Spectral Colour Changes: A Promising Parameter for Smart Dew Retting Monitoring Solutions

Suvajit Mukherjee, Dmitry Galinousky, Anne-Sophie Blervacq, Lionel Buchaillot, S. Arscott, Sébastien Grec

Spectral Colour Changes: A Promising Parameter for Smart Dew Retting Monitoring Solutions

Suvajit Mukherjee [1 *], Dmitry Galinousky [1], Anne- Sophie Blervacq [1], Lionel Buchaillot[2], Steve Arscott [2], Sébastien Grec [1*]

[1]*Univ. Lille, CNRS, UMR 8576 - UGSF - Unité de Glycobiologie Structurale et Fonctionnelle, F-59000 Lille, France*

[2]*Univ. Lille, CNRS, Centrale Lille, Univ. Polytechnique Hauts-de-France, UMR 8520 - IEMN - Institut d'Electronique de Microélectronique et de Nanotechnologie, F-59000 Lille, France*

* Author for correspondence: Dr Suvajit Mukherjee, suvajit.mukherjee@inrae.fr,
Dr Sébastien Grec, sebastien.grec@univ-lille.fr,

Present address :

Dr Suvajit Mukherjee : *INRAE, UR 1268 - BIA - Unite Biopolymeres, Interactions, Assemblages, F-44316 Nantes, France*

Dr Dmitry Galinousky : *Univ. Lille, INSERM UMR 1011 - Integrated molecular analysis of liver diseases, F-59045 Lille, France*

Abstract

Natural fibres from plants such as flax (*Linum usitatissimum*) are gaining interest as possible green replacements of synthetic fibres to reduce textile pollution footprints. However, the extraction-facilitating process of bast fibres from the flax stem, i.e. dew retting, is still empirically evaluated, resulting in either over or under retted fibre, both of which decrease the resulting fibre quality and introduce batches variability. There is currently no "in-the-field" efficient device available for farmers to monitor retting of straws. Expert farmers generally use a 'Fried-test', a manual colour marking method, to define dew retting progress. In our study, we have objectively measured this optical colour-change evaluation using a portable spectrophotometer to acquire spectral data. In this work, spectra in 400 to 700 nm range were acquired during 3 retting campaigns, carried out in different retting years, involving 4 fields in total. Cluster analysis of all the spectra showed that retting could be divided into 4 distinct phases of stem decolouration (assigned as Beginning, Early, Middle and Advanced). Moreover,

we established that measurements of 4 selected wavelengths (480, 490, 600, and 610 nm) are sufficient to predict the retting stage with a high degree of confidence. This work demonstrates the relevant use of colour-related wavelengths in monitoring flax dew retting on fields. Future development of colour sensors, possibly carried by drones, is therefore possible and might represent a convenient tool for farmers to identify optimal retting end.

Keywords: flax dew retting, colour scale, spectro-colourimetry, smart sensor, agriculture 4.0

1. Introduction

Natural fibers from various plant species have been used by humans for centuries to create textiles, and in recent years, there has been a significant increase in their use as a sustainable alternative to synthetic fibers in composite materials (Tariq et al., 2022). Among natural fiber, bast fibres of flax (*Linum usitatissimum* L.) are long and dense in crystalline cellulose (*i.e.* strong and flexible) and particularly appreciated to make high-quality textiles and various composite products (Mohanty et al., 2000; Chand and Fahim, 2020). The fibres are extracted by a process starting from retting, followed by mechanical scutching, combing, spinning, and eventually weaving (Gomez-Campos et al., 2021). Retting is a biological-dynamics where the stem tissue surrounding the bast-fibre bundles are disaggregated by the enzymatic action of micro-organisms from soil and phyllosphere and moisture (Djemiel et al., 2017). This biological process facilitates the separation of the bast fibre bundles from the stem and from each other (Meijer et al., 1995; Bourmaud et al., 2019). In Europe, dew retting is mainly practiced. Uprooted Flax plants lie in fields for several weeks submitted to dew falls during night-time and warm summer days and frequent rainfall. These conditions promote microorganisms' growth on and within the flax stems and so that the enzymatic partial degradation of the cell wall polymers such as pectin and hemicellulose can occur (Wang et al., 2017; Mukherjee et al., 2024, 2025). Degradation of these cementing materials during dew retting facilitates the release of the cellulosic fibres from other stem tissues and from each other, as a consequence, fibre becomes easier to separate during further mechanical steps so that their high tensile strength is preserved (Baley, 2002; Akin, 2013; Bourmaud et al., 2019). Nonetheless, a prolonged dew-retting can cause the degradation of cellulosic fibres thus impacting the mechanical quality of fibres, a phenomenon known as over-retting (Brown and Sharma, 1984; Bratt et al., 1988; Meijer et al., 1995). On the other hand, under-retted fibres are not easily separated from the stem tissue, often causing cuticle/woody particles contamination reducing the fibre quality in fineness and homogeneity (Akin, 2013). Despite major advances in our understanding of the chemical and biological mechanisms of dew retting these recent years (Djemiel et al., 2017; Bourmaud et al., 2019; Chabbert et al., 2020; Mukherjee et al., 2024), there is no real standardisation of flax retting assessment, leading farmers to decide when to stop retting based on their experience. This means that management of one of the key stages in obtaining consistent fibre quality is still based on empirical assessment. Farmers often rely on the straw colour change or manual in-hand evaluation of ease of fibre separation (Fried's test) to decide to stop retting (Akin et al., 1996). Several studies, attempting to produce a tool to test the degree of retting using both physical (Seaby and Mercer, 1984; Brown et al., 1986; Donaghy et al., 1992; Meijer et al., 1995; Sharma and Faughey, 1999) and biochemical properties (Meijer et al., 1995; Sharma and Faughey, 1999), suggested that a more sensitive and multimodal method is needed to determine initial capacity of flax stems to be retted and to monitor the degree of

retting in the field. But until now none of them has demonstrated its efficiency in the field (Meijer et al., 1995). The monitoring of stem surface colour changes during dew-retting could represent a very promising parameter that could be easily measured with a reliable instrument in a non-destructive way. Although systematically evaluated by farmers to define retting stages, this physical parameter had been poorly scientifically studied. Studies of colour changes during retting were reduced within the CIE Lab colour space (Commission Internationale de l'éclairage – L*a*b*) (Bleuze et al., 2018; Chabbert et al., 2020; Peyrache et al., 2024) and standard statistical analysis. They reveal tendencies, but did not evaluate the ability of colour changes to predict the retting stage and achievement.

Therefore, in this study, we carry out a detailed investigation into the retting-induced colour changes of flax stems not only within CIE Lab colour space but for all the colour spectra. A dataset collection has been constituted by measuring colour spectra at different retting stages and for several retting cycles (four different fields, 3 different years and 3 different geographical places). An original statistical pipeline combining multivariate analyses such as clustering, PCA and PLS-DA has been designed with a view to establish a statistical model able to predict the evolution and then the end of the retting process. This work demonstrates the concept that integrating simple colour spectroscopic data along with support from statistical predictive models, is a prerequisite that can lead to the development of effective predictive tools for monitoring dew retting directly on-site. This approach paves the way for the future creation of an on-board device, such as a drone, capable of measuring spectra directly in the field in a non-destructive and automated manner.

2. Materials and Methods

2.1. Plant sample collection

Dew-retted flax stems were sampled from three years (2014, 2021 and 2022), across four different experimental fields (names A to D) in the North of France (**Fig. 1(a)**). Samples were collected from 3 - 5 sampling spots (150×100 cm) representing biological replicates of each field/point (named T1-T3 or T5). Each sample was a bundle of stems (~50-60 stems of each) grouped to acquire colour spectra (**Fig. 1(b)**). Extra-retted samples were also collected from field D in 2022 representing days 69 and 90 after flax uprooting. In that case, the retting process finished on day 55 after plants uprooted when farmers decided to collect the straw from the field to perform scutching. Samples were kept on the field D for an extra-retting period and then collected 69 and 90 days after plants were uprooted. Details about sampling and data sets are provided in **Table 1** and **Fig. 1(a).**

All collected samples were dried and stored in room temperature, labelled according to the day of retting and stored in the dark-for further spectra acquisition.

2.2. Collection of Spectral Readings

Colour changes in flax stem surface during retting were measured for all samples all at the same time using a portable spectrophotometer (CM-23d, Konica Minolta). Measurements were taken at five points along each flax bundle (following a contact-based approach): two measurements at the top and bottom, one at the middle (**Fig. 1(b), (c)**). The Specular Component Included (SCI) mode of the spectrophotometer was used to capture readings across the visual spectrum (400-700 nm) with 10 nm intervals. SCI data captures total reflectance, including specular reflections, providing a comprehensive assessment, especially for glossy or textured surfaces (Sanderson, 2019). CIE Lab values were also acquired for an alternative variant of the optical data representation. Data was processed using SpectraMagicNX software (www.konicaminolta.fr, 2021). Photographs provided in **Supplementary Fig. 3**., illustrating the colour changes, were taken by a NIKON D5000 camera (24 megapixels, f/4.8 aperture) using auto settings (ISO 400, exposure time 1/90 second) under indoor lighting at a 90° angle.

2.3. Statistical analysis

Statistical analysis was carried out in R software (https://www.r-project.org/, version 4.4.2, 2024-10-31) using specialised packages and libraries. The raw spectral SCI data were normalized using the 'TotInt' method of the *Chemospec* (v.6.1.10) package. The Shapiro-Wilk test was used to check the normality of data distribution and to check for equal variances among groups the Levene's test was used. Outliers identified by Inter Quartile Range (IQR) were removed. A one-way Analysis of Variance (ANOVA) along with Tukey's HSD (Honestly Significant Difference) test was performed using the *Stats* package (v.4.4.2) under R software.

Unsupervised (PCA-analysis) and supervised Partial Least Squares Discriminant Analysis (PLS-DA) methods were used for data reduction. Both methods were performed using the mixOmics R package (v.6.30.0) (https://www.r-project.org/) (Rohart et al., 2017). Number of sample groups for supervised approach was estimated by K-means clustering using inbuilt *Stats* (v.4.4.2) package in R. The elbow plot was used to plot within-cluster sum of squares (WSS) to evaluate the optimal number of clusters and the resulting cluster assignments were visualized with the *factoextra* R package.

Model robustness was assessed using 10-fold repeated cross-validation (10 repeats) with the *caret* (v.7.0.1) package (*method = "pls"*). Class separation and feature contribution were examined through score plots, loadings, and variable importance plots, while confusion matrices and ROC curves were generated using *caret* and *pROC* (v.1.18.5) R packages. The model was tested on independent yearly datasets that were not used during training. Predicted class labels were compared to the actual labels using a confusion matrix, which summarized correct classifications (True Positives) and

misclassifications (False Positives and False Negatives). From this, sensitivity and specificity were calculated for each class to evaluate the model's classification performance. Sensitivity (or True Positive Rate) is the proportion of actual positive cases that are correctly identified as positive, whereas specificity assesses the model's ability to correctly exclude other classes—such as avoiding mislabelling advanced-stage samples as beginning-stage (Florkowski, 2008).

These metrics were used to generate ROC - curves, where the Y-axis represents sensitivity and the X-axis represents specificity.

3. Results

3.1. Clustering of reflectance data (throughout the visual spectrum) reveals stage specific stem colour patterns during retting.

In total, we collected 630, 320, 320, and 200 raw spectral SCI readings from Datasets A, B, C, and D, respectively from 2014, 2021a, 2021b, and 2022 campaigns (**Supplementary Table 1, Supplementary Data 1a**). All recordings were firstly grouped based on their stem position (Top, Middle, Bottom) across the biological replicates (T_1 to T_3 or T_5) to be averaged (**Fig. 1(b)**). This approach yielded 9 to 15 averaged readings per retting time point, depending on dataset and sampling structure (**Supplementary Data 1b**). Averaged recordings were then normalized by the use of the TotIn procedure of the chemospec R package (**Supplementary Fig. 1, Supplementary Data 2**).

A pairwise comparison of the different time-points of the different data acquired during each retting campaign would suggest that retting is a linear process. However, this is not the case, particularly as it depends on external factors such as the weather, which can be highly variable (**Supplementary Fig. 2, Supplementary Data 3**). Rather than considering individual retting points, we proposed to group the different points trying to define retting stages. This grouping of retting points in retting stages will make it possible to compare datasets obtained over different retting campaigns.

To explore the presence of inherent groupings within the four colour datasets, K-means clustering was performed on the normalized values of each dataset individually (**Fig. 2**., **Supplementary Fig. 3, Supplementary Data 4**). In datasets A, B, and C, K-means consistently identified four distinct clusters (**Fig. 2(a)**, **(b)** and **(c)**). In contrast, six clusters were identified in dataset D (**Fig. 2(d)**). In all cases, the clustering forms, group of data from specific retting days or combinations of multiple days. Nonetheless, in the 2022 dataset (Dataset D), four of the clusters were analogous to those observed in the other datasets, while a fifth cluster (named cluster 6 on **Fig. 2**) was uniquely associated with an extended retting point (day 90). The sixth cluster (named cluster 2 on **Fig. 2**) comprised data points from various

days, which were categorized as outliers or atypical values. This can be due to biological heterogeneity profoundly seen in case of prolonged retting (unfavourable weather). Despite differences in retting years and field locations, the clusters reveal consistent patterns of retting across all four datasets, allowing the consolidation of the data into four distinct retting stages—Beginning, Early, Middle, and Advanced—(**Table 2**) Extra retting points measured in 2022 were splitted between a new cluster (day90) and Advanced stage (day55). These extra points will be considered separately.

3.2. CIE Lab colour metrics reveal consistent retting-stage patterns in flax stems across multiple years

Visual assessment of flax stems during retting revealed a progressive colour transition from green-yellow in the early stages to grey in the middle stages, and to blue-grey in advanced stages. Over-retted stems displayed dark brown to black colouration (**Supplementary Fig. 4**). We quantified these colour changes using the CIE Lab colour system (**Supplementary Data 5**), which is widely used in research and industry. Across all four retting data sets collected over three years, the **b*** values exhibited a more than 50% reduction between the beginning and the advanced stages of retting (**Fig. 3**), following a gradual trend. The **a*** values increased during early retting stages (B to E) and then decreased towards the later stages, showing a modest overall variation. Changes in **L*** values were more pronounced in longer retting durations, such as in the dataset A (2014) and D (2022), than in shorter durations like that of B and C (2021) (**Fig. 3**). In the extra-retting time points, **b*** and **L*** values continued to decrease, while **a*** values increased.

3.3. Multivariate analysis of spectral data reveals stage specific wavelength contribution

To address the complexity of the acquired colour data, in which each wavelength acted as a variable with respect to retting time, multivariate analyses were conducted. Normalized spectral data were first subjected to unsupervised Principal Component Analysis (PCA) for each retting dataset individually. PCA is a dimensionality reduction technique that simplifies complex multivariate data by projecting it onto a smaller set of principal components (PCs) while retaining most of the original variability. The number of principal components (PCs) was determined using 'scree plots'. In all datasets, the first two components (PC1 and PC2) consistently captured over 90% of the variability, enabling clear visualization and interpretation of spectral differences among samples. For instance, in dataset A (44-day retting period), the PC1 and PC2 together explained more than 90% of the variance (**Fig. 4(a)**).

Subsequently, supervised Partial Least Squares Discriminant Analysis (PLS-DA) confirmed that X-variate 1 and X-variate 2 effectively differentiated the stages of colour change during retting (**Fig. 4(b)**).

In dataset A, the beginning stage (day 1) was predominantly explained by X-variate 2, accounting for 30% of the variance. The early (day 6), middle (days 12–26), and advanced stages (from day 33 onwards) were primarily associated with X-variate 1, which explained 69% of the variance.

Component loading plots (**Fig. 4(c)**) for X-variates 1 and 2 identified specific wavelengths contributing to retting-stage differentiation. Positive and negative loadings reflected correlations between specific wavelengths and retting time points. Heatmaps or Clustered Image Maps (CIMs) based on X-variates 1 and 2 further resolved distinct spectral patterns across stages (**Fig. 5**). In dataset A, the beginning stage (day 1) was characterized by strong positive contributions (>1.6, colour scale) from wavelengths in the 510–630 nm range (green, yellow, orange). By early stage (day 6), spectral contributions shifted to the 610–700 nm range. During the middle stage (days 12–26), contributions across the 400–700 nm spectrum remained uniformly low. In the advanced stage (days 33 and 45), spectral contributions were segmented into three distinct blocks: (i) very low (560–660 nm and 690–700 nm; 0 to –3.2), (ii) low (670–680 nm; ≥0), and (iii) high (400–530 nm; 0 to +3.2) on the colour scale.

Similar patterns were observed in the remaining datasets. In the 2021 datasets B and C (25-day retting period), PLS-DA results showed that the beginning stage (day 1) was explained by X-variate 2, accounting for 23% (B) and 28% (C) of the variance, while the early (day 4), middle (days 6–13 in B, day 6 in C), and advanced (days 20–25 in B, day 13, 18–25 in C) stages were associated with X-variate 1, explaining 74% (B) and 70% (C) of the variance (**Supplementary Fig. 5(c), 6(c)**). CIM analysis of dataset B revealed that the beginning stage (day 1) exhibited positive values at 520–630 nm and 690–700 nm (green, yellow, orange). By day 4, the early stage showed increased contributions at 650–680 nm (red). During the middle stage (days 6–13), spectral contributions remained low across the entire 400–700 nm range. In the advanced stage (days 20–25), spectral patterns were again divided into three blocks: (i) very low (520–640 nm and 690–700 nm; 0 to –2.74), (ii) low (650–680 nm; ≥0), and (iii) high (400–510 nm; 0 to +2.74), corresponding to violet-blue regions (**Supplementary Fig. 5(d)**). Similar clustering and colour-shift patterns were observed in dataset C (**Supplementary Fig. 6(d)**).

In the 2022, dataset D (55-day retting period), a consistent stage-specific colour trend was identified through both PLS-DA and CIM analysis (**Supplementary Fig. 7**). X-variate 1 explained 66% and X-variate 2 explained 32% of the variance. The beginning stage (day 1) showed high positive values in the 520–630 nm and 700 nm range (green, yellow, orange). By day 13 (end of early stage), an increase in contributions at 640–690 nm was observed. During the middle stage (days 23–34), colour values across the spectrum remained consistently low. In the advanced stage (days 45–55), typical patterns were observed with increased contributions in the 400–510 nm range (violet-blue), very low values in the

530–630 nm and 690–700 nm range (0 to –3.23), and low contributions in the 650–680 nm range (≥0), thereby recapitulating the colour shift trends observed in the previous datasets.

The loading values of the first X-variate (X-variate 1), which accounts for 65–74% of the total covariance across all four datasets, are presented in **Table 3** along with their corresponding variable weights (**Supplementary Data 6**). The two variables with the highest positive and the two with the highest negative average loading weights were identified. Normalized data for these variables were pooled from all four datasets (**Supplementary Data 7**), and a confusion matrix was constructed to evaluate classification performance, including the generation of ROC curves and calculation of AUC scores.

3.4. Confusion matrix, ROC Curve and AUC scores identifies key wavelengths sufficient to distinguish retting stages

PLS-DA was applied separately to four datasets to identify key discriminatory variables. From component 1, the two most positively and two most negatively contributing variables were selected based on average loading values for further analysis. These selected variables were then pooled across all datasets to construct a unified feature matrix for comparative evaluation. To ensure data uniformity, three biological replicates were randomly selected from each group where more than three were available.

The model was trained on the pooled dataset, allowing it to learn distinctions between classes (B, E, M, A) based on measured input features on selected wavelength variables (480, 490, 600 and 610) (Szymańska et al., 2012) and then tested on independent yearly datasets that were not used during training. In our graph, specificity starts at 1 (**Fig. 6(a), 6(d)**), representing perfect classification with no false positives, and decreases toward 0 as the false positives begin to occur (**Fig. 6(c), Supplementary Fig. 8**). The Area Under the Curve (AUC) for each ROC plot quantifies the model's overall ability to distinguish between classes.

Training the model on pooled data using datasets B, C, and D, and testing it on dataset A yielded strong AUC-ROC values ranging from 0.97 to 1, demonstrating the model's robustness in distinguishing beginning, early, and advanced retting stages (**Fig. 6(a), (d)**). Similar performance was observed when other datasets were used for testing, confirming both the model's consistency and the reproducible nature of retting-associated colour changes over different years (**Supplementary Fig. 8**, **Supplementary Table 2**).

We also build linear classifiers for each of the considered retting stages (**Supplementary Data 8**). Linear classifiers allowed us to calculate ratings of the retting stages which were transformed to the retting stage probability. The model, which had 13 predictors, correctly classified retting stages like Beginning, Early, Middle and Advanced, 86% of the time (**Supplementary Data 9**). Two variables from thirteen were most informative in the classification models: SCI of 490 and 600 nm. Bivariable model for four class classification gave 77% of right classification on a validation dataset (**Supplementary Data 9**).

3.5. Variable loadings revel colour shift trend during flax extra-retting

The additional retting time points (d69 and d90) from the Dataset D samples (retting 2022) were analysed along with the last retting time point of the advanced stage (d55) to determine whether extra-retting can be distinguished based on colorimetric parameters. Principal Component Analysis (PCA) showed that the first two components accounted for the majority of the variability in the dataset (**Fig. 7(a)**). Partial Least Squares Discriminant Analysis (PLS-DA) revealed that X-variate 1 explained 55% of the covariance, clearly separating d90 from overlapping d55 and d69 (**Fig. 7(b)**). In contrast, X-variate 2, which explained 44% of the covariance, demonstrated potential for separating d55 from both d69 and d90. Analysis of loading weights on X-variate 1 indicated that the 510–590 nm wavelength range (green-yellow) was associated with d55 and d69, whereas the 660–700 nm range (red) was characteristic of d90 (**Fig. 7(c)**). For X-variate 2, the 580–620 nm range (yellow-orange) was associated with d55, while the 400–470 nm range (violet-blue) defined d69.

4. Discussions

4.1. Visual monitoring of Flax retting aligns with well-documented colour changes

Our visual observation of the flax retting process was in accordance with what has already been observed in previous studies on hemp (Bleuze et al., 2018; Mazian et al., 2019; Bou Orm et al., 2024). For example, during hemp dew retting, it has been shown that stem samples initially present a vivid green colour, which turns bright yellow within the first week. By the second week, black spots begin to appear, and by the fourth week, the appearance shifts to pale grey with additional black spots. Finally, By the sixth week, the stems become dark grey with numerous black spots. These colour changes during dew retting might result from a mix of biological and chemical processes: degradation of plant pigments, oxidation, biodegradation, and microbial colonization (Lee and Gould, 2002; Sousa, 2022; Lu et al., 2023). First of all, chlorophyll, the pigment that gives plants their green aspect, breaks down (releasing the pyrrole rings) as a plant goes into senescence and finally dies, causing a noticeable change to yellow, orange, or brown. Enzymatic processes and environmental constraints such as UV exposure cause oxidation which turns chlorophyll into non-coloured molecules and makes other pigments like carotenoids more visible (Hu et al., 2021; Sousa, 2022). Some proteins related to protection during UV

stress such as UvrA or UrvC proteins were revealed in flax retting metaproteome (Mukherjee et al., 2024). Retting results from microbial actions could speed up these modifications. The emergence of black spots can be linked to microbial activity, as noted in previous hemp studies (Jankauskienė and Gruzdevienė, 2013; Bleuze et al., 2018, 2020). Phenolics could be released when bacteria and fungi break down structural elements such as lignin, further oxidizing and darkening the surface of the stem (Chowdhary et al., 2021). Reactive oxygen species produced during this breakdown add to oxidative browning, which intensifies the darkening over time (Hu et al., 2021). Moreover, fungi or bacteria might contribute to colour changes by producing pigments, such as melanin (Eisenman and Casadevall, 2012). *Aspergillu*s spp. known to produce melanin have been identified during retting in previous studies of both flax (Fila et al., 2001; Djemiel et al., 2020) and hemp (Nykter et al., 2008). These interplaying biological and chemical processes can explain the gradual colour change in the stems of retted plants but certainly need to be studied in the context of flax retting. As visible colour changes are the result of a series of chemical transformations of the stem, the simple perception of the eye is not sufficient to measure their complexity.

4.2. CIE Lab data shows colour changes during retting can be monitored with portable devices

By instrumental measurement, we captured surface light reflectance across the visual spectrum and expressed them in the CIE Lab colour space. This system uses three values to represent colour, i.e. blue/yellow (b*), green/red (a*) and lightness (L). Colour differences in Lab space closely match human perception and ensure consistent representation across devices (Fairchild, 2013). The CIE Lab colour space is commonly used and measured by industrial spectro-colorimeters. Some of them used this approach to characterize and classify already retted extracted fibres batches from flax (Akin et al., 2000; Peyrache et al., 2024). Previous studies have already reported the CIE Lab colour space for the hemp retting (Bleuze et al., 2018). Other studies have shown the value of spectral analysis in assessing physiological processes in plants, such as fruit ripening (Merzlyak et al., 2003). Our study observed the a* and b* values decreased by over 50% from the start to the end of retting, reflecting a significant shift in colour metrics. This reduction corresponds to a shift from green to achromatic grey tones. While the L* value exhibited a general downward trend, this pattern varied across the three years of samples, reflecting a transition from brightness to darker shades (**Fig. 3**). Obtained data are consistent with prior findings on hemp. CIE Lab colour analysis of hemp stems during retting further highlights decreases of a* (red-green axe) and b* (yellow-blue axe) values by 32.2% and 32.4%, respectively, approaching zero. Concurrently, the L* value, representing lightness, decreases by 18.5% (Bleuze et al., 2018). The CIE Lab values for extra-retted flax samples display the lowest L* value, indicating a duller and darker appearance, along with the lowest a* and b* values, which further represent this bluish shift. However,

CIE Lab is not precise enough to detect subtle colour changes because it can show significant deviations from actual perception, especially in certain regions of colour space, leading to poor agreement with visual experience (Sharma and Rodríguez-Pardo, 2012; Seymour, 2022). Multivariate analysis using the full visible spectrum would be a more efficient approach to capture these subtle and complex changes, and so far, build an infield tool that can monitor the dew retting.

4.3. Multivariate analysis of colour change patterns during retting identifies key wavelengths for monitoring and may provide a predictive model of retting stage.

Initial k-means clustering of the four datasets revealed that, despite variations in retting duration across different years, the process follows a consistent sequence of spectral changes, allowing samples to be grouped into common retting stages. Multivariate statistical analyses—particularly PCA and PLS-DA—further confirmed these shared progression patterns and led to the identification of key wavelength variables that can be used to discriminate these stages. This tends to show that retting, however long it lasts, whatever the year and wherever it is practised, is a universal process. This has also been shown from a biological point of view during the enzymatic characterisation of the process in the course of other studies in flax (Chabbert et al., 2020; Mukherjee et al., 2024).

In our study, PLS-DA proved especially valuable as both a feature selector and a classifier, enabling effective differentiation of retting stages based on spectral data. This approach has been firstly applied to Raman spectral data in stressed flax (Blervacq et al., 2023), or FT-IR spectral data during flax dew retting (Mukherjee et al., 2025). Although limitations have been noted when class separation relies on complex linear or nonlinear relationships (Ruiz-Perez et al., 2020), PLS-DA performs robustly in scenarios like ours where classes exhibit clear clustering based on spectral features.

In the context of retting, PLS-DA loadings revealed a reproducible shift in dominant wavelength contributions over time. Initially, spectral variables between 520–630 nm were most influential, indicating a green-yellow to orange coloration characteristic of freshly harvested flax stems (**Supplementary Fig. 4**). This was followed by increased contributions from the 650–680 nm range, suggesting an orange-red transition as retting progressed. In later stages, a marked shift occurred: variables in the 400–510 nm range gained prominence, corresponding to a violet-blue hue, while contributions from the red region (650–680 nm) diminished, and the initial green-yellow-orange tones were no longer present.

More specifically, analysis of the first latent variable (X-variate 1) across all datasets revealed a consistent trend: reflectance values at 480 and 490 nm increased, while values at 600 and 610 nm decreased from the beginning to the advanced stages of retting. This spectral evolution suggests a

progressive colour shift from orange to blue on the stem surface, which corresponds well with the visual transformation we observed during microbial and enzymatic degradation of outer tissues.

To evaluate whether a reduced set of wavelengths could suffice for stage discrimination, we tested classification performance using selected spectral bands (*e.g.*, 480, 490, 600, 610 nm) via confusion matrix analysis and ROC AUC curves (**Fig.6, Supplementary Fig. 8**). Results from these metrics confirmed that a small subset of wavelengths is indeed capable of distinguishing between retting stages with high accuracy, demonstrating the feasibility of simplified, targeted spectral monitoring for process control.

Together, these findings underscore the advantage of full-spectrum multivariate analysis over traditional colour spaces such as CIE Lab and highlight its potential to select wavelength for monitoring in real-time and in a non-invasive way the retting progression.

4.4. Colour variables effectively discriminate prolonged retting from the advanced retting stage

The advanced stage of retting, as defined in each dataset, includes the endpoint of retting determined by experienced flax farmers based on field observation. While these endpoints serve as practical guidelines to define the end of retting and so potentially the optimal retting point, it remains unclear what would be the next transformation if this optimal retting is exceeded (*i.e.* over retting). To explore this, dataset D included two additional time points—day 69 and day 90—that extend beyond the endpoint at day 55, enabling a closer examination of surface colour evolution during extended retting (see section 2.1).

PLS-DA analysis demonstrated that the first latent variable (X1) effectively separated day 90 from days 55 and 69 **(Fig. 7)**. Day 90 exhibited high loading contributions in the 660–700 nm range, indicative of a red-shifted surface colour, whereas both day 55 and day 69 showed dominant contributions in the 510–590 nm region, corresponding to green-yellow hues. This spectral distinction suggests that day 90 represents a more advanced stage of retting compared to the earlier time points.

Furthermore, the second latent variable (X2) provided additional resolution, revealing that day 55 was primarily associated with wavelengths between 580–620 nm (yellow-green), while day 69 shifted toward the 400–460 nm range, characteristic of violet-blue tones. This shift implies continued colour transformation beyond the conventionally defined endpoint, highlighting that retting may still be progressing chemically and visually even after the point considered complete by field observation.

5. Conclusions

In this study, we explored the evolution of flax stem surface colour during dew retting by acquiring visible spectra reflectance, an objective measurement of colour changes. By the use of a portable spectrophotometer (cm23d), we successfully recorded colour changes both in CIE Lab space colour and visible spectra from 400 to 700 nm along 3 different retting campaigns (involving 4 fields). Thanks to the development of an original multivariate statistical pipeline based on unsupervised (PCA) and supervised (PLSDA) methods, we identified four wavelengths—480, 490, 600, and 610 nm—as consistent markers of retting progression. Moreover, based on these wavelengths, a statistical model was built and showed efficiency to predict the retting stage with a high degree of confidence. Of course, additional years of retting colour data, will increase the robustness of the predictive model. Additionally, exploring wavelengths beyond the visible range—such as ultraviolet or near-infrared regions—to capture deeper chemical and structural changes might be a way to add other markers in the model, such as biological (Mukherjee et al., 2024) or chemical (Chabbert et al., 2020; Mukherjee et al., 2025). Nevertheless, this work has proved the concept that, by using a simple measurement that can be easily automated by the use of drones and tuning them to measure colour from a distance, it is possible to monitor the progress of retting reliably. This opens the door to the development of digital tools which, coupled with artificial intelligence, will help farmers to determine the best time to stop retting, thereby increasing fibre quality and batch homogeneity and bring flax growing into the era of Agriculture 4.0.

Acknowledgments

Dr. Suvajit Mukherjee would like to thank the I-Site ULNE foundation for funding and M. Alexis Boulet for coordinating the PEARL (Programme for Early-stage Researchers in Lille) program. The authors thank Dr. Ali Reda from IEMN (UMR 8520), and Dr. Christophe Djemiel for their valuable help in collecting samples. Our industrial partners Van Robaeys Frères, F-59122 Killem, are thanked for making the flax fields B, C and D available to us and for carrying out the industrial characterization of flax fibres, as well as the C.A.L.I.R.A (Coopérative Agricole linière de la région d'Abbeville" F- 80140 Martainneville) for making their flax fields available to us.

Fundings

This work was performed within the framework of the PEARL FlaxTronic project co-funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement N°847568, and the I-SITE ULNE Fondation (ANR), DG was granted by College de France PAUSE program (https://www.college-de-france.fr/programme-pause) and the University of Lille.

Author's contributions

Conceptualization: SM., SG.

Formal analysis, Validation: SM., SG., DG.

Methodology: SM., SG.

Investigation, Visualization, Interpretation: SM., SG., DG.

Project Administration: SG.

Writing-original draft: SM.

Writing-review and editing: SM., SG., DG., LB., SA., A.-S. B.

Declaration of competing interest

The authors declare no competing financial interest.

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

Authors did not use generative AI and AI-assisted technologies in the writing process of this manuscript.

Ethical approval

We confirm that all the research meets ethical guidelines and adheres to the legal requirements of the study country. The research does not involve any human or animal welfare related issues.

References

Akin, D.E., 2013. Linen Most Useful: Perspectives on Structure, Chemistry, and Enzymes for Retting Flax. ISRN Biotechnology 2013, 1–23. https://doi.org/10.5402/2013/186534

Akin, D.E., Epps, H.H., Archibald, D.D., Sharma, H.S.S., 2000. Color Measurement of Flax Retted by Various Means. Textile Research Journal 70, 852–858. https://doi.org/10.1177/004051750007001002

Akin, D.E., Gamble, G.R., Morrison III, W.H., Rigsby, L.L., Dodd, R.B., 1996. Chemical and Structural Analysis of Fibre and Core Tissues from Flax. Journal of the Science of Food and

Agriculture 72, 155–165. https://doi.org/10.1002/(SICI)1097-0010(199610)72:2<155::AID-JSFA636>3.0.CO;2-X
Baley, C., 2002. Analysis of the flax fibres tensile behaviour and analysis of the tensile stiffness increase. Composites Part A: Applied Science and Manufacturing 33, 939–948. https://doi.org/10.1016/S1359-835X(02)00040-4
Blervacq, A.-S., Moreau, M., Duputié, A., Hawkins, S., 2023. Comparative Analysis of G-Layers in Bast Fiber and Xylem Cell Walls in Flax Using Raman Spectroscopy. Biomolecules 13, 435. https://doi.org/10.3390/biom13030435
Bleuze, L., Chabbert, B., Lashermes, G., Recous, S., 2020. Hemp harvest time impacts on the dynamics of microbial colonization and hemp stems degradation during dew retting. Industrial Crops and Products 145, 112122. https://doi.org/10.1016/j.indcrop.2020.112122
Bleuze, L., Lashermes, G., Alavoine, G., Recous, S., Chabbert, B., 2018. Tracking the dynamics of hemp dew retting under controlled environmental conditions. Industrial Crops and Products 123, 55–63. https://doi.org/10.1016/j.indcrop.2018.06.054
Bou Orm, E., Sutton-Charani, N., Bayle, S., Benezet, J.-C., Bergeret, A., Malhautier, L., 2024. Influence of field retting on physicochemical and biological properties of “Futura 75” hemp stems. Industrial Crops and Products 214, 118487. https://doi.org/10.1016/j.indcrop.2024.118487
Bourmaud, A., Siniscalco, D., Foucat, L., Goudenhooft, C., Falourd, X., Pontoire, B., Arnould, O., Beaugrand, J., Baley, C., 2019. Evolution of flax cell wall ultrastructure and mechanical properties during the retting step. Carbohydrate Polymers 206, 48–56. https://doi.org/10.1016/j.carbpol.2018.10.065
Bratt, R.P., Mercer, P.C., Brown, A.E., 1988. Degradation of flax stems by Botrytis cinerea. Transactions of the British Mycological Society 90, 537–544. https://doi.org/10.1016/S0007-1536(88)80004-4
Brown, A.E., Sharma, H. s. s., 1984. Production of polysaccharide-degrading enzymes by saprophytic fungi from glyphosate-treated flax and their involvement in retting. Annals of Applied Biology 105, 65–74. https://doi.org/10.1111/j.1744-7348.1984.tb02803.x
Brown, A.E., Sharma, H.S.S., Black, D.L.R., 1986. Relationship between pectin content of stems of flax cultivars, fungal cell wall-degrading enzymes and pre-harvest retting. Annals of Applied Biology 109, 345–351. https://doi.org/10.1111/j.1744-7348.1986.tb05326.x
Chabbert, B., Padovani, J., Djemiel, C., Ossemond, J., Lemaître, A., Yoshinaga, A., Hawkins, S., Grec, S., Beaugrand, J., Kurek, B., 2020. Multimodal assessment of flax dew retting and its functional impact on fibers and natural fiber composites. Industrial Crops and Products 148, 112255. https://doi.org/10.1016/j.indcrop.2020.112255
Chand, N., Fahim, M., 2020. Tribology of Natural Fiber Polymer Composites. Woodhead Publishing.
Chowdhary, V., Alooparampil, S., Pandya, R.V., Tank, J.G., 2021. Physiological function of phenolic compounds in plant defense system. Phenolic compounds-chemistry, synthesis, diversity, non-conventional industrial, pharmaceutical and therapeutic applications.
Donaghy, J.A., Boomer, J.H., Haylock, R.W., 1992. An assessment of the quality and yield of flax fiber produced by the use of pure bacterial cultures in flax rets. Enzyme and Microbial Technology 14, 131–134. https://doi.org/10.1016/0141-0229(92)90170-S
Eisenman, H.C., Casadevall, A., 2012. Synthesis and assembly of fungal melanin. Appl Microbiol Biotechnol 93, 931–940. https://doi.org/10.1007/s00253-011-3777-2
Fairchild, M.D., 2013. Color appearance models. John Wiley & Sons.
Fila, G., Manici, L. M., & Caputo, F. (2001). In vitro evaluation of dew-retting of flax by fungi from southern Europe. *Annals of applied biology*, *138*(3), 343-351.
Florkowski, C.M., 2008. Sensitivity, Specificity, Receiver-Operating Characteristic (ROC) Curves and Likelihood Ratios: Communicating the Performance of Diagnostic Tests. Clin Biochem Rev 29, S83–S87.
Gomez-Campos, A., Vialle, C., Rouilly, A., Sablayrolles, C., Hamelin, L., 2021. Flax fiber for technical textile: A life cycle inventory. Journal of Cleaner Production 281, 125177. https://doi.org/10.1016/j.jclepro.2020.125177

Hu, X., Gu, T., Khan, I., Zada, A., Jia, T., 2021. Research Progress in the Interconversion, Turnover and Degradation of Chlorophyll. Cells 10, 3134. https://doi.org/10.3390/cells10113134

Jankauskienė, Z., Gruzdevienė, E., 2013. Physical parameters of dew retted and water retted hemp (Cannabis sativa L.) fibres. Zemdirbyste-Agriculture 100, 71–80. https://doi.org/10.13080/z-a.2013.100.010

Lee, D.W., Gould, K.S., 2002. Why Leaves Turn Red: Pigments called anthocyanins probably protect leaves from light damage by direct shielding and by scavenging free radicals. American Scientist 90, 524–531.

Li, L., Lin, D., Wang, J., Yang, L., Wang, Y., 2020. Multivariate Analysis Models Based on Full Spectra Range and Effective Wavelengths Using Different Transformation Techniques for Rapid Estimation of Leaf Nitrogen Concentration in Winter Wheat. Front. Plant Sci. 11. https://doi.org/10.3389/fpls.2020.00755

Lu, X., Li, W., Wang, Q., Wang, J., Qin, S., 2023. Progress on the Extraction, Separation, Biological Activity, and Delivery of Natural Plant Pigments. Molecules 28, 5364. https://doi.org/10.3390/molecules28145364

Mazian, B., Cariou, S., Chaignaud, M., Fanlo, J.-L., Fauconnier, M.-L., Bergeret, A., Malhautier, L., 2019. Evolution of temporal dynamic of volatile organic compounds (VOCs) and odors of hemp stem during field retting. Planta 250, 1983–1996. https://doi.org/10.1007/s00425-019-03280-6

Meijer, W.J.M., Vertregt, N., Rutgers, B., van de Waart, M., 1995. The pectin content as a measure of the retting and rettability of flax. Industrial Crops and Products 4, 273–284. https://doi.org/10.1016/0926-6690(95)00041-0

Merzlyak, M.N., Gitelson, A.A., Chivkunova, O.B., Solovchenko, A.E., Pogosyan, S.I., 2003. Application of Reflectance Spectroscopy for Analysis of Higher Plant Pigments. Russian Journal of Plant Physiology 50, 704–710. https://doi.org/10.1023/A:1025608728405

Mohanty, A.K., Misra, M., Hinrichsen, G., 2000. Biofibres, biodegradable polymers and biocomposites: An overview. Macromolecular Materials and Engineering 276–277, 1–24. https://doi.org/10.1002/(SICI)1439-2054(20000301)276:1<1::AID-MAME1>3.0.CO;2-W

Mukherjee, S., Goulas, E., Creach, A., Krzewinski, F., Galinousky, D., Blervacq, A.-S., D'Arras, P., Ratahiry, S., Menuge, A., Soulat, D., Saliou, J.-M., Lacoste, A.-S., Hawkins, S., Grec, S., 2024. Metaproteomics identifies key cell wall degrading enzymes and proteins potentially related to inter-field variability in fiber quality during flax dew retting. Industrial Crops and Products 222, 119907. https://doi.org/10.1016/j.indcrop.2024.119907

Mukherjee, S., Goulas, E., De Waele, I., Créach, A., Hawkins, S., Grec, S., Blervacq, A.-S., 2025. Tracking flax dew retting by infrared vibrational spectroscopy combined with a powerful multivariate statistical analysis. Industrial Crops and Products 227, 120798. https://doi.org/10.1016/j.indcrop.2025.120798

Peyrache, T., Chabbert, B., Aguié-Béghin, V., Delattre, F., Kurek, B., Gainvors-Claisse, A., 2024. Multiscale assessment of the heterogeneity of scutched flax fibers. Industrial Crops and Products 220, 119260. https://doi.org/10.1016/j.indcrop.2024.119260

Rohart, F., Gautier, B., Singh, A., Cao, K.-A.L., 2017. mixOmics: An R package for 'omics feature selection and multiple data integration. PLOS Computational Biology 13, e1005752. https://doi.org/10.1371/journal.pcbi.1005752

Ruiz-Perez, D., Guan, H., Madhivanan, P., Mathee, K., Narasimhan, G., 2020. So you think you can PLS-DA? BMC Bioinformatics 21, 2. https://doi.org/10.1186/s12859-019-3310-7

Sanderson, K. (2015). A comparative study of handheld reflectance spectrophotometers. In *proceedings of the AIC annual meeting topics in photographic preservation* (Vol. 16, pp. 47-62).

Seaby, D.A., Mercer, P.C., 1984. Development of a hand tool to test the degree of retting of flax straw. Annals of Applied Biology 104, 567–573. https://doi.org/10.1111/j.1744-7348.1984.tb03040.x

Seymour, J., 2022. Color inconstancy in CIELAB: A red herring? Color Research & Application 47, 900–919. https://doi.org/10.1002/col.22782

Sharma, G., Rodríguez-Pardo, C.E., 2012. The dark side of CIELAB, in: Color Imaging XVII: Displaying, Processing, Hardcopy, and Applications. Presented at the Color Imaging XVII: Displaying, Processing, Hardcopy, and Applications, SPIE, pp. 94–103. https://doi.org/10.1117/12.909960

Sharma, H.S.S., Faughey, G.J., 1999. Comparison of subjective and objective methods to assess flax straw cultivars and fibre quality after dew-retting. Annals of Applied Biology 135, 495–501. https://doi.org/10.1111/j.1744-7348.1999.tb00879.x

Sousa, C., 2022. Anthocyanins, Carotenoids and Chlorophylls in Edible Plant Leaves Unveiled by Tandem Mass Spectrometry. Foods 11, 1924. https://doi.org/10.3390/foods11131924

Szymańska, E., Saccenti, E., Smilde, A.K., Westerhuis, J.A., 2012. Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. Metabolomics 8, 3–16. https://doi.org/10.1007/s11306-011-0330-3

Wang, Q., Chen, H., Fang, G., Chen, A., Yuan, P., Liu, J., 2017. Isolation of Bacillus cereus P05 and Pseudomonas sp. X12 and their application in the ramie retting. Industrial Crops and Products 97, 518–524. https://doi.org/10.1016/j.indcrop.2016.12.047

Table 1

Datasets	Year	Locations	GPS Coordinates	Retting duration (days)*	Extra Retting (days)**	Sampling points number
A	2014	Martainneville	50°00'03''N, 1°42'27''E	44	None	6
B	2021	Bavinchove	50°47'05.5"N, 2°27'07.4"E	25	None	6
C	2021	Bavinchove	50°47'11''N 2°27'1''E	25	None	6
D	2022	Killem	50°57'04.18"N, 2°34'55.59"E	55	35	6+2

Table 2

Datasets	Retting time points (days) grouping in stages				Extra-points
	Beginning (B)	Early (E)	Middle (M)	Advanced (A)	
A	1	6	12, 26	33, 44	-
B	1	4	6, 13	20, 25	-
C	1	4	6	13, 18, 25	-
D	1	13	23, 34	44, 55	69, 90

Table 3

Wavelengths	Comp1 Loadings Dataset A	Comp1 Loadings Dataset B	Comp1 Loadings Dataset C	Comp1 Loadings Dataset D	Average Loading
480	-0.214484427	-0.2039411	-0.208338215	-0.20182853	-0.207148068
490	-0.212426754	-0.203853267	-0.208704063	-0.203982974	-0.207241764
….					
600	0.194851123	0.210820797	0.213832036	0.24700028	0.216626059
610	0.206585489	0.209678795	0.206770846	0.244134064	0.216792299
…					

Legends of figures

Figure_1. Experimental design of flax stem surface colour analysis during retting.

(a) Geographical location of experimental fields, schematic representation of retting mains events. (b) Surface colour acquisition of flax stems during the retting, with the cm23d portable spectrophotometer (Konika-Minolta). (c) Spectral data acquisition and (d) data analysis pipeline.

Figure_2. Clustering of retting days by the K-means method (Supplementary Data 4).

The representation of clusters based on stem surface colour reflectance during retting is illustrated in figures a, b, c, and d, corresponding to the dataset A, B, C, and D, respectively. Tables in each figure parts show sample numbers assigned to clusters.

Figure_3. CIE L*a*b* values of stem surface colour compared between datasets (fields) in relative retting groups and extra retting points (Supplementary Data 5).

Figure a, b and c represents b*, a* and L* values respectively, where b* indicates blue-yellowness, a* indicates green-redness and L* shows lightness. Retting stages: B = beginning, E = early, M = Middle, A = Advanced. The letters represent statistical significance between days (ANOVA, Tukey's HSD test, $P \leq 0.05$).

Figure_4. Multivariate analysis of spectral data of Dataset A (Supplementary Data 1, 2 and 6).

(a) Scree Plot of first ten principal components of PCA showing the amount of explained variance; (b) dot plot of samples projected into the X-variate space (after PLS-DA), samples are coloured according to the retting stages defined in table 2. X-variate 1 explains 69%, and X-variate 2 explains 30% of the variance in the data variables; (c) loading plots of X-variate 1 and 2. Colours indicate the retting stages (B = Beginning, E = Early, M = Middle and A = Advanced). The scale bar indicates the weight of each variable to the corresponding X-variate.

Figure_5. The heatmap of respective PLS-DA loading weights of wave-length variables (Dataset A).

Colour scale represents the loading weights of respective variables. On the left side of the heatmap, individual samples names are encoded as X.Y.Z.K where X represents Relative groups, Y represents day of sampling, Z represents Replicate number, K represents the position of measurement along the flax stem (bottom = 1, middle =2 and top =3).

Figure_6. Testing statistical predictive model built with data from datasets B to C, with data from Dataset A.

ROC curves with Area under the curve (AUC) score of (a) B= beginning retting stage, (b) E = early retting stage, (c) M = middle stage of retting and (d) A = advanced retting stage from dataset A (retting 2014) with prediction model trained from pooled data including values from datasets B, C and D (supplementary Data 7 and 10). (AUC score of 1 indicates perfect performance of prediction and 0.5 indicates random assignment).

Figure_7. Multivariate analysis of Dataset D (Supplementary Data 1, 2 and 6) focusing on advanced stages and extra retting points.

(a) Scree Plot of first ten principal components of PCA showing the amount of explained variance; (b) dot plot of samples projected into the X-variate space (after PLS-DA), samples are coloured according to the retting stages defined in table 2. X-variate 1 explains 55%, and X-variate 2 explains 44% of the variance in the data variables; (c) loading plots of X-variate 1 and X-variate 2. Colour indicates the retting points (d55, d69 and d90). The scale bar indicates the weight of each variable to the corresponding X-variate.

Legends of tables

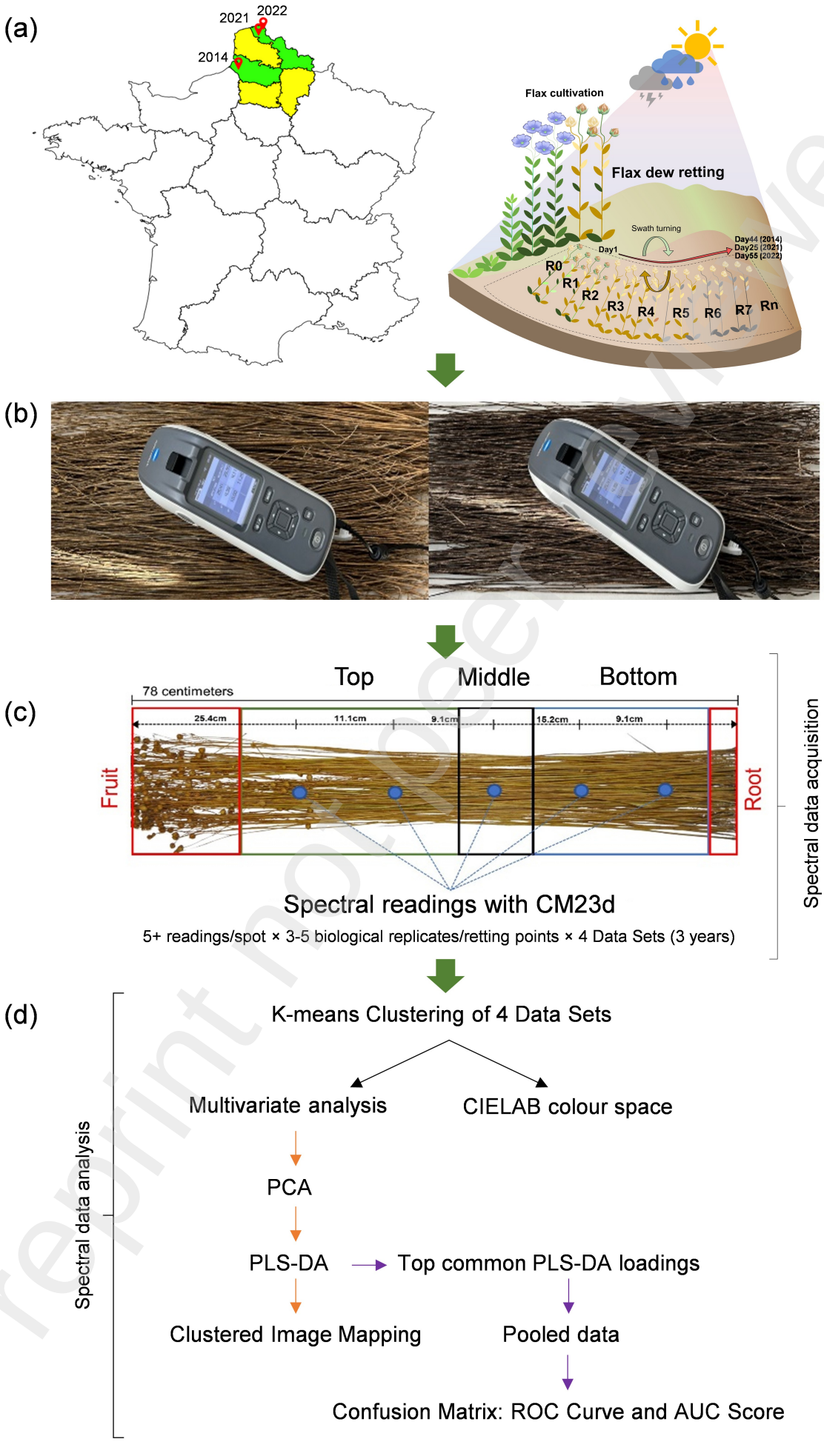
Table 1. Four data sets metadata.

* Retting duration corresponds to the number of days between the uprooting of flax and harvesting of retted straw for scutching.

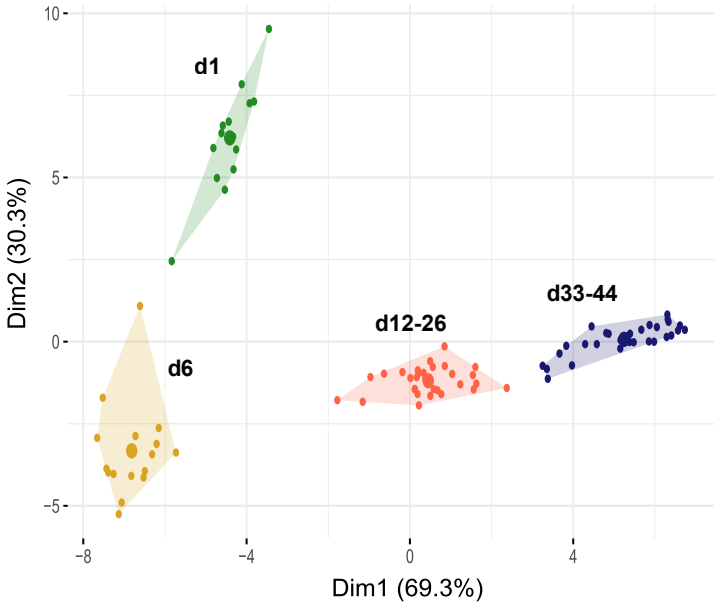
** Extra retting days correspond to sampling points collected on the field after the farmer harvested the flax straw for scutching.

Table 2. Relative retting stages based on the time points grouping.

Table 3. Summary of PLS-DA loading values from X-variate 1 (explaining 65-74% of covariance) for dataset A-D, highlighting wavelengths with the highest and lowest average loadings. Full table in Supplementary data 6.



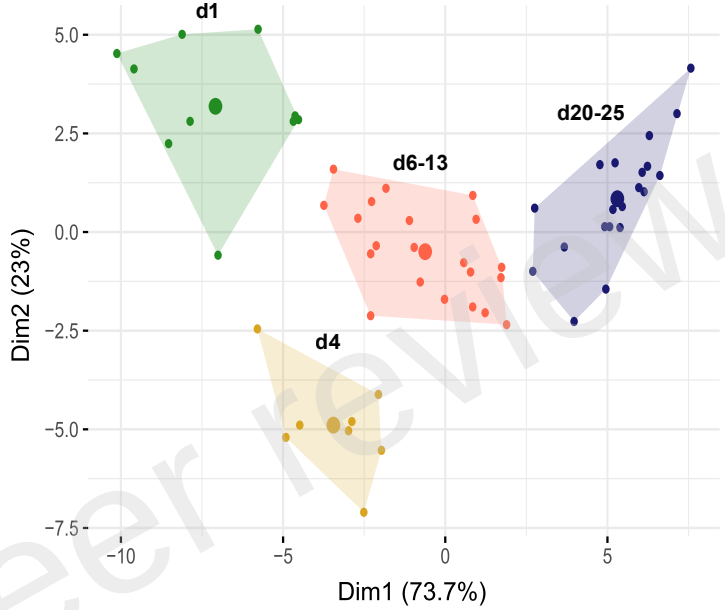
(a)



Cluster assignments of data points from dataset A based on K-Means

Cluster	1	2	3	4
day1	0	14	1	0
day6	0	0	15	0
day12	15	0	0	0
day26	15	0	0	0
day33	0	0	0	15
day44	0	0	0	15

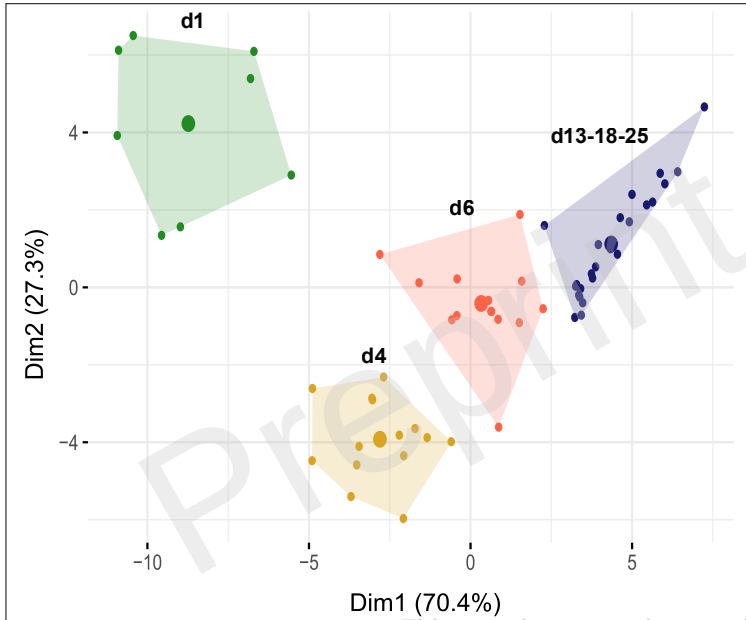
(b)



Cluster assignments of data points from dataset B based on K-Means

Cluster	1	2	3	4
day1	0	0	0	9
day4	7	2	0	0
day6	1	13	0	1
day13	0	6	3	0
day20	0	0	9	0
day25	0	0	9	0

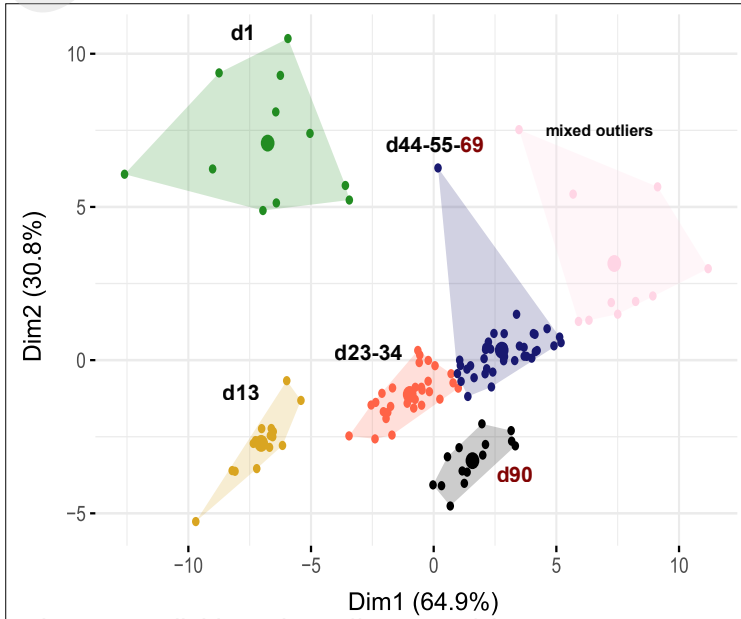
(c)



Cluster assignments of data points from dataset C based on K-Means

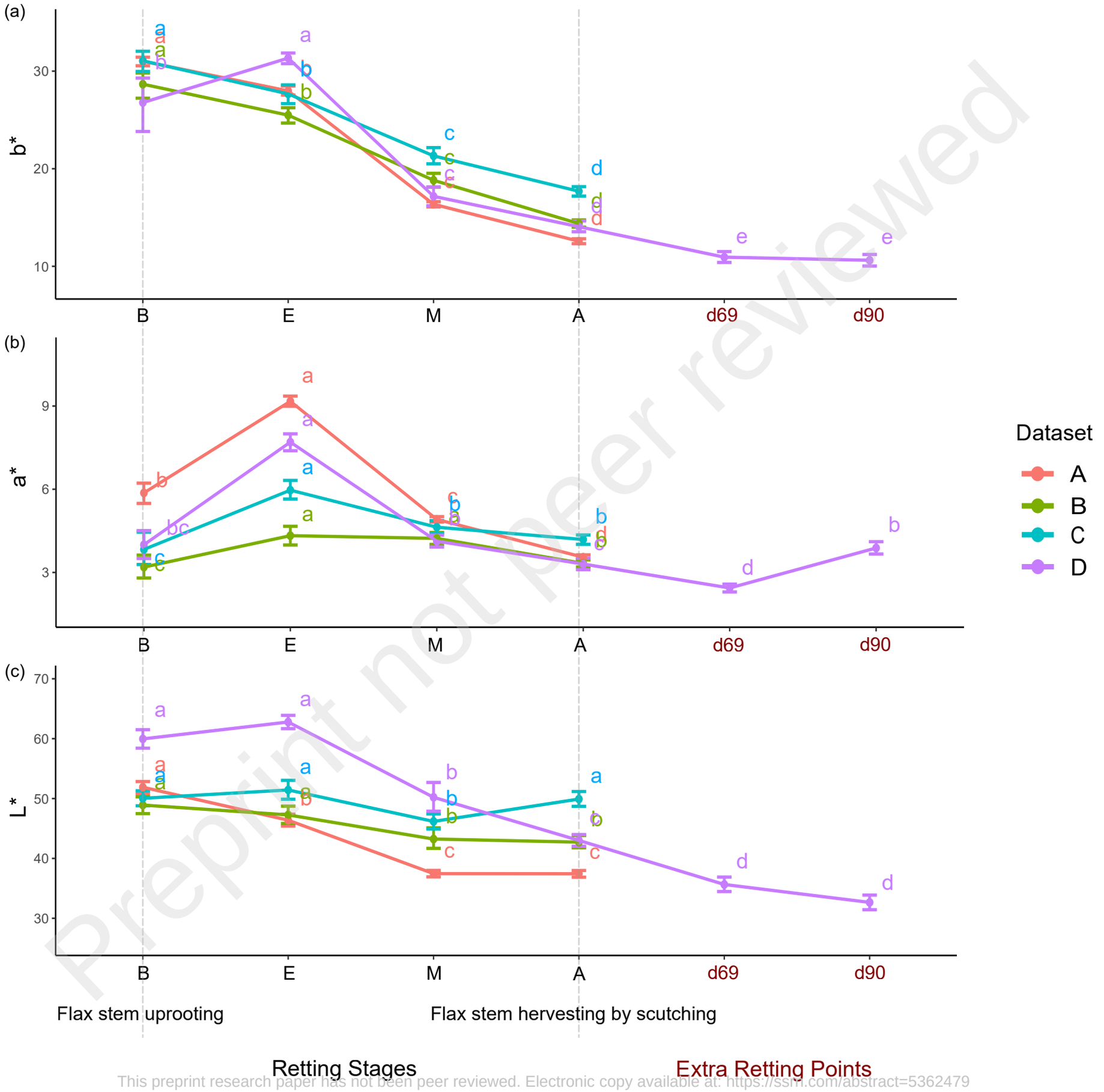
Cluster	1	2	3	4
day1	0	8	0	1
day4	0	0	0	9
day6	0	0	11	4
day13	7	0	2	0
day18	8	0	1	0
day25	9	0	0	0

(d)



Cluster assignments of data points from dataset D based on K-Means

Cluster	1	2	3	4	5	6
day1	11	3	0	0	1	0
day13	0	0	15	0	0	0
day23	0	2	0	7	6	0
day34	0	0	0	15	0	0
day44	0	1	0	2	12	0
day55	0	2	0	5	8	0
day69	0	2	0	0	13	0
day90	0	0	0	0	0	15



Dataset A (2014)