need to be modeled and thus reduces the model's complexity while increasing its accuracy. The successive projections algorithm (SPA) is an algorithm that has been successfully used for this purpose (Xie et al., 2022).

However, to achieve more conclusive results and to better compare the effects of different preprocessing techniques when applied to large datasets, the spectral variable selection was not utilized in this study and the spectra were modeled without any spectra omission. Additionally, possible outliers were not removed from the dataset to further test the applicability of the different preprocessing methods. The models' performance was measured using the correlation coefficient (R^2) and root mean squared error (RMSE) (Xie et al., 2022) values for the calibration and validation sets. This study does not utilize feature or interval selection, and none of the spectral variables were removed (Guo et al., 2021).

Table 3
The results of PLSR for different preprocessing techniques (sesame oil adulterated with sunflower oil).

NO.	Preprocessing	Factor	R_c^2	RMSEC (%)	R_p^2	RMSEP (%)
1	None	13	0.978	4.743	0.816	15.502
2	OSC	10	0.978	4.776	0.884	12.307
3	SNV	16	0.997	1.716	0.880	12.544
4	EMSC	13	0.991	3.055	0.889	12.064
5	Detrending	12	0.983	4.211	0.866	13.221
6	Median filter	13	0.986	3.814	0.860	13.501
7	Savitzky-Golay D^1	16	0.998	1.586	0.514	25.223
8	Savitzky-Golay D^2	2	0.212	28.541	0.104	34.233
9	OSC, Detrending	10	0.985	3.875	0.910	10.844
10	SNV, Median filter	13	0.990	3.197	0.877	12.684
11	SNV, Savitzky-Golay D^1	13	0.990	3.199	0.773	17.234
12	SNV, Savitzky-Golay D^2	2	0.224	28.317	0.097	34.357
13	EMSC, Median filter	12	0.990	3.276	0.894	11.794
14	EMSC, Savitzky-Golay D^1	16	0.999	0.998	0.799	16.211
15	EMSC, Savitzky-Golay D^2	2	0.246	27.914	0.088	34.533

3. Results and discussion

The results align with the previous studies, and FTIR combined with PLS regression can discriminate mixed vegetable oil samples (Bunaciu et al., 2020). The overall appearance of the spectra of all four vegetable oils are quite similar and adulterated samples cannot be distinguished with the naked eye. The spectra of samples containing sesame and canola oil and their similarity can be seen in Fig. 2a. The spectra show variation in peak absorbance values and are most dissimilar in wavenumbers below 1500 cm^{-1}. The most significant peak is observed at around 2800 cm^{-1} and belongs to the stretching vibration of the aliphatic CH_2 groups (Ozulku et al., 2017). Information about the models developed by PLSR for sesame oil adulterated with canola oil, corn oil, and sunflower oil can be seen in Tables 1 to 3, respectively.

3.1. Sesame oil adulteration detection

Duo to the extensive range of adulteration (0–100%), the values of root mean squared error of calibration (RMSEC) and prediction (RMSEP) are substantial, and the coefficient of calibration (R_c^2) and prediction (R_p^2) are more descriptive of the accuracy of the models and their ability to predict the adulteration level. OSC magnifies the differences between the spectra (Fig. 2.c) and OSC with detrending (Fig. 3.a) has given the most accurate prediction results for sesame oil adulterated with canola oil with a value of $R_p^2 = 0.971$ and can predict the level of canola oil adulteration in sesame oil with an error of about 6% (Table 1). OSC and OSC in combination with detrending, yielded in quite similar results for sesame oil adulterated with corn oil with 0.963 and 0.963 for R_p^2, respectively (Table 2). In the case of sesame oil adulterated with sunflower oil, OSC combined with detrending featured the most accurate prediction result with $R_p^2 = 0.910$, EMSC with Median filter had the second-best results ($R_p^2 = 0.894$), subsequently, EMSC, OSC, and SNV all showed similar results (Table 3).

The goal of preprocessing is to improve the model performance. However, for each of the three adulterants, a number of preprocessing methods proved to be less accurate in predicting adulteration than the

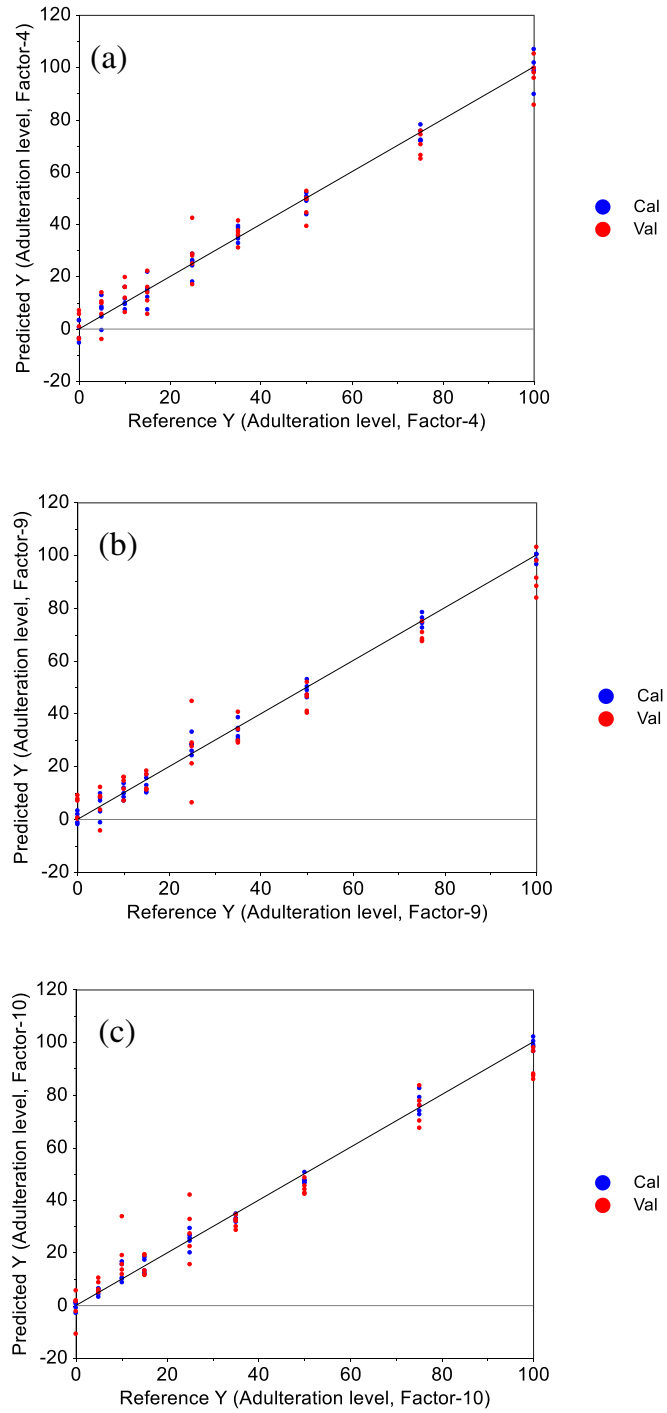

Fig. 3. Prediction results of a preprocessing combinations developed by PLSR of OSC with Detrending (a), EMSC with Median filter (b), and SNV with Savitzky-Golay D^1 (c).

non-preprocessed data. Detrending, SNV with Savitzky-Golay D^2, and EMSC, Savitzky-Golay D^2 under-performed compared to non-preprocessed spectra of adulterated sesame oil and canola oil (Table 1). About half of the preprocessing methods did not improve the RMSEP of the PLSR model in predicting the level of corn oil adulteration (Table 2). In samples adulterated with sunflower oil, all of the methods that included Savitzky-Golay (D^1 and D^2) derivatives lowered the R_p^2 value of PLSR models (Table 3).

The combination of preprocessing methods featured a heterogenous picture, from improving to worsening the prediction results. Combining EMSC with median filter (Fig. 2g and Fig. 3.b) and SNV with Savitzky-Golay D^1 (Fig. 2e and Fig. 3.c) reduced RMSEP of detecting canola oil by only one percent. However, derivatives and especially Savitzky-Golay $D^{2,}$ performed rather poorly and, combined with SNV and EMSC, significantly reduced the accuracy of the primary preprocessing method in detecting corn oil and sunflower oil in sesame oil. Additionally, combining different preprocessing methods did not always affect the robustness of the model and was sometimes redundant as is the case in combining Median filter with SNV. The other preprocessing methods, though not delivering the best results, had an acceptable performance with maximum RMSEP value of around 15%. Preprocessing steps of spectra should be applied with great care, or valuable information might be removed from the spectra and classification and the robustness of modeling could be impaired.

3.1.1. *Evaluation of the PLSR model*

The most valuable result is the prediction result because ultimately, a chemometricians goal is a robust model that can accurately predict the results. Having a lower number of factors cannot determine the efficiency of a PLSR model alone, and all of the model parameters must be evaluated as a whole. For instance, a lower number of optimal factors in comparison to the non-preprocessed data does not necessarily translate to a better model. An example of this can be seen in Table 1, where even though the optimal number of factors for detrending is smaller than that of the non-preprocessed data, the latter features better performance in both calibration and prediction. Similarly, a high R_c^2 accompanied by a low R_p^2 is a sign of a model that is over-fitted to the calibration data. Overall, because the calibration model is fitted to calibration spectra, R_c^2 remained above 0.98 and 0.99 in most cases.

3.1.2. *Comparison of the preprocessing methods*

Overall, OSC with detrending and OSC alone showed promising and the most accurate PLSR models in predicting the adulteration level of sesame oil with a prediction range of 6–12% for different adulterants. Savitzky-Golay D^1 and Savitzky-Golay D^2 (both alone and combined with SNV and EMSC) did not show the same level of robustness as the other preprocessing methods and, in most cases, even performed worse than non-preprocessed data. It should be noted that the results of this study are limited to the vegetable oil samples and the preprocessing methods that were chosen. Regardless of preprocessing method, extending the number of samples and the size of calibration data can improve the prediction results (Schoot et al., 2020). It is essential to preprocess spectra with care and to avoid relying solely on previous studies. Moreover, to avoid negatively affecting the accuracy of modeling and classification, it is recommended to forego preprocessing and work with non-preprocessed spectra in the absence of methodological preprocessing.

4. Conclusion

The methodological approach to preprocessing of the FTIR spectra of sesame oil adulterated with three vegetable oils took into account and visualized both the favorable or unfavorable effect of some preprocessing methods on the spectra and classification accuracy. It can be concluded that preprocessing affects the calibration and prediction accuracy of the PLSR model. Out of all the preprocessing methods investigated in this study, OSC showed the best results for all of the three adulterants (canola oil, corn oil, and sunflower oil), with R^2 of over 0.97 for calibration and prediction of sesame oil adulterated with canola oil. The results show that while some combinations of preprocessing methods increase the accuracy of the PLSR model, other combinations hinder the model and perform poorly compared to the non-processed spectra. Choosing a suitable preprocessing method is therefore crucial to developing a robust model. This study provides an insight into the effects of preprocessing FTIR spectra of vegetable oils and introduces a new concept for researchers dealing with spectra to develop standard

preprocessing procedures.

CRediT authorship contribution statement

Rasool Khodabakhshian: Conceptualization, Methodology, Validation, Supervision, Writing – review & editing. **Hajarsadat Seyedalibeyk Lavasani:** Conceptualization, Formal analysis, Writing – original draft, Visualization. **Philipp Weller:** Conceptualization, Methodology, Writing – review & editing.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Rasool Khodabakhshian reports financial support was provided by Parand Oil.

Data availability

Data will be made available on request.

Acknowledgements

This research was funded by Ferdowsi University of Mashhad (grant number 2/54630) and Parand Oil, Mashhad, Iran (grant number 3270). The funding sources had no involvement that could have appeared to influence the work reported in this paper.

References

Aboulwafa, M. M., Youssef, F. S., Gad, H. A., Sarker, S. D., Nahar, L., Al-Azizi, M. M., & Ashour, M. L. (2019). Authentication and discrimination of green tea samples using UV–vis, FTIR and HPLC techniques coupled with chemometrics analysis. *Journal of Pharmaceutical and Biomedical Analysis, 164*, 653–658. https://doi.org/10.1016/j.jpba.2018.11.036

Biney, J. K. M., Blöcher, J. R., Borůvka, L., & Vašát, R. (2021). Does the limited use of orthogonal signal correction pre-treatment approach to improve the prediction accuracy of soil organic carbon need attention? *Geoderma, 388*, Article 114945. https://doi.org/10.1016/j.geoderma.2021.114945

Biney, J. K. M., Borůvka, L., Chapman Agyeman, P., Němeček, K., & Klement, A. (2020). Comparison of field and laboratory wet soil spectra in the Vis-NIR range for soil organic carbon prediction in the absence of laboratory dry measurements. *Remote Sensing, 12*(18), 3082. https://doi.org/10.3390/rs12183082

Brendel, R., Schwolow, S., Rohn, S., & Weller, P. (2020). Comparison of PLSR, MCR-ALS and Kernel-PLSR for the quantification of allergenic fragrance compounds in complex cosmetic products based on nonlinear 2D GC-IMS data. *Chemometrics and Intelligent Laboratory Systems, 205*, Article 104128. https://doi.org/10.1016/j.chemolab.2020.104128

Bunaciu, A. A., Aboul-Enein, H. Y., & Hoang, V. D. (2020). Edible oil analysis. In *Vibrational spectroscopy applications in biomedical, pharmaceutical and food sciences* (pp. 167–187). Elsevier. https://doi.org/10.1016/B978-0-12-818827-9.00009-3.

Dogruer, I., Uyar, H. H., Uncu, O., & Ozen, B. (2021). Prediction of chemical parameters and authentication of various cold pressed oils with fluorescence and mid-infrared spectroscopic methods. *Food Chemistry, 345*, Article 128815. https://doi.org/10.1016/j.foodchem.2020.128815

Fadzlillah, N. A., Che Man, Y. B., & Rohman, A. (2014). FTIR Spectroscopy Combined with Chemometric for Analysis of Sesame Oil Adulterated with Corn Oil. *International Journal of Food Properties, 17*(6), 1275–1282. https://doi.org/10.1080/10942912.2012.689409

Fath, V., Lau, P., Greve, C., Weller, P., Kockmann, N., & Röder, T. (2021). Simultaneous self-optimisation of yield and purity through successive combination of inline FT-IR spectroscopy and online mass spectrometry in flow reactions. *Journal of Flow Chemistry, 11*(3), 285–302. https://doi.org/10.1007/s41981-021-00140-x

Gerhardt, N., Schwolow, S., Rohn, S., Pérez-Cacho, P. R., Galán-Soldevilla, H., Arce, L., & Weller, P. (2019). Quality assessment of olive oils based on temperature-ramped HS-GC-IMS and sensory evaluation: Comparison of different processing approaches by LDA, kNN, and SVM. *Food Chemistry, 278*, 720–728. https://doi.org/10.1016/j.foodchem.2018.11.095

Guo, Z., Barimah, A. O., Yin, L., Chen, Q., Shi, J., El-Seedi, H. R., & Zou, X. (2021). Intelligent evaluation of taste constituents and polyphenols-to-amino acids ratio in matcha tea powder using near infrared spectroscopy. *Food Chemistry, 353*, Article 129372. https://doi.org/10.1016/j.foodchem.2021.129372

Hofmann, J., Masuch, R., Wolf, A., Steinkämper, A., & Biener, R. (2017). Acetatmessung mit MIR-Transmissionsspektroskopie bei der E. *coli-Kultur. BIOspektrum, 23*(3), 273–275. https://doi.org/10.1007/s12268-017-0795-3

Horn, B., Esslinger, S., Pfister, M., Fauhl-Hassek, C., & Riedl, J. (2018). Non-targeted detection of paprika adulteration using mid-infrared spectroscopy and one-class classification – Is it data preprocessing that makes the performance? *Food Chemistry, 257*, 112–119. https://doi.org/10.1016/j.foodchem.2018.03.007

Jamwal, R., Amit, Kumari, S., Balan, B., Kelly, S., Cannavan, A., & Singh, D. K. (2021). Rapid and non-destructive approach for the detection of fried mustard oil adulteration in pure mustard oil via ATR-FTIR spectroscopy-chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 244*. https://doi.org/10.1016/j.saa.2020.118822, 118822.

Khodabakhshian, R., Bayati, M. R., & Emadi, B. (2021). An evaluation of IR spectroscopy for authentication of adulterated turmeric powder using pattern recognition. *Food Chemistry, 364*, Article 130406. https://doi.org/10.1016/j.foodchem.2021.130406

Khodabakhshian, R., Bayati, M. R., & Emadi, B. (2022). Adulteration detection of Sudan Red and metanil yellow in turmeric powder by NIR spectroscopy and chemometrics: The role of preprocessing methods in analysis. *Vibrational Spectroscopy, 120*, Article 103372. https://doi.org/10.1016/j.vibspec.2022.103372

Khodabakhshian, R., Emadi, B., Khojastehpour, M., Golzarian, M. R., & Sazgarnia, A. (2017). Non-destructive evaluation of maturity and quality parameters of pomegranate fruit by visible/near infrared spectroscopy. *International Journal of Food Properties, 20*(1), 41–52. https://doi.org/10.1080/10942912.2015.1126725

Kumar, K. (2021). Competitive adaptive reweighted sampling assisted partial least square analysis of excitation-emission matrix fluorescence spectroscopic data sets of certain polycyclic aromatic hydrocarbons. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 244*, Article 118874. https://doi.org/10.1016/j.saa.2020.118874

Li, S., Xing, B., Lin, D., Yi, H., & Shao, Q. (2020). Rapid detection of saffron (Crocus sativus L.) Adulterated with lotus stamens and corn stigmas by near-infrared spectroscopy and chemometrics. *Industrial Crops and Products, 152*, Article 112539. https://doi.org/10.1016/j.indcrop.2020.112539

Manley, M. (2014). Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. *Chemical Society Reviews, 43*(24), 8200–8214. https://doi.org/10.1039/C4CS00062E

Mishra, P., Rutledge, D. N., Roger, J.-M., Wali, K., & Khan, H. A. (2021). Chemometric pre-processing can negatively affect the performance of near-infrared spectroscopy models for fruit quality prediction. *Talanta, 229*, Article 122303. https://doi.org/10.1016/j.talanta.2021.122303

Moghimi, A., Aghkhani, M. H., Sazgarnia, A., & Sarmad, M. (2010). Vis/NIR spectroscopy and chemometrics for the prediction of soluble solids content and acidity (pH) of kiwifruit. *Biosystems Engineering, 106*(3), 295–302. https://doi.org/10.1016/j.biosystemseng.2010.04.002

Mousa, M. A. A., Wang, Y., Antora, S. A., Al-qurashi, A. D., Ibrahim, O. H. M., He, H.-J., … Kamruzzaman, M. (2022). An overview of recent advances and applications of FT-IR spectroscopy for quality, authenticity, and adulteration detection in edible oils. *Critical Reviews in Food Science and Nutrition, 62*(29), 8009–8027. https://doi.org/10.1080/10408398.2021.1922872

Munnaf, M. A., & Mouazen, A. M. (2022). Removal of external influences from on-line vis-NIR spectra for predicting soil organic carbon using machine learning. *Catena, 211*, Article 106015. https://doi.org/10.1016/j.catena.2022.106015

Nørgaard, M., Ganz, M., Svarer, C., Frokjaer, V. G., Greve, D. N., Strother, S. C., & Knudsen, G. M. (2020). Different preprocessing strategies lead to different conclusions: A [11 C]DASB-PET reproducibility study. *Journal of Cerebral Blood Flow & Metabolism, 40*(9), 1902–1911. https://doi.org/10.1177/0271678X19880450

Obeidat, S. M., Hammoudeh, A. Y., & Alomary, A. A. (2018). Application of FTIR spectroscopy for assessment of green coffee beans according to their origin. *Journal of Applied Spectroscopy, 84*(6), 1051–1055. https://doi.org/10.1007/s10812-018-0585-9

Ozulku, G., Yildirim, R. M., Toker, O. S., Karasu, S., & Durak, M. Z. (2017). Rapid detection of adulteration of cold pressed sesame oil adultered with hazelnut, canola, and sunflower oils using ATR-FTIR spectroscopy combined with chemometric. *Food Control, 82*, 212–216. https://doi.org/10.1016/j.foodcont.2017.06.034

Parsaei-Khomami, A., Badiei, A., Ghavami, Z. S., & Ghasemi, J. B. (2022). A new fluorescence probe for simultaneous determination of Fe2+ and Fe3+ by orthogonal signal correction-principal component regression. *Journal of Molecular Structure, 1252*, Article 131978. https://doi.org/10.1016/j.molstruc.2021.131978

Rinnan, Å., van den Berg, F., & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry, 28*(10), 1201–1222. https://doi.org/10.1016/j.trac.2009.07.007

Santos, D. I., Neiva Correia, M. J., Mateus, M. M., Saraiva, J. A., Vicente, A. A., & Moldão, M. (2019). Fourier Transform Infrared (FT-IR) Spectroscopy as a Possible Rapid Tool to Evaluate Abiotic Stress Effects on Pineapple By-Products. *Applied Sciences, 9*(19), 4141. https://doi.org/10.3390/app9194141

Schoot, M., Kapper, C., van Kollenburg, G. H., Postma, G. J., van Kessel, G., Buydens, L. M. C., & Jansen, J. J. (2020). Investigating the need for preprocessing of near-infrared spectroscopic data as a function of sample size. *Chemometrics and Intelligent Laboratory Systems, 204*, Article 104105. https://doi.org/10.1016/j.chemolab.2020.104105

Shi, H., Lei, Y., Louzada Prates, L., & Yu, P. (2019). Evaluation of near-infrared (NIR) and Fourier transform mid-infrared (ATR-FT/MIR) spectroscopy techniques combined with chemometrics for the determination of crude protein and intestinal protein digestibility of wheat. *Food Chemistry, 272*, 507–513. https://doi.org/10.1016/j.foodchem.2018.08.075

Upadhyay, R., Gupta, A., Mishra, H. N., & Bhat, S. N. (2022). At-line quality assurance of deep-fried instant noodles using pilot scale visible-NIR spectroscopy combined with deep-learning algorithms. *Food Control, 133*, Article 108580. https://doi.org/10.1016/j.foodcont.2021.108580

Wang, Q., Wu, G., Pian, F., Shan, P., Li, Z., & Ma, Z. (2021). Simultaneous detection of glucose, triglycerides, and total cholesterol in whole blood by Fourier-Transform

Raman spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 260*, Article 119906. https://doi.org/10.1016/j.saa.2021.119906

Xie, L., Zhu, J., Wang, Y., Wang, N., Liu, F., Chen, Z., … Shen, X. (2022). Rapid and accurate determination of prohibited components in pesticides based on near infrared spectroscopy. *Infrared Physics & Technology, 121*, Article 104038. https://doi.org/10.1016/j.infrared.2022.104038

Yang, X., Li, G., Song, J., Gao, M., & Zhou, S. (2018). Rapid discrimination of Notoginseng powder adulteration of different grades using FT-MIR spectroscopy combined with chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 205*, 457–464. https://doi.org/10.1016/j.saa.2018.07.056

Yang, Y., Hettinga, K. A., Erasmus, S. W., Pustjens, A. M., & van Ruth, S. M. (2020). Opportunities for fraudsters: When would profitable milk adulterations go unnoticed by common, standardized FTIR measurements? *Food Research International, 136*, Article 109543. https://doi.org/10.1016/j.foodres.2020.109543

Yao, L., Lyu, N., Chen, J., Pan, T., & Yu, J. (2016). Joint analyses model for total cholesterol and triglyceride in human serum with near-infrared spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 159*, 53–59. https://doi.org/10.1016/j.saa.2016.01.022