Evaluating the impact of data pre-processing methods on classification of ATR-FTIR spectra of bituminous binders

Khalighi, Sadaf; Ma, Lili; Ren, Shisong; Varveri, Aikaterini

DOI
10.1016/j.fuel.2024.132701

Publication date
2024

Document Version
Final published version

Published in
Fuel

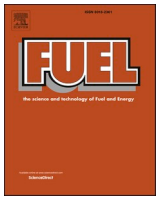
Citation (APA)
Khalighi, S., Ma, L., Ren, S., & Varveri, A. (2024). Evaluating the impact of data pre-processing methods on classification of ATR-FTIR spectra of bituminous binders. *Fuel*, *376*(132701), Article 132701. https://doi.org/10.1016/j.fuel.2024.132701

Fuel 376 (2024) 132701

Contents lists available at ScienceDirect

Fuel

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



Full Length Article

Evaluating the impact of data pre-processing methods on classification of ATR-FTIR spectra of bituminous binders

Sadaf Khalighi *, Lili Ma, Shisong Ren, Aikaterini Varveri

Delft University of Technology, Netherlands

ARTICLE INFO

Keywords:
ATR-FTIR
PLSDA
Normalization
Baseline correction
Peak areas
Indices

ABSTRACT

Attenuated total reflectance Fourier-transform infrared spectroscopy (ATR-FTIR) is an essential tool for the analysis of bituminous binders due to its cost-effectiveness, user-friendliness, and non-destructive nature. However, its effectiveness is often hampered by challenges such as non-informative regions, lack of standardized analysis methods, and inconsistent baselines in spectral data. Addressing these challenges, this study aims to comprehensively evaluate the impact of various data pre-processing (DP) methods on ATR-FTIR spectra from diverse bituminous binder types, sources, and aging conditions. Using partial least squares-discriminant analysis (PLSDA) classification, the study assesses the effectiveness of baseline correction, normalization, and their combinations. The methodology involves analyzing peak areas, indices, entire spectra, and first derivative spectra to determine the most effective pre-processing strategies. Key findings reveal that the effectiveness of DP methods is influenced by the classification goals, characteristics of the spectral dataset, and the specific methods employed for input data preparation. The study demonstrates that using entire spectra or their first derivatives leads to higher classification accuracy compared to indices or specific spectral peak areas. The choice between peak area and indices calculation methods should align with the study's objectives. For efficient and rapid selection of DP methods, tools like PLSDA are recommended. Among the normalization methods, normalization to constant vector length (NCV), normalization to change the maximum to 1 (NMO), robust scaling (RS), and normalization to sum (NTS) are suitable for peak area or indices-based classification. For entire spectra and their first derivatives, NTS, NCV, autoscaling (AS), pareto scaling (PS), and standard normal variate (SNV) methods are recommended. Regarding baseline correction, Adaptive Smoothness Penalized Least Squares (aspls) is suitable for studies focusing on gradual material changes, such as multi-level aging studies, but not for additive detection studies. The findings of this study provide valuable insights and practical recommendations for selecting appropriate DP methods, thereby enhancing the classification accuracy and reliability of ATR-FTIR spectral analysis of bituminous binders. This contributes significantly to the design of experiments, reduces operational risks, and optimizes resource utilization in the field.

1. Introduction

Bituminous binders are essential materials in pavement construction, derived from crude oil residues and rich in functional groups, including aromatic and aliphatic hydrocarbons. These materials' chemical complexity, influenced by various elements such as sulfur and oxygen-containing groups, affects their stability and adhesive properties [1]. Understanding these functional groups is critical for explaining the behavior of bituminous binders under diverse conditions and optimizing their application in fields like pavement construction and waterproofing. A widely used method to analyze the chemical complexity of bituminous binders is attenuated total reflectance Fourier-transform infrared spectroscopy (ATR-FTIR) [2]. ATR-FTIR is valued for its cost-effectiveness, user-friendly nature, and non-destructive analysis capabilities [3,4]. This method can track changes in functional groups due to oxidative aging and the introduction of new materials like polymers or rejuvenators [5,6]. Such changes are observable in ATR-FTIR spectra, facilitating comparisons between chemically altered and unaltered samples [7].

While ATR-FTIR offers several advantages and can quickly produce extensive data, there are still challenges, especially when applied to bituminous binders. One specific issue is the complexity in analysing the

* Corresponding author.
E-mail address: s.khalighi@tudelft.nl (S. Khalighi).

https://doi.org/10.1016/j.fuel.2024.132701
Received 18 May 2024; Received in revised form 29 July 2024; Accepted 2 August 2024

spectra results of binders. Generally, values like the area under spectral curves or indices are used to characterize samples and compare various samples with different degrees of modifications and aging [2,8]. Unfortunately, the lack of a standardized analysis method makes it difficult to compare findings across different studies. Aside from a lack of a standardized analysis method for FTIR spectra, presence of both informative and uninformative regions in the spectra leads to reduced performance of post-processing steps [9]. Moreover, inconsistent baseline and noise cause systematic variation in the spectra [4,9]. Sample conditions, particle sizes, chemical interferences, and how the data is collected can change the intensity of both wanted and unwanted parts of the spectra [10]. Sources of unwanted variation could arise from inherent limitation of instrument (e.g. instrument drifts) or samples (e.g. particle size or homogeneity level). An additional source of systematic variations induced by different sample quantity is especially pronounced in case of solid sample [10,11].

These challenges necessitate effective data pre-processing (DP) methods to improve the accuracy and reliability of spectral analysis. The DP methods can be divided in to two main categories: first, normalization data-pre-processing (NDP) and second, baseline-correction data-pre-processing (BDP).

Searching through literature, it is not conclusive what the best Data Pre-processing (DP) method is because many works do not explain why a specific DP method is chosen perhaps because they rely on others' choices when selecting DP methods for their own work [9]. Most studies seem to select DP methods based on literature, as only a few methods are consistently highlighted by researchers. This practice may lead researchers to overlook the most suitable DP methods for their specific data. Additionally, from the available literature, it is noted that most researchers have not compared clearly how their chosen DP method performs compared to the raw, untreated IR spectra [9]. They have assumed that the selected DP method would improve the results without further investigation.

When using ATR-FTIR to evaluate bituminous binders, analyses fall into two categories: using the entire spectral data or relying on peak areas/indices. In both categories, the selection and discussion of data pre-processing (DP) methods are often unclear. Ma et al. [12] used standard normal variate (SNV) and Savitzky–Golay (SG) methods before applying PCA-linear discriminant analysis (LDA) to classify different aging states of bituminous binders. Primerano et al. [13] applied standardized scaling to FTIR data and used random forest, PCA, and PLSDA to differentiate between various aged samples. Garmarudi et al. [14] employed Mean Centering followed by PCA, hierarchical cluster analysis (HCA), and SIMCA to classify crude oils geologically. Weigel and Stephan [15] used SNV transformation and the first derivative of FTIR data in LDA and PLSR algorithms to distinguish bituminous binder samples by refinery.

In studies focusing on peak areas and indices, DP methods are also limited. Hofko et al. [2] evaluated ATR-FTIR reliability and sensitivity, recommending normalized spectra with an absolute baseline and peak area integration. Weigel and Stephan [16] applied SNV transformation to calculate peak areas, creating PLS regression models for various bituminous binder parameters. Wieser et al. [17] optimized DP methods for aging analysis, finding exponential baseline correction, vector 2 normalization, and peak-fit integration as the best methods for studying aging based on the carboxyl peak.

2. Objectives and research structure

In this study, our motivation to evaluate the effects of different DP methods on ATR-FTIR spectra analysis in bituminous binder research stems from the lack of comprehensive references guiding DP method selection. We aim to incorporate all pre-treatment methods from the literature into a unified framework and compare standard Pressure Aging Vessel (PAV) aging with moisture-aged samples, crucial for future laboratory aging protocols.

To ensure dataset diversity, we use samples from various bituminous binder types, sources, and aging conditions, including unmodified binder from different sources, polymer modified binder, and rejuvenated binder, both pre- and post-aging. Using the partial least squares-discriminant analysis (PLSDA) algorithm, we assess how pre-treatment methods affect FTIR spectral data classification. The classification goals include grouping by binder source, modification type, or aging state. While binder modification and source significantly influence FTIR spectra, aging conditions result in less pronounced differences. Therefore, we focus on aging studies to ensure optimized DP methods can detect even minor discrepancies.

Specifically, aging under PAV and humid conditions was examined, ensuring the findings are applicable to various aging protocols. As a result, our findings become more reliable and are capable of distinguishing between fresh samples and those aged under different conditions. Given the consideration that the chemical differences among samples originated from different sources or modified by various materials are larger than that among samples at all kinds of ageing levels, our hypothesis suggests that the DP method most effective for aging classification should also perform well for other classification purposes. We aim to examine and confirm this hypothesis through an example in the last section that focuses on the classification of samples before and after rejuvenations.

This study aims to address the following research questions:

1. Can pre-treatment methods, such as normalization, baseline correction, or a combination of both, enhance classification accuracy compared to using raw input data?
2. Which pre-treatment method is most effective for the aging classification of bituminous binders?
3. What is the most suitable method for calculating areas or aging indices in aging studies?
4. What region of FTIR spectra are more important for classification? Does the importance of regions change with pre-treatment of spectra?
5. When considering either the entire spectra or the first derivative instead of peak areas or indices, which method proves superior for future aging-related chemometric analysis?
6. Does the optimal method for aging classification demonstrate effectiveness for other classification purposes?

By addressing these questions, this study aims to provide guidelines for selecting DP methods, thereby enhancing the accuracy and reliability of ATR-FTIR spectral analysis for bituminous binders.

3. Materials and methods

3.1. Data collection (sample preparation, aging conditions, and FTIR measurement)

The first set of data in this research was extracted from a study conducted by Ren et al. [18] (Table 1S, supporting information). For the investigation of rejuvenated binders, a 70/100 virgin bituminous binder was utilized. The study encompassed four types of rejuvenators: bio-oil (BO), engine-oil (EO), naphthenic-oil (NO), and aromatic-oil (AO). The rejuvenated binders comprised 20 h-PAV aged samples mixed with 10 % of rejuvenators, 40 h-PAV aged samples mixed with 5 %, 10 %, 15 % of rejuvenators, and 80 h-PAV aged samples mixed with 10 % of rejuvenators. These samples were collectively treated as a fresh group. All these rejuvenated samples were then conditioned by 1PAV ageing, which were considered as the aged group. An ATR-FTIR device (Waltham, MA, USA) was used to detect the distribution of functional groups in the samples, within the wavenumber range of 600–4000 cm^{-1}, with 12 scans conducted for each sample. To ensure data reliability, a minimum of three parallel tests were conducted for each specimen. In total, 120 samples were measured at room temperature for this dataset.

The second set of data was obtained from the study by Ma et al. [12] (Table 1S, supporting information). They aimed to explore the chemical properties of bituminous binder derived from diverse crude oil sources, exhibiting various penetration grades, polymer modifiers (Styrene–butadiene–styrene (SBS)), and aging states. A total of 16 binder samples, representing eight different types, were prepared at two aging states—fresh and PAV aged. These samples included bituminous binder sourced from Q, featuring penetration grades 40/60, 70/100, and SBS modifier (using 70/100 as the base bituminous binder). Additionally, bituminous binder from source T included penetration grades 70/100, 100/150, and 160/220. Another variation was bituminous binder from source V, with penetration grade 70/100 and SBS modifier (using 70/100 as the base bituminous binder). The long-term aged samples were crafted through a combination of short-term aging simulated by the thin-film oven test (TFOT) at 163 °C for five hours, followed by extended aging in the 1PAV. The ATR-mode FTIR spectroscopy was utilized, with a wavelength range of 4000–600 cm^{-1}, a resolution of 1 cm^{-1} and 32 scans. Three independent measurements were conducted for each sample, resulting in a total of 48 samples (16 × 3 repetitions) measured at room temperature for this group.

The final section of our input data set was sourced from the works by Tarsi [19] and Mocetti [20] (Table 1S, supporting information). Three types of bituminous binder were employed, including two unmodified binders and one polymer-modified bituminous binder. The two pure binders have penetration grades of 40/60 and 70/100, respectively, both derived from source Q. The polymer-modified bituminous binder was achieved through the addition of SBS polymers into a Q-sourced bituminous binder. The binder films had a diameter of 27.50 mm, and a thickness of 2 mm. Five aging methods were applied to these binders, namely aging at room temperature, oven aging, aging through the Pressure Aging Vessel (PAV), a protocol combining Rolling Thin Film Oven Test (RTFOT) and PAV aging, and moisture aging. Room temperature aging lasting for 5, 15, and 25 days were performed, with an average room temperature of 24.6 °C. Oven aging was conducted at 135 °C for three durations, namely, 60 h, 10 days, and 15 days. The PAV test adhered to standard conditions, aging samples at a temperature of 100 °C and a pressure of 2.1 MPa for 20 h, following the NEN-EN 14769:2005 European Standard. PAV testing was also conducted at the same temperature and pressure but for twice the standard aging time (40 h). The combined short-term and long-term aging procedures involved initial RTFOT aging according to the NEN-EN 12607–1:2014 European Standard [21], followed by PAV aging under the standard condition [22]. For moisture aging, all bituminous binder types underwent two different conditions, i.e., liquid water and water vapor with a RH=88 %, at temperatures of 20 and 40 °C for 5 and 15 days. The FTIR spectrum was obtained in the spectral range between 4,000 and 600 cm^{-1}, with a scanning resolution of 4 cm^{-1}. Five scans were obtained and averaged for each measurement, and four repetitions were conducted for each sample. In total, 216 FTIR spectra were obtained.

These three groups were combined and examined collectively to determine the optimal pre-treatment method. The spectra from all groups were gathered, resulting in a comprehensive data frame comprising 384 spectra.

3.2. Data pre-processing (DP) methods

3.2.1. Baseline correction methods

The BDPs are crucial for minimizing irrelevant variations linked to elevated baselines which are created due to reduction of reflection with decrease of the wavenumbers [9,23]. In this research, nine different data pre-processing techniques for baseline correction (BDP) were chosen from past bituminous binder studies [24] and the pybaselines python library [25]. We specifically picked methods from the pybaselines python library because they align with the inherent characteristics of the binder's FTIR spectra, based on the explanations provided in the library. The descriptions and relevant equations for these BDP methods are provided in Table 2S in supporting information.

Whittaker-smoothing-based (WSB) algorithms, often known as weighted least squares, Penalized Least Squares, or Asymmetric Least Squares, aim to align the baseline with the measured data while penalizing its roughness. The general function minimized to determine the baseline is outlined in Table 2S. Asymmetric Least Squares (asls), Adaptive Iteratively Reweighted Penalized Least Squares (airpls), and Adaptive Smoothness Penalized Least Squares (aspls) fall under the principles of the WSB algorithms. The baseline is determined iteratively through a linear system. This process involves solving for the baseline, updating the weights, solving for the baseline using the updated weights, and repeating these steps until the exit criteria are met. The distinction among WSB algorithms lies in the choice of weights [26].

Additionally, three approaches employing polynomial fitting were employed for baseline correction: regular polynomial (poly), modified polynomial (modpoly), and improved modified polynomial (imodpoly). Firstly, regular polynomial fitting utilizes least-squares polynomial fitting along with selective masking, where a custom weight array is introduced into the fitting function. This array has values set to 0 in peak regions and 1 in baseline regions, facilitating its use for baseline fitting. Secondly, modified polynomial employs thresholding to iteratively fit a polynomial baseline to the data. Thresholding, an iterative method, initially fits the data using traditional least-squares and then sets the next iteration's fit data as the element-wise minimum between the current data and the ongoing fit. Thirdly, improved modified polynomial is an enhancement of the modpoly algorithm for noisy data, incorporating the standard deviation of the residual (data − baseline) during thresholding [26].

Another approach for baseline correction involves using splines (refers to a flexible, piecewise polynomial function), specifically basis splines (B-splines), which are emphasized due to their prevalent use in pybaselines. To regulate the smoothness of the fitting spline, a penalty is introduced on the finite-difference between spline coefficients, resulting in penalized B-splines known as P-splines. P-splines share similarities with Whittaker smoothing; setting the number of basic functions, (M, Table 2S, supporting information), equal to the number of data points, (N, Table 2S, supporting information), and the spline degree to 0 makes the identity matrix, rendering the equation identical to that used for Whittaker smoothing. Consequently, Penalized Spline Asymmetric Least Squares (pspline_asls) and Penalized Spline Asymmetric Least Squares (pspline_airpls) are penalized versions of Asymmetric Least Squares (asls) and Adaptive Iteratively Reweighted Penalized Least Squares (airpls) [26]. The final method is the eight-point baseline correction (8points) proposed by RILEM work [24].

3.2.2. Normalization methods

NDPs can address challenges related to varying sample size or quantity, flaws or limitations intrinsic to the instrument, for instance, low signal intensity or scattering, and variations caused by different mean or standard deviations. In this investigation, the impacts of nine distinct normalization data pre-processing (NDP) techniques on ATR-FTIR spectra were explored. Namely, the NDP methods include Normalization to sum (NTS), Normalization to constant vector length (NCV), Normalization to change the maximum to 1 (NMO) [27], Mean centring (MC), Autoscaling (AS), Pareto scaling (PS), Robust scaling (RS), Standard normal variate (SNV), Multiplicative scatter correction (MSC) [28,29]. Table 3S (supporting information) lists these NDP methods alongside related equations. In this study, the reference spectrum required for MSC is set as the average spectral value of the entire dataset. Regarding the NMO method, the minimum point of the curve in the range 2800–3200 cm^{-1} is set at zero, and then the maximum point is adjusted to 1.

Among the selected NDP methods, normalization to sum (NTS), normalization to constant vector (NCV), and standard normal variate (SNV) are referred to as one-way methods, where each spectrum is pre-processed individually, meaning that normalized results of a spectrum

are determined solely on by its own characteristics. In contrast, the other NDP methods are essentially two-way methods, where the results of normalization are affected by all spectra. Reviews detailing the theoretical foundations of the NDP methods are widely available [29,30]. To prevent incorrect analyses for one versus two-way methods, this study independently pre-processed the test set from the training set [28,31].

3.2.3. Combined normalization and baseline-correction methods

It remains uncertain whether normalization, baseline-correction, or a combination of both is optimal for analysing FTIR spectra in bituminous binder-related studies. When both methods are considered, the common practice in existing FTIR studies in the bituminous binder field involves first applying baseline correction and then normalization to the raw spectra. This approach is to ensure that baseline irregularities do not influence the normalization process, allowing us to focus on normalizing the actual spectral features without being hindered by baseline artifacts [17,24].

In line with literature, the combined pre-processing was achieved by initially conducting baseline correction followed by normalization in this work. Consequently, employing 9 baseline correction methods with 9 normalization correction methods results in 81 combined data pre-processing (CDP) methods. Ultimately, 100 pre-processed datasets were generated from the initial input data frame.

3.3. Transformation of spectra to peak areas and indices

In a FTIR spectrum of a bituminous binder sample, informative peaks describing the carbonyl, sulfoxide, aromaticity, aliphatic, sulfone, long chain, and hydroxyl groups can be observed, with their wavenumber ranges detailed in Table 1. Two common approaches employed in pavement engineering and bituminous binder studies, peak area calculation and index calculation, were used to characterize these functional groups following the pre-processing of the input data. Two peak area calculation methods were utilized, i.e., area-to-base, and tangential area. For index calculation, Equation (1) was applied for each functional group, using the vertical limits outlined in Table 1. Four distinct types of indices were derived by varying the methods for peak area calculation and the selection of reference area. The definitions and applications of these indices and peak areas are presented in Table 4S (supporting information).

The computation of indices followed this equation:

$$index = A_x / A_{ref} \quad (1)$$

The symbol A_x denotes the peak area under the curve within specific ranges outlined in Table 1. A_{ref} represents either the sum of all peak areas (A_{TOTAL}) or exclusively the aliphatic peak areas (A_{ALI}).

The calculation can be performed in two distinct manners: either as peak area to the baseline (A_B) or tangential peak area (A_T), as illustrated in Fig. 1S (supporting information).

The methods for calculating peak areas and indices in ATR-FTIR spectra analysis are varied and defined by specific criteria in Table 4S. The "Peak area to base" method (A_B) calculates the peak area underneath the curve and above the x-axis in the specified range, without computing any indices. The "Tangential peak area" method (A_T) also calculates the peak area underneath the curve but above a tangential line in the specified range, similarly without indices. The "Peak area to base-all peak area" method (A_B / A_{TOTAL}) involves calculating the peak area to base and then dividing each calculated peak area by the summation of all the calculated peak areas (A_{TOTAL}) to form indices. The "Peak area to base-aliphatic peak areas" method (A_B / A_{ALI}) follows a similar approach but the indices are formed by dividing each calculated peak area by the summation of all the calculated peak areas specifically for aliphatic groups (A_{ALI}). The "Tangential peak area-all peak areas" method (A_T/ A_{TOTAL}) calculates the tangential peak area and then divides each by the summation of all peak areas to create indices. Lastly, the "Tangential peak area-aliphatic peak areas" method (A_T/ A_{ALI}) calculates the tangential peak area and forms indices by dividing each peak area by the summation of all peak areas for aliphatic groups (A aliphatic). The details of peak areas and indices calculation are presented in Table 4S.

3.4. Partial least squares-discriminant analysis

Three different approaches for assessing DP methods are recognized: spectrum-comparison, clustering of samples projected on a PCA score plot, and modelling accuracy/error [33]. The spectrum-comparison method is subjective, requiring laborious work, and is qualitative, while clustering of samples on a PCA score plot is also subjective but somewhat quantitative. Evaluating modelling accuracy/error is goal-oriented and demands time and computational resources. The accuracy assessment can be performed using chemometric approaches like Partial Least Squares Discriminant Analysis (PLSDA) [9,34].

Over the last two decades, the PLSDA has become widely acknowledged and extensively utilized in previous research [35–37]. Theoretically, PLSDA combines two processes – dimensionality reduction and discriminant analysis – into a single algorithm, which is particularly effective for handling complex, high-dimensional (HD) data [38]. Importantly, PLSDA does not assume that the data follows a specific distribution, making it more adaptable compared to other discriminant algorithms, such as Fisher's linear discriminant analysis. Spectral data usually have high dimensionality, and the variables (like wavenumbers) are often interconnected (collinear). As a result, traditional discriminant methods like LDA are not suitable for spectral data [34,39]. Additionally, while PCA could address the challenges associated with LDA in HD data, PLSDA has frequently exhibited superior performance compared to PCA-LDA [40,41]. The combination of spectral data and PLSDA has exhibited notable success across various domains, such as food analysis [42] and forensic science [43]. Therefore, in this study, PLSDA will be employed to assess the impact of different pre-processing methods.

PLSDA relies on PLS multivariate calibration, enabling the simultaneous decomposition of matrix X(n, m) (indices/spectral data-frame) containing n samples and m variables, and a vector y, which denotes the numerical labelling of each sample based on its class, such as fresh or aged bituminous binder. The introduction of new axes termed latent variables establishes a coordinate system that captures the most valuable information from the original variables. Equations 1S and 2S (supporting information) describe the decomposition of X and y into this novel coordinate system [44].

The decision rule employs a fixed point-based approach, specifically considering only the model's predicted values for the response variable y and incorporating a single fixed point, i.e., the cut-off point set at 0.5.

Table 1
Main functional groups of bituminous binder in FTIR spectra with their respective vertical bands and their molecular information [32].

Peak area	Vertical band limit(cm^{-1})	Functional groups
A810	710–734	Hydrocarbon chain, $(CH_2)_n$, C–H in isolated/two/
	734–783	four adjacent hydrogen aromatic rings or C–CH_2
	783–833	rocking in alkyl side chains with more than four
	833–912	carbons
A1030	984–1047	Oxygenated function-sulfoxide, S=O
A1200	1100–1180	Tertiary alcohol C–C–O, C–O in carboxylic acid,
	1280–1330	C–C–C in diaryl ketones, C–N secondary amides, O=S=O in sulfone
A1376	1350–1395	Branched aliphatic structures bending, CH_3
A1460	1395–1525	Aliphatic structures bending, CH_3 and CH_2
A1600	1535–1670	Aromatic structure, C=C
A1700	1660–1800	Oxygenated function-carbonyl, C=O
A2953	2820–2880	Aliphatic structures, Symmetric, Asymmetric
	2880–2990	stretching, CH
A3400	3100–3800	Hydroxyl stretching, OH, NH

A_{TOTAL} = A 810 + A1030 + A1376 + A1460 + A1600 + A 1700 + A2953 + A 3400
A_{ALI} = A1376 + A1460

3.5. Partial least squares-discriminant analysis model validation

In practical terms, the selection of a model validation method is influenced by the dataset's size. For smaller datasets, internal validation methods like cross-validation (CV) are often preferred [34]. Additionally, in v-fold CV, reducing the value of v increases the associated bias but reduces both variance and computational load [34]. Hence, in this study, a 10-fold CV approach was employed. Despite its computational expense, it minimized the risk of bias compared to 5 or 7-fold CVs.

In this investigation, a 10-fold cross-validation strategy was applied to conduct PLSDA. This approach entails dividing the dataset into 10 equal segments, employing 9 segments for model training and reserving one segment for testing. The process iterated 10 times, using a different segment for testing each time, and the accuracy outcomes were averaged to yield a robust evaluation of the model's effectiveness. Given the uneven distribution of 282 aged samples and 102 fresh samples in the input data-frame, a potential bias could arise from simple random sampling [28]. To address this, stratified k-fold cross-validation was employed instead of the standard random k-fold method. This ensures that the class distribution in each data split aligns with the distribution in the complete training dataset and the target variable (y), thus controlling the sampling process [45].

In multilabel classification, the accuracy score provides the subset accuracy. A subset accuracy of 1.0 is achieved when the entire set of predicted labels for a sample exactly matches the true set of labels; otherwise, it is 0.0. The accuracy score is calculated based on Equation 3S (supporting information). The accuracy values derived from this method were utilized to assess the suitability of various pre-processing techniques and their combinations in distinguishing between binary classifications.

3.6. Variable importance in projection scores (VIP scores)

VIP scores in PLSDA quantify the importance of each predictor variable in distinguishing between predefined classes [46]. These scores are calculated by evaluating the contribution of each variable to the PLS components and the explained variance in the classification model, involving a weighted sum of the PLS weights, adjusted for the discriminative power of each variable. Specifically, the calculation involves a weighted sum of squared correlations between the PLSDA components and the original variable, with the weights representing the percentage of variation explained by the respective PLSDA component. The formula for calculating the VIP score for a variable j can be expressed as Equation 4S (supporting information).

Variables with VIP scores greater than 1 are typically considered significant contributors to the classification model, while those with scores less than 1 are deemed less influential [47]. In some studies, threshold values of 1 and 0.7 are used to categorize features into highly, moderately, and lowly important groups. In summary, VIP scores in PLSDA provide a valuable measure of variable importance, highlighting which predictors are most influential in the classification process, and were computed for the preprocessing methods demonstrating the highest accuracy in each section, particularly for aging classification.

3.7. Hierarchical cluster analysis (HCA)

HCA utilizes two main strategies: agglomerative and divisive. In the agglomerative approach, each sample starts as an independent cluster. The algorithm progressively merges pairs of clusters based on a predefined metric describing sample distance (commonly Euclidean, Mahalanobis, or Manhattan distance) and the selected linkage criterion (single, complete, average, and Ward's linkage). These linkage criteria offer different ways to define the distance between clusters, contributing to cluster formation. In this study, we adopted the agglomerative approach, employing Euclidean distance and Ward's linkage criterion. Agglomerative approach is often computationally more efficient and allows for the flexibility of choosing different linkage methods. Ward's linkage minimizes variance within each cluster and is chosen for its effectiveness in optimizing a specific target function. HCA results are typically depicted as dendrograms, providing a visual representation of sample organization within a hierarchical tree-like structure [48].

In a dendrogram (Fig. 2S, supporting information), each branch is referred to as a clade, and the end of each clade is called a leaf. The arrangement of these clades communicates the level of similarity among individual leaves. The point where branches intersect indicates the extent of likeness or dissimilarity, with higher intersections signifying more significant differences. Dendrograms can be interpreted in two ways: firstly, concerning broad-scale groupings, starting from the top and emphasizing high-level branch points (forming cluster X and cluster Y). Secondly, to discern which specific components are most similar, reading from the bottom and identifying the earliest converging clades as we move upwards. The length of the vertical lines in the dendrogram reflects the degree of divergence between branches, where longer lines denote greater distinctions. The horizontal alignment of dendrograms is not crucial, seeing it as a flexible structure where the arms may shift while maintaining consistent vertical height and subgroup arrangement [49].

4. Methodological approach

In this section, we describe the analysis steps undertaken in this study, illustrated in Fig. 1.

Initially, pre-processing was applied to a data frame containing 384 raw FTIR spectra measured at 3400 wavenumbers, resulting in 100 pre-processed data frames of size 384 × 3400. These frames included 9 normalized, 9 baseline-corrected, 81 combined-methods (BDPs + NDPs), and 1 raw data frame. For each pre-processed frame, 6 new frames of indices and peak areas (Table 4S) were generated, yielding a total of 600 data frames of size 384 × 14. Additionally, the entire spectra and their first derivatives were considered, resulting in 200 frames of size 384 × 3400 after applying the Savitzky–Golay (SG) method for derivative calculation.

Subsequently, for PLSDA analysis (Fig. 1), 600 frames of size 384 × 14 and 200 frames of size 384 × 3400 were considered. Two classification systems, aging and rejuvenation classification, were employed, and results were presented separately. Our primary focus has been on aging classification. We consider it to be a more complex task than identifying or classifying other materials like rejuvenators or polymers, based on the assumption that aging introduces fewer alterations to FTIR spectra compared to the introduction of new substances. The idea is to optimize the selection of DP methods initially for the more challenging scenario, namely, the classification of fresh and ageing states. Subsequently, we extend this comparison of DP methods to other simpler cases, such as the classification of rejuvenation. This is to test if the effectiveness of DP methods observed in the case of ageing remains consistent or varies for other classification objectives. To evaluate the suitability of the optimized DP methods for different types of classifications, the complete dataset was divided into two separate groups. For aging classification, the dataset was split into fresh and aged samples of various binder types. For rejuvenation classification, the dataset was divided into two parts: one containing FTIR spectra of both fresh and aged binders with rejuvenators, and the other comprising binders without any rejuvenators, including both fresh and aged samples. For each classification, the following steps were undertaken and discussed individually, and subsequently, the results were compared.

Each data frame was divided into 10 parts, with 9 parts used for PLSDA training and 1 part reserved for testing. This process iterated 10 times, using a different segment for testing each time, and accuracy outcomes were averaged and presented in the result and discussion section. Subsequently, HCA was conducted on the accuracies of classification for DP modified spectra's indices and peak areas. This was undertaken to assess the similarities and differences among these methods

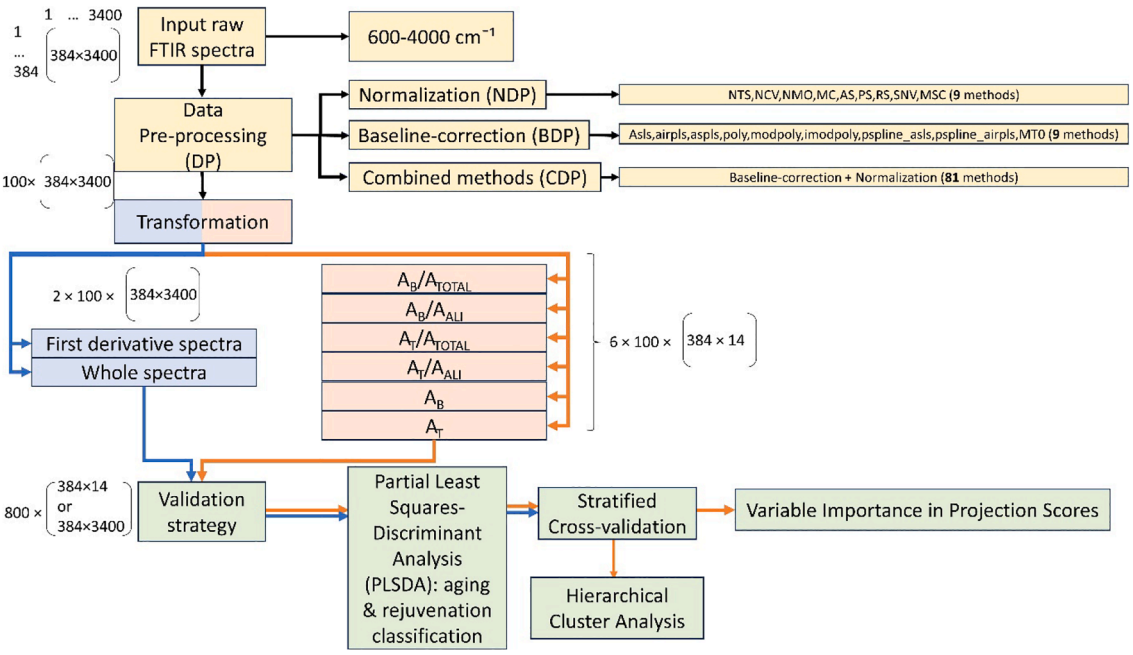


Fig. 1. The schematic diagram depicts the sequential steps involved in the pre-processing and post-processing analysis conducted in this study. It offers a comprehensive overview of each stage, including details about the size and quantity of the data frames considered. The process begins with reading the data, applying pre-processing methods, and transforming the pre-processed data into area, indices, entire spectra, or first derivative spectra. This is followed by applying PLSDA classification with stratified cross-validation and HCA. Additionally, VIP scores were analyzed.

for calculating peak areas and indices.

In the Results and Discussion section, classification accuracies for NDP and BDP modified peak areas/indices were visualized in a heatmap with an HCA clustering (cluster map). Classification accuracies for combined methods and entire spectra with their first derivatives were presented in separate tables. For methods with the highest accuracy, VIP scores, which are indicators of the most influential predictors in the classification process, were obtained to identify important regions, especially for aging classification.

The PLSDA and HCA implementation used Python programming libraries, and the entire statistical analysis process is outlined in Fig. 1. The Python code for the entire procedure, covering normalization, baseline correction, combined methods, and PLSDA classification, will be made available upon request from the authors.

5. Results and discussion

5.1. IR spectra

The complete dataset utilized in this study is compiled from three distinct investigations, detailed in the materials and methods section. Fig. 2 illustrates the spectra for all samples, measured in the wavenumber range of 600 to 4000 cm^{-1} through ATR-FTIR spectroscopy. While the spectra in Fig. 2 exhibit similarities, characterized by peaks in the regions specified in Table 1, variations in absorbance intensity are evident due to differences in binder type, source, and aging level. This difference is particularly noticeable in the carbonyl and hydroxyl regions upon visual examination. The differently coloured spectrum groups consist of both fresh samples and their corresponding aged counterparts, as explained earlier. Additionally, all spectra display a slope in the lower wavelength region, though the slope varies across groups, with the last part of the data showing the highest rise. It is important to emphasize that these observations are solely based on visual inspection of the spectra. The classification potential of the spectra will be further assessed through PLSDA modelling, as discussed in the subsequent sections.

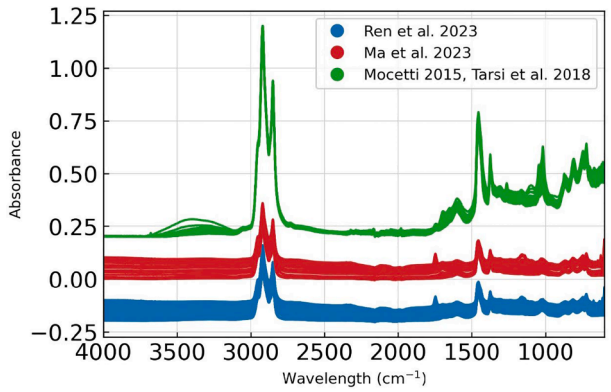


Fig. 2. ATR-FTIR spectra derived from three distinct studies: the initial section sourced from [18], the second portion from [12], and the final segment obtained from [19,20]. Details of each study are presented in Table 1S. For better visualization, the data from the Ren et al. study and the Mocetti and Tarsi studies are shifted by −0.25 and +0.25 in their absorbance, respectively.

5.2. Effect of DPs on aging classification of areas and indices

This section examines the effects of nine normalization methods, nine baseline correction methods, and 81 combinations of both methods in comparison to the raw data through binary classification of fresh and aged samples using PLSDA.

5.2.1. Effect of different baseline correction methods on ageing classification of peak areas/indices

Fig. 3 illustrates the impact of 9 baseline correction methods on a specific spectrum. The baseline-corrected spectra exhibit variations depending on the correction formula employed in Table 2S in supporting information. The disparity between the corrected and original spectra is more noticeable for certain methods.

To evaluate the impact of different baseline correction methods on aging classification accuracy, PLSDA and HCA were applied to various

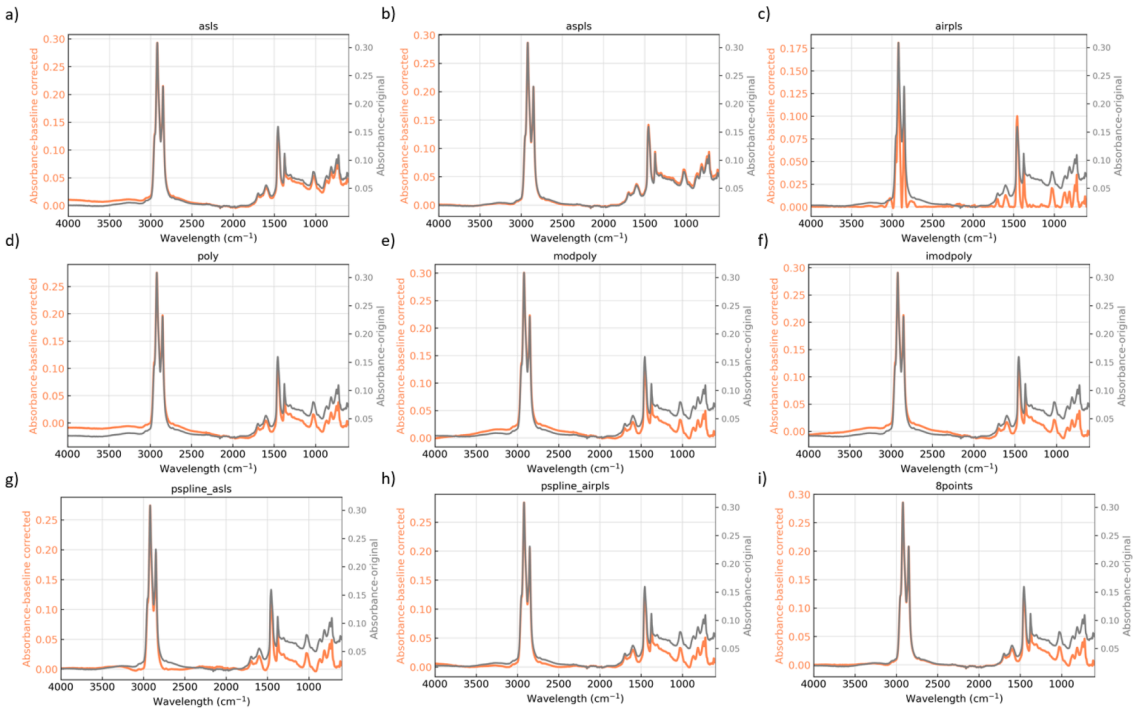


Fig. 3. FTIR spectra before (Gray spectra, right vertical axis) and after (coral spectra, left vertical axis) the implementation of BDPs: a) asls, b) aspls, c) airpls, d) poly, e) modpoly, f) imodpoly, g) pspline_asls, h) pspline_airpls, i) 8points.

FTIR-related peak areas and indices. Classification accuracies and clustering of peak areas and indices were presented in Fig. 4.

FTIR-related peak areas and indices, calculated as per Table 4S, showed both positive and negative influences after applying BDP methods. HCA clustering indicated that peak areas and indices respond differently to BDPs. A_{ALI}-based methods are more closely grouped compared to A_{TOTAL}-based indices. A_B and A_T varied in accuracy after baseline corrections, ranging from 0.70 to 0.83 for A_B and from 0.71 to 0.82 for A_T. Indices were less affected due to internal normalization by A_{ALI} or A_{TOTAL}.

Among baseline correction methods, asls and aspls significantly impacted classification accuracies, with asls showing the lowest and aspls the highest performance. Asls may not adapt well to local variations in classification tasks due to its emphasis on handling asymmetric errors in regression [50], while aspls's penalty term for smoothing coefficients effectively reduces noise, enhancing classification accuracy by improving the signal-to-noise ratio in the data [51 52]. This aligns well with previous findings that the importance of noise reduction in spectroscopic data analysis were emphasized [53,54].

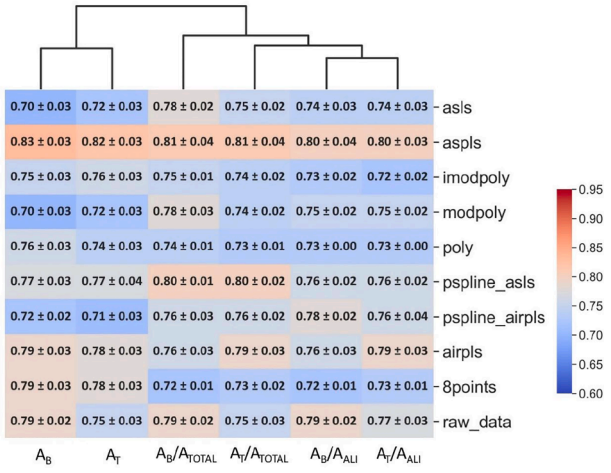


Fig. 4. Cluster map illustrating aging classification accuracies associated with BDPs and the HCA dendrogram. Each column displays accuracies for various BDPs concerning a specific peak area or index calculation method, while each row represents the classification accuracies for a distinct BDP method but with different approaches to peak area and index calculations.

Polynomial fitting and 8points methods also negatively affected classification accuracies. Modpoly [55] and imodpoly [56], despite being advanced methods, sometimes reduced accuracy due to overfitting [57,58]. In our scenario, where data was collected from different studies with varying baseline characteristics, simpler methods outperform the modified versions. This suggests that simpler baseline correction methods might be more effective since data-expantion is needed for complicated models to fine-tune the hyper-parameters [58]. The pspline_airpls method performed worse than airpls, which adapts better to diverse baseline patterns [59]. The 8points method notably reduced accuracies, particularly for indices, probably due to linear shifts in spectra causing information loss. While indices calculation stabilizes datasets, it can reduce valuable information for specific baseline-correction methods before classification.

In conclusion, only aspls consistently improved classification accuracy for all peak area or index calculations, mainly by effectively eliminating noise from the spectra. The ability of aspls to enhance signal quality suggests its potential application in other spectroscopic studies requiring high classification accuracy.

5.2.2. Effect of different normalization methods on aging classification of peak areas/indices

Fig. 5 illustrates the impact of 9 normalization methods on a group of 20 spectra. Each normalization method uniquely alters the raw spectra, as indicated by the formula outlined in Table 3S in supporting information. The application of the nine normalization methods to the FTIR spectra in Fig. 5 results in spectra that retain the same overall shape but exhibit differences in absorbance values and relative positions. Although the peaks and their positions within each spectrum remain consistent, different normalization methods alter the relative positioning and

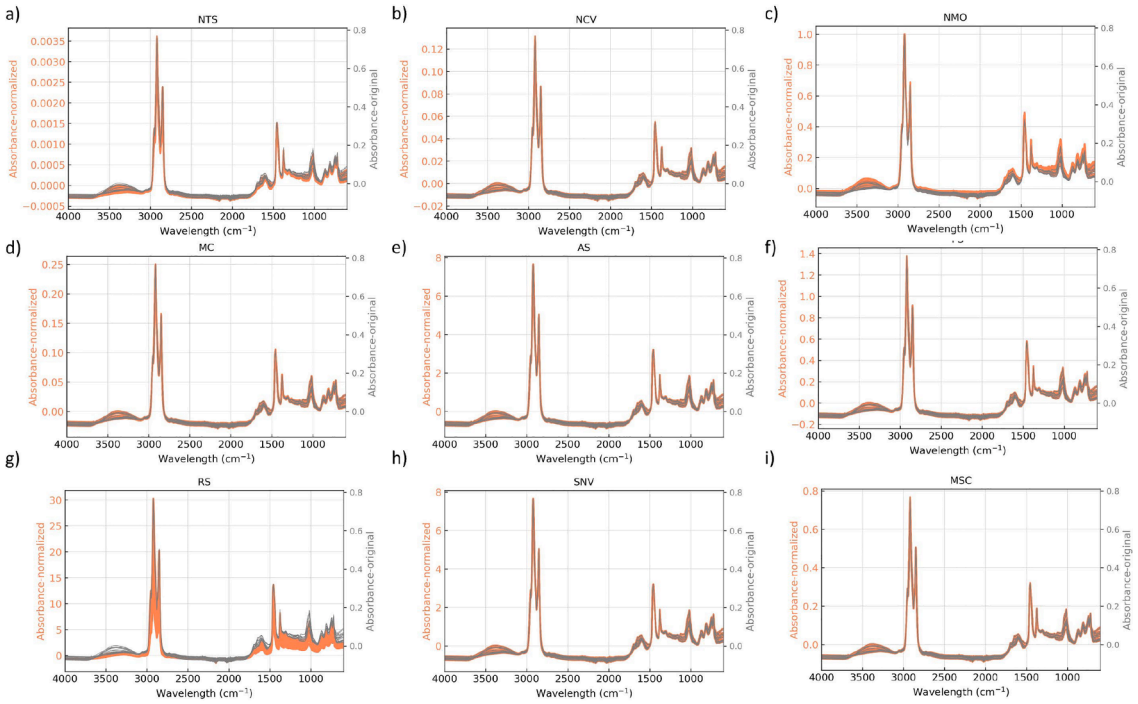


Fig. 5. FTIR spectra of 20 samples before (Gray spectra, right vertical axis) and after (coral spectra, left vertical axis) the implementation of NDPs: a) NTS, b) NCV, c) NMO, d) MC, e) AS, f) PS, g) RS, h) SNV, and i) MSC.

spacing of the spectra in relation to each other. Consequently, each normalization technique modifies the variation within the dataset.

Fig. 6 presents a cluster map and HCA dendrogram showing the classification accuracies of aging states using index and peak area parameters. The HCA analysis reveals two distinct clusters for A_T-based and A_B-based methods. Indices using A_{ALI} as the reference are more distant than those with A_{TOTAL} or peak areas. Classification accuracy for raw data ranges from 0.75 to 0.79, with A_B and A_B-based indices consistently yielding higher accuracies than A_T and A_T-related indices. This indicates the robustness of A_B-based methods in capturing relevant features for classification aligns well with previous findings [2].

Among normalization methods, NDP is ineffective for A_T/A_{ALI}, while

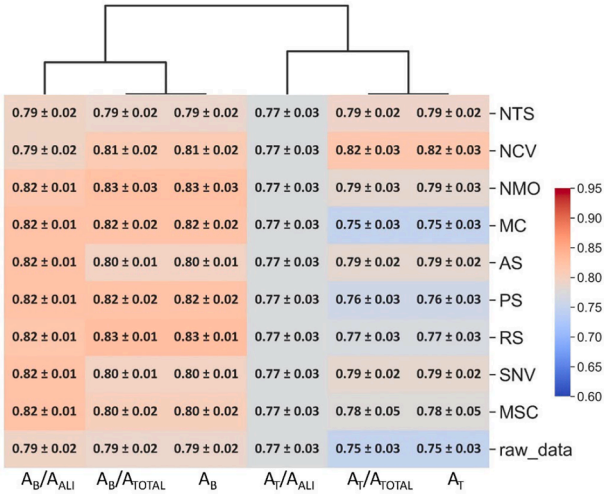


Fig. 6. Cluster map illustrating aging classification accuracies associated with NDPs and HCA dendrogram. Each column displays accuracies for various NDPs concerning a specific peak area or index calculation method, while each row represents the classification accuracies for a distinct NDP method but with different approaches to peak area and index calculations.

NCV is most effective for A_T and A_T/A_{TOTAL}, increasing classification accuracy from 0.75 to 0.82 by normalizing vector length [17]. For A_B and A_B-based indices, NMO and RS show the highest improvement. NMO preserves proportionality between features, and RS standardizes data, reducing sensitivity to outliers [60]. Outlier detection was not performed as all data points were integral and previous analyses showed no unusual findings [12,18,19], highlighting that extreme points impact full peak area classifications more significantly.

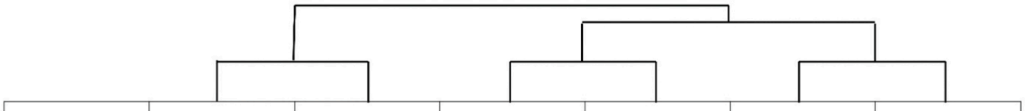
Integrating MC with A_T-based calculations (A_T, A_T/A_{TOTAL}, and A_T/A_{ALI}) shows no improvement in classification accuracy, consistent with previous studies [28]. Applying NTS to A_B-based calculations also proves ineffective. This suggests that eliminating the mean effect (MC) or focusing on relative contributions (NTS) is less effective than preserving absolute values (NMO) or standardizing data (RS).

In summary, the superior performance of NMO and RS for A_B and A_B/A_{TOTAL} index, and NCV for A_T and A_T/A_{TOTAL} index for enhancing classification accuracy, provides valuable insights for future research and practical applications in aging studies. These findings provide a foundation for optimizing data preprocessing techniques in similar spectroscopic analyses.

5.2.3. Effect of combined BDP and NDP methods on aging classification of peak areas/indices

To evaluate the combined impact of baseline correction and normalization, the original data frame was transformed into 81 modified data frames using combinations of NDP and BDP methods. Peak areas and indices were computed for each modified data frame, resulting in 486 accuracy values from PLSDA analysis, with their standard deviations from a 10-fold CV. These values are provided in the supporting information, Table 5S. This paper focuses on the highest, lowest, and raw data accuracy values for each FTIR parameter calculation method, as shown in Table 4S. Additionally, an HCA dendrogram of all 486 values is presented in Table 2, revealing three primary groups: one for peak area variables (A_B and A_T), another for aliphatic-based indices ($A_{B,T}/A_{ALI}$), and the last for summation of all peak area-based indices ($A_{B,T}/A_{TOTAL}$). This illustrates the unique impact of peak area or index calculation methods on the effectiveness of different DP methods.

Table 2
Upper and lower aging classification accuracy limits for the combination of NDPs and BDPs are provided, alongside raw data accuracies. The HCA dendrogram for the entire accuracy dataset (Table 5S) is included for the clustering analysis of peak areas and indices.



Accuracy rank	A_B/A_{ALI}	A_T/A_{ALI}	A_T/A_{TOTAL}	A_B/A_{TOTAL}	A_B	A_T
Max	aspls_NMO 0.81 ± 0.02	aspls_NMO 0.80 ± 0.03	aspls_NMO 0.81 ± 0.02	aspls_NMO 0.81 ± 0.02	aspls_NMO 0.83 ± 0.03	aspls_NMO 0.83 ± 0.03
Raw data	0.79 ± 0.02	0.77 ± 0.03	0.75 ± 0.03	0.79 ± 0.02	0.79 ± 0.02	0.75 ± 0.03
Min	asls_MC 0.71 ± 0.01	asls_MC 0.72 ± 0.01	8points_NMO 0.72 ± 0.02	8points_NMO 0.72 ± 0.01	modpoly_MSC 0.73 ± 0.04	modpoly_MSC 0.70 ± 0.06

The accuracy values indicate that the application of DP methods does not consistently enhance post-processing quality. The belief that randomly selecting DP methods or using multiple ones can be beneficial is disproved, as some DP methods result in lower accuracy than raw data. However, certain combinations can positively impact PLSDA classification. For example, combining aspls with NMO for A_B and A_T enhances accuracies to levels exceeding 0.80, although similar high accuracies are achieved by using either NMO or aspls alone, indicating no additional benefit from the combination.

Conversely, the lowest accuracies in each HCA cluster are from different combined methods. For datasets based on peak areas (A_B or A_T), the modpoly_MSC combination shows the lowest accuracy. MSC normalizes variations across the entire spectrum, which, while eliminating scatter effects, may diminish variations crucial for class discrimination in aging classification. Modpoly and imodpoly, designed to eliminate baseline variations, might lose discriminatory information when followed by MSC. Additionally, combining DP methods can result in an overly complex model, leading to overfitting and loss of important classification information.

For indices based on A_{ALI}, the asls_MC combination results in the lowest accuracy. Individually, asls decreased accuracy, and MC was either ineffective or slightly improved accuracy, but their combination significantly reduced accuracies. For indices based on A_{TOTAL}, the 8points_NMO method exhibited the lowest accuracy. The 8points method performed poorly individually, and its combination with NMO led to reduced final accuracy. These findings highlight that the negative impact of baseline correction methods can outweigh the positive effect of normalization methods.

To sum up, the impact of combining BDPs and NDPs on classification depends on the specific techniques used and can either enhance or weaken post-processing quality. It is advisable to use appropriate NDPs, such as NMO and RS for A_B and A_B/A_{TOTAL} TAL, and NCV for A_T and A_T/A_{TOTAL}, or BDPs like aspls, rather than combined methods, to minimize unnecessary computational expenses.

5.3. *Variable importance in projection scores (VIP scores) for aging classification of peak areas/indices*

To determine the influential regions in bituminous binder spectra affecting PLSDA classification of aging, VIP scores for each peak region were calculated. VIP scores, which indicate the contribution of various features to the PLSDA modeling, were categorized into highly, moderately, and lowly important groups using thresholds of 1 and 0.7 [47]. This comparative analysis identified key regions for aging classification. Fig. 7 shows VIP scores derived from classification using all peak areas and indices calculated from raw spectral data, revealing significant variability in scores based on the calculation method and region selection. Notably, the carbonyl region consistently attained the highest VIP score for aging classification when analyzed using A_B, A_T, and the A_{ALI}-based indices. High VIP scores were also observed in the hydroxyl and fingerprint regions for indices-based classification, whereas the sulfoxide region demonstrated lower discriminatory power, particularly for A_B/A_{ALI} and A_T/A_{ALI} indices.

Fig. 8 illustrates VIP scores for each peak region after applying DP methods, using classification scenarios with the highest accuracy: NMO and RS for A_B and A_B/A_{TOTAL}, NCV for A_T and A_T/A_{TOTAL}, and aspls for both A_B and A_T. For NMO-based A_B (Fig. 8 a), the regions 1280–1330, 1350–1395, and 1535–1670 cm^{-1} exhibit the highest VIP scores,

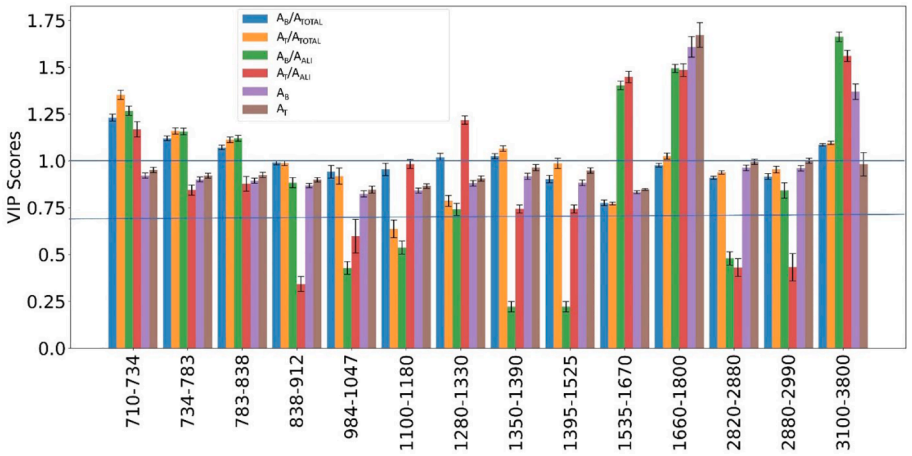


Fig. 7. VIP scores of different regions in the PLSDA classification of ageing on peak areas and indices of raw spectral. This study employs threshold values of 1 and 0.7 to categorize features into highly, moderately, and lowly important groups.

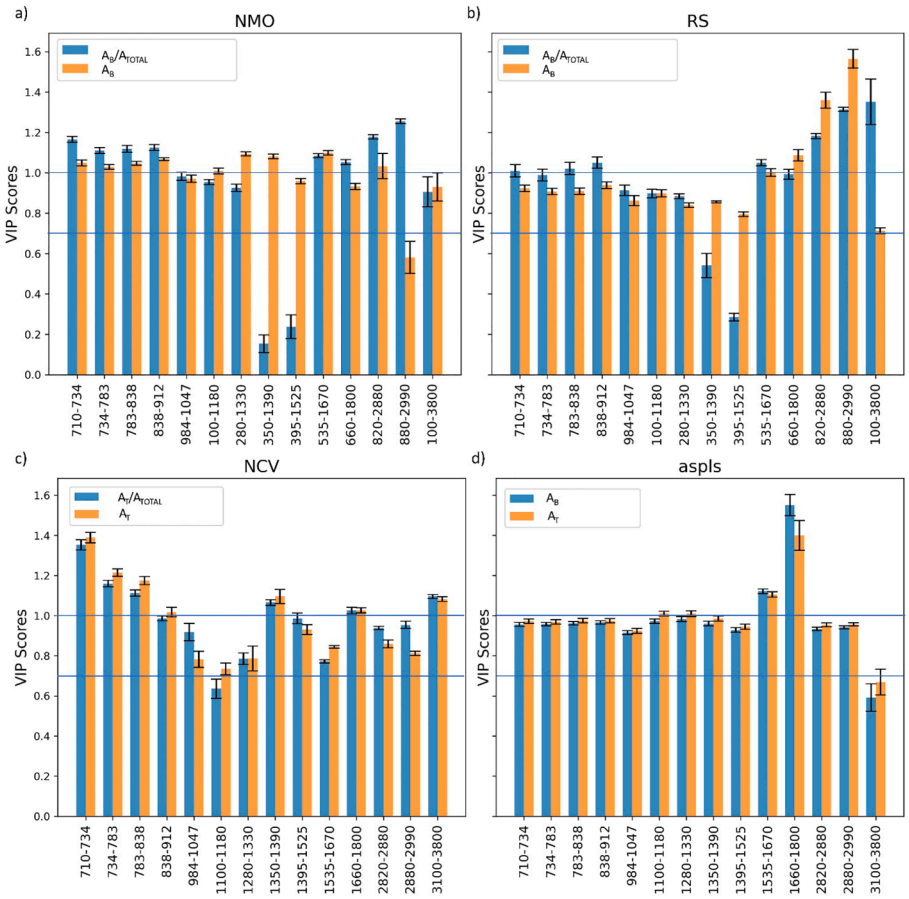


Fig. 8. VIP scores of different regions in the PLSDA classification of ageing on, a) NMO-modified A_B and A_B/A_{TOTAL} indices, b) RS-modified A_B and A_B/A_{TOTAL} indices, C) NCV-modified A_T and A_T/A_{TOTAL}, d) aspls-modified A_B and A_T. This study employs threshold values of 1 and 0.7 to categorize features into highly, moderately, and lowly important groups.

indicating their significant contribution to aging classification. These regions correspond to various functional groups (Table 1), including branched aliphatic structures (CH_3) and aromatic structures (C=C bonds), which play crucial roles in bituminous binder's aging behaviour [1]. The regions 710–912, 1100–1180, and 2820–2880 cm^{-1} also show high VIP scores, while the 2880–2990 cm^{-1} region has the least impact. Other regions, such as sulfoxide, carbonyl, and hydroxyl, have moderate influence.

For NMO-based A_B/A_{TOTAL} indices, the regions 2820–2880 and 2880–2990 cm^{-1} (aliphatic structure stretching) demonstrate the highest scores, followed by the fingerprint region (710–912 cm^{-1}) and aromatic and carbonyl regions. Conversely, the regions 1350–1390 and 1395–1525 cm^{-1} (aliphatic structure bending) exhibit the lowest contribution. This difference highlights the sensitivity of stretching vibrations to molecular alterations due to aging processes like oxidation. Stretching vibrations involve changes in bond length and are more responsive to these modifications, whereas bending vibrations, which entail changes in bond angles, are generally less influenced by chemical alterations compared to stretching vibrations [61]. Other regions, such as sulfoxide, sulfones, and hydroxyl regions, hold a moderate impact on aging classification.

In RS-based A_B (Fig. 8 b), the regions 2820–2880 and 2880–2990 cm^{-1} show the highest VIP scores, with aromatic and carbonyl regions also presenting high scores. The hydroxyl region shows the least contribution, with other regions having moderate impact. Similar to NMO patterns are observed for RS-based A_B/A_{TOTAL} indices. For A_T/A_{TOTAL} based on NCV-modified spectra (Fig. 8 c), the fingerprint region exhibits the highest VIP score, followed by the segments 1350–1390 cm^{-1}, hydroxyl, and carbonyl regions. The sulfone region shows the lowest contribution, while other regions such as sulfoxide, aromatic, 2820–2990, and 1280–1330 cm^{-1} have moderate impact. Similar results were obtained for NCV-modified A_T.

For aspls-modified A_B and A_T (Fig. 8 d), only the carbonyl and aromatic regions have VIP scores above 1, demonstrating their significant contribution, while the hydroxyl region shows the lowest contribution. Other regions, including sulfoxide, sulfone, fingerprint, and 2820–2990 and 1280–1525 cm^{-1}, have moderate VIP scores. This finding aligns with the VIP scores observed for raw data. Thus, it can be inferred that implementing the aspls modification has the potential to enhance the accuracy of peak area-based classification while maintaining the region's importance similar to the original data. The VIP scores for A_B and A_B/A_{TOTAL} indices indicate that calculating indices diminishes the significance of the aliphatic region (1350–1525 cm^{-1}) compared to A_B calculation. For A_T and A_T/A_{TOTAL} indices, there is no observable difference in scores for all spectral regions. Normalization methods and index calculations can alter the significance of various regions, highlighting the importance of methodological choices in aging classification.

In summary, the key regions for aging classification depend on the chosen normalization or baseline correction method, the approach to peak area or index calculation, and the underlying chemistry of aging. While data analysis provides initial insights, a comprehensive evaluation requires integrating both data-driven analysis and fundamental chemical knowledge.

5.4. Effect of DPs on entire and first derivative spectra for aging classification

As previously mentioned in the introduction, research of bituminous binders using chemometrics often concentrates on entire spectra or their first derivatives. Consequently, this work investigated the impact of DP methods on the classification utilizing both the entire spectra and its first derivative (Fig. 3S, supporting information).

Table 3 lists the classification accuracies for the entire spectra with and without DPs. Results show that the accuracy of 80 % for the entire raw spectra is nearly identical to the value of 79 % when using peak area or index parameters calculated based on raw spectra, as shown in Table 4S. However, the influence of DP methods is more pronounced for the entire spectra compared to the classification based on peak area or index.

The baseline correction methods applied to the entire spectra did not significantly improve PLSDA performance, likely due to the minimal impact of baseline distortions on the overall spectral shape and pattern. Only the pspline_asls method showed a slight increase in accuracy for the entire spectra by correcting baseline distortions and improving peak measurement accuracy [62]. This method's improvement in spectra quality can enhance classification performance. However, when pspline_asls was used for peak area and index calculation, a lower accuracy was observed, as shown in Fig. 4. This is because pspline_asls focuses on correcting distortions rather than the peak shapes crucial for classification.

For the entire spectra, the NDPs with the highest accuracies were autoscaling (AS), Pareto scaling (PS), and standard normal variate (SNV) methods, which effectively normalize data and reduce variations in absolute intensity levels [63]. Autoscaling and Pareto scaling [64] also minimize noise impact and highlight relevant spectral features by dividing by the standard deviation. However, these methods had less impact on peak area/indices-based classification, as shown in Fig. 6, possibly due to their specific effectiveness in normalizing intensity levels rather than affecting peak regions significantly. When combined DP methods were applied to the entire spectra, the highest accuracy reached 0.87 ± 0.02, comparable to normalization methods alone. However, the poly_MC method decreased PLSDA accuracy for the entire spectra, showing limited individual impact and poorer results when combined.

Table 4 shows classification accuracies for the first derivatives of FTIR spectra affected by DP methods. Baseline correction applied to the entire spectra, or their first derivatives did not significantly enhance PLSDA performance. For first derivative spectra classification, autoscaling (AS), robust scaling (RS), and standard normal variate (SNV) methods again showed the highest accuracies. The poly_MC and asls_MC methods reduced PLSDA accuracy when applied to the entire spectra and its first derivative, respectively, indicating limited individual influence and even poorer results when combined.

In light of these findings, it is advisable to exclusively employ appropriate normalization methods (SNV, AS, PS, and RS) for both entire spectra and their first derivative datasets.

5.5. Effect of DPs on rejuvenation classification

The main objective of the following sections is to evaluate whether the use of DP methods can improve the efficiency of distinguishing between rejuvenated and unrejuvenated binders. The characteristic peaks of the utilized rejuvenators [18,65] (Fig. 4S, supporting information) fall within the same ranges as those of bituminous binder, as specified in Table 1. This observation indicates that none of the rejuvenators introduce new range of consideration to the bituminous binder spectra. In the bituminous binder mixed with rejuvenators spectra, only bio oil exhibits two significant peaks at 1750 and 1160 cm^{-1}. These peaks fall within the relevant range according to Table 1. While this was somewhat expected due to the shared aromatic and aliphatic components between rejuvenators and bituminous binder, only variations in peak shape and area may occur due to peak overlap from both the binder and the added oil.

5.5.1. Effect of different baseline correction methods on rejuvenation classification of peak areas/indices

Fig. 9 illustrates the impact of baseline correction methods on rejuvenation classification and HCA clustering results for different peak area and indices calculation methods.

The HCA dendrogram reveals two clusters: one with A_B, A_T and their A_{ALI}-based indices, and the other with A_{TOTAL}-based indices. These clusters illustrate how BDPs influence accuracy. BDPs negatively impact A_B, A_T, A_{ALI}-based indices, as shown in the right cluster of Fig. 9, while positively affecting A_{TOTAL}-based indices in the left cluster.

For raw data without BDPs, high classification accuracy indicates the original dataset effectively distinguishes between rejuvenated and unrejuvenated binders. BDPs have mixed effects on rejuvenation classification accuracy. All BDPs reduced accuracies for A_B and A_T classifications to 0.88, suggesting that rejuvenated binders possess unique features detectable in raw data. BDPs make the data more uniform, complicating classification, particularly for A_B and A_T −based classifications without additional steps like division by A_{ALI} or A_{TOTAL}.

Baseline corrections like airpls, poly, and 8points decreased classification accuracy, and the aspls method reduced accuracy across all peak area or indices calculations. The aspls method is not beneficial for distinguishing rejuvenated binders, possibly due to constraints applied to the aspls model coefficients. However, other BDPs enhance classifications for A_{TOTAL}-based indices.

To summarize, baseline correction methods varied in effectiveness, often negatively impacting classification accuracies by eliminating authentic spectral features or introducing artifacts that disturbed essential discriminative information.

Table 3
PLSDA aging classification accuracies for the entire spectra in their raw form, after normalization, baseline correction, and using the top three accuracies along with the lowest accuracy achieved through the combination of NDPs and BDPs.

Baseline-correction method	Accuracy	Normalization method	Accuracy	Combined BDP and NDP methods Max accuracy	Accuracy
asls	0.81 ± 0.02	NTS	0.82 ± 0.02	poly_RS	0.87 ± 0.02
aspls	0.80 ± 0.02	NCV	0.83 ± 0.02		
imodpoly	0.82 ± 0.01	NMO	0.84 ± 0.02	pspline_asls_RS	0.87 ± 0.02
modpoly	0.82 ± 0.02	MC	0.82 ± 0.02		
poly	0.78 ± 0.02	**AS**	0.86 ± 0.01	8points_RS	0.87 ± 0.02
pspline_asls	0.84 ± 0.02	**PS**	0.85 ± 0.02		
pspline_airpls	0.81 ± 0.02	RS	0.84 ± 0.02	**Min accuracy**	**Accuracy**
airpls	0.81 ± 0.02	**SNV**	0.86 ± 0.01	poly_MC	0.77 ± 0.02
8points	0.82 ± 0.02	MSC	0.84 ± 0.02		
raw_data: 0.80 ± 0.02					

Table 4
PLSDA aging classification accuracies for first derivative spectra in their raw form, after normalization, baseline correction, and using the top three accuracies along with the lowest accuracy achieved through the combination of NDPs and BDPs.

Baseline-correction method	Accuracy	Normalization method	Accuracy	Combined BDP and NDP methods Max accuracy	Accuracy
asls	0.82 ± 0.02	NTS	0.82 ± 0.02	asls_NTS	0.87 ± 0.02
aspls	0.82 ± 0.02	NCV	0.86 ± 0.02		
imodpoly	0.83 ± 0.02	NMO	0.86 ± 0.02	asls_NCV	0.87 ± 0.02
modpoly	0.83 ± 0.02	MC	0.82 ± 0.02		
poly	0.82 ± 0.02	**AS**	0.87 ± 0.02	modpoly_NTS	0.87 ± 0.02
pspline_asls	0.82 ± 0.02	PS	0.83 ± 0.02		
pspline_airpls	0.82 ± 0.02	**RS**	0.87 ± 0.02	Min accuracy	Accuracy
airpls	0.82 ± 0.02	**SNV**	0.87 ± 0.02	asls_MC	0.82 ± 0.02
8points	0.82 ± 0.02	MSC	0.86 ± 0.02		
raw data: 0.82 ± 0.02					

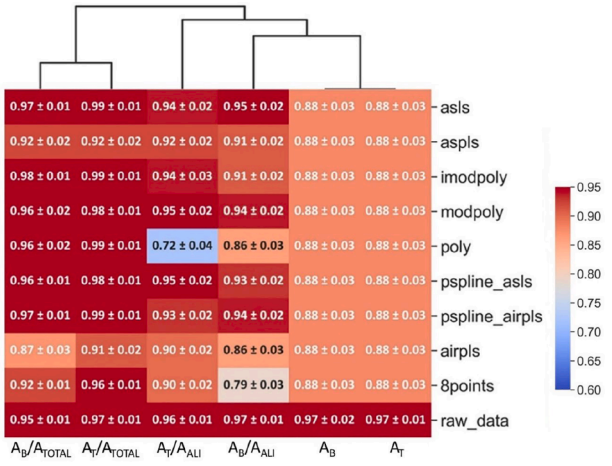


Fig. 9. Cluster map illustrating rejuvenation PLSDA classification accuracies associated with BDPs and the clustering of peak area or indices methods. Each column displays accuracies for various BDPs concerning a specific peak area or index calculation method, while each row represents the classification accuracies for a distinct BDP method but with different approaches to peak area and index calculations.

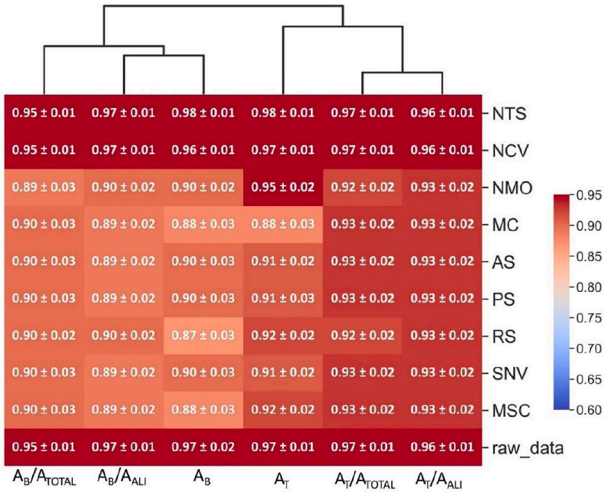


Fig. 10. Cluster map illustrating rejuvenation classification accuracies associated with NDPs and clustering of peak area or indices methods. Each column displays accuracies for various NDPs concerning a specific peak area or index calculation method, while each row represents the classification accuracies for a distinct NDP method but with different approaches to peak area and index calculations.

5.5.2. Effect of different normalization methods on rejuvenation classification of peak areas/indices

The classification accuracies for the PLSDA classification of rejuvenation conditions using normalized data and the clustering based on HCA analysis are illustrated in Fig. 10. The HCA dendrogram yields two main clusters. One cluster contains A_B and A_B-based indices, while the second group involve A_T calculations.

Most normalization methods do not enhance classification accuracy compared to raw spectral data, and many even reduce it. Exceptions are the NTS and NCV methods, which achieve the same accuracy as raw data. These methods address normalization by focusing on proportionality and Euclidean length, crucial for identifying rejuvenation oils in binders, thereby facilitating comparison between rejuvenated and unrejuvenated groups. Given the dataset, normalization methods are not advisable for classifying rejuvenated binders, as raw data exhibits the highest accuracies.

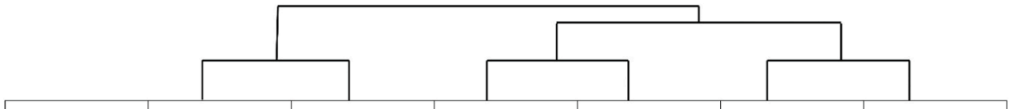
5.5.3. Effect of combined BDP and NDP methods on rejuvenation classification of peak areas/indices

Table 5 shows the best and worst accuracies achieved by combining NDPs and BDPs. Detailed results are in the supporting information (Table 6S). The table also includes an HCA dendrogram, illustrating how peak areas or indices calculation methods are grouped based on their accuracies. Notably, A_{ALI}-based indices cluster differently compared to other peak areas and indices. For these A_{ALI}-based indices, combined NDPs and BDPs did not improve accuracy; all combined methods yielded lower accuracy than raw data. For A_B/A_{ALI} and A_T/A_{ALI} indices, 8points and poly baseline corrections yielded the lowest classification accuracies when applied individually. Interestingly, combining 8points and poly with NMO and RS surpassed other NDPs in accuracy.

For A_B, A_T, and their A_{TOTAL}-based indices, combining NDPs and BDPs enhances classification. The imodpoly method, combined with NTS or NCV, achieves the highest accuracies for A_B and A_T-based classification. While applying imodpoly independently significantly reduces accuracy, combining it with appropriate normalization methods proves beneficial. Conversely, if a normalization method negatively impacts accuracy when applied alone, its combination with BDPs, such as RS for A_B-based classification, further reduces accuracy. RS yields the lowest accuracy for A_B-based classification when applied alone, and when combined with 8points and airpls, it again results in the lowest accuracy. For A_T-based classification, the combined method with MC yields the lowest accuracies. For A_{TOTAL}-based indices, the airpls method produces the lowest classification accuracy both individually and in combination with NDPs.

The efficacy of DP methods depends on the specific peak area or indices method and the classification purpose. Positive individual outcomes of DP methods increase the likelihood of achieving high accuracy through their combination with other DP methods. Conversely, if a DP method performs poorly when applied individually, its combination with others will likely also reduce overall performance.

Table 5
Upper and lower rejuvenation classification accuracy limits for the combination of NDPs and BDPs are provided, alongside raw data accuracies. The HCA dendrogram for the entire accuracy dataset (Table 6S) is included for the clustering analysis of peak areas and indices.



Accuracy rank	A_B/A_{ALI}	A_T/A_{ALI}	A_T/A_{TOTAL}	A_B/A_{TOTAL}	A_B	A_T
Max	raw data	raw data	imodpoly_NTS/NCV 0.99 ± 0.01	imodpoly_NTS/NCV 0.98 ± 0.01	imodpoly_NTS/NCV 0.98 ± 0.01	imodpoly_NTS/NCV 0.99 ± 0.01
Raw data	0.97 ± 0.01	0.96 ± 0.01	0.97 ± 0.01	0.95 ± 0.01	0.97 ± 0.02	0.97 ± 0.01
Min	8points_NDPs except NMO/RS 0.79 ± 0.03	poly_NDPs except NMO/RS 0.72 ± 0.03	airpls_NDPs 0.91 ± 0.02	airpls_NDPs 0.87 ± 0.03	8points/airpls_RS 0.79 ± 0.02	BDPs_MC 0.88 ± 0.03

5.6. Effect of DPs on the entire and first derivative spectra for rejuvenation classification

Examining the entire spectral range for rejuvenation classification yields an accuracy of 0.95 (Table 6). Normalization methods NTS and NCV significantly enhance accuracy, while MC negatively impacts it. Baseline correction techniques do not improve accuracy, likely due to the elimination of crucial features. NDPs outperform BDPs in classifying based on entire spectra and can mitigate accuracy declines from BDPs. However, combining poorly performing NDP methods with BDPs further diminishes accuracy, making MC the least favourable method.

For first derivative spectra, NDP methods generally perform better, achieving accuracies of 1, except for MC and PS (Table 7). BDP methods are ineffective and can reduce accuracy, particularly poly and pspline_asls/airpls. Combined NDPs and BDPs show high accuracies, except when involving MC and PS. Combining individual underperforming methods from NDPs, such as MC, and BDPs, like poly, further reduced accuracy in the combined approach. Effective NDP methods like NTS and NCV are sufficient for high-quality classification using either entire spectra or first derivative spectra. Utilizing random DPs and an unoptimized combined method can suffer from high computational costs and, more importantly, result in lower classification accuracies.

5.7. Comparison of DP's effect on different classification goals

In rejuvenation classification, raw data PLSDA accuracy surpasses aging classification due to distinctive features in rejuvenated binders. BDPs impact peak area and index classification variably. The aspls method improves aging classification accuracy when combined with peak areas and indices, while asls, poly, and 8points decrease it. BDPs do not enhance aging classification accuracy based on entire spectra or their first derivative; poly even reduces it. For rejuvenation, BDPs generally reduce classification quality, with airpls slightly improving the first derivative. This underscores the necessity of selecting suitable pre-processing methods aligned with specific objectives and input data.

Among NDP methods, aging classification achieves highest accuracies with NMO and RS for A_B and A_B/A_{TOTAL}, NCV for A_T and A_T/A_{TOTAL}, and AS, PS, and SNV for entire spectra and their first derivative. For rejuvenation, NTS and NCV maintain PLSDA accuracy equivalent to raw data for peak areas and indices and enhance accuracy for entire spectra and their first derivative. The MC method reduces accuracies in both classifications and is not recommended for FTIR spectra studies of bituminous binders. Combining NDP and BDP methods does not enhance classification accuracies more than using only NDPs for peak areas and indices. For entire spectra or their first derivative, the combined DP methods' performance depends on individual methods' effectiveness. Low-performing methods like MC and poly result in low performance when combined.

HCA clustering of peak areas and indices based on BDP-modified spectra shows that different classification objectives influence clustering outcomes. Irrespective of the clustering objective, calculations involving A_B and A_T tend to cluster together. HCA dendrograms based on NDP-modified spectra reveal consistent clustering patterns for both aging and rejuvenation classifications, with one cluster encompassing A_B and A_B-based indices and another consisting of A_T calculations. Changes in the dataset impact the arrangement and proximity of samples in clusters. Similar clustering patterns for aging and rejuvenation classifications based on the accuracies of combined DP methods were also observed.

6. Conclusions

This study investigates the efficacy of various data pre-processing (DP) methods, including normalization (NDPs), baseline correction (BDPs), and their combinations (CDPs), in enhancing the classification accuracy of ATR-FTIR spectra from different bituminous binder types, sources, and aging conditions. The primary focus is on evaluating these pre-treatment methods using Partial Least Squares Discriminant

Table 6
PLSDA rejuvenation classification accuracies for entire spectra in their raw form, after normalization, baseline correction, and using the top three accuracies along with the lowest accuracy achieved through the combination of NDPs and BDPs.

Baseline-correction method	Accuracy	Normalization method	Accuracy	Combined BDP and NDP methods Max accuracy	Accuracy
asls	0.92 ± 0.03	**NTS**	0.97 ± 0.01	pspline_asls_NTS/NCV	0.99 ± 0.01
aspls	0.90 ± 0.02	**NCV**	0.98 ± 0.01		
imodpoly	0.88 ± 0.03	NMO	0.90 ± 0.02	pspline_airpls_NTS/NCV	0.99 ± 0.01
modpoly	0.85 ± 0.02	MC	0.88 ± 0.03		
poly	0.88 ± 0.03	AS	0.95 ± 0.02	airpls_NTS/NCV	0.99 ± 0.01
pspline_asls	0.87 ± 0.03	PS	0.93 ± 0.01		
pspline_airpls	0.89 ± 0.03	RS	0.90 ± 0.01	**Min accuracy**	Accuracy
airpls	0.88 ± 0.03	SNV	0.95 ± 0.02	BDPs + MC	0.88 ± 0.03
8points	0.90 ± 0.03	MSC	0.93 ± 0.02		
raw data: 0.95 ± 0.01					

Table 7
PLSDA rejuvenation classification accuracies for first derivative spectra in their raw form, after normalization, baseline correction, and using the top three accuracies along with the lowest accuracy achieved through the combination of NDPs and BDPs.

Baseline-correction method	Accuracy	Normalization method	Accuracy	Combined BDP and NDP methods Max accuracy	Accuracy
asls	0.93 ± 0.03	NTS	1.00 ± 0.00	All except methods with MC &PS	> 0.97 ± 0.01
aspls	0.92 ± 0.03	NCV	1.00 ± 0.00		
imodpoly	0.93 ± 0.03	NMO	0.99 ± 0.01		
modpoly	0.93 ± 0.03	MC	0.93 ± 0.03		
poly	0.88 ± 0.02	AS	1.00 ± 0.00		
pspline_asls	0.89 ± 0.03	PS	0.96 ± 0.01		
pspline_airpls	0.93 ± 0.03	RS	0.99 ± 0.01	**Min accuracy**	Accuracy
airpls	0.96 ± 0.01	SNV	1.00 ± 0.00	poly_MC	0.88 ± 0.02
8points	0.93 ± 0.03	MSC	1.00 ± 0.00		
raw data: 0.93 ± 0.03					

Analysis (PLSDA) classification, examining areas under peaks, indices, entire spectra, and first derivative spectra. This research addresses a significant gap in the literature concerning optimal pre-processing strategies for spectral data analysis in the context of bituminous binder rejuvenation.

The impact of BDPs on classification accuracy is diverse, with both positive and negative outcomes. The selection of BDPs should be guided by the dataset's specific characteristics and the objectives of the classification task. Therefore, it is advisable to thoroughly weigh the costs, risks, and benefits of employing baseline correction before incorporating it into future studies. Moreover, the efficacy of normalization methods proves to be dependent on the specific objectives of the study and the calculation method for input data (whether it involves peak areas, indices, entire spectra, or their first derivative). Furthermore, the combination of normalization and baseline correction methods can yield diverse outcomes, dependent on the specific DP methods and input datasets employed.

Analyzing the significance of different spectral regions for classification, the variable importance in projection (VIP) scores indicated that regions like carbonyl and sulfoxide are crucial, although their importance can change with different pre-treatment methods. Normalization methods and index calculations can alter the significance of various regions, highlighting the importance of methodological choices in aging classification. This underscores the need for careful selection of spectral regions tailored to the study's objectives and the pre-processing methods employed. Relying solely on common practices of focusing on carbonyl and sulfoxide may not provide sufficient discrimination.

Based on the findings of this study, the comparison between the performance of entire spectra/first derivative and indices/peak areas showed that using entire/first derivative spectra yields higher classification accuracy, indicating greater informativeness. This finding supports the use of these approaches for future aging-related chemometric analyses. Moreover, regarding the most suitable method for calculating peak areas or aging indices, the findings indicate that indices based on A_{TOTAL} outperform A_{ALI}-based indices and peak areas (either A_T or A_B) when subjected to different DP methods. The choice between A_T/A_{TOTAL} and A_B/A_{TOTAL} depends on the study's goal. For investigations aiming to use diverse data sources and recognize gradual changes across all sources (like aging), A_B/A_{TOTAL} indices are recommended. On the other hand, for studies focused on distinguishing a specific data source from the rest, it is suggested to use A_T/A_{TOTAL}.

Furthermore, the investigation into whether a single pre-processing method could be universally effective showed that the effectiveness of pre-processing methods is dependent on various factors. These factors include the specific classification goal (e.g., aging studies or the identification of additional materials like rejuvenators), the characteristics of the dataset, and the methods employed (peak areas, indices, entire spectra, or their first derivative). It is crucial to align the choice of DP methods with specific characteristics of the input dataset and the objectives of the classification study. Additionally, to choose the best DP methods rapidly and efficiently for a given study, multivariate analysis tools like PLSDA are necessary. Among the methods employed in this study, it is advisable to include NCV, NMO, RS, and NTS methods for peak area or indices-based classification. For entire spectra and first derivative, NTS, NCV, AS, PS, and SNV methods are recommended. On the other hand, MC and poly methods showed poor performance in both classification analyses, so they are recommended to be excluded from future analyses. Lastly, aspls is a suitable option for studies focusing on gradual material changes, such as multi-level aging studies. However, for studies focused on additive detection, like rejuvenator studies, it is likely better to exclude the aspls method.

In conclusion, this study provides a comprehensive evaluation of DP methods for spectral data classification, offering valuable insights into selecting appropriate pre-treatment strategies to enhance classification accuracy. By identifying the most effective methods for different classification tasks, researchers can design more accurate and efficient experiments, optimizing resource allocation and reducing operational risks. This targeted approach ensures reliable and robust classification outcomes, contributing to more efficient research and development processes in the field of bituminous binder analysis.

CRediT authorship contribution statement

Sadaf Khalighi: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Lili Ma:** Writing – review & editing, Investigation. **Shisong Ren:** Writing – review & editing. **Aikaterini Varveri:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

This paper/article is created under the research program Knowledge-based Pavement Engineering (KPE, funded by Rijkswaterstaat, contract number = 31164321). KPE is a cooperation between Rijkswaterstaat, TNO and TU Delft in which scientific and applied knowledge is gained about asphalt pavements and which contributes to the aim of Rijkswaterstaat to be completely climate neutral and to work according to the circular principle by 2030. The opinions expressed in this paper is solely from the authors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fuel.2024.132701.

References

[1] Mirwald J, et al. Understanding bitumen ageing by investigation of its polarity fractions. Constr Build Mater 2020;250:118809.
[2] Hofko B, et al. Repeatability and sensitivity of FTIR ATR spectral analysis methods for bituminous binders. Mater Struct 2017;50:1–15.
[3] Brereton RG. Chemometrics for pattern recognition. John Wiley & Sons; 2009.
[4] Fringeli U. ATR and reflectance IR spectroscopy, applications. Encyclopedia Spectroscopy Spectrometry 2000:115–29.
[5] Hofko B et al. *Alternative approach toward aging of bitumen and asphalt mixes*. In: *Proceedings of the transport research board 94th annual meeting*; 2015.
[6] Eberhardsteiner L, et al. Towards a microstructural model of bitumen ageing behaviour. Int J Pavement Eng 2015;16(10):939–49.
[7] Lamontagne J, et al. Comparison by Fourier transform infrared (FTIR) spectroscopy of different ageing techniques: application to road bitumens. Fuel 2001;80(4):483–8.
[8] Nivitha M, Prasad E, Krishnan J. Ageing in modified bitumen using FTIR spectroscopy. Int J Pavement Eng 2016;17(7):565–77.
[9] Lee LC, Liong C-Y, Jemain AA. A contemporary review on Data Preprocessing (DP) practice strategy in ATR-FTIR spectrum. Chemom Intel Lab Syst 2017;163:64–75.
[10] Chalmers JM, Edwards HG, Hargreaves MD. Vibrational spectroscopy techniques: basics and instrumentation. Infrared Raman Spectroscopy Forensic Sci 2012:9–44.
[11] Sun D-W. Infrared spectroscopy for food quality analysis and control. Academic press; 2009.
[12] Ma L, et al. Chemical characterisation of bitumen type and ageing state based on FTIR spectroscopy and discriminant analysis integrated with variable selection methods. Road Mater Pavement Des 2023:1–15.
[13] Primerano K, et al. Characterization of long-term aged bitumen with FTIR spectroscopy and multivariate analysis methods. Constr Build Mater 2023;409:133956.
[14] Garmarudi AB, et al. Origin based classification of crude oils by infrared spectrometry and chemometrics. Fuel 2019;236:1093–9.
[15] Weigel S, Stephan D. The prediction of bitumen properties based on FTIR and multivariate analysis methods. Fuel 2017;208:655–61.
[16] Weigel S, Stephan D. Bitumen characterization with Fourier transform infrared spectroscopy and multivariate evaluation: prediction of various physical and chemical parameters. Energy Fuel 2018;32(10):10437–42.
[17] Wieser M, et al. Assessment of aging state of bitumen based on peak-area evaluation in infrared spectroscopy: Influence of data processing and modeling. Constr Build Mater 2022;326:126798.
[18] Ren S, et al. Aging and rejuvenation effects on the rheological response and chemical parameters of bitumen. J Mater Res Technol 2023;25:1289–313.
[19] Tarsi G, et al. Effects of different aging methods on chemical and rheological properties of bitumen. J Mater Civ Eng 2018;30(3):04018009.
[20] Mocetti F. Characterization of moisture susceptibility of asphaltic bitumen. Context-Sensitive Design in Transportation Infrastructures. University of Bologna and TUDelft; 2015.
[21] En C. 12607–1: Bitumen and Bituminous Binders—Determination of the Resistance to Hardening under Influence of Heat and Air—Part 1: RTFOT Method. Brussels, Belgium: European Committee for Standardization; 2014.
[22] En C. 14769: Bitumen and Bituminous Binders—Accelerated Long-Term Ageing Conditioning by a Pressure Ageing Vessel (PAV). Brussels, Belgium: European Committee for Standardization; 2012.
[23] Smith BC. Fundamentals of Fourier transform infrared spectroscopy. CRC Press; 2011.
[24] Porot L, et al. Fourier-transform infrared analysis and interpretation for bituminous binders. Road Mater Pavement Des 2023;24(2):462–83.
[25] Erb D. *Pybaselines: A Python library of algorithms for the baseline correction of experimental data*; 2024.
[26] Erb D. *Pybaselines Documentation, Release 1.1.0*; 2024; Available from: https://pybaselines.readthedocs.io/en/latest/introduction.html.
[27] Khalighi S, Erkens S, Varveri A. Exploring the impact of humidity and water on bituminous binder aging: a multivariate analysis approach (TI CAB). Road Mater Pavement Des 2024:1–25.
[28] Lee LC, Liong C-Y, Jemain AA. Effects of data pre-processing methods on classification of ATR-FTIR spectra of pen inks using partial least squares-discriminant analysis (PLS-DA). Chemom Intel Lab Syst 2018;182:90–100.
[29] Rinnan Å, Van Den Berg F, Engelsen SB. Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends Anal Chem 2009;28(10):1201–22.
[30] Lasch P. Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging. Chemom Intel Lab Syst 2012;117:100–14.
[31] Hastie T, Tibshirani R, Friedman J. The wrong and right way to do cross-validation. Elements Stat Learn: Data Mining, Inference, Predict 2009:245–7.
[32] Jing R, et al. Ageing effect on chemo-mechanics of bitumen. Road Mater Pavement Des 2021;22(5):1044–59.
[33] Engel J, et al. Breaking with trends in pre-processing? TrAC Trends Anal Chem 2013;50:96–106.
[34] Lee LC, Liong C-Y, Jemain AA. Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps. Analyst 2018;143(15):3526–39.
[35] Kumar N, et al. Chemometrics tools used in analytical chemistry: An overview. Talanta 2014;123:186–99.
[36] Trevisan J, et al. Extracting biological information with computational analysis of Fourier-transform infrared (FTIR) biospectroscopy datasets: current practices to future perspectives. Analyst 2012;137(14):3202–15.
[37] Serrano-Cinca C, Gutiérrez-Nieto B. Partial least square discriminant analysis for bankruptcy prediction. Decis Support Syst 2013;54(3):1245–55.
[38] Barker M. Partial least squares for discrimination: statistical theory and implementation. LAP Lambert Academic Publishing; 2010.
[39] Yang J, Yang J-Y. Why can LDA be performed in PCA transformed space? Pattern Recogn 2003;36(2):563–6.
[40] Barker M, Rayens W. Partial least squares for discrimination. J Chemometrics: J Chemometrics Soc 2003;17(3):166–73.
[41] Nocairi H, et al. Discrimination on latent components with respect to patterns. Application to multicollinear data. Comput Stat Data Anal 2005;48(1):139–47.
[42] Wu L, et al. Recent advancements in detecting sugar-based adulterants in honey–A challenge. TrAC Trends Anal Chem 2017;86:25–38.
[43] Manheim J, et al. Forensic hair differentiation using attenuated total reflection Fourier transform infrared (ATR FT-IR) spectroscopy. Appl Spectrosc 2016;70(7):1109–17.
[44] Custódio MF, et al. Identification of synthetic drugs on seized blotter papers using ATR-FTIR and PLS-DA: Routine application in a forensic laboratory. J Braz Chem Soc 2021;32:513–22.
[45] He H, Ma Y. *Imbalanced learning: foundations, algorithms, and applications;* 2013.
[46] Kubinyi H. *3D QSAR in drug design: volume 1: theory methods and applications*, vol. 1. Springer Science & Business Media; 1993. p. 523–50.
[47] Zheng R, et al. Variable importance for projection (VIP) scores for analyzing the contribution of risk factors in severe adverse events to Xiyanping injection. Chin Med 2023;18(1):15.
[48] Siroma RS, et al. Clustering aged bitumens through multivariate statistical analyses using phase angle master curve. Road Mater Pavement Des 2021;22(sup1):S51–68.
[49] Saraf R, Patil SP. Study Paper on How to Read a Dendrogram. Int J Comput Appl 2014;103(6):8–11.
[50] Newey WK, Powell JL. Asymmetric least squares estimation and testing. Econometrica 1987:819–47.
[51] Zhang F, et al. Baseline correction for infrared spectra using adaptive smoothness parameter penalized least squares method. Spectrosc Lett 2020;53(3):222–33.
[52] Zhang F, et al. An automatic baseline correction method based on the penalized least squares method. Sensors 2020;20(7):2015.
[53] Butler HJ, et al. Optimised spectral pre-processing for discrimination of biofluids via ATR-FTIR spectroscopy. Analyst 2018;143(24):6121–34.
[54] Yu H-G et al. Noise reduction for improving the performance of gas detection algorithms in the FTIR spectrometer. In: Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XXIV. SPIE; 2018.
[55] Lieber CA, Mahadevan-Jansen A. Automated method for subtraction of fluorescence from biological Raman spectra. Appl Spectrosc 2003;57(11):1363–7.
[56] Zhao J, et al. Automated autofluorescence background subtraction algorithm for biomedical Raman spectroscopy. Appl Spectrosc 2007;61(11):1225–32.
[57] Hu H, et al. Improved baseline correction method based on polynomial fitting for Raman spectroscopy. Photonic Sensors 2018;8:332–40.
[58] Ying X. An overview of overfitting and its solutions. Journal of physics: Conference series. IOP Publishing; 2019.
[59] Zhang Z-M, Chen S, Liang Y-Z. Baseline correction using adaptive iteratively reweighted penalized least squares. Analyst 2010;135(5):1138–46.
[60] Iglewicz B, Hoaglin DC. How to detect and handle outliers, vol. 16. Quality Press; 1993.
[61] Dutta A. Fourier transform infrared spectroscopy. Spectroscopic Methods Nanomater Charact 2017:73–93.
[62] Eilers P, Marx B. Splines, knots, and penalties. Wiley Interdiscip Rev Comput Stat 2010;2:637–53.
[63] Xu Y, et al. Raman spectroscopy coupled with chemometrics for food authentication: A review. TrAC Trends Anal Chem 2020;131:116017.
[64] van den Berg RA, et al. Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics 2006;7:142.
[65] Ren S, et al. Chemical characterizations and molecular dynamics simulations on different rejuvenators for aged bitumen recycling. Fuel 2022;324:124550.