*polysaccharides*

MDPI

*Article*

# Shedding Light on Carob Seeds: A Non-Destructive Approach to Assess Dehusking Efficiency Using Diffuse Reflectance Spectroscopy and Kubelka–Munk Theory

**Rui Guerra [1,2,*], António Brázio [1], Sandra Gonçalves [3], Anabela Romano [3] and Bruno Medronho [3,4,*]**

[1] CEOT—Centre for Electronics, Optoelectronics and Telecommunications, Faculdade de Ciências e Tecnologia, Edf. 2, Campus Gambelas, 8005-139 Faro, Portugal; ambrazio@ualg.pt
[2] Departamento de Física, Faculdade de Ciências e Tecnologia, Universidade do Algarve, Edf. 2, Campus Gambelas, 8005-139 Faro, Portugal
[3] MED—Mediterranean Institute for Agriculture, Environment and Development, CHANGE—Global Change and Sustainability Institute, Faculdade de Ciências e Tecnologia, Universidade do Algarve, Ed. 8, Campus de Gambelas, 8005-139 Faro, Portugal; smgoncalves@ualg.pt (S.G.); aromano@ualg.pt (A.R.)
[4] Surface and Colloid Engineering, FSCN Research Centre, Mid Sweden University, SE-851 70 Sundsvall, Sweden
* Correspondence: rguerra@ualg.pt (R.G.); bfmedronho@ualg.pt (B.M.)

## Abstract

The carob tree (*Ceratonia siliqua* L.) is receiving growing attention for its agro-industrial potential, particularly due to its seeds, which are the source of locust bean gum (LBG), a galactomannan-rich polysaccharide with wide applications in food and pharmaceutical industries. Efficient dehusking of carob seeds is critical to maximize LBG purity and yield, yet current industrial methods pose environmental concerns and lack robust quality control tools. In this study, we demonstrate the use of Diffuse Reflectance Spectroscopy (DRS) and Kubelka–Munk (KM) modeling as a rapid, non-destructive technique to assess dehusking efficiency. By combining spectral data from four complementary spectrometers (450–1800 nm), we identified key reflectance and absorbance features capable of distinguishing raw, industrially treated, and laboratory-dehusked seeds. Notably, our laboratory-treated seeds exhibited a considerably lower reflectance in the NIR plateau (800–1400 nm) compared to raw and industry-treated seeds, and their KM-reconstructed skin showed enhanced absorption bands at 960, 1200, and 1400 nm, consistent with more complete husk removal and improved light penetration. Principal Component Analysis revealed tighter clustering and lower variability in lab-processed seeds, indicating superior process reproducibility. These results establish DRS as a scalable, green analytical tool to support quality control and optimization in carob processing.

**Keywords:** carob; seeds dehusking; locust beam gum; diffuse reflectance spectroscopy; Kubelka–Munk; galactomannan-based polysaccharides



Academic Editor: Monika Szymańska-Chargot

Received: 6 August 2025
Revised: 19 September 2025
Accepted: 1 October 2025
Published: 20 October 2025

**Citation:** Guerra, R.; Brázio, A.; Gonçalves, S.; Romano, A.; Medronho, B. Shedding Light on Carob Seeds: A Non-Destructive Approach to Assess Dehusking Efficiency Using Diffuse Reflectance Spectroscopy and Kubelka–Munk Theory. *Polysaccharides* **2025**, *6*, 95. https://doi.org/10.3390/polysaccharides6040095



## 1. Introduction

The carob tree (*Ceratonia siliqua* L.), a species native to the Mediterranean basin, has long been appreciated for both its ecological resilience and economic importance. Naturally adapted to semi-arid and low-input environments, it grows well in poor soils and under drought conditions, making it a strategic crop for sustainable agriculture, particularly in marginal regions, such as the Moroccan hinterland or the Algarve *barrocal* [1]. In recent decades, the carob has received renewed interest due to the nutritional, functional, and

industrial potential of its fruit, positioning it as a valuable asset in both traditional and emerging value chains [2–4].

The carob pod is composed primarily of a naturally sweet pulp, widely processed into powders and syrups that serve as natural sweeteners and cocoa alternatives. Its high sugar content has also triggered growing interest in its potential for bioethanol production, placing carob pulp as a promising renewable energy source [5,6]. The seeds, which represent 10–15% of the pod's weight, mainly comprise three distinct parts: the husk (also called shell or peel), the embryo, and the endosperm. The husk, a tough brown outer layer, constitutes 30–35% of the seed's dry weight. The embryo, centrally located, accounts for 15–30% and is rich in proteins, enzymes, dietary fiber, minerals, and polyphenols. Most notably, the endosperm, which makes up 40–50% of the seed, is the source of locust bean gum (LBG), a highly valued galactomannan-based polysaccharide also designated as E410 [7]. Chemically, LBG consists of a linear β-(1→4)-linked mannose backbone with α-(1→6)-linked galactose side groups, where the mannose-to-galactose ratio is typically around 4:1 [8]. This structural composition strongly influences its solubility and functional properties, such as viscosity and synergistic gel formation with other hydrocolloids (e.g., xanthan gum, carrageenan) [9,10]. Beyond its well-known thickening and stabilizing effects, LBG has been studied for its prebiotic potential [11], controlled drug delivery systems [12], and as a sustainable alternative to synthetic polymers [8,13]. LBG is highly appreciated for its thickening, stabilizing, emulsifying, and gelling capabilities [14]. Previous studies have highlighted the rheological behavior of LBG in aqueous solutions and its role in improving texture and water retention in food products [7]. Its biodegradable and non-toxic nature makes it suitable for diverse in pharmaceutical and biomedical formulations [15] as well as in cosmetics and textiles [7,16]. With growing global demand for clean-label, plant-based ingredients, carob seed-based products are increasingly recognized as valuable agro-industrial resources.

Extracting LBG with high purity and functionality strongly depends on the effective removal of the seed coat, a process known as dehusking. Conventional dehusking methods generally rely on mechanical milling, roasting, or thermal treatments, but these approaches often result in incomplete husk removal, contamination of the endosperm, and even thermal degradation of galactomannans [7]. Thermal roasting at high temperatures facilitates husk brittleness but can negatively affect gum viscosity and solubility by partial depolymerization of the polysaccharide chains [7]. As a consequence, conventional approaches often yield products of heterogeneous quality, requiring additional purification steps to achieve food- or pharmaceutical-grade gums [17].

To overcome these drawbacks, several alternative techniques have been proposed to improve efficiency and gum quality. Among them, water-dehulling (boiling the seeds followed by manual peeling) softens the husk, thereby reducing contamination, whereas acid-peeling—typically using sulfuric acid at elevated temperatures—carbonizes the husk and produces whiter gum with higher polysaccharide purity, better solubility, and greater intrinsic viscosity [18]. However, acid-based methods are environmentally hazardous, generating toxic $SO_x$ gases and requiring specialized waste management, which limits their industrial sustainability.

To address these issues, alternative sustainable dehusking solutions have been explored. One promising direction involves novel aqueous acid systems, such as methanesulfonic acid, which allow efficient peel removal under milder and safer conditions [19]. Despite the progress, evaluating the effectiveness of these dehusking methods remains a technical challenge.

Current assessment techniques are often subjective, destructive, or lack the sensitivity needed to detect subtle variations in surface properties and seed integrity. In this context,

Diffuse Reflectance Spectroscopy (DRS) emerges as a powerful and non-destructive analytical tool. Based on the measurement of reflected light across a broad spectral range, DRS provides unique spectral fingerprints that capture both chemical composition and physical surface characteristics [20]. It has been widely applied in agriculture and food sciences for different purposes, such as monitoring fruit firmness [21], ripening [22], evaluating leaf pigmentation [23], plant stress [20], assessing post-harvest quality [24], and detecting moisture-related changes in biological materials [25].

Building upon our earlier findings, where DRS was introduced as rapid, non-invasive, and highly sensitive method, the present study explores its remarkable potential as a robust and reliable diagnostic tool for monitoring dehusking efficiency. This technique offers a unique opportunity to objectively compare processing methods, ensure consistency and quality in LBG production, and minimize material loss during seed treatment. By benchmarking spectral data against commercial standards and analyzing changes in seed surface properties, we demonstrate the feasibility of integrating DRS into industrial-scale carob processing workflows. To refine the analysis, we employed the Kubelka–Munk (KM) formalism. Originally developed by Paul Kubelka and Franz Munk to describe the behavior of light in turbid media like paint films, this theory has become a cornerstone for interpreting reflectance data across numerous fields, including textiles, color science, and paper manufacturing [26]. In biological and agronomical sciences, the KM model is widely applied to deconvolve the complex interplay of light absorption and scattering, enabling the non-destructive analysis of plant leaves and fruit quality [27,28]. In this study, we leverage the KM two-layer model to isolate the optical properties of the carob seed's skin from those of the underlying endosperm. This approach enables the derivation of the spectrum attributable to the skin alone, providing a robust and quantitative metric to evaluate the efficacy of different dehusking treatments.

DRS ability to monitor both surface reflectivity and internal compositional features offers significant advantages for standardization, quality control, and sustainable process optimization in carob value chains. Beyond carob, the versatility of DRS suggests broad applicability across other agricultural and food systems.

## 2. Materials and Methods

### *2.1. Plant Material and Chemicals*

The company "A Industrial Farense, Lda" (Faro, Portugal) kindly provided *Ceratonia siliqua* L. raw seeds as well as samples of dehusked seeds produced using their in-house standard $H_2SO_4$-based approach. The model seeds (i.e., dehusked seeds from the company) were used as a benchmark for comparison with the dehusked seeds obtained via an optimized mild extraction procedure recently developed by us, using methanesulfonic acid ($CH_4O_3S$ Sigma-Aldrich St. Louis, MO, USA, 99%) and sulfuric acid ($H_2SO_4$, Panreac, Barcelona, Spain, 96%) [19].

### *2.2. Seed Husk Extraction*

The general process for extracting the carob seed husk was recently described by us [19]. In brief, the optimized extraction conditions used in this work were as follows: 24.5 g of seeds were treated in 50 mL of a solvent mixture (41% $CH_4O_3S$ and 59% $H_2SO_4$) at 90 °C for 10 min, followed by washing by 5 min with water (87 mL).

### *2.3. Diffuse Reflectance Spectroscopy*

Diffuse reflectance spectra of carob seeds were collected across the 450–1800 nm range using four spectrometers, each selected for optimal sensitivity within specific spectral intervals. To ensure continuity and maximize signal quality, the spectra from each device

were carefully aligned at overlapping borders and concatenated. The 450–680 nm region (referred to hereafter as the Visible or Vis band) was acquired using an Avantes AvaSpec-Mini2048CL-VI25 spectrometer (Avantes B.V., Apeldoorn, The Netherlands, full range: 360–1100 nm). The 680–1035 nm range (Visible–Near Infrared, Vis/NIR band) was captured with an Avantes AvaSpec HS2048XL-EVO (Avantes B.V., Apeldoorn, The Netherlands, full range: 600–1100 nm). Spectral data from 1035 to 1650 nm were obtained using an Avantes AvaSpecNIR256-1.7 TEC (Avantes B.V., Apeldoorn, The Netherlands, full range: 900–1700 nm), while the 1650–1800 nm interval was recorded with an Avantes AvaSpec NIR256-2.5 HSC EVO (Avantes B.V., Apeldoorn, The Netherlands, full range: 1000–2500 nm). The combined 1035–1800 nm region is henceforth referred to as the Near Infrared (NIR) band. Due to excessive noise, data beyond 1800 nm was excluded from analysis. For spectral acquisition in the Vis and Vis/NIR bands, a tungsten–halogen AvaLight-HAL-S-Mini (Avantes B.V., Apeldoorn, The Netherlands) light source was used. In contrast, the NIR measurements were conducted under illumination from a 150 W Philips PAR38 infrared lamp. It is important to note that the four spectrometers described above were combined to cover the full spectral range with the best possible signal-to-noise ratio (SNR). Although some redundancy existed between instruments, this strategy minimizes the use of spectral regions near detector limits, where SNR typically deteriorates. Attempts to reduce the number of spectrometers led to noisier spectra, which would compromise the proof of principle. For industrial applications, however, this limitation can be overcome with commercially available full-range instruments, which integrate both silicon and InGaAs detectors and thus avoiding the need for multiple spectrometers.

The experimental setup is schematically illustrated in Figure 1.

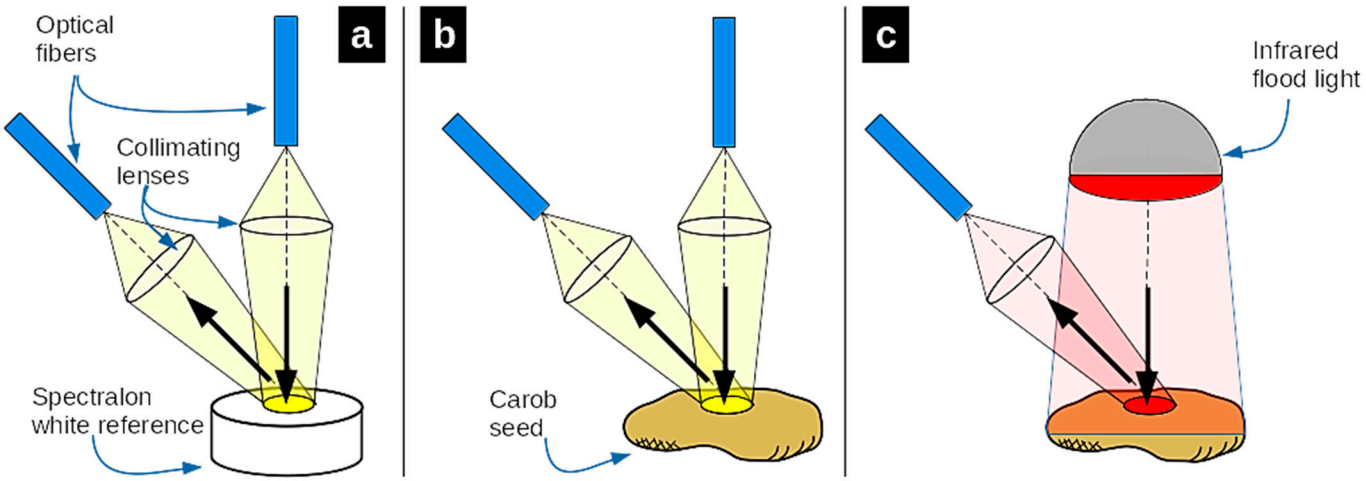


**Figure 1.** The schematic representation of the setup for the spectroscopic measurements. In all cases, the incident light is positioned vertically while the light collection is performed at 45°, to avoid direct specular reflection. In the Vis and Vis/NIR measurements, an optical fiber plus a collimator lens were used for both irradiation and collection, while for the NIR measurements, the light source is a flood infrared light. (**a**) reference measurement, with a Spectralon white tile; (**b**) measurement of the reflected spectra from the carob seed in the Vis/NIR; (**c**) measurement of the reflected spectra from the carob seed in the NIR.

For all Vis, Vis/NIR, and NIR measurements, the incident light was directed vertically onto the sample surface, while detection was performed at a 45° angle to minimize direct specular reflection. In the Vis and Vis/NIR setups, light irradiation and collection were achieved using optical fibers coupled with collimator lenses, ensuring that the illuminated spot and the detected area were spatially aligned. For the NIR measurements, irradiation was provided by a flood-infrared light source, ensuring uniform exposure across the entire

seed surface. Nevertheless, the detection spot remained consistent with that used in the Vis/Vis-NIR setups. Each measurement included three standard acquisitions: dark signal (*D*), recorded with the spectrometer entrance closed; reference signal (*Ref*), obtained from a calibrated white standard (Spectralon tile, WS-1, Ocean Optics, Orlando, FL, USA); and sample signal (*S*), collected from the carob seed surface. The reflectance (*R*) is then calculated as:

$$R = \frac{S - D}{Ref - D} \tag{1}$$

The classical definition of absorbance used in transmission spectrophotometry for clear media remains useful in the context of diffuse reflectance measurements. Although it no longer exhibits a linear relationship with true absorption, it still effectively highlights absorption peaks. Moreover, applying an absorbance transformation, followed by spectral derivation, helps to partially eliminate background and scattering effects, enhancing peak resolution. In this context, absorbance is expressed as:

$$A = -log_{10}\left(\frac{1}{R}\right) \tag{2}$$

*2.4. Kubelka–Munk Calculations*

The Kubelka–Munk (KM) theory models the propagation of diffuse light through a scattering and absorbing medium using two fundamental coefficients: the absorption coefficient (*K*) and the scattering coefficient (*S*). The ratio of these coefficients (*K*/*S*) for an infinitely thick, opaque sample is related to its diffuse reflectance, traditionally represented by $R_\infty$, through the KM function:

$$F(R_\infty) = \frac{(1 - R_\infty)^2}{2R_\infty} = \frac{K}{S} \tag{3}$$

In this work, there are two types of reflectance measurements: on the dehusked seeds and on the intact seeds. In the former case, the medium may be assumed homogeneous because light does not reach the embryo and probes only the superficial regions of the endosperm. Globally, we will refer to this as the flesh reflectance $R_f$. Therefore, for the flesh reflectance, the KM equation holds:

$$\frac{K_f}{S_f} = \frac{\left(1 - R_f\right)^2}{2R_f} \tag{4}$$

For simplicity, the explicit dependence on wavelength is omitted but is implicit; for example, $K_f \equiv K_f(\lambda)$. The raw seeds exhibit a dual-layer geometry, with the thin skin layer overlying the flesh layer, which is, for all purposes of light propagation, considered infinite. In this case, the KM theory states that the skin + flesh reflectance, $R_{sf}$, is given by

$$R_{sf} = \frac{1 - R_f[a_s - b_s \coth(b_s S_s d)]}{a_s - R_f + b_s \coth(b_s S_s d)} \tag{5}$$

where $d$ is the thickness of the skin layer, $S_s$ is the scattering coefficient of the skin, $a_s = 1 + (K_s/S_s)$, $K_s$ is the absorption coefficient of the skin, and $b_s = \sqrt{a_s^2 - 1}$. In this equation, $R_f$ and $R_{sf}$ are known from the spectroscopic measurements. The width of the skin was determined from photos of seeds' cross sections, with an average value of 0.13 mm. This leaves two unknowns, $K_s$ and $S_s$, and only one equation. Therefore, the only way to proceed is to assume one of the parameters and solve the equation for the other. Fortunately, it is well known that scattering coefficient in biological media is a slowly varying function

of the wavelength and can be modeled by a fractal power law [28,29], making it a suitable candidate for the guessed parameter. Furthermore, the reduced scattering coefficient $\mu_s'$ is related to KM's scattering coefficient by $S \approx (3/4)\mu_s'$, and has been determined in several fruit tissues [30], typically yielding values on the order of 10 $mm^{-1}$. Although there are no published values in the literature for $\mu_s'$ in carob seed skin, we have assumed as starting point $S_s = 10\ mm^{-1}$, consistent with values obtained for apple skin [31] and onion skin [32]. To incorporate the fractal power law, the guess input for $S_s$ was of the form

$$S_s(\lambda) = a\left(\frac{\lambda}{\lambda_0}\right)^{-b} \quad (6)$$

where $\lambda_0$ is a reference wavelength, $S(\lambda_0) = a$ and $b$ is the scattering power, which determines how rapidly the scattering decreases with wavelength. The starting trial values were $a = 10\ mm^{-1}$, $\lambda_0 = 400$ nm and $b = 0.1$. Once $S_s$ is chosen, for a specific wavelength, through (6), Equation (5) is then solved numerically for $K_S$ and the process is repeated for all the wavelengths. To obtain the reflectance of the skin (as if it was infinite), $R_{s,\infty}$, (4) is inverted to obtain

$$R_{s,\infty} = 1 + \frac{K_s}{S_s} - \sqrt{\left(\frac{K_s}{S_s}\right)^2 + 2\left(\frac{K_s}{S_s}\right)} \quad (7)$$

The solution for $R_{s,\infty}$ is then checked for consistency. For example, it should not exceed 1 across all wavelength range, should be larger than $R_f$ in the NIR plateau and smaller than $R_f$ in the pigments' region. The tentative values of $a$ and $b$ are changed until all the consistency checks are satisfied.

The results for $R_{s,\infty}$ are physically plausible but there is room for significant uncertainty in view of the procedure's limitations and assumptions: (1) $S_s$ is guessed (although plausible and subject to consistency check); (2) illumination is not diffuse, but one assumes that it is quickly randomized by scattering after some microns of penetration. This is also plausible, due to the high scattering coefficient; (3) the reflectance is not integrated on all the angles (as in an integrating sphere), but assuming isotropic reflectance from the surface, the acquisition at 45° is representative of the whole.

### *2.5. Statistical and Numerical Analysis*

The spectroscopic analysis was performed in a script written by us in R (version 4.4.3, 2025-02-28 ucrt) [33]. The PCA analysis was performed through the R base function prcomp. The Savitzky–Golay filtering was performed through the function sgolayfilt of the signal package [34]. Inversion of Equation (5) was performed using the R base function uniroot().

## 3. Results

Colorimetric analysis offers a quick and straightforward approach to evaluate visual differences in seed appearance, providing an initial indication of the effectiveness of the extraction process. However, if the treatment changes the internal composition of the endosperm, such as affecting polysaccharides integrity, these variations may not be visually perceptible. Thus, a more robust and sensitive analytical method is required to fully capture the solvent's impact on seed chemistry. To this end, DRS coupled with KM analysis was employed, for the first time, as a non-destructive tool to infer potential compositional changes induced by different extraction treatments [19].

As described in the Materials and Methods Section, four spectrometers with distinct optical configurations (e.g., diffraction gratings, detectors) were used to cover the full spectral range. These differences led to slight mismatches at the spectral boundaries, which were corrected using a custom matching and smoothing algorithm to generate continuous, artifact-free spectra. The initial results obtained are presented in Figure 2. For

each seed type, five replicates were recorded. The raw seed spectra are shown in black, the spectra from industrially treated seeds appear in red, and those from seeds treated with our optimized method are depicted in blue. This visualization clearly highlights the spectral distinctions across treatments, offering insights into both surface and compositional changes. The main spectroscopic absorption bands in this spectral range are overlayed in the plot. They are represented in the form $n\gamma(X)$, with $n - 1$ the order of the overtone, $\gamma = \nu$ for stretching and $\gamma = \delta$ for bending oscillations, and $X$ = O-H, C-H, $CH_2$ or $CH_3$, the chemical bonds. The main spectroscopic bands highlighted in Figure 2 are related to the chlorophyll absorption peaks (Chl), second and first overtones of O-H stretching ($3\nu$(O-H) and $2\nu$(O-H), respectively), second overtone of C-H stretching ($3\nu$(C-H)), combination band of first overtone of stretching and fundamental of bending of C-H ($2\nu$(C-H) + $\delta$(C-H)) and first overtone of $CH_2$ and $CH_3$ stretching ($2\nu(CH_2,CH_3)$).

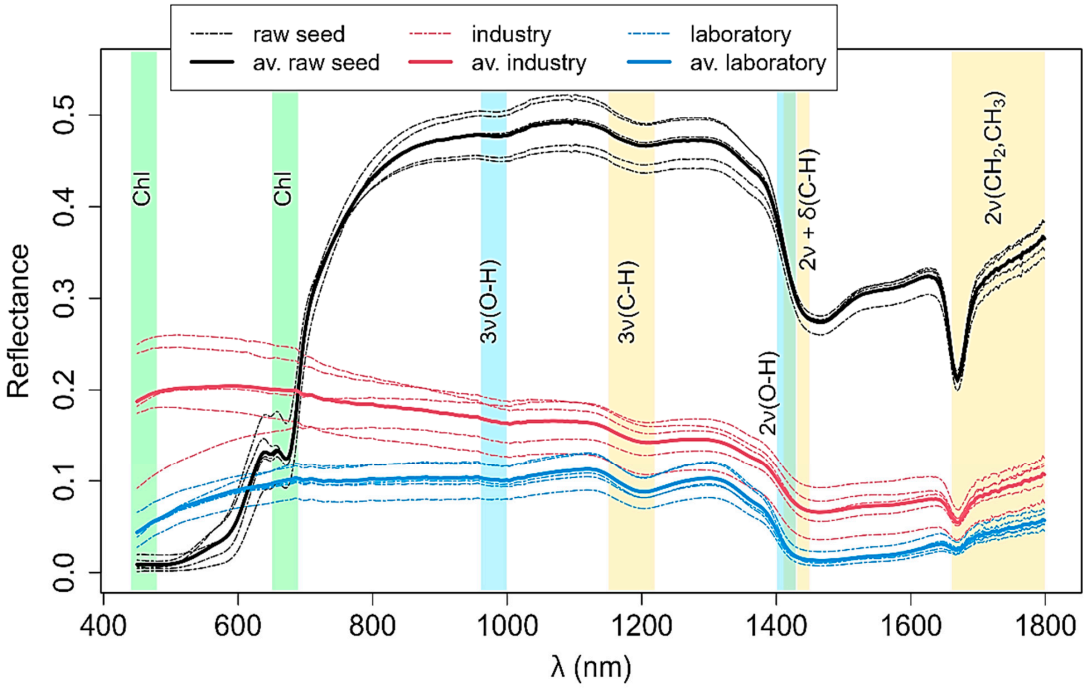


**Figure 2.** Diffuse reflectance spectra of carob seeds: native (black curves), commercially dehusked via the industrial process (red curves), and dehusked using the optimized method (blue curves). For each seed type, five spectra are shown as dashed lines and their average values in thick solid line. The vertical bars mark the main spectroscopic bands (see text for details).

The first striking interpretation of the results is that the carob seed husk, like fruit peels and most plant leaves, functions as a broadband optical reflector. This reflective effect is interrupted at specific wavelengths where absorption is significant, most notably in the visible range, due to high pigment concentrations, and at water absorption bands in the NIR region. Thus, the Vis-NIR reflectance spectra of seeds and fruits can be broadly described as a reflectance plateau with distinct absorption dips at characteristic bands, that correspond essentially with those shown in the plot bars. This framework explains the main features observed in the raw carob seed Vis-NIR spectra shown in Figure 2.

A pronounced reflectance plateau appears between 800 and 1400 nm at reflectance values of approximately 0.4 to 0.5, where the optical influence of the husk dominates due to minimal absorption in this region. We refer to this as the NIR plateau. The notable dip in the visible region is attributed to pigments, such as phenolic compounds, including tannins, flavonoids, anthocyanins, and lignin. Particularly, the local dip near 680 nm aligns with traces of chlorophyll presence. Above 1400 nm, the reflectance decreases to around 0.3,

reflecting absorption by water (specifically the first overtone of O-H vibrations, $2\nu$(O-H)) and C-H bonds (combination bands of the first overtone stretching with fundamental bending, $2\nu + \delta$(C-H)).

Additional minor spectral features further complement this interpretation: for example, a subtle dip near 1200 nm is linked to the second overtone of C-H stretching vibrations, $3\nu$(C-H), while a depression around 980 nm corresponds to the second overtone of O-H vibrations, $3\nu$(O-H). An unusually strong dip at approximately 1700 nm is observed, whose origin remains unclear but coincides with the first overtone of C-H stretching in $CH_2$ and $CH_3$ groups, although this feature is typically less pronounced in other samples.

This spectral characterization of raw seeds serves as a baseline for comparing the effects of dehusking. The husk's principal role as a reflector means that its removal leads to a global decrease in seed reflectance, as evidenced by the data: reflectance curves for both industrially and laboratory-dehusked seeds are significantly lower than those of raw seeds. Focusing on the NIR plateau, reflectance drops from 0.4 to 0.5 in raw seeds to 0.1–0.2 in industrially treated seeds and further to 0.07–0.1 in laboratory-processed seeds. The lower reflectance in the laboratory samples suggests a more efficient husk removal process, whereas industrially dehusked seeds likely retain residual husk fragments, explaining the differences between the two methods.

One might initially expect industrially dehusked seeds to reflect less light, since residual seed coat should retain absorbing pigments. However, our results show the opposite trend. This apparent discrepancy can be explained by considering the two main optical roles of seed skin: absorption by pigments and scattering due to structural refractive index mismatches. Industrial treatment removes most of the pigment-rich outer layers, thereby reducing absorption, but often leaves behind residual inner layers. These remnants contain few pigments yet maintain a strong refractive index contrast with the underlying tissue, which enhances scattering. As a result, industrially dehusked seeds exhibit higher reflectance than lab-treated seeds, where the seed coats are removed more completely.

Figure 3 shows the absorbance, calculated form Equation (3). As expected, higher reflectance corresponds to lower absorbance; thus, raw seeds exhibit the lowest absorbance values. The laboratory-treated seeds display higher absorbance than the industry-treated ones, which, according to our hypothesis, is due to their cleaner surface (i.e., more effectively stripped of husk remnants) allowing greater light penetration and absorption. It is also interesting to note that the convolution of the $2\nu$(O-H) and $2\nu + \delta$(C-H) bands now emerge clearly as the main absorption feature not caused by the pigments.

Figure 4 shows the result of the KM procedure, applied to the average spectra of the three groups. Therefore, for the industry-treated seeds, $R_f$ = average of reflectance from industry-treated seeds and $R_{sf}$ = average of reflectance from raw seeds (see Equation (5)). For the laboratory-treated seeds $R_{sf}$ is the same and $R_f$ = average of reflectance from laboratory-treated seeds. The value of the skin reflectance, $R_s$, is obtained from (7), after finding $K_s$ from (5) and the self-consistent guess of $a$ and $b$ in (6) (which determines $S_s$).

The spectroscopic features are enhanced in the husk: the absorption dips caused by O-H and C-H vibrations are much more visible than in the raw or dehusked seeds. As expected, skin reflectance, $R_s$, is larger than that of the flesh in the NIR and lower in the visible, $R_f$, due to the pigments. The raw seed reflectance, $R_{sf}$, is a combination of the two, yielding intermediate values in the NIR and approximately the same values as the husk in the visible. The KM calculated $R_s$ in the NIR is higher for the laboratory treated samples than for the industry-treated samples. This is easily understandable, considering that the laboratory treatment removed the husk more efficiently. Therefore, the mathematical reconstruction of $R_s$ from the laboratory data preserved all the reflective power from the

husk. On the contrary, the husk residuals left in the flesh on industry-treated seeds, were missing in the mathematical reconstruction of $R_s$, leading to a calculated lower reflectance.

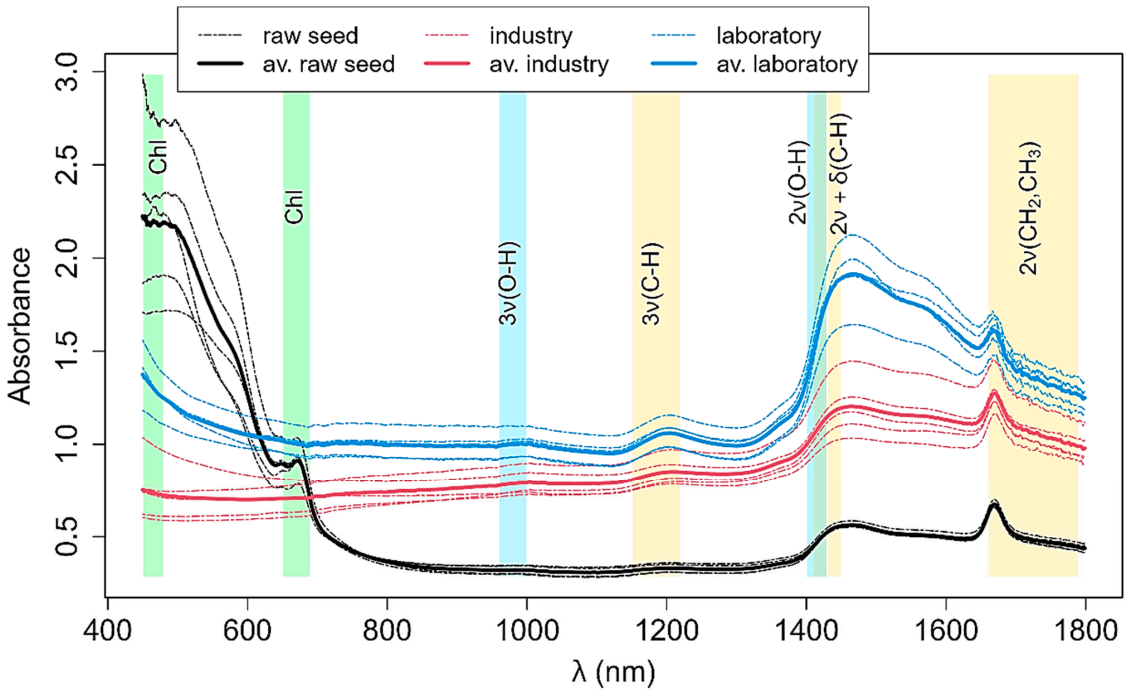


**Figure 3.** Absorption spectra of carob seeds: native (black curves), commercially dehusked via the industrial process (red curves), and dehusked using the optimized method (blue curves). For each seed type, five spectra are shown as dashed lines and their average values as thick solid lines. The vertical bands show the principal absorption bands (see Figure 2 for details).

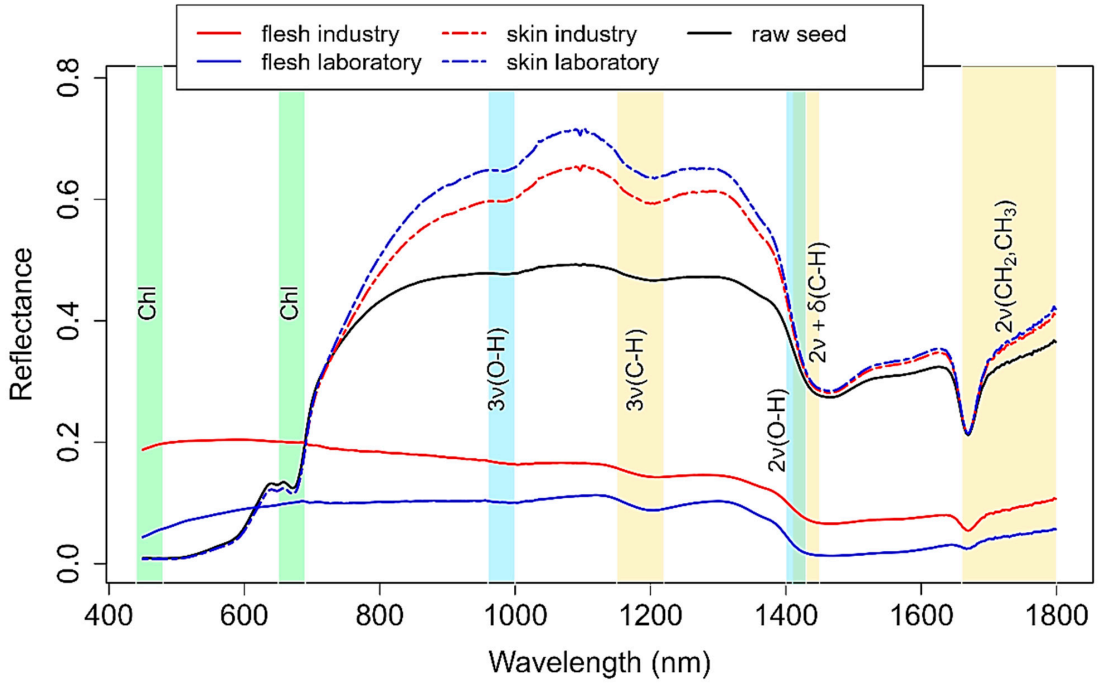


**Figure 4.** Average diffuse reflectance spectra of carob seeds and the reflectance of the husk, calculated by the KM method. Raw, industrially dehusked and laboratory-dehusked average seed spectra in black, red and blue, respectively. The husk reflectance spectra calculated from the KM method are plotted in the same color, in dashed lines. The vertical bands show the principal absorption bands (see Figure 2 for details).

In practice, the results may be interpreted as $R_s$ (industry) $< R_s$ (laboratory) because the width of the skin removed by the industry process is effectively smaller than that removed by our laboratory approach. Figure 5 shows the calculated absorption coefficient of the skin (from the numerical resolution of Equation (5)) and its assumed scattering coefficient (from the self-consistent guess of a and b in (6)).

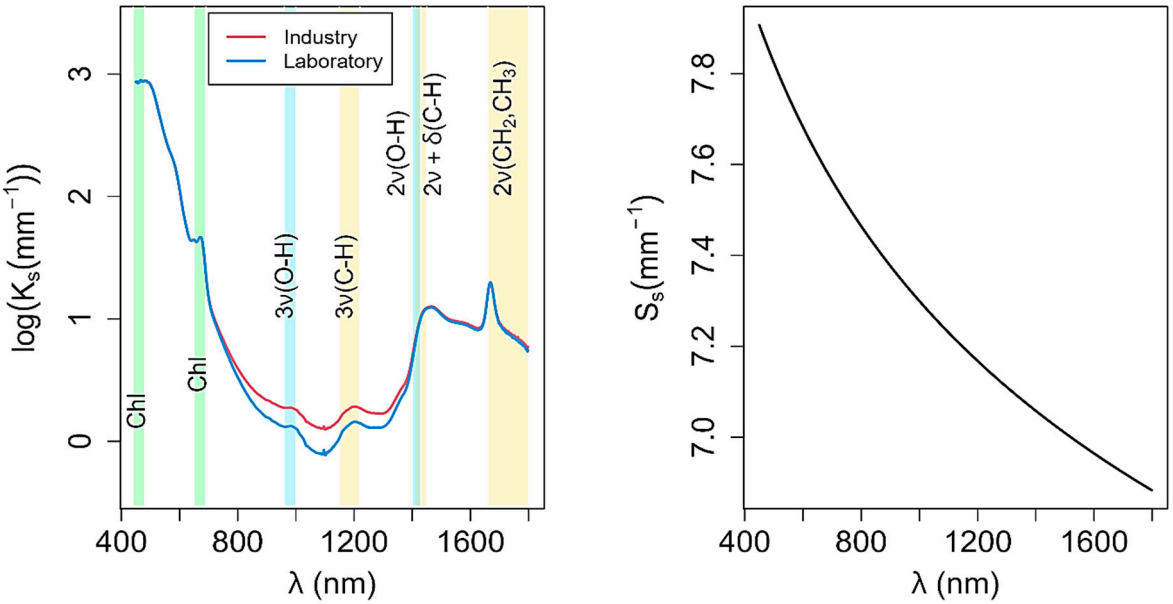


**Figure 5.** Left plot: logarithm (base 10) of the skin absorption coefficient. The vertical bands show the principal absorption spectral regions (see Figure 2 for details). Right plot: the best guessed scattering coefficient, obtained with a = 8 $mm^{-1}$ and b = 0.1 (Equation (6)).

On the left, note that the scale is logarithmic, showing the largely dominant effect of the pigments extending from 400 nm to nearly 800 nm. On the other hand, there is a NIR optical window of extremely low absorption in the range 800–1400 nm. This window is limited to the right by the water absorption peak at 1400 nm. The KM reconstruction of $K_s$ delivers high absorption in the NIR window for the industry-treated seeds. As before, the best explanation seems to be the fact that the husk tissue removed by the industrial treatment missed some reflecting residues. Therefore, the tissue effectively removed by the industrial treatment reflects less, that is, absorbs more, as calculated by the KM method.

It is well established in chemometrics that applying an absorbance transformation followed by spectral derivation can effectively eliminate multiplicative effects, particularly those related to baseline offsets and scattering [35]. Specifically, the first derivative removes linear baselines, while the second derivative corrects for sloping baselines and further reduces multiplicative scattering effects [26]. A key advantage of the second derivative is its ability to sharpen and highlight absorption features, producing well-defined dips at the location of absorption peaks, particularly in the case of isolated bands. This enhances the spectral resolution and facilitates more accurate interpretation.

Figure 6 illustrates the second derivative of the absorbance spectra of carob seeds after applying a Savitzky–Golay filter with a polynomial order of 4 and a window width of 25 points. In the visible region, the raw seed spectra exhibit prominent variations associated with pigment absorption. In contrast, the dehusked seeds display relatively flat curves, consistent with the removal of pigment-rich outer layers.

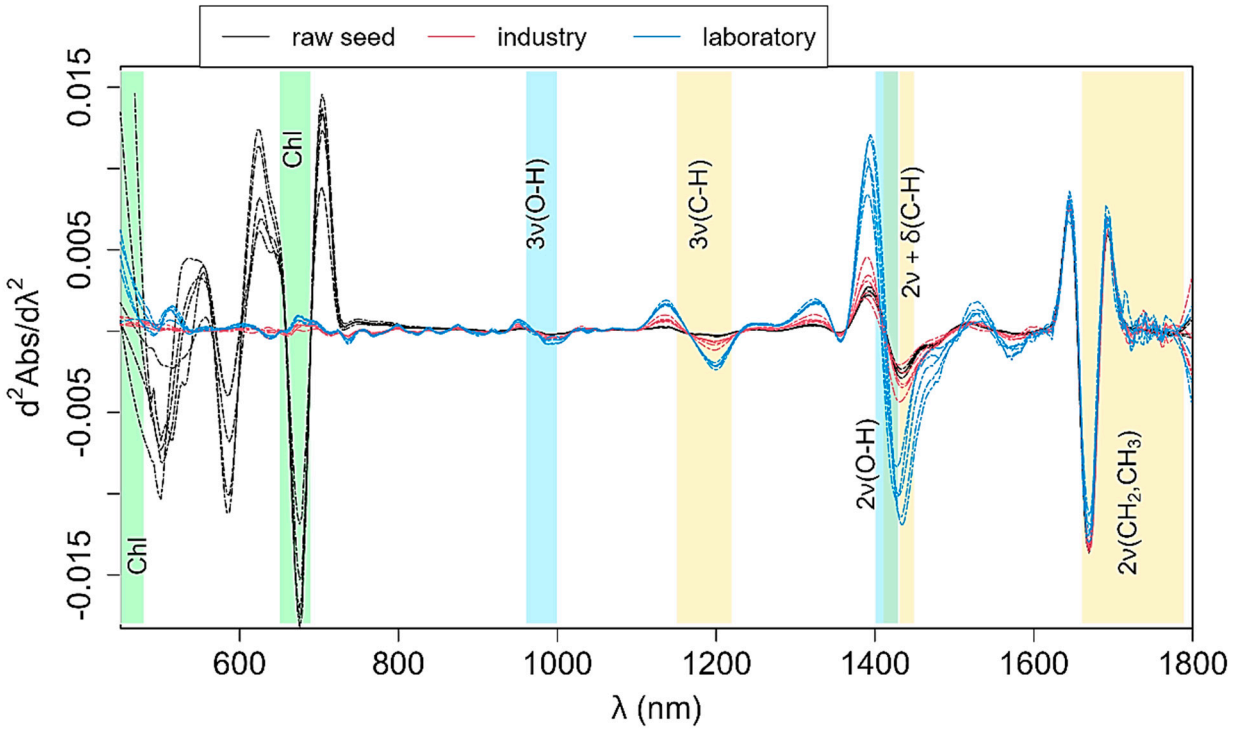


**Figure 6.** Second derivative of the absorbance spectra for native seeds (black curves), commercially dehusked seeds obtained through the industrial process (red curves), and seeds dehusked using the method developed by us (blue curves). Five spectra per seed type are shown. The vertical bands show the principal absorption bands (see Figure 2 for details).

In the Vis/NIR and NIR regions, dehusked seeds, particularly those treated using the laboratory process, exhibit enhanced absorption features around 960 nm (second overtone of O-H vibration), 1200 nm (second overtone of C-H stretching), and 1400 nm (first overtone of O-H stretching). This enhancement is attributed to increased light penetration through the seed surface, allowing more light to interact with the underlying kernel. The greater absorbance observed in the lab-treated seeds supports the conclusion that their husk removal was more effective, providing less obstruction to light and improving access to the kernel.

Notably, the shape of these absorption features remains consistent across all treatments, suggesting that the chemical composition of the seed kernel remains unaltered by the dehusking process. However, a deviation from this trend is observed at 1700 nm. Here, no enhancement is detected, and all curves are nearly superimposed. A plausible explanation is that absorption at this wavelength, associated with the first overtone of C-H stretching in $CH_2$ and $CH_3$ groups, arises from compounds equally present in both the husk and the kernel, resulting in similar spectral responses regardless of dehusking process. An optical artifact could also explain the peak (second order diffraction effects, for example), but this has not been detected in other samples, for example, in orange peel. Overall, the second derivative analysis reinforces the interpretation that the laboratory dehusking method is more efficient at removing the husk, while preserving the chemical integrity of kernel's polysaccharides.

The obtained spectra appear clearly distinguishable, both in the raw and derivative plots. In this respect, a PCA plot to discriminate the samples would seem redundant. However, the PCA loadings may reveal finer details about the spectral differences between the treatments, and this is the main motivation to proceed for a PCA. Figure 7 shows the scores plot of the first two principal components obtained from the raw reflectance data (on the left) and from the first derivative of the absorbance data (on the right). It is worth

noticing that the second derivative spectra were presented in Figure 6 because they provide a more intuitive visualization of absorbance features, with peaks appearing as valleys at the same positions. However, for PCA, the first derivative was used since its loadings are less complex and therefore easier to interpret than those from second derivative spectra, which contain more peaks and valleys. Importantly, both approaches yield similar results in the PCA scores plot (Figure 7), so the choice of first derivative data for PCA was made primarily for clarity of interpretation. In this regard, the Supplementary Information contains the figures derived from the derivative processing that are not presented here, specifically the first derivative spectra, the PCA scores plot on the second derivative data and the PCA loadings plot from the second derivative data (Figures S1–S3).

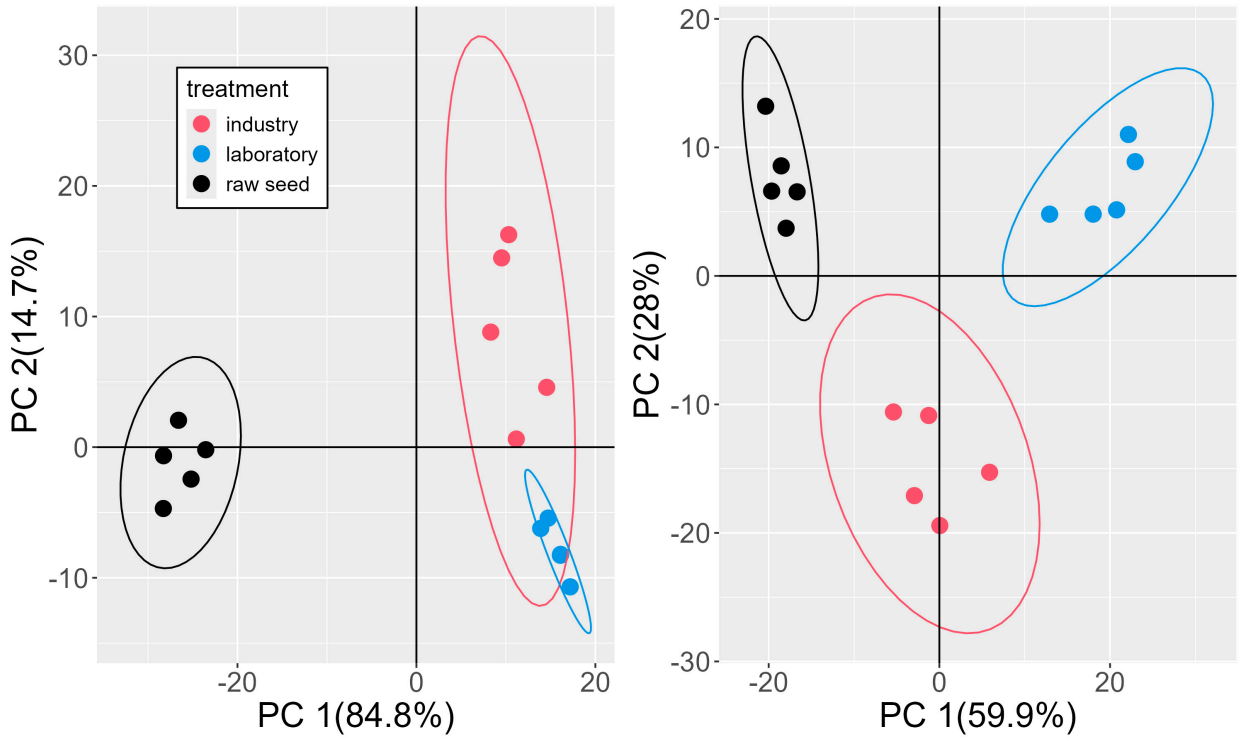


**Figure 7.** PCA of reflectance data. (**Left**) raw reflectance; (**Right**) first derivative of absorbance. Clusters correspond to native seeds (black), industrially dehusked seeds (red), and seeds dehusked with our method (blue). Data were mean-centered and normalized to the standard deviation.

As expected, both absorbance and first order derivative spectra exhibit a distinct clustering of the samples. In both plots, raw seeds are clearly separated from dehusked seeds along the first principal component (PC1), which accounts for most of the data variance. The separation is clearer in the raw reflectance data. The dehusked seeds also form separate clusters, and their separation is clearer in the derivative data. Additionally, seeds dehusked by the industrial process show larger confidence ellipses, which means results less homogeneous than with the laboratory process.

Figure 8 shows the loadings plots of the first (PC1, top) and second (PC2, bottom) principal components for the PCA decomposition of the raw reflectance (left) and first derivative of absorbance (right) datasets.

The loadings for PC1 of raw data (top left) reveal a fundamental distinction along the dichotomy between the NIR plateau and the pigments in the visible. The dehusked seeds have smaller NIR plateaus and higher reflectances in the visible, which is the picture captured in the PC1 loadings plot.

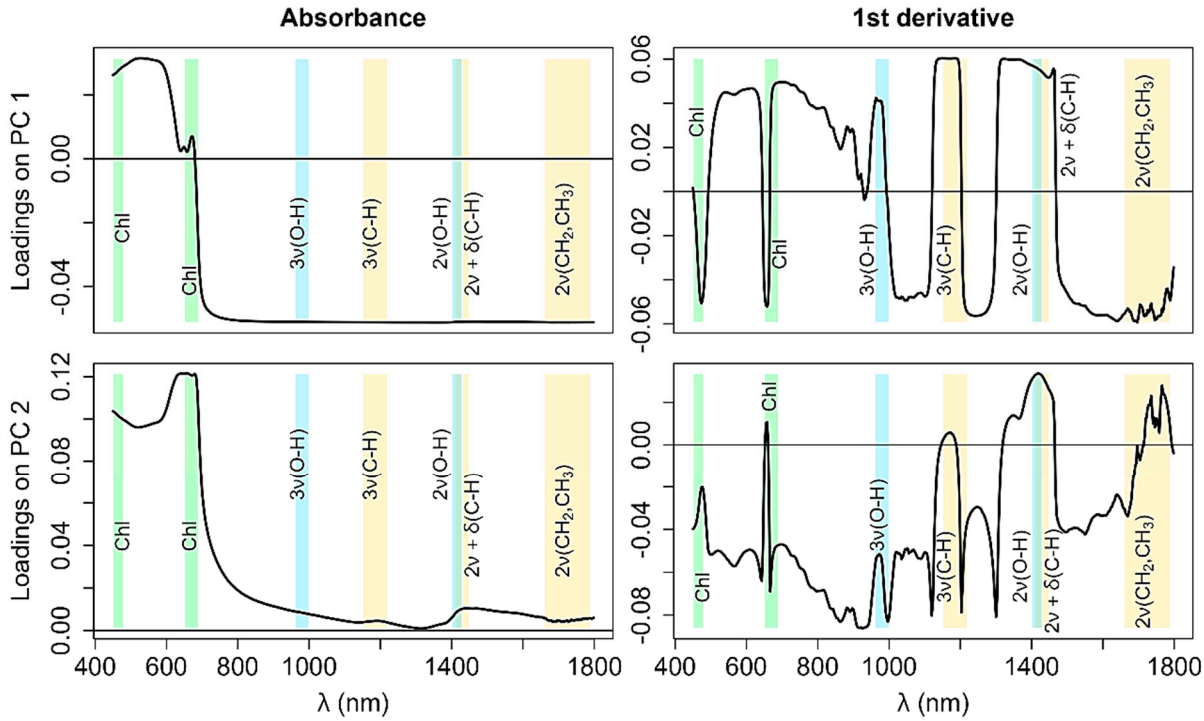


**Figure 8.** Loadings of the first (**top**) and second (**bottom**) principal components for the PCA decomposition of the raw reflectance (**left**) and first derivative of absorbance (**right**) datasets.

The loadings for PC2 of raw data (bottom left) put again the weight on the visible range and a much smaller weight on the features around 1400 nm. Interestingly, PC1 attributes the highest loadings near the Chlorophyll blue peak, while PC2 attributes the highest loadings to the red peak.

The loadings associated with the derivative of the absorbance are more difficult to interpret, but essentially, they align with the main spectroscopic bands. The interpretation must be made in terms of first derivative peaks = wavelengths with more pronounced change. For example, for PC1, the raw seeds have the more negative loadings and the laboratory-treated the more positive. Therefore, the negative peaks ("dips", or "valleys") in PC1 mean more intense changes in the raw seeds and positive peaks mean more intense changes in laboratory-treated seeds. There are two narrow dips in the Chlorophyll bands meaning, as expected, large variations in the raw seeds' spectra in that region. On the other hand, positive peaks at $3\nu$(O-H), $3\nu$(C-H), $2\nu$(O-H) and $2\nu + \delta$(C-H) mean stronger variations in these bands on the laboratory-treated spectra, which is also evident from Figure 3.

PC1 describes a continuous transformation raw seed → industry → laboratory, but PC2 describes the difference between raw seed + laboratory, on one side, and industry treated seeds, on the other. The main source of differences is in negative loadings and in the band 950–1300 nm. This is also the core range for the spectral differences in $K_s$ (Figure 5). Globally, this suggests that the main difference in the chemical treatments between industry and laboratory treated seeds is to be found in this band.

As stressed in the introduction, the main objective of this work was to demonstrate the proof of principle of using DRS to assess dehusking efficiency. However, we recognize such a method is only relevant if accompanied by a plausible route to industrial application. One possible implementation would involve a spectrometer positioned above a conveyor belt carrying the seeds, acquiring their reflectance spectra in real time. From the data, the average dehusking efficiency could be estimated using a pre-calibrated model. As with all spectroscopic models, calibration requires a set of reference ("golden standard") sam-

ples with known dehusking efficiencies, determined individually by microscopic and/or photographic methods that quantify residual seed coat material. The calibration model would then establish a relationship between spectral features and dehusking efficiency. This could rely on a simple linear approach, such as the ratio between reflectance of raw and dehusked seeds at selected wavelengths, or on more advanced multivariate methods using all or part of the spectra, for example, partial least squares (PLS).

## 4. Conclusions

This work highlights the strong potential of DRS coupled to KM analysis as a rapid and non-destructive method for assessing dehusking efficiency in carob seeds. The spectral data revealed that the seed husk acts essentially as a broadband optical reflector, with its removal leading to significant decreases in reflectance, especially in the NIR range. Laboratory-treated seeds showed the lowest reflectance and highest absorbance, indicating more effective husk removal than the well-implemented industrial process. KM calculations allowed us to estimate the contribution to the reflectance of the husk tissues removed and confirmed this picture. PCA analysis confirmed clearer clustering and lower variability in lab-processed seeds, suggesting improved process control. Moreover, the second-derivative analysis further highlighted key absorption bands (i.e., 960, 1200, and 1400 nm), supporting enhanced light penetration and better kernel exposure in lab-treated samples, without signs of chemical alteration of endosperm polysaccharides. The stability of spectral features across treatments confirms that the kernel's integrity is preserved. Altogether, these findings establish DRS as a powerful, scalable platform for non-invasive diagnostics in seed processing. Its sensitivity to subtle changes in surface and internal structure makes it ideal for evaluating processes such as peeling, roasting, drying, or fermentation in fruits, seeds, and grains. Moreover, DRS could serve as a valuable tool in plant breeding programs and as a real-time quality control method for sustainable extraction processes in agro-industrial operations.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/polysaccharides6040095/s1, Figure S1. First derivative of the absorbance spectra for native seeds (black curves), commercially dehusked seeds obtained through the industrial process (red curves), and seeds dehusked using the method developed by us (green curves). Five spectra per seed type are shown. The vertical bands show the principal absorption bands (see Figure 1 for details). Figure S2. PCA analysis of the reflectance data. Left: from raw reflectance. Right: from the second derivative of absorbance data. The three clusters highlighted correspond to the native seeds (black cluster), commercial dehusked seeds obtained with the industrial process (red cluster) and dehusked seeds with our developed method (blue cluster). Each spectral variable was mean-centered and normalized to the standard deviation. Figure S3. Loadings of the first (top) and second (bottom) principal components for the PCA decomposition of the second derivative of absorbance.

**Author Contributions:** Conceptualization, B.M. and R.G.; methodology, A.B. and R.G.; software, R.G.; validation, B.M. and R.G.; formal analysis, R.G.; investigation, A.B. and R.G.; resources, R.G., S.G. and A.R.; writing—original draft preparation, R.G. and B.M.; writing—review and editing, B.M., S.G., A.R. and R.G.; supervision, B.M. and R.G.; project administration, R.G. and A.R.; funding acquisition, R.G., S.G. and A.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by National Funds through FCT—Foundation for Science and Technology under the Project "BIG opportunities in a small SEED: valorization of carob byproducts into novel hydrogels for sustainable agriculture" (2022.07519.PTDC). FCT is further acknowledged for the projects MED (https://doi.org/10.54499/UIDB/05183/2020 In Proceedings of the 5183/2020), CEOT (UIDB/00631/2020 CEOT BASE), CHANGE (https://doi.org/10.54499/LA/P/0121/2020) and researcher grants (https://doi.org/10.54499/CEECIND/01014/2018/CP1540/CT0002; CEECINST/00052/2021).

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** The original contributions presented in this study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

**Acknowledgments:** The company "A Industrial Farense, Lda" (Faro, Portugal) is acknowledged for its support during the development of this work.

**Conflicts of Interest:** The company "A Industrial Farense, Lda" (Faro, Portugal) is acknowledged for its support during the development of this work. The company was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| LBG | Locust bean gum |
| DRS | Diffuse Reflectance Spectroscopy |
| KM | Kubelka–Munk |
| PCA | Principal Component Analysis |
| $H_2SO_4$ | Sulfuric acid |
| $CH_4O_3S$ | Methanesulfonic acid |
| Vis | Visible |
| NIR | Near Infrared |
| $D$ | Dark signal |
| $Ref$ | Reference signal |
| $R$ | Reflectance |
| $A$ | Absorbance |
| $S$ | Scattering coefficient |
| $R_\infty$ | Diffuse reflectance |
| $R_f$ | Flesh reflectance |
| $R_{sf}$ | Cumulative reflectance of seed skin and flesh |
| $d$ | Skin thickness |
| $S_s$ | Scattering coefficient of the seed skin |
| $K_s$ | Absorption coefficient of the seed skin |
| $\lambda_0$ | Reference wavelength |
| $b$ | Scattering power |
| $R_{s,\infty}$ | Reflectance of the seed skin |

## References

1. Martins-Loução, M.A.; Correia, P.J.; Romano, A. Carob: A Mediterranean Resource for the Future. *Plants* **2024**, *13*, 1188. [CrossRef] [PubMed]
2. Rodríguez-Solana, R.; Romano, A.; Moreno-Rojas, J.M. Carob Pulp: A Nutritional and Functional By-Product Worldwide Spread in the Formulation of Different Food Products and Beverages. A Review. *Processes* **2021**, *9*, 1146. [CrossRef]
3. Basharat, Z.; Afzaal, M.; Saeed, F.; Islam, F.; Hussain, M.; Ikram, A.; Pervaiz, M.U.; Awuchi, C.G. Nutritional and Functional Profile of Carob Bean (*Ceratonia siliqua*): A Comprehensive Review. *Int. J. Food Prop.* **2023**, *26*, 389–413. [CrossRef]
4. Correia, P.J.; Pestana, M. Sugars and Phenols in Carob Tree Fruits from Different Producing Countries: A Short Review. *Heliyon* **2024**, *10*, e30922. [CrossRef]
5. Mazaheri, D.; Shojaosadati, S.A.; Mousavi, S.M.; Hejazi, P.; Saharkhiz, S. Bioethanol Production from Carob Pods by Solid-State Fermentation with Zymomonas Mobilis. *Appl. Energy* **2012**, *99*, 372–378. [CrossRef]
6. Sánchez, S.; Lozano, L.J.; Godínez, C.; Juan, D.; Pérez, A.; Hernández, F.J. Carob Pod as a Feedstock for the Production of Bioethanol in Mediterranean Areas. *Appl. Energy* **2010**, *87*, 3417–3424. [CrossRef]

7. Barak, S.; Mudgil, D. Locust Bean Gum: Processing, Properties and Food Applications—A Review. *Int. J. Biol. Macromol.* **2014**, *66*, 74–80. [CrossRef]
8. Prajapati, V.D.; Jani, G.K.; Moradiya, N.G.; Randeria, N.P.; Nagar, B.J. Locust Bean Gum: A Versatile Biopolymer. *Carbohydr. Polym.* **2013**, *94*, 814–821. [CrossRef]
9. Jo, W.; Bak, J.H.; Yoo, B. Rheological Characterizations of Concentrated Binary Gum Mixtures with Xanthan Gum and Galactomannans. *Int. J. Biol. Macromol.* **2018**, *114*, 263–269. [CrossRef]
10. Camacho, M.M.; Martínez-Navarrete, N.; Chiralt, A. Rheological Characterization of Experimental Dairy Creams Formulated with Locust Bean Gum (LBG) and λ-Carrageenan Combinations. *Int. Dairy J.* **2005**, *15*, 243–248. [CrossRef]
11. Modrackova, N.; Makovska, M.; Mekadim, C.; Vlkova, E.; Tejnecky, V.; Bolechova, P.; Bunesova, V. Prebiotic Potential of Natural Gums and Starch for Bifidobacteria of Variable Origins. *Bioact. Carbohydrates Diet. Fibre* **2019**, *20*, 100199. [CrossRef]
12. Soumya, R.S.; Raghu, K.G.; Abraham, A. Locust Bean Gum – A Potential Drug Delivery Carrier. In *Polysaccharide-based Biomaterials*; The Royal Society of Chemistry: London, UK, 2022; pp. 247–268.
13. Chen, Y.; Guo, J.; Alamri, A.S.; Alhomrani, M.; Huang, Z.; Zhang, W. Recent Research Progress on Locust Bean Gum (LBG)-Based Composite Films for Food Packaging. *Carbohydr. Polym.* **2025**, *348*, 122815. [CrossRef]
14. Petitjean, M.; Isasi, J.R. Locust Bean Gum, a Vegetable Hydrocolloid with Industrial and Biopharmaceutical Applications. *Molecules* **2022**, *27*, 8265. [CrossRef] [PubMed]
15. Dionísio, M.; Grenha, A. Locust Bean Gum: Exploring Its Potential for Biopharmaceutical Applications. *J. Pharm. Bioallied Sci.* **2012**, *4*, 175. [CrossRef] [PubMed]
16. Brassesco, M.E.; Brandão, T.R.S.; Silva, C.L.M.; Pintado, M. Carob Bean (*Ceratonia siliqua* L.): A New Perspective for Functional Food. *Trends Food Sci. Technol.* **2021**, *114*, 310–322. [CrossRef]
17. Lopes da Silva, J.A.; Gonçalves, M.P. Studies on a Purification Method for Locust Bean Gum by Precipitation with Isopropanol. *Food Hydrocoll.* **1990**, *4*, 277–287. [CrossRef]
18. Wielinga, W. Seed Gums. In *Food Stabilisers, Thickeners and Gelling Agents*; Wiley: Hoboken, NJ, USA, 2009; pp. 275–292.
19. Medronho, B.; Boutoub, O.; Duarte, H.; Aliaño-González, M.J.; Guerra, R.; Brázio, A.; Gonçalves, S.; Romano, A. On the Optimization of Carob Seed Peel Extraction Using Aqueous-Based Acidic Systems. *Molecules* **2025**, *30*, 1397. [CrossRef]
20. Zahir, S.A.D.M.; Omar, A.F.; Jamlos, M.F.; Azmi, M.A.M.; Muncan, J. A Review of Visible and Near-Infrared (Vis-NIR) Spectroscopy Application in Plant Stress Detection. *Sensors. Actuators A Phys.* **2022**, *338*, 113468. [CrossRef]
21. Cavaco, A.M.; Pinto, P.; Antunes, M.D.; da Silva, J.M.; Guerra, R. 'Rocha' Pear Firmness Predicted by a Vis/NIR Segmented Model. *Postharvest Biol. Technol.* **2009**, *51*, 311–319. [CrossRef]
22. Wang, M.; Xu, Y.; Yang, Y.; Mu, B.; Nikitina, M.A.; Xiao, X. Vis/NIR Optical Biosensors Applications for Fruit Monitoring. *Biosens. Bioelectron. X* **2022**, *11*, 100197. [CrossRef]
23. Huang, Q.; Yang, M.; Ouyang, L.; Wang, Z.; Lin, J. Vis/NIR Spectroscopy and Chemometrics for Non-Destructive Estimation of Chlorophyll Content in Different Plant Leaves. *Sensors* **2025**, *25*, 1673. [CrossRef] [PubMed]
24. Walsh, K.B.; Blasco, J.; Zude-Sasse, M.; Sun, X. Visible-NIR 'Point' Spectroscopy in Postharvest Fruit and Vegetable Assessment: The Science behind Three Decades of Commercial Use. *Postharvest Biol. Technol.* **2020**, *168*, 111246. [CrossRef]
25. Ibrahim, A.; Alghannam, A.; Eissa, A.; Firtha, F.; Kaszab, T.; Kovacs, Z.; Helyes, L. Preliminary Study for Inspecting Moisture Content, Dry Matter Content, and Firmness Parameters of Two Date Cultivars Using an NIR Hyperspectral Imaging System. *Front. Bioeng. Biotechnol.* **2021**, *9*, 720630. [CrossRef] [PubMed]
26. Kubelka, P.; Munk, F. Ein Beitrag Zur Optik Der Farbanstriche. *Z. Tech. Phys.* **1931**, *12*, 593–601.
27. Budiastra, I.W.; Ikeda, Y.; Nishizu, T. Optical Methods for Quality Evaluation of Fruits (Part 1). *J. Jpn. Soc. Agric. Mach.* **1998**, *60*, 117–128. [CrossRef]
28. Vogelmann, T.C. Plant Tissue Optics. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1993**, *44*, 231–251. [CrossRef]
29. Passos, D.; Hebden, J.C.; Pinto, P.N.; Guerra, R. Tissue Phantom for Optical Diagnostics Based on a Suspension of Microspheres with a Fractal Size Distribution. *J. Biomed. Opt.* **2005**, *10*, 064036. [CrossRef]
30. Hu, D.; Fu, X.; Wang, A.; Ying, Y. Measurement Methods for Optical Absorption and Scattering Properties of Fruits and Vegetables. *Trans. ASABE* **2015**, *58*, 1387–1401. [CrossRef]
31. Saeys, W.; Velazco-Roa, M.A.; Thennadil, S.N.; Ramon, H.; Nicolaï, B.M. Optical Properties of Apple Skin and Flesh in the Wavelength Range from 350 to 2200 Nm. *Appl. Opt.* **2008**, *47*, 908. [CrossRef]
32. Wang, W.; Li, C. The Optical Properties of Onion Dry Skin and Flesh at the Wavelength 632.8 Nm. *Proc. SPIE—Int. Soc. Opt. Eng.* **2012**, *8369*, 107–115.
33. R Core Team. *R: A Language and Environment for Statistical Computing*; R Core Team, R Foundation for Statistical Computing: Vienna, Austria, 2022; Available online: https://www.R-project.org/ (accessed on 6 August 2025).

34. Signal Developers. Signal: Signal Processing. Available online: https://cran.r-project.org/web/packages/signal/index.html (accessed on 6 August 2025).
35. Tong, P.; Du, Y.; Zheng, K.; Wu, T.; Wang, J. Improvement of NIR Model by Fractional Order Savitzky–Golay Derivation (FOSGD) Coupled with Wavelength Selection. *Chemom. Intell. Lab. Syst.* **2015**, *143*, 40–48. [CrossRef]