Trends in Analytical Chemistry 180 (2024) 117957

Contents lists available at ScienceDirect

Trends in Analytical Chemistry

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

Data fusion strategies for the integration of diverse non-destructive spectral sensors (NDSS) in food analysis

Lorenzo Strani [a], Caterina Durante [a], Marina Cocchi [a,*], Federico Marini [b], Ingrid Måge [c], Alessandra Biancolillo [d]

[a] Department of Chemical and Geological Sciences, University of Modena and Reggio Emilia, Via 4 Campi 103, 41125 Modena, Italy
[b] Department of Chemistry, Sapienza University of Rome, P.le Aldo Moro 5, Rome 00185, Italy
[c] Nofima AS - Norwegian Institute of Food, Fisheries and Aquaculture Research, Tromsø, Norway
[d] Department of Physical and Chemical Sciences, University of L'Aquila, Via Vetoio 67100, Coppito, L'Aquila, Italy

ARTICLE INFO

Keywords:
Spectroscopic sensors
Image fusion
Data fusion
Food analysis
Multiblock methods
In-field
On-line

ABSTRACT

The evolving landscape of agri-food systems, driven by climate change and population growth, necessitates innovative approaches to ensure food integrity, safety, and sustainability. This review explores the role of data fusion strategies, particularly focusing on non-destructive spectroscopic sensors (NDSS) in three key application contexts: in-field monitoring, on/in-line food processing, and food quality authentication. Various data fusion scenarios, including fusing spectra from different spectroscopic platforms, integrating images and spectra, and combining non-spectroscopic sensor data with spectroscopic ones are reviewed. Focus is set on practical considerations, such as selecting the level of data fusion, defining blocks, variable selection, and validation methods, highlighting the importance of tailored approaches based on research aims and data characteristics.

While combining information from diverse sensors generally enhances information extraction and modelling performance, their implementation in routine applications is still limited and especially studies focused on data fusion models' performance over time and their maintenance are lacking.

Abbreviations	
ATR-FTIR	Attenuated Total Reflection - Fourier-Transform Infrared spectroscopy
CV	Cross-Validation
DF	Data Fusion
EEM	Excitation-Emission Matrix
FAAS	Flame Atomic Absorption Spectrometry
FT-NIR	Fourier-Transform Near Infrared spectroscopy
GLCM	Grey Level Co-Occurrence Matrix
GS	Gram-Schmidt
IHS	Intensity Hue Saturation transform
^{1}H NMR	Proton Nuclear Magnetic Resonance
HS	Low-Spatial High-Spectral resolution image
HSI	Hyperspectral Imaging
HSS	High-Spatial -Spectral resolution image
LIDAR	Light Detection And Ranging
MB-PLS	Multi-block - Partial Least Squares
MCR	Multivariate Curve Resolution
MIR	Mid-Infrared
MSI	Multispectral Images
NDSS	Non-Destructive Spectral Sensors
NIR	Near Infrared
P	Panchromatic image
PARAFAC	Parallel Factor analysis
PAT	Process Analytical Technologies
PCA	Principal Component Analysis
PLS	Partial Least Squares regression
PLS-DA	Partial Least Squares - Discriminant Analysis
PDO	Protected Designation of Origin
QbD	Quality by Design
ROSA	Response-Oriented Sequential Alternation
ROI	Region Of Interest
SMB-PLS	Sequential Multi-block - Partial Least Squares
SO-CovSel	Sequential and Orthogonalized - Covariance Selection
SO-PLS	Sequential and Orthogonalized - Partial Least Squares
SO-PLS-LDA	Sequential and Orthogonalized - Partial Least Squares - Linear Discriminant Analysis
SVM	Support Vector Machine
UAV	Unmanned Aerial Vehicle
VIP	Variable Importance in Projection
Vis/NIRS	Visible/Near Infrared Spectroscopy

* Corresponding author.
E-mail address: marina.cocchi@unimore.it (M. Cocchi).

https://doi.org/10.1016/j.trac.2024.117957
Received 6 May 2024; Received in revised form 6 August 2024; Accepted 2 September 2024
Available online 4 September 2024

1. Introduction

The research and application scenarios in the agri-food context have changed significantly in recent years due to the new challenges posed by climate change and population growth. These put increased pressure on the agri-food chain, requiring adjustments to reduce waste, adapt to climate change and increase resilience, while ensuring food integrity, safety, and health [1]. Concepts such as digital agriculture, digital food and big data are now being used to drive the evolution that food systems need to meet these challenges [2], encompassing both crop production and industrial food processing.

Nevertheless, the implementation of new technologies and paradigms, such as QbD/PAT tools in the food industry and precision agriculture, as well as in-field monitoring tools in agronomy, is progressing slowly. On the other hand, the technologies are quite mature; in particular, non-destructive spectroscopic sensors (NDSS), combined with data analysis, offer cost-effective, high-value solutions to challenges such as integrity verification [3] and better understanding of products and ingredients. Furthermore, their in-line implementation [4–6], combined with data provided by process sensors installed to monitor the plant, can provide a real-time solution to ensure complete process control throughout the food chain.

The impasse may be due to critical issues such as the high complexity and variability of agri-food matrices, the need for long-term instrument stability, robustness of calibrations, integration of sensors into production environments, and the creation of real-time decision systems. Therefore, appropriate strategies for data handling, analysis, and integration (fusion) are a key aspect for further progress.

Near Infrared (NIR), Fluorescence, Raman, thermal and/or time-resolved spectroscopy are among the NDSS technologies that are becoming increasingly available, in smaller and more affordable instrumentation [7]. Most of these techniques can be combined together and, with multispectral imaging, allow extremely powerful sampling of the entire surface of a product/product stream.

Recent reviews [8–11] have comprehensively revised data fusion approaches and methodologies from a data structure and methodological point of view.

A comprehensive survey reporting application of data fusion strategies for food quality authentication has been recently published [9], it highlights that in about 80 % improved models could be obtained by fusing information from diverse sensors. However, most of the surveyed studies were conducted at a laboratory scale and concerned a limited number of samples, also the performance comparison was limited to accuracy (classification tasks) and coefficient of determination (multivariate calibration tasks) often solely in cross-validation.

In this work, in reviewing data analysis strategies that enable data fusion of different NDSS, also integrated with imaging and other sensors (focusing on recent advances), we will adopt a practitioner's perspective with the aim of discussing why and when it might be beneficial to fuse more spectroscopic sensors, and what the specific challenges are in different contexts. In particular, three different implementation scenarios will be considered: in-the-field, in-line monitoring and adulteration detection. Furthermore, we will discuss the aspects of model implementation and validation when dealing with fused data, also from the perspective of model maintenance.

2. Applicative contexts

2.1. *agrifood: in-field monitoring*

The integration of diverse NDSS on in-field monitoring applications plays a crucial role in ensuring the quality and healthiness of agricultural products. In-field monitoring refers to the continuous assessment of agricultural parameters and the quality of crops directly at the production site. By utilizing various sensors mounted on satellites, drones, or aircraft, remote sensing technologies offer a non-invasive and efficient means of collecting data on crop conditions, including factors such as vegetation vigor, stress levels, and disease outbreaks. A comprehensive picture of crop health can be obtained by fusing data from multiple spectroscopic sensors, ranging from nutrient levels and water stress to disease detection and yield prediction.

In agricultural monitoring, researchers frequently utilize spectral data from HSI to differentiate between different materials by analyzing their reflectance values [12]. The integration of hyperspectral data with light detection and ranging (LIDAR) shape profiling on an Unmanned Aerial Vehicle (UAV) platform is employed to assess the photosynthetic processes in forest vegetation [13]. Through data fusion, the 3D LIDAR point cloud data and hyperspectral reflectance data are merged to forecast the canopy-level biochemical traits that exhibit strong correlations. Data fusion can also benefit plant disease detection, as shown by Mahlein et al. [14]. In this work, an examination of thermal, fluorescence, and HSI supports the adoption of a multi-sensor data fusion strategy for monitoring plant health. A specific study on wheat head blight delineated the significant advantages and limitations of each system, further exploring the particular sensor combinations. These applications show how reliable HSI is when applied to heterogeneous samples/matrices.

Sagan et al. [15] fused multispectral, RGB and thermal images acquired by both satellite and UAV for early stress detection, crucial for proactive field management and predicting terminal yield accurately. The results indicate that early stress can be efficiently identified through the utilization of multi-temporal and multi-scale observations from UAVs and satellites. Aerial systems can also be used to phenotype crops in the field. Bartlett et al. [16] used hyperspectral visible and near-infrared fused together for the identification of phenotypes like biomass and chlorophyll content. Multispectral satellite images have also been fused together to evaluate the impact of the growth stage on the amalgamation of spectral bands associated with wheat grain nitrogen uptake [17].

2.2. *food processing: on/in-line monitoring*

On/In-line NDSS have the potential to measure numerous quality attributes of food products, directly in the processing lines without any manual sampling [7,18]. Such measurements are important to improve quality control, safety, and efficiency in the food industry. Typical use cases are monitoring and control of key quality parameters, detection of contamination and defects, process optimization, and process control and simulation through digital twins.

In industrial quality monitoring applications, fusion of multiple NDSS may be needed to obtain a high precision of the predictions or to be able to measure a suite of different quality attributes. Several examples of the fusion of NDSS for food quality assessment may be found in Refs. [19–22]. Sometimes, the monitoring is based on soft sensors, i.e. models that predict hard-to-measure product quality characteristics based on easily available inline sensors. In such cases, it has been shown that a combination of spectroscopic and process sensors often may be needed to obtain good models [6,23].

When NDSS are integrated into closed-loop control systems or digital twins, they need to be combined with other important process variables. These types of systems have the potential to significantly improve efficiency and resource utilization, but they are still in their infancy and industrial implementations remain to be demonstrated [19,24]. There is currently a huge amount of research on this topic, and several sources point out that sensors for measuring physiochemical food quality attributes are key elements in automatic control systems and digital twins [19,24–27].

2.3. food quality and authentication

Combining different types of non-destructive sensors allows for capturing different aspects of food properties such as chemical composition, aroma, taste, texture, and visual appearance, achieving a more accurate understanding of food properties, and enhancing the effectiveness of authentication protocols.

By employing multiple sensors, it becomes possible to analyze several aspects of food authenticity, including geographical origin, production methods, and the presence of contaminants/adulterants.

NDSS facilitate the rapid screening of large volumes of food products, eliminating the need for sample preparation. This streamlined approach maintains the integrity of the products while expediting decision-making in quality control processes. Consequently, NDSS play a pivotal role in ensuring the efficiency and effectiveness of food quality assessment and authentication across different sectors of the food industry [28]. In this regard, significant attention has been devoted to the NDSS analysis of meat, vegetables, or fats. Sanchez and collaborators have reported that hyperspectral Imaging (HSI) and Visible/Near Infrared Spectroscopy (Vis/NIRS) are considered the cornerstone techniques for the analysis of both beef and pork [29]. On the other hand, Silva et al. have highlighted that for ovine meat, optical technologies are gaining importance for monitoring and evaluating the quality and safety of carcasses and meat [30]. Among these technologies, visible and infrared reflectance spectroscopy, hyperspectral imaging, and Raman spectroscopy deserve particular attention. These analytical methods have been also successfully used for adulteration detection, as discussed by Alamprese et al., who achieved the detection of minced beef adulteration with turkey meat by UV–vis, NIR and MIR spectroscopy [31]. A further relevant application in this context is the one proposed by Pu and collaborators, where visible and near-infrared hyperspectral imaging and textural analysis are used to classify fresh and frozen-thawed pork muscles [32]. Concerning the analysis of vegetables, NDSSs are often used to assess firmness and ripeness. This practice has been exemplified by Orlandi et al., who devised a methodology for evaluating grape ripeness by electronic nose (E-nose) and tongue (E-tongue) integrating the original signals by means of DF. Eventually, latent variables extracted through iterative PLS were combined in a Mid-level DF which exhibited superior performance compared to Low-level DF and traditional methods reliant on individual sensors [33]. A further example has been provided by Mendoza and colleagues who explored HSI to assess apple firmness by PLS achieving notable results [34].

For the analysis of edible fats, like virgin olive oils, Weesepoel and co-authors have highlighted that a good strategy could be the combined use of visible, fluorescence, and near-infrared spectroscopy, along with data fusion methods that allow for the concerted processing of all information [35]. Low- and Mid-level DF approaches were also applied on Headspace-Mass Spectrometry, FT-MIR and UV–vis instrumental responses by Borràs and co-authors to predict olive oil sensory descriptors [36].

It is worth mentioning that in the field of food quality and authentication, obtaining authentic samples with full traceability is essential for building reliable statistical models. Authentic samples serve as the benchmark for developing models that can accurately identify and verify food products. This is essential for addressing issues such as food fraud, contamination, and mislabeling.

3. Data structure and fusion scenarios

3.1. fusion levels

There are in general three different levels at which data can be fused [8,19,37], as schematically illustrated in Fig. 1. Here they will be only briefly recalled.

Low-level DF involves modelling the different blocks of data directly, allowing the results to be interpreted in terms of the original variables. This can be done using a variety of methods, which fall under the umbrella of multiblock methods [11], even if in the literature the case of simple concatenation (data augmentation in the shared mode, most often the samples mode) is somehow distinguished as the augmented data set is modeled as it were a single block, without posing attention on the analysis of shared or common information carried by the single blocks.

Mid-level DF implies a first modelling step, before fusion, aimed at extracting features from each data block using data decomposition techniques, such as PCA. It is worth mentioning that using variable

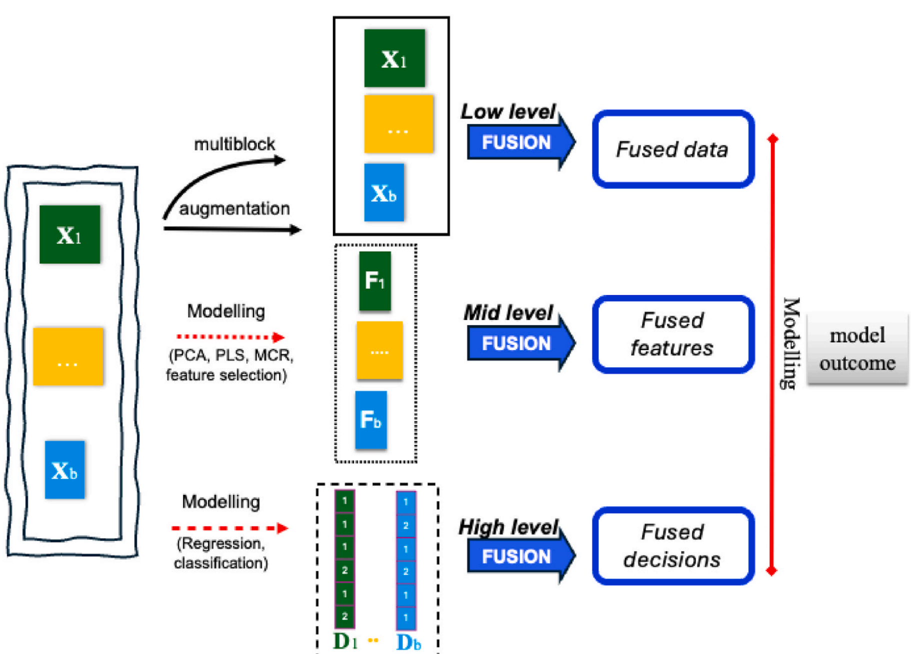

Fig. 1. Different levels of data fusion. The notation **X** indicates a single data block, **F** a data set holding features extracted from a single data block, **D** a vector holding the decision obtained by modelling a single data block.

selection methods (to extract features) can be considered either a low-level or mid-level process, as there is ongoing debate in the literature [19] regarding the appropriate classification of these techniques. The features are then fused into a new dataset, which is finally modeled to produce the desired outcomes. Model interpretation in terms of original variables can be done when variables selection is used in the first step, whereas it requires to interpret the contribution/role of the original variables in the model used to obtain the features, such as inspecting the loadings/weights (PCA/PLS) or pure/independent spectra profile (MCR/ICA).

High-level DF involves fusion of decisions obtained from separate modelling of each data block, typically by supervised models performing regression or classification tasks. Fusing the decisions is expected to produce more accurate predictions. Thus, the focus is on final outcomes, with almost no investigation into the role of each data block's variables because a fused model in strict sense is not obtained but only a fused decision.

3.2. fusing spectra from different spectroscopic platforms

Fusing profiles from different spectroscopic platforms involves dealing with two or more blocks of data with a high number of correlated variables so it is not surprising that the large majority of the applications involve mid-level data fusion strategy as the best compromise between increasing the information content, avoiding dealing with too much irrelevant sources of variability and keeping the models still reasonably interpretable. When looking at the way mid-level fusion is implemented, the large majority of the research reported in the literature favours the use of latent variables extracted from the different blocks as features to be combined (concatenated) to build the matrix to be subjected to the modelling of choice, principal component scores being the latent variables most commonly used. For instance, by applying SVM on the concatenated PC scores from ATR-FTIR and FT-NIR spectra, Zheng et al. [38] were able to authenticate the geographical origin of Gastrodia elata from Zhaotong with an accuracy higher than 97 % on test data. On the other hand, the possibility of concatenating the scores of MCR [39] or PLS/PLS-DA [40] on individual matrices for achieving mid-level fusion has also proved highly effective. Mid-level fusion approaches can also help overcome the limitations of low-level strategies when dealing with the fusion of two- and three- or in general multi-way spectral data: indeed, scores can be extracted from the different blocks by means of PCA and, e.g., PARAFAC or Tucker3, and then concatenated. This strategy was adopted by Ríos-Reina et al. who performed mid-level fusion to classify PDO wine vinegars [41]. More specifically PCA scores extracted from MIR and NIR blocks, PARAFAC scores from EEM fluorescence landscapes and the peak areas of MCR components from ^{1}H NMR were concatenated and subjected to PLS-DA allowing the correct classification of all the samples in the test set.

One of the main limitations of mid-level approaches is that the models are not immediately interpretable in terms of the original variables and block contributions, but these limitations are mostly overcome by approaches like SO-PLS, SO-CovSel and their discriminant counterparts, where the sequential and orthogonal nature of feature extraction allows to retain model coefficients which are directly relatable to the original variables and to identify the incremental contribution of the different blocks. The latter property can also translate to the possibility of identifying cases in which the inclusion of additional blocks doesn't result in any improvement in the predictive ability of the models. This was for instance the case of a study by Rocha Baqueta et al. [42] where molecular and atomic spectroscopic techniques (^{1}H NMR, portable NIR, benchtop NIR, ATR-FTIR-MIR, and FAAS) were used to characterize and discriminate Brazilian Canephora coffees with geographical indication, and also to differentiate them from Arabica. The use of SO-PLS-LDA suggested the best model to be the one built only on the benchtop NIR data, leading to 100 % correct classification on both training and test data.

3.3. fusing images and spectra

Imaging techniques are powerful tools in all three application contexts discussed in section 2 [43]. In general, there can be two different motivations for the use of imaging systems: one is the interest in recovering morphological or textural information, e.g. for assessing the compositional heterogeneity or defects of crops, fruits/vegetables [44–46] or food products during food processing; the other is to have a fast, non-destructive at/in-line or in-field instrumentation that can be used for monitoring over time [47,48] and may be easier to implement than a single-spot spectroscopic sensor, or may cover a larger area of the product to be monitored. RGB images are widely used as fast scanning in-line systems in food production [49], e.g. to detect defects in product sorting, and more recently multispectral and hyperspectral imaging systems are gaining increasing attention, offering the possibility of capturing the distribution of different food components, ingredients [50]. At the same time, there is interest in combining imaging data with other spectroscopic sensors that can be installed at other stages of the process line to monitor more homogeneous intermediates or the finished product. There is also an interest in integrating in-situ spectroscopic imaging data with environmental information, e.g. from humidity and temperature sensors or remote sensing devices.

Strategies are therefore needed for both the fusion of different image types and the fusion of images and spectra or other punctual variables.

Fusion of images may deal with fusing imaging of the same scene/sample acquired with different spatial/spectral resolutions, or with different spectroscopic modalities, which eventually may show different dimensions, or with the combination of spectroscopic (MSI or HSI) and colour information (RGB), as well as images referring to different samples (e.g. at different process stages).

Pan-sharpening [51–63] is a general term used when the objective is to retrieve a high-spatial and spectral image from a low-spectral/high-spatial resolution image and a low-spatial/high-spectral resolution image, and this can be achieved by different methodologies (as illustrated in Table 1) such as component substitution [54], regression [55], multi-resolution analysis [56,57], bayesian pan-sharpening [58], non-negative matrix factorization [59], multivariate curve resolution [60,61], deep learning [62,63], etc. Methodologies, which take directly into account the 3D structure of the hyperspectral image have been also used for this aim, e.g. multiple co-inertia analysis or multivariate inter-battery Tucker analysis [64,65] and tensor decomposition [66]. Multivariate curve resolution offers a general and flexible approach also to fuse images acquired with different modalities [61]. In these cases, the focus is on exploration, and generally a single sample/scene is analysed at time, at a second analysis stage some salient parameters may be obtained by post-processing of the retrieved information and then fed to further multivariate analysis.

In the case images have to be fused with spectra (or other variables), there is the need to reconduct the image data in a set of features per single/sample so that these can be handled as a data block in a data fusion perspective (Fig. 2), and several strategies may be employed to this aim, which can be distinguished based on the capability to retain solely spatial or spectral information or, explicitly or implicitly, both of them. The most diffuse approach, when dealing with colour images and multispectral images (MS), with a limited number of spectral channels, is to analyze the single-channel image and either extract bulk indices, e.g. statistics from the distribution, or textural features [10,67], e.g. using the grey level co-occurrence matrix (GLCM) approach and calculating Haralick features, thus obtaining for each sample a vector comprising all the features extracted for each spectral channel. In this way, the spatial information is explicitly retained, and spectral information can be indirectly retrieved by assessing variables' importance (e.g. which feature at which spectral channel) when analysing the data fused model. When dealing with MSI it is more common to apply the feature extraction step on the scores image (thus reducing the number of overall features) obtained by PCA applied on a data matrix holding textural

Table 1
Main employed methods for pan-sharpening.

Method	Principle	Preprocessing	Transform	Ref.
Component Substitution (CS)	HS^{up} *Transform*→ HS' replace a HS' band with P^{1}[a] $Transform^{-1}$→ HSS^{2}[c] Extend to all wavelengths	Register, spatially upsample HS[b] to P (HS^{up})	^{3}[d]IHS; PCA; GS; etc.	[54]
Regression	Divide HS and MS in small areas (patches) then learn regression from MS to HS for each patch. The HSS is finally restored by mapping MSI using the learnt relationship	Register, spatially downsample MSI to HS (MS) Pixelwise unfolding		[55]
Multi-Resolution Analysis (MRA)	Multiscale decomposition of P, the detail sub-image is injected in HS then inverse transform is applied	Register, spatially upsample HS to P	DWT; UDWT; “à-trous” WT.	[55–57]
Bayesian	Bayesian inference framework.			[57]
Coupled Non-negative Matrix Factorization (CNMF)	HSS=HU H holds pure spectral components (endmembers) and U the relative pixel abundances MS = H_{MS}U HS=HU_{HS} H and U can be retrieved by ALS	Pixelwise unfolding		[58]
Multivariate Curve Resolution	Spectral unmixing. HSS is recovered by incomplete multiset MCR analysis (augmentation of MSI column-wise with binned MSI, and sample-wise with HS)	Register, spatially downsample MSI to HS Pixelwise unfolding.		[59–61]
Deep Learning	Learn HSS from MSI by using a deep CNN architecture.			[61–63]

[a] P stands for panchromatic image, i.e. high-spatial, low-spectral resolution, either true panchromatic, only one band, or RGB/multispectral (MS N_xN_y x λ) in this case 1 to 3 bands are taken from P and a transform is applied.
[b] HS stands for hyperspectral, i.e. low-spatial, high-spectral resolution (HS n_xn_y x Λ, with $n << N$ and »λ).
[c] HSS stands for high-spatial, high-spectral resolution image.
[d] HIS: Intensity Hue Saturation transform; PCA: principal component analysis; GS: Gram-Schmidt.

information (e.g. by image unfolding and data augmentation with neighbouring pixels, or by multiresolution analysis) [68,69].

In the case of hyperspectral images (HSI), especially for a huge number of samples, it is common to consider a region of interest in the hyperspectral image, unfold it pixel wise and calculate the average spectrum, this way per each sample the HSI is compressed to a 1D spectrum, losing spatial information [70,71]. Spectral unmixing approaches, such as MCR [60,61], non-negative factor matrix decomposition [59], independent component analysis [72] and coupled tensor matrix decomposition methods [73–75], allow recovering the spatial signature (endmember) of each constituent and the relative concentration (abundance) map from which different features can be calculated, such as homogeneity indices [47,76], Haralick features, etc. Thus, obtaining a set of features for each constituent, this way direct information on constituent spatial distribution is retained and will be fed in the data fusion, and spectral profiles are available for putative identification of the constituent and interpretation. Depending on the unmixing method, the spatial structure can be considered during the decomposition (tensor decomposition) or not (bilinear decomposition).

Another approach applied to either RGB, MSI or HSI is to calculate and fit the frequency histogram for each single channel image thus obtaining for each sample a vector holding the histogram profile at each spectral wavelength, alternatively histogram can be calculated on scores after PCA compression [77]. In this way both spatial and spectral information are indirectly considered, e.g. pixels falling in bins found to be relevant (after fusion and multivariate analysis) may be repositioned on the image areas they belong to, and salient wavelengths can be depicted by interpreting the variables' contribution.

A different perspective is the one introduced by the deep learning framework when convolutional neural networks are employed [63,78]. In this case, the feature extraction process takes place during convolution and features can be extracted either directly from the HSI (3D convolution) or distinctly from spatial and spectral dimension and fused afterward, e.g. this approach was applied by Al-Sarayreh to detect meet adulteration [79]. However, it must be underlined that several open issues need deep evaluation such as the huge number of tuneable parameters that require lots of input data, hence significance of the data augmentation procedure for NDSS has to be carefully considered; significance of 3D convolution for HSI; model transparency and interpretability.

3.4. fusing non-spectroscopic sensors data with spectroscopic ones

This task typically involves fusing high-dimensional spectroscopic data with a limited number of additional variables. For example, hyperspectral images often need to be combined with location and weather sensors in agricultural applications, while in the food industry the NDSS are most often combined with readily available sensors for temperatures, pressures, torques and flow rates [80,81]. Based on a large number of pharmaceutical PAT applications, Casian et al. [82] have shown that mid-level data fusion is the preferred choice in such situations for monitoring, classification and regression tasks. Other examples from food and related process industries also show good results using mid-level data fusion [6,83,84]. The most straightforward mid-level strategy is to replace the spectra with predictions of one or more physiochemical attributes. The advantage of this approach is that the variables are easy to interpret. The disadvantage is that good calibration models are needed, which may be time-consuming and expensive to build and maintain. Also, there might be more information in the spectra than what is reflected in the predicted parameters.

An alternative is to replace the spectra with a set of factors, for instance derived from PCA or MCR. The advantage of this approach is that no calibration models are needed, and the approach is easy to implement in practice.

The last mid-level modelling alternative is to use multiblock methods that are specifically developed to handle blocks of different types of variables, such as SO-PLS, ROSA [11], and SMB-PLS [85]. All of them are dealing with the predictive context, and are characterised by sequentially extracting PLS components, which are constrained to be orthogonal, while they differ in the way they use information embedded in each block. SO-PLS focuses on the complementary information carried by each block; in ROSA there is no possibility to explicitly assess to which extent the blocks contribute to the final model with common or distinctive/complementary information; in SMB-PLS the information in subsequent blocks that is collinear with the first one is pooled in the

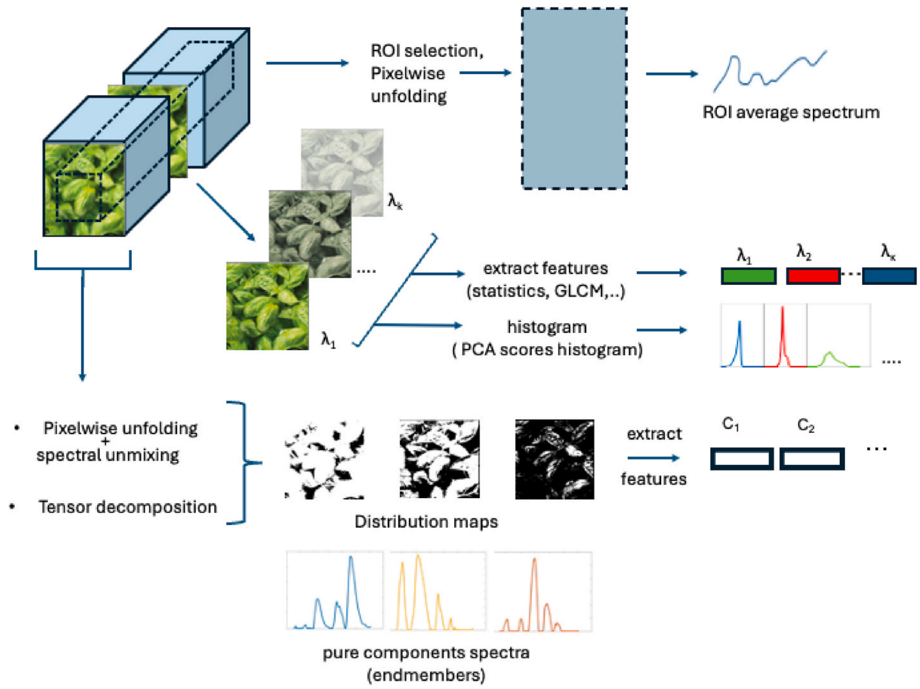

Fig. 2. Different paths to obtain a 1D vector (holding representative spectra or features per sample) from the HSI data cube. ROI = Region Of Interest.

same component and once y-related information from the first block is exhausted the same holds for the second and so on. Analysis of block explained variance by each component can highlight common and distinctive information contributed by each block. In SO-PLS and SMB-PLS the block order may have an influence on the model; this does not happen with ROSA in which, however, the identification of the first winning block may have an impact on the selection of the successive components. From an applicative point of view, the SO-PLS approach is particularly advantageous when the aim is to identify possible extra benefits from the inclusion of successive block(s) of data (and the information they carry) into the model; for instance, in the context of the present review, to understand whether it is really worth to add a spectroscopic sensor to non-spectroscopic ones, or to combine different types of spectroscopy. ROSA is computationally very fast, so it is useful when the number of sensors is huge. SMB-PLS can be suited for situations where it is important to assess common and distinctive contributions by each block, e.g., in industrial process monitoring to improve interpretability and feedback actions.

The advantage of these methods is that they extract latent variables that are directly related to the response, but it comes at the cost of more complicated algorithms that need a skilled data scientist to apply.

It is not recommended to do simple concatenation of variables in these situations as the results will depend heavily on the weighing of spectra relative to the other variables, and the models will be difficult to interpret.

4. Practical issues

4.1. preprocessing prior to data fusion

The preprocessing step is critical for any DF approach. Generally, for low-level DF, it goes over three stages [86]: the first, which we can refer to as signal preprocessing, aims to correct artefacts and uninteresting variations such as noise, multiplicative effects, scaling, baseline drift, peak shift, etc.; the second deals with intra-block offset and variability/scale correction aiming at ensuring equal contribution of the variables within a block; finally, the third is the inter-block preprocessing, such as block scaling, having the objective to correct differences in the scales, number of variables and the pseudo rank of different blocks. The latter step ensures that no single block of data contributes more than others simply because it contains a higher number of variables or because the way the signals are defined results in a higher total variance with respect to the other blocks.

In mid- and high-level DF, where features or decisions, respectively, are fused, preprocessing only involves the second and the third steps described above. In fact, possible signal preprocessing is performed in the modelling step, before the feature/decision extraction.

4.2. selecting the level of data fusion and the fusion method

Many papers have compared different data fusion methods [41, 87–90]. They usually focus on comparing predictive ability, and the results show that there is no universally best method. This indicates that the choice of method should depend on the properties of the data set, as well as the research aim and the intended use of the model [37].

For example, in food quality control, several analytical platforms can be used, and the most important question is what redundant and unique information each provides, so that it might be possible to select the platform(s) that are really needed to achieve the most effective monitoring set-up at the lowest cost. To this aim, low level DF methods are more appropriate [11].

Other considerations concern the type of data to be fused, for instance in NMR spectroscopy feature extraction by a spectral unmixing technique, such as MCR, could be very effective while NIR data may be more informative as such, thus suggesting a combination of low and mid-level DF as a choice.

Regarding the research aims and intended use of the models, when it comes to regression tasks it is important to distinguish between prediction and interpretation. The main advantage of some methods (e.g. SO-PLS and SMB-PLS) is that they give a better interpretation of the contributions from each block. If the model is going to be used for prediction only, more flexible machine learning methods may give more precise predictions [91], as well as a high-level DF framework. If the model will be used for routine predictions, it is necessary to monitor the model's performance over time and do maintenance and adjustment when needed. In these cases, it is important to choose a method that is robust and maintained. Also, with respect to the DF level, when more modelling steps are involved, as in mid-level but especially in high-level, routine implementation might become troublesome.

4.3. block definition and variables selection

The process of defining blocks and selecting variables holds paramount importance in data analysis; hence, in constructing a multi-block model, the sequence of blocks is pivotal. While including each spectrum type into distinct blocks can look natural, it prompts critical inquiries into the optimal strategy, and block definition can either enhance or impede the efficacy of subsequent analyses. Decisions regarding block order can be guided by different factors, including prior knowledge, research objectives, and the inherent structure of the dataset [92]. Approaches that rely on predictive performance to guide the selection of block sequences have been devised [93]. Although the block order will not sway the models' predictions, it significantly influences the interpretation of their findings. In certain scenarios, such as when dealing with blocks gathered at various stages of an online process or a mix of categorical and continuous variables, the order may be straightforward [94]; however, in other situations, is not as clear-cut. In particular cases, such as when employing sequential multi-block methods, it is worth emphasizing the advantageous aspect that these methods allow the recognition and interpretation of distinct and common components among blocks [84]. This underscores the importance of considering block order, as it directly impacts this ability. Moreover, beyond the delineation of blocks, the selective inclusion or exclusion of variables becomes imperative. This process necessitates a judicious balance between retaining informative features and mitigating noise, thereby refining the analytical framework for enhanced predictive performance and interpretability.

In the literature, various types of variable selection methods are

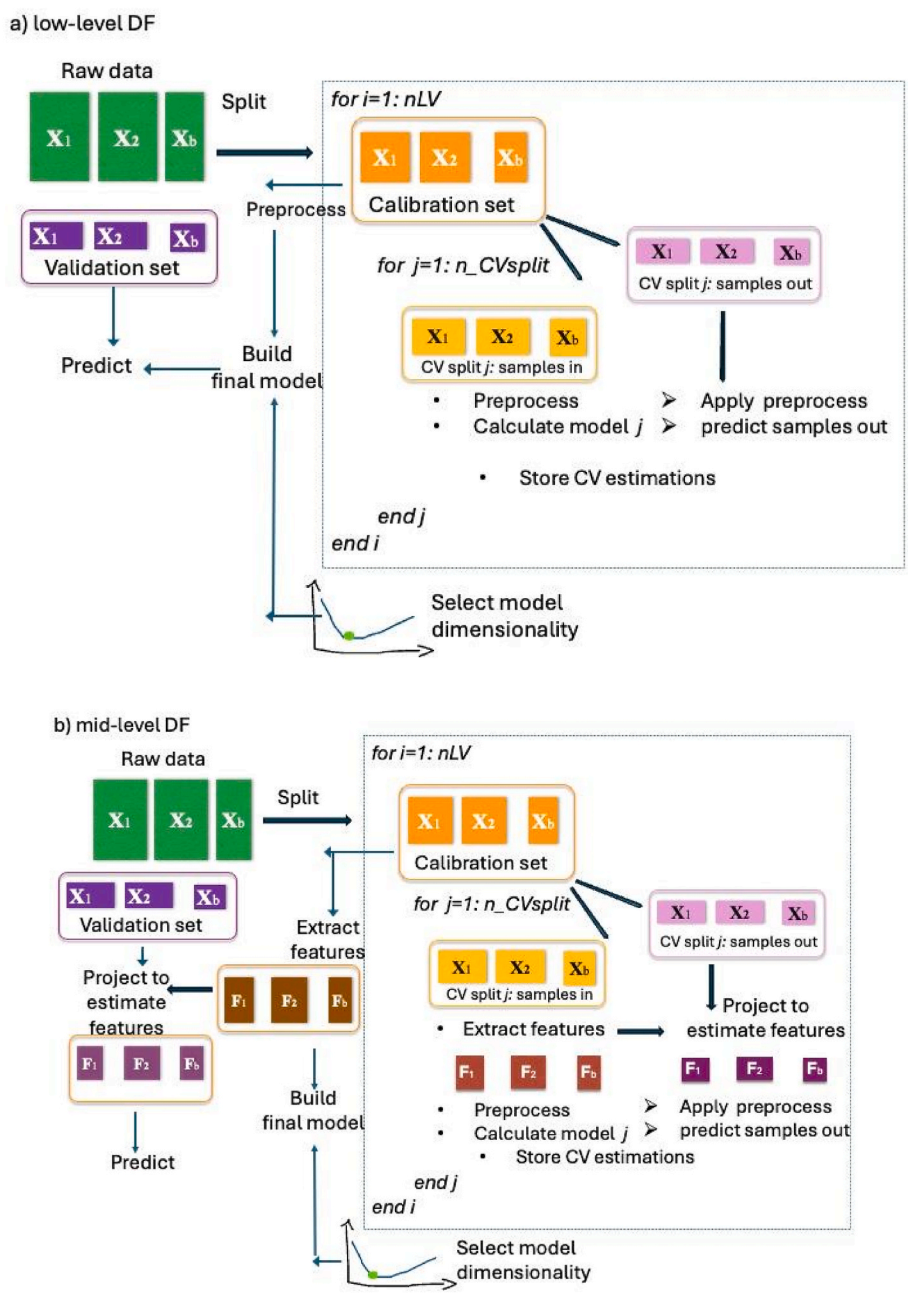

Fig. 3. Validation workflow in low-level (a) and mid-level (b) DF scenarios.

available, often categorized as filter, wrapper, or embedded methods [95]. However, this classification is not specifically tailored to multi-block approaches. Customized variable selection strategies for multi-block methods are relatively rare, and there is a lack of comprehensive discussion in the literature regarding which strategy is the most appropriate. An attempt was made to address this issue, focusing on MB-PLS and SO-PLS [96]. In this instance, it has been determined that, for spectroscopic data, when the objective is to attain a parsimonious set of features, it is preferable to integrate the sequential approach with forward selection. Conversely, when the emphasis is placed on interpretation and there is no imperative need to significantly reduce the number of variables, VIP can be incorporated into the creation of the SO-PLS model. However, it was concluded that it is challenging to determine upfront which strategy would be the most suitable. This difficulty is common in variable selection methodologies, as outcomes are inherently reliant on the specific dataset under examination. Alternatively, solutions of diverse nature, stemming from the hyphenation of multi-block and variable selection approaches, have been proposed as viable resolutions to this issue [97].

4.4. validation

Data fusion, as with any modelling approach, has some tuneable parameters which depend on the level at which DF operates and, on the method, adopted. Cross-validation (CV), or resampling methods in general, are widely used to select the optimal setting of tunable parameters [98]. However, the independence of the data splits must be carefully considered.

Fig. 3 represents the data analysis pipeline for each DF scenario highlighting at which stage the data splitting must take place.

In low-level DF scenario, the single data blocks are directly considered, either by data augmentation or coupling, and the number of components (overall or per block) is the main parameter to tune. In this case, the critical step is preprocessing, which has been described in section 4.1., and distinguished in three levels: 1) signal preprocessing; 2) inter-block and 3) intra-block preprocessing. A frequently encountered mistake is that often preprocessing at levels 1 and 2 is done before the data is split, and only the third level preprocessing enters the CV loop. However, this way the training data, in each CV iteration, are already informed about the left-out samples, e.g. because the mean or variance (level 2) or the average reference spectrum considered for multiplicative scatter correction (level 1) are calculated across the whole data block.

In mid-level DF scenario, features are extracted by each data block and afterward these features are fused, in this case it is critical that single block data are split before feature extraction and obviously also before inter-block preprocessing as already discussed above. In addition, there are more model parameters to be tuned, e.g. if the extracted features are components obtained by a decomposition model (PCA, PLS, etc.) their number is to be tuned too. Then, there are also the parameters of the fused model built on these selected features to be set. Thus, a double CV scheme is more appropriate [99].

Analogous reasoning applies to high-level DF scenarios, in which each single block undergoes a complete modelling step, and the models' decisions are fused.

5. Concluding remarks

In the area of in-field monitoring, remote sensing technologies combined with NDSS offer a comprehensive assessment of crop health and quality, aiding in early disease detection and yield prediction. Similarly, in food processing, on/in-line NDSS facilitate real-time monitoring of quality parameters, contamination detection, and process optimization. The integration of NDSS with other sensors enhances the precision and scope of monitoring, paving the way for efficient quality control and resource utilization. Furthermore, data fusion enables a more comprehensive understanding of food properties and enhances the effectiveness of authentication protocols.

In this paper, different data fusion scenarios have been reviewed, such as the fusion of spectra from different spectroscopic platforms, the combination of non-spectroscopic sensor data with spectroscopic data, and the integration of images and spectra. In the first two cases, both available methods and application examples, especially in food authentication, are abundant; in the latter, most of the literature focuses on image fusion, while a limited number of applications concern the integration of imaging, especially HSI and spectral data.

The focus of in-field monitoring is on fusing the results of remote sensing technologies, such as LIDAR and UAVs, with HSI to provide a comprehensive picture of the crop, from nutrient levels and water stress to disease detection and yield prediction. On/in-line NDSS are attractive because they can measure food quality attributes directly in the processing line, eliminating the need for manual sampling. Combining multiple NDSS between them and with process variables seems promising to obtain soft sensors for predicting hard-to-measure product quality attributes. It is likely that both areas will grow in the near future and more research is needed to improve the available DF methods to the point of routine implementation.

Coming to practical issues the following points can be highlighted.

- To profit from data fusion and selecting appropriate DF levels and methods the research questions and objectives must be clear, such as to which extent model interpretation is critical? Is common or distinctive information among blocks, or both, sought? Is the DF model to be implemented for real-time predictions? and so on.
- Block definition and data structure are closely interconnected, and block definition can either enhance or impede the efficacy of subsequent analyses.
- Implementing data fusion requires some expertise since the available methods are based on different assumptions. When combining block of different order, e.g. two dimensional and three-dimensional tensor decomposition methods are more suitable. Multiblock methods may offer several advantages on interpretative ground. However, they are influenced (to which extent depends on the specific method) by the block order which is not always straightforward to define.
- Variables selection may improve model predictive performance and enhance interpretability but a judicious balance between retaining informative features and mitigating noise is needed. It is worth to underline that variable selection strategies specific for multi-block methods are a few as illustrated in section 4.2.
- The level at which data are fused involves a different number of modelling steps which translate in different efforts in terms of validation as discussed in section 4.3.

Finally, we emphasize the message that seen from the perspective of going from research and development to routine implementation in-situ and in-line the state of art of DF approaches is still in his infancy.

CRediT authorship contribution statement

Lorenzo Strani: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Conceptualization. **Caterina Durante:** Writing – review & editing, Methodology, Conceptualization. **Marina Cocchi:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Conceptualization. **Federico Marini:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Formal analysis, Conceptualization. **Ingrid Måge:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Conceptualization. **Alessandra Biancolillo:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used DeepL Write in order to check English grammar. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Co-guest editor for the special issue Chemometrics If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

This research has been developed in the frame of the COST Action CA19145"European Network for Assuring Food Integrity using Non-Destructive Spectral Sensors (SENSORFINT).

References

[1] J. Chapman, A. Power, M.E. Netzel, Y. Sultanbawa, H.E. Smyth, V. Khanh Truong, D. Cozzolino, Challenges and opportunities of the fourth revolution: a brief insight into the future of food, Crit. Rev. Food Sci. Nutr. 62 (2022) 2845–2853, https://doi.org/10.1080/10408398.2020.1863328.
[2] A. Hassoun, S. Jagtap, H. Trollman, G. Garcia-Garcia, N.A. Abdullah, G. Goksen, F. Bader, F. Ozogul, F.J. Barba, J. Cropotova, P.E.S. Munekata, J.M. Lorenzo, Food processing 4.0: current and future developments spurred by the fourth industrial revolution, Food Control 145 (2023) 109507, https://doi.org/10.1016/j.foodcont.2022.109507.
[3] C. McVey, C.T. Elliott, A. Cannavan, S.D. Kelly, A. Petchkongkaew, S.A. Haughey, Portable spectroscopy for high throughput food authenticity screening: advancements in technology and integration into digital traceability systems, Trends Food Sci. Technol. 118 (2021) 777–790, https://doi.org/10.1016/j.tifs.2021.11.003.
[4] P. Bhatt, S. Mutturi, M.S. Thakur, Spectroscopy based in-line monitoring and control of food quality and safety, in: H.U. Hebbar, R. Sharma, R.S. Chaurasiya, S. Ranjan, K. Raghavarao (Eds.), Engineering Aspects of Food Quality and Safety, Food Engineering Series. Springer, Berlin, 2023, pp. 339–382, https://doi.org/10.1007/978-3-031-30683-9_12.
[5] D. Tanzilli, A. D'Alessandro, S. Tamelli, C. Durante, M. Cocchi, L. Strani, A feasibility study towards the on-line quality assessment of pesto sauce production by NIR and chemometrics, Foods 12 (2023) 1679, https://doi.org/10.3390/foods12081679.
[6] I. Måge, S. Gizachew W.J.P. Wold, L.E. Solberg, U. Böcker, K. Dankel, T. A. Lintvedt, B. Kafle, M. Cattaldo, J. Matić, L. Sorokina, N.K. Afseth, The role of biospectroscopy and chemometrics as enabling technologies for upcycling of raw materials from the food industry, Anal. Chim. Acta 1284 (2023) 342005, https://doi.org/10.1016/j.aca.2023.342005.
[7] S.O. Fakayode, G.A. Baker, D.K. Bwambok, N. Bhawawet, B. Elzey, N. Siraj, S. Macchi, D.A. Pollard, R.L. Perez, A.V. Duncan, I.M. Warner, Molecular (Raman, NIR, and FTIR) spectroscopy and multivariate analysis in consumable products analysis1, Appl. Spectrosc. Rev. 55 (2020) 647–723, https://doi.org/10.1080/05704928.2019.1631176.
[8] S.M. Azcarate, R. Ríos-Reina, J.M. Amigo, H.C. Goicoechea, Data handling in data fusion: methodologies and applications, TrAC, Trends Anal. Chem. 143 (2021) 116355, https://doi.org/10.1016/j.trac.2021.116355.
[9] L. Zhou, C. Zhang, Z. Qiu, Y. He, Information fusion of emerging non-destructive analytical techniques for food quality authentication: a survey, TrAC, Trends Anal. Chem. 127 (2020) 115901, https://doi.org/10.1016/j.trac.2020.115901.
[10] M. Guo, K. Wang, H. Lin, L. Wang, L. Cao, J. Sui, Spectral data fusion in nondestructive detection of food products: strategies, recent applications, and future perspectives, Compr. Rev. Food. Sci. 23 (2024) 1–23, https://doi.org/10.1111/1541-4337.13301.
[11] P. Mishra, J.M. Roger, D.J. Rimbaud-Bouveresse, A. Biancolillo, F. Marini, A. Nordon, D.N. Rutledge, Recent trends in multi-block data analysis in chemometrics for multi- source data integration, TrAC, Trends Anal. Chem. 137 (2021) 116206, https://doi.org/10.1016/j.trac.2021.116206.
[12] W. Li, E.W. Tramel, S. Prasad, J.E. Fowler, Nearest regularized subspace for hyperspectral classification, IEEE Geosci. Remote Sens. 52 (2013) 477–489, https://doi.org/10.1109/TGRS.2013.2241773.
[13] X. Shen, L. Cao, N.C. Coops, H. Fan, X. Wu, H. Liu, G. Wang, F. Cao, Quantifying vertical profiles of biochemical traits for forest plantation species using advanced remote sensing approaches, Remote Sens. Environ. 250 (2020) 112041, https://doi.org/10.1016/j.rse.2020.112041.
[14] A.K. Mahlein, E. Alisaac, A. Al Masri, J. Behmann, H.-W. Dehne, E.C. Oerke, Comparison and combination of thermal, fluorescence, and hyperspectral imaging for monitoring fusarium head blight of wheat on spikelet scale, Sensors 19 (2019) 2281, https://doi.org/10.3390/s19102281.
[15] V. Sagan, M. Maimaitijiang, P. Sidike, M. Maimaitiyiming, H. Erkbol, S. Hartling, K.T. Peterson, J. Peterson, J.G. Burken, F. Fritschi, UAV/satellite multiscale data fusion for crop monitoring and early stress detection, Int. J. Rem. Sens. 42 (2019) 715–722, https://doi.org/10.5194/isprs-archives-XLII-2-W13-715-2019.
[16] P.W. Bartlett, L. Coblenz, G. Sherwin, A. Stambler, A. van der Meer, A custom multi-modal sensor suite and data analysis pipeline for aerial field phenotyping, in: Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping II, 10218 A, 2017, pp. 19–28, https://doi.org/10.1117/12.2262858.
[17] F. Castaldi, A. Castrignanò, R. Casa, R. A data fusion and spatial data analysis approach for the estimation of wheat grain nitrogen uptake from satellite data, Int. J. Rem. Sens. 37 (2016) 4317–4336, https://doi.org/10.1080/01431161.2016.1212423.
[18] S. Grassi, C. Alamprese, Advances in NIR spectroscopy applied to process analytical technology in food industries, Curr. Opin. Food Sci. 22 (2018) 17–21, https://doi.org/10.1016/j.cofs.2017.12.008.
[19] E. Borràs, J. Ferré, R. Boqué, M. Mestres, L. Aceña, O. Busto, Data fusion methodologies for food and beverage authentication and quality assessment – a review, Anal. Chim. Acta 891 (2015) 1–14, https://doi.org/10.1016/j.aca.2015.04.042.
[20] M. Asachi, M. Alonso Camargo-Valero, Multi-sensors data fusion for monitoring of powdered and granule products: current status and future perspectives, Adv. Powder Technol. 34 (2023) 104055, https://doi.org/10.1016/J.APT.2023.104055.
[21] M. Guo, K. Wang, H. Lin, L. Wang, L. Cao, J. Sui, Spectral data fusion in nondestructive detection of food products: strategies, recent applications, and future perspectives, Compr. Rev. Food Sci. Food Saf. 23 (2024) e13301, https://doi.org/10.1111/1541-4337.13301.
[22] E. Hayes, D. Greene, C. O'Donnell, N. O'Shea, M.A. Fenelon, Spectroscopic technologies and data fusion: applications for the dairy industry, Front. Nutr. 9 (2023) 1074688, https://doi.org/10.3389/FNUT.2022.1074688/BIBTEX.
[23] V. Brunner, M. Siegl, D. Geier, T. Becker, Challenges in the development of soft sensors for bioprocesses: a critical review, Front. Bioeng. Biotechnol. 9 (2021) 722202, https://doi.org/10.3389/FBIOE.2021.722202/BIBTEX.
[24] P. Verboven, T. Defraeye, A.K. Datta, B. Nicolai, Digital twins of food process operations: the next step for food process models? Curr. Opin. Food Sci. 35 (2020) 79–87, https://doi.org/10.1016/J.COFS.2020.03.002.
[25] A. Hassoun, A. Aït-Kaddour, A.M. Abu-Mahfouz, N.B. Rathod, F. Bader, F.J. Barba, A. Biancolillo, J. Cropotova, C.M. Galanakis, A.R. Jambrak, J.M. Lorenzo, I. Måge, F. Ozogul, J. Regenstein, The fourth industrial revolution in the food industry—Part I: industry 4.0 technologies, Crit. Rev. Food Sci. Nutr. 63 (2023) 6547–6563, https://doi.org/10.1080/10408398.2022.2034735.
[26] W.Y. Chai, K.T.K. Teo, M.K. Tan, H.J. Tham, Model predictive control in fermentation process – a review, AIP Conf. Proc. 2610 (2022), https://doi.org/10.1063/5.0101066/2830914.
[27] A. Koulouris, N. Misailidis, D. Petrides, Applications of process and digital twin models for production simulation and scheduling in the manufacturing of food ingredients and products, Food Bioprod. Process. 126 (2021) 317–333, https://doi.org/10.1016/J.FBP.2021.01.016.
[28] L. Rodriguez-Saona, D.P. Aykas, K.R. Borba, A. Urtubia, Miniaturization of optical sensors and their potential for high-throughput screening of foods, Curr. Opin. Food Sci. 31 (2020) 136–150, https://doi.org/10.1016/j.cofs.2020.04.008.
[29] P.D.C. Sanchez, H.B.T. Arogancia, K.M. Boyles, A.J.B. Pontillo, M.M. Ali, Emerging nondestructive techniques for the quality and safety evaluation of pork and beef: recent advances, challenges, and future perspectives, Appl. Food Res. 2 (2022), https://doi.org/10.1016/j.afres.2022.100147.
[30] S. Silva, C. Guedes, S. Rodrigues, A. Teixeira, Non-destructive imaging and spectroscopic techniques for assessment of carcass and meat quality in sheep and goats: a review, Foods 9 (8) (2020) 1074, https://doi.org/10.3390/foods9081074.
[31] C. Alamprese, M. Casale, N. Sinelli, S. Lanteri, E. Casiraghi, Detection of minced beef adulteration with Turkey meat by UV–vis, NIR and MIR spectroscopy, LWT–Food Sci. Technol. 53 (2013) 225–232, https://doi.org/10.1016/j.lwt.2013.01.027.
[32] H. Pu, D.-W. Sun, J. Ma, J.-H. Cheng, Classification of fresh and frozen-thawed pork muscles using visible and near infrared hyperspectral imaging and textural analysis, Meat Sci. 99 (2015) 81–88, https://doi.org/10.1016/j.meatsci.2014.09.001.
[33] G. Orlandi, R. Calvini, G. Foca, L. Pigani, G. Vasile Simone, A. Ulrici, Data fusion of electronic eye and electronic tongue signals to monitor grape ripening, Talanta 195 (2019) 181–189, https://doi.org/10.1016/j.talanta.2018.11.046.
[34] F. Mendoza, R. Lu, D. Ariana, H. Cen, B. Bailey, Integrated spectral and image analysis of hyperspectral scattering data for prediction of apple fruit firmness and soluble solids content, Postharvest Biol. Technol. 62 (2011) 149–160, https://doi.org/10.1016/j.postharvbio.2011.05.009.

[35] Y. Weesepoel, M. Alewijn, M. Wijtten, J. Müller-Maatsch, Detecting food fraud in extra virgin olive oil using a prototype portable hyphenated photonics sensor, J. AOAC Int. 104 (2021) 7–15, https://doi.org/10.1093/jaoacint/qsaa099.

[36] E. Borràs, J. Ferré, R. Boqué, M. Mestres, L. Aceña, A. Calvo, O. Busto, Prediction of olive oil sensory descriptors using instrumental data fusion and partial least squares (PLS) regression, Talanta 155 (2016) 116–123, https://doi.org/10.1016/j.talanta.2016.04.040.

[37] M. Cocchi, Chapter 1 Introduction: ways and Means to deal with data from multiple souces, in: M. Cocchi (Ed.), Data Handling in Science and Technology, vol. 31, Elsevier, 2019, pp. 311–344, https://doi.org/10.1016/B978-0-444-63984-4.00001-6.

[38] C. Zheng, J. Li, H. Liu, Y. Wang, Data fusion of FT-NIR and ATR-FTIR spectra for accurate authentication of geographical indications for Gastrodia elata Blume, Food Biosci. 56 (2023) https://doi.org/103308.10.1016/j.fbio.2023.103308.

[39] M. Silvestri, A. Elia, D. Bertelli, E. Salvatore C. Durante, M. Li Vigni, A. Marchetti, M. Cocchi, A mid level data fusion strategy for the Varietal Classification of Lambrusco PDO wines, Chemometr. Intell. Lab. Syst. 137 (2014) 181–189, https://doi.org/10.1016/j.chemolab.2014.06.012.

[40] A. Biancolillo, R. Bucci, A.L. Magrì, A.D. Magrì, F. Marini, Data-fusion for multiplatform characterization of an Italian craft beer aimed at its authentication, Anal. Chim. Acta 820 (2014) 23–31, https://doi.org/10.1016/j.aca.2014.02.024.

[41] R. Ríos-Reina, R.M. Callejón, F. Savorani, J.M. Amigo, M. Cocchi, Data fusion approaches in spectroscopic characterization and classification of PDO wine vinegars, Talanta 198 (2019) 560–572, https://doi.org/10.1016/j.talanta.2019.01.100.

[42] M.R. Baqueta, P. Valderrama, M. Mandrone, F. Poli, A. Coqueiro, A.C. Costa-Santos, A.P. Rebellato, G.M. Luz, R.B. Rocha, J.A.L. Pallone, F. Marini, ^{1}H NMR, FAAS, portable NIR, benchtop NIR, and ATR-FTIR-MIR spectroscopies for characterizing and discriminating new Brazilian Canephora coffees in a multi-block analysis perspective, Chemometr. Intell. Lab. Syst. 240 (2023) 104907, https://doi.org/10.1016/j.chemolab.2023.104907.

[43] J.A. Fernández Pierna, P. Vermeulen, D. Eylenbosch, J. Burger, B. Bodson, P. Dardenne, V. Baeten, Chemometrics in NIR hyperspectral imaging: theory and applications in the agricultural crops and products sector, in: S. Brown, R. Tauler, B. Walczak (Eds.), Comprehensive Chemometrics, second ed., Elsevier, 2020, pp. 361–379, https://doi.org/10.1016/B978-0-12-409547-2.14811-5. ISBN 9780444641663.

[44] F. Abdelghafour, S.R. Sivarajan, I. Abdelmeguid, M. Ryckewaert, J.M. Roger, R. Bendoula, E. Alexandersson, Including measurement effects and temporal variations in VIS-NIRS models to improve early detection of plant disease: application to Alternaria solani in potatoes, Comput. Electron. Agric. 211 (2023) 107947, https://doi.org/10.1016/j.compag.2023.107947.

[45] D. Lorente, N. Aleixos, J. Gómez-Sanchis, S. Cubero, O.L. García-Navarrete, J. Blasco, Recent advances and applications of hyperspectral imaging for fruit and vegetable quality assessment, Food Bioprocess Technol. 5 (2012) 1121–1142, https://doi.org/10.1007/s11947-011-0725-1.

[46] S. Vignati, A. Tugnolo, V. Giovenzana, A. Pampuri, A. Casson, R. Guidetti, R. Beghi, Hyperspectral imaging for fresh-cut fruit and vegetable quality assessment: basic concepts and applications, Appl. Sci. 13 (2023) 9740, https://doi.org/10.3390/app13179740.

[47] R.R. de Oliveira, A. de Juan, A. SWiVIA–Sliding window variographic image analysis for real-time assessment of heterogeneity indices in blending processes monitored with hyperspectral imaging, Anal. Chim. Acta 1180 (2021) 338852, https://doi.org/10.1016/j.aca.2021.338852.

[48] S. Munera, A. Rodríguez-Ortega, N. Aleixos, S. Cubero, J. Gómez-Sanchis, J. Blasco, Detection of invisible damages in 'rojo brillante' persimmon fruit at different stages using hyperspectral imaging and chemometrics, Foods 10 (2021) 2170, https://doi.org/10.3390/foods10092170.

[49] J. Blasco, N. Aleixos, N. Computer, Vision in agriculture, in: Q. Zhang (Ed.), Encyclopedia of Smart Agriculture Technologies, Springer, Berlin, 2023, https://doi.org/10.1007/978-3-030-89123-7_45-1.

[50] T.A. Lintvedt, P.V. Andersen, N.K. Afseth, K. Heia, S.K. Lindberg, J.P. Wold, Raman spectroscopy and NIR hyperspectral imaging for in-line estimation of fatty acid features in salmon fillets, Talanta 254 (2023) 124113, https://doi.org/10.1016/j.talanta.2022.124113.

[51] Hyperspectral pansharpening: a review, IEEE Geosci. Remote Sens. Mag. 3 (2015) 2746, https://doi.org/10.1109/MGRS.2015.2440094.

[52] N. Yokoya, C. Grohnfeldt, J. Chanussot, Hyperspectral and Multispectral Data Fusion. A comparative review of the recent literature, IEEE Geosci. Remote Sens. Mag. 5 (2017) 29–56, https://doi.org/10.1109/MGRS.2016.2637824.

[53] F.D. Javan, F. Samadzadegan, S. Mehravar, A. Toosi, R. Khatami, A. Stein, A review of image fusion techniques for pan-sharpening of high-resolution satellite imagery, ISPRS J. Photogrammetry Remote Sens. 171 (2021) 101–117, https://doi.org/10.1016/j.isprsjprs.2020.11.001.

[54] S. Ashraf, L. Brabyn, B.J. Hicks, Image data fusion for the remote sensing of freshwater environments, Appl. Geogr. 32 (2012) 619–628, https://doi.org/10.1016/j.apgeog.2011.07.010.

[55] L. Song, Y. Ful, H. Huang, Y. Chen, Fast HSI super resolution using linear regression, IET Image Process. 13 (2019) 1671–1679, https://doi.org/10.1049/iet-ipr.2018.5475.

[56] R. Gomez, A. Jazaeri, M. Kafatos, Wavelet-based hyperspectral and multi-spectral image fusion, Proc. SPIE 4383 (2001) 36–42, https://doi.org/10.1117/12.428249.

[57] J. Nunez, X. Otazu, O. Fors, A. Prades, V. Pala, R. Arbiol, Multiresolution-based image fusion with additive wavelet decomposition, IEEE Trans. Geosci. Rem. Sens. 37 (1999) 1204–1211, https://doi.org/10.1109/36.763274.

[58] D. Fasbender, J. Radoux, P. Bogaert, Bayesian data fusion for adaptable image pan-sharpening, IEEE Trans. Geosci. Rem. Sens. 46 (2008) 1847–1857, https://doi.org/10.1109/TGRS.2008.917131.

[59] N. Yokoya, T. Yairi, A. Iwasaki, Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion, IEEE Trans. Geosci. Rem. Sens. 50 (2012) 528–537, https://doi.org/10.1109/TGRS.2011.2161320.

[60] A. De Juan, R.R. de Oliveira, A. Gómez-Sánchez, Chapter 6 - multiset analysis by multivariate curve resolution: the unmixing methodology to handle hyperspectral image fusion scenarios, in: A.C. Olivieri, G.M. Escandar, H.C. Goicoechea, A. Muñoz de la Peña (Eds.), Data Handling in Science and Technology, vol. 33, Elsevier, 2024, pp. 111–132. ISBN 9780443132612.

[61] A. De Juan, A. Gowen, l. Duponchel, C. Ruckebusch, Chapter 8 image fusion, in: M. Cocchi (Ed.), Data Handling in Science and Technology, vol. 31, Elsevier, 2019, pp. 311–344, https://doi.org/10.1016/B978-0-444-63984-4.21001-6.

[62] H. Al-Mekhlafi, S. Liu, Single image super-resolution: a comprehensive review and recent insight, Front. Comput. Sci. 18 (2024) 181702, https://doi.org/10.1007/s11704-023-2588-9.

[63] A. Signoroni, M. Savardi, A. Baronio, S. Benin, Deep learning meets hyperspectral image analysis: a multidisciplinary review, J. Imaging 5 (2019) 52, https://doi.org/10.3390/jimaging5050052.

[64] F. Allouche, M. Hanafi, F. Jamme, P. Robert, C. Barron, F. Guillon, M.F. Devaux, Coupling hyperspectral image data having different spatial resolutions using Multiple Co-inertia Analysis, Chemometr. Intell. Lab. Syst. 117 (2012) 200–212, https://doi.org/10.1016/j.chemolab.2012.04.004.

[65] F. Allouche, M. Hanafi, F. Jamme, P. Robert, C. Barron, F. Guillon, M.F. Devaux, Coupling hyperspectral image data having different spatial resolutions by extending multivariate inter-battery Tucker analysis, Chemometr. Intell. Lab. Syst. 113 (2012) 43–51, https://doi.org/10.1016/j.chemolab.2011.08.010.

[66] N. Liu, W. Li, Y. Wang, Q. Du, J. Chanussot, A survey on hyperspectral image restoration: from the view of low-rank tensor approximation, Sci. China Inf. Sci. 66 (2023) 140302, https://doi.org/10.1007/s11432-022-3609-4.

[67] A.F. Ahsani, Y.A. Sari, P.P. Adikara, Food image retrieval with gray level Co-occurrence matrix texture feature and CIE L*a*b* color moments feature. Proceedings of 4th International Conference on Sustainable Information Engineering and Technology, SIET 2019, art. no. 8985990, pp. 130 – 134. DOI: 10.1109/SIET48054.2019.8985990.

[68] J.M. Prats-Montalbán, A. de Juan, A. Ferrer, Multivariate image analysis: a review with applications, Chemometr. Intell. Lab. Syst. 107 (2011) 1–23, https://doi.org/10.1016/j.chemolab.2011.03.002.

[69] M. Li Vigni, J.M. Prats-Montalbán, A. Ferrer, M. Cocchi, Coupling 2D-wavelet decomposition and multivariate image analysis (2D WT-MIA), J. Chemom. 32 (2018) e2970, https://doi.org/10.1002/cem.2970.

[70] J.M. Amigo, B.M. Jespersen, F. van den Berg, J.J. Jensen, S.B. Engelsen, Batch-wise versus continuous dough mixing of Danish butter cookies. A near infrared hyperspectral imaging study, Food Chem. 414 (2023) 135731, https://doi.org/10.1016/j.foodchem.2023.135731.

[71] H. Priyashantha, A. Höjer, K.H. Saedén, A. Lundh, M. Johansson, G. Bernes, P. Geladi, M. Hetta, Use of near-infrared hyperspectral (NIR-HS) imaging to visualize and model the maturity of long-ripening hard cheeses, J. Food Eng. 264 (2019) 109687, https://doi.org/10.1016/j.jfoodeng.2019.109687.

[72] P. Mishra, C.B.Y. Cordella, D.N. Rutledge, P. Barreiro, J.M. Roger, B. Diezma, Application of independent components analysis with the JADE algorithm and NIR hyperspectral imaging for revealing food adulteration, J. Food Eng. 168 (2016) 7–15, https://doi.org/10.1016/j.jfoodeng.2015.07.008.

[73] C. Chatzichristos, S. Van Eyndhoven, E. Kofidis, S. Van Huffel, Chapter 10 - coupled tensor decompositions for data fusion, in: Y. Liu (Ed.), Tensors for Data Processing, Academic Press, 2022, pp. 341–370, https://doi.org/10.1016/B978-0-12-824447-0.00016-9. ISBN 9780128244470.

[74] C.I. Kanatsoulis, X. Fu, N.D. Sidiropoulos, W.-K. Ma, Hyperspectral super-resolution via coupled tensor factorization: identifiability and algorithms, Proc IEEE Int Conf Acoust Speech Signal Process (2018) 3191–3195.

[75] J.E. Cohen, R. Bro, P. Comon, Tensor decompositions: principles and application to food sciences. In C. Jutten, L. Tomazeli Duarte, S. Moussaoui (Eds.). Source Separation in Physical Chemical Sensing, Wiley (2023) 255-324, 978-1-119-13722-1. https://hal.science/hal-03367935v2.

[76] L.M. França, J.M. Amigo, C. Cairós, M. Bautista, M.F. Pimentel, Evaluation and assessment of homogeneity in images. Part 1: unique homogeneity percentage for binary images, Chemometr. Intell. Lab. Syst. 171 (2017) 26–39, https://doi.org/10.1016/j.chemolab.2017.10.002.

[77] R. Calvini, G. Foca, A. Ulrici, Data dimensionality reduction and data fusion for fast characterization of green coffee samples using hyperspectral sensors, Anal. Bioanal. Chem. 408 (2016) 7351–7366, https://doi.org/10.1007/s00216-016-9713-7.

[78] P. Mishra, D. Passos, F. Marini, J. Xu, J.M. Amigo, A. Gowen, J.J. Jansen, A. Biancolillo, J.M. Roger, D.N. Rutledge, A. Nordon, Deep learning for near-infrared spectral data modelling: hypes and benefits, TrAC, Trends Anal. Chem. 157 (2022) 116804, https://doi.org/10.1016/j.trac.2022.116804.

[79] M. Al-Sarayreh, M. Reis, W.Q. Yan, R. Klette, Detection of red-meat adulteration by deep spectral–spatial features in hyperspectral images, J. Imaging 4 (2018) 63, https://doi.org/10.3390/jimaging4050063.

[80] N.N. Misra, Y. Dixit, A. Al-Mallahi, M.S. Bhullar, R. Upadhyay, A. Martynenko IoT, Big data, and artificial intelligence in agriculture and food industry, IEEE Internet Things J. 9 (2022) 6305–6324, https://doi.org/10.1109/JIOT.2020.2998584.

[81] I. Kumar, J. Rawat, N. Mohd, S. Husain, Opportunities of artificial intelligence and machine learning in the food industry, J. Food Qual. (2021) 4535567, https://doi.org/10.1155/2021/4535567.

[82] T. Casian, B. Nagy, B. Kovács, D.L. Galata, E. Hirsch, A. Farkas, Challenges and opportunities of implementing data fusion in process analytical technology—a review, Molecules 27 (2022) 4846, https://doi.org/10.3390/MOLECULES27154846.
[83] R.R. De Oliveira, C. Avila, R. Bourne, F. Muller, A. De Juan, Data fusion strategies to combine sensor and multivariate model outputs for multivariate statistical process control, Anal. Bioanal. Chem. 412 (2020) 2151–2163, https://doi.org/10.1007/S00216-020-02404-2.
[84] L. Strani, R. Vitale, D. Tanzilli, F. Bonacini, A. Perolo, E. Mantovani, A. Ferrando, M. Cocchi, A multiblock approach to fuse process and near-infrared sensors for on-line prediction of polymer properties, Sensors 22 (2022) 1436, https://doi.org/10.3390/S22041436.
[85] J. Lauzon-Gauthier, P. Manolescu, C. Duchesne, The Sequential Multi-block PLS algorithm (SMB-PLS): comparison of performance and interpretability, Chemometr. Intell. Lab. Syst. 180 (2018) 72–83, https://doi.org/10.1016/j.chemolab.2018.07.005.
[86] F. Westad, F. Marini, Validation of chemometric models – a tutorial, Anal. Chim. Acta 893 (2015) 14–24, https://doi.org/10.1016/j.aca.2015.06.056.
[87] H.D. Yu, L.W. Qing, D.T. Yan, G. Xia, C. Zhang, Y.H. Yun, W. Zhang, Hyperspectral imaging in combination with data fusion for rapid evaluation of tilapia fillet freshness, Food Chem. 348 (2021) 129129, https://doi.org/10.1016/J.FOODCHEM.2021.129129.
[88] Y. Li, Y. Xiong, S. Min, Data fusion strategy in quantitative analysis of spectroscopy relevant to olive oil adulteration, Vib. Spectrosc. 101 (2019) 20–27, https://doi.org/10.1016/J.VIBSPEC.2018.12.009.
[89] C. Assis, H.V. Pereira, V.S. Amador, R. Augusti, L.S. de Oliveira, M.M. Sena, Combining mid infrared spectroscopy and paper spray mass spectrometry in a data fusion model to predict the composition of coffee blends, Food Chem. 281 (2019) 71–77, https://doi.org/10.1016/J.FOODCHEM.2018.12.044.
[90] J. Lauzon-Gauthier, P. Manolescu, C. Duchesne, The Sequential Multi-block PLS algorithm (SMB-PLS): comparison of performance and interpretability, Chemometr. Intell. Lab. Syst. 180 (2018) 72–83, https://doi.org/10.1016/J.CHEMOLAB.2018.07.005.
[91] P. Mishra, D. Passos, Deep multiblock predictive modelling using parallel input convolutional neural networks, Anal. Chim. Acta 1163 (2021) 338520, https://doi.org/10.1016/j.aca.2021.338520.
[92] A.K. Smilde, T. Næs, K.H. Liland, Multiblock Data Fusion in Statistics and Machine Learning: Applications in the Natural and Life Sciences, John Wiley & Sons, New York, 2022, https://doi.org/10.1002/9781119600978.
[93] M.P. Campos, R. Sousa, M.S. Reis, Establishing the optimal blocks' order in SO-PLS: stepwise SO-PLS and alternative formulations, J. Chemom. 32 (2018) e3032, https://doi.org/10.1002/cem.3032.
[94] K. Jørgensen, T. Næs, A design and analysis strategy for situations with uncontrolled raw material variation, J. Chemom. 18 (2004) 45–52, https://doi.org/10.1002/cem.835.
[95] J.M. Roger, B. Palagos, D. Bertrand, E. Fernandez-Ahumada, CovSel: variable selection for highly multivariate and multi-response calibration: application to IR spectroscopy, Chemometr. Intell. Lab. Syst. 106 (2011) 216–223, https://doi.org/10.1016/j.chemolab.2010.10.003.
[96] A. Biancolillo, K.H. Liland, I. Måge, T. Næs, R. Bro, Variable selection in multi-block regression, Chemometr. Intell. Lab. Syst. 156 (2016) 89–101, https://doi.org/10.1016/j.chemolab.2016.05.016.
[97] A. Biancolillo, F. Marini, J.-M. Roger, SO-CovSel: a novel method for variable selection in a multiblock framework, J. Chemom. 34 (2020) e3120, https://doi.org/10.1002/cem.3120.
[98] M.P. Campos, M.S. Reis, Data preprocessing for multiblock modelling–a systematization with new methods, Chemometr. Intell. Lab. Syst. 199 (2020) 103959.
[99] P. Filzmoser, B. Liebmann, K. Varmuza, Repeated double cross validation, J. Chemometr. 23 (2009) 160–171, https://doi.org/10.1002/cem.1225.