



Near-infrared hyperspectral imaging for polymer particle size estimation

Lukasz Pieszczek, Michal Daszykowski*

Institute of Chemistry, University of Silesia in Katowice, 9 Szkolna Street, 40-006 Katowice, Poland

ARTICLE INFO

Keywords:

Particle size distribution
Hyperspectral imaging
Polymer detection
Polymer processing
Chemometrics

ABSTRACT

This study examines the potential of near-infrared hyperspectral imaging for assessing the size of polymer particles in model fractions based on the scattering phenomena. Different fractions of ground polymers, either polymethyl methacrylate or polypropylene, were characterized by near-infrared spectra collected between 900 and 1700 nm. The possibility to estimate the size of polymer particles using hyperspectral images was confronted with a basic single spot near-infrared measurement. Hyperspectral imaging, in addition to the standard spectral data dimension, provides information about the spatial distribution of sample components and reveals changes in physical properties. Therefore, one can gain a better insight into the scattering phenomena and study the physical inhomogeneity of a sample in terms of particle size distribution. The partial least-squares models constructed to estimate particle size of polymers that were characterized by hyperspectral images (a pixel-based approach) outperforms models built for mean spectra regardless of the considered powdered polymer.

1. Introduction

Light scattering is a natural phenomenon that occurs in spectroscopic signals. Like any physical interference, it decreases the intensity of the chemically relevant analytical signal. Generally, these undesired interferences are generated by the physical inhomogeneity of a sample. Surface roughness is a good example of physical inhomogeneity. It is caused by droplets, dust grains, bubbles, fibers, fluctuations in density, defects in the crystal structure and any cells or microorganelles that are present in biological samples [1]. All of the physical features contribute to the inherent characteristics of solid samples, the analysis of which is becoming increasingly popular. An analysis of sample composition is facilitated by using advanced high-quality near-infrared imaging, which operates in the so-called near-infrared (NIR) spectral region (between 800 and 2500 nm), provides non-destructive, rapid measurements and requires very little sample preparation. The NIR spectra are mainly formed through combinations of the overtones of the fundamental bond vibrations, which are primarily present in the mid-infrared spectral range and the scattering component often affects the shape of the spectra.

In contrast to other spectroscopic signals, scattering can substantially mask the analytical information in the NIR range. This is why many preprocessing methods have been developed to suppress the scattering effect [1,2]. Multiplicative scatter correction (MSC) [3] and inverted/inverse scatter correction (ISC) [4], as well as their modifications, i.e.,

the extended multiplicative scatter correction (EMSC) [5] and the extended inverted/inverse scatter correction (EISC) [6], represent the well-established applied methods for scattering correction.

However, in the last years, some articles have been published, indicating that basic spectral preprocessing techniques can negatively affect the prediction performance of a model [7–9]. Therefore, novel and more robust techniques aiming at reducing undesired sources of spectral variability have been proposed, including normalization variants such as, for instance, variable sorting for normalization (VSN) [10] – a weighted modification of standard normal variate (SNV) [11]. More comprehensive techniques are based on the concept of orthogonalized spectral pretreatment, for instance, sequential preprocessing through orthogonalization (SPORT) [12] and parallel preprocessing through orthogonalization (PORTO) [13].

Although the scattering phenomenon represents an undesired part of the analytical signal in standard applications and most efforts are focused on reducing it, scattering can also be used to characterize a sample. As is discussed in the literature, NIR reflectance spectroscopy can be used to determine the size of the particles in powdered and ground samples by quantifying the scattering effect [14]. This approach has large implementation potential in the process analytical technology (quality control of tablets and granules) [8,15–18].

A novel possibility for estimating the physical properties of bulk powders that is based on the reflectance spectra that are collected in the Vis/near-infrared region has recently been developed [19,20]. A smart

* Corresponding author.

E-mail address: michal.daszykowski@us.edu.pl (M. Daszykowski).

double-integrating sphere setup [21] and the concept of measuring a sample in both the reflectance and transmittance modes simultaneously provide more in-depth information about a powder sample and even enables prediction of particles distribution [19]; however, a sample has to be suspended in a solvent before being measured.

To date, many studies that associate the particle size of powdered samples with the scattering component of the NIR spectra have been undertaken. However, in these studies, a sample is usually described by a single spectrum that cannot reflect surface heterogeneity and particle size distribution. In standard NIR applications, the reflectance spectra are collected to reduce a solid sample's heterogeneity, for instance, by spinning a sample to obtain a representative spectrum, but then information about the heterogeneity is inevitably lost. The near-infrared hyperspectral imaging (NIR-HSI) is an extension of the classic NIR, which enables the sample surface to be characterized and the homogeneity fluctuations to be visualized. A sample is described by many reflectance spectra. They are recorded at measurement points (pixels) that are distributed over the entire sample surface. Consequently, a single spectrum can be replaced by hundreds of reflectance spectra that resemble the physical variability of a sample surface. NIR-HSI has found many applications in analytical chemistry and process analytical technology, primarily for solving quality control issues [22–24].

Adequately designed experiments with hyperspectral imaging detection provide insight into the effect of scattering on the predictive potential of multivariate models. In this study, we examined how the scattering affects the calibration models estimating particle size of powdered polymethyl methacrylate or polypropylene. The different physicochemical parameters of these two thermoplastic polymers, including the particle size, affect the melting process. Polymethyl methacrylate is an amorphous resin that has a randomly ordered molecular structure and melts steadily as the temperature rises, even below the melting point. Polypropylene is a semi-crystalline resin and remains hard until the melting temperature is reached. The observations that were made during the experiment lead to more general conclusions regarding the usefulness of the NIR-HSI approach for online applications. Estimating particle size is essential not only for pharmaceutical technology. In the polymer industry, many products are obtained by different compression molding methods (e.g., injection molding, compression molding, powder injection molding, etc.). The feedstock that is used for these processes is usually supplied as cylindrical pellets or resin powders [25,26]. The geometry and size of the raw material affect the uniformity of melted resin and determine the final quality of a product. In different industrial branches, fine-powder grade polymer feedstock is preferred for compression molding and it is used to avoid any granular boundary fragments in the molded resin [26]. Therefore, developing new remote and online applications for estimating powder particle size can enhance quality control in the polymer industry.

2. Theory

2.1. Popular techniques for estimation of particle size

The dynamic light scattering (DLS) method [27,28] is often used to determine particle size and evaluate scattering. DLS is based on the diffusion behavior of the particles in a suspension that is illuminated with a laser beam. During the measurement, any changes in the scattering intensity are recorded. Later, the Stokes-Einstein relationship is used to estimate the size of the particles. Despite the many applications for DLS, it has limitations. The multiple scattering affects the measurements for large particles and particles that are characterized by a high refractive index. Additionally, DLS is imprecise when samples are represented by particles with a wide range of diameters. Therefore, before analysis, any polydisperse mixtures have to be separated and a sample must be dispersed in a liquid. Laser diffraction (LD) [29,30] overcomes this drawback. The intensity of particle scattering is recorded as a function of the scattering angle, which enables examining the

suspensions and solid powders by dry dispersion. However, dry dispersion is not readily available in standard LD instruments and requires an additional sample flow accessory (see, e.g., LA-960 PowderJet, Horiba or Mastersizer 3000 with the Areo S disperser, Malvern Panalytical). The LD technique was initially developed to analyze particles in suspensions. The measurements of dry powders are less accurate compared to the results that are obtained for samples suspended in a liquid. Moreover, because of the requirement of expensive detectors and a laser source, all laser diffraction analyzers are relatively expensive. These shortcomings are partially solved in dynamic image analysis (DIA) [31–34]. It is usually based on analyzing images of particles that have been separated using the free-fall method. The falling samples are illuminated with powerful light pulses that reduce blurring and support the registering of sharp images of the particles. DIA also offers the possibility to estimate the size and shape of a particle. This technique is not based on the light-scattering phenomena and can analyze particles with diameters that range from a few micrometers to several millimeters (e.g., Analysette 28 ImageSizer, Fritsch, Camsizer X2, Retsch Technology GmbH or QICPIC particle size and shape analyzer, Sympatec GmbH).

Size estimation of tiny particles (nanometer scale) can also be done using transmission electron microscopy (TEM) followed by image analysis. Microscopy is frequently selected to examine nanoparticles and even can assist in examining particle size distribution [35]. Unfortunately, TEM cannot be considered a high-throughput technique, and thus, due to the effective measurement range of particle sizes, it finds many applications in nanotechnology.

A broader application potential arises from instrumental configurations based on inexpensive and widely available digital cameras. The possibility of particle size estimation of sand-sized materials, e.g., sediments [36] and soils [37] has been shown in different studies. Such an approach to a sample examination requires image analysis methods based on pattern spectra descriptors or other image statistics. To this day, the potential of standard digital cameras for examining particle size in the range of millimeters has been investigated [37].

The main direction of development in estimating particle size based on the scattering phenomena involves lowering the price and reducing the size of the instruments. Over the past few years, CMOS and CCD sensors have replaced the expensive detectors in LD analyzers, and highly precise lasers have been substituted with inexpensive light-emitting diodes (LED). Examples of new devices that are equipped with these accessories include a smartphone camera sensor [38] or an angular spatial filter (ASF) [39]. In addition, chemometric methods are being used to predict particle size because they enable complex scattering correction models to be replaced and decrease the computation time [39].

2.2. Near-infrared spectroscopy and near-infrared hyperspectral imaging

NIR spectroscopy is a different approach to estimating particle size. NIR spectra have two components: absorption and scattering. Usually, the primary spectral information is related to absorption (and thereby, to the chemical content of a sample). Thus, NIR has been ignored as a technique for predicting particle size [27]. However, the NIR spectra reflect scattering and can help to estimate the diameter of particles in solid samples [1,14–18]. In order to extract particle scattering information from the NIR spectra, a reference set of samples with different particle sizes must be measured. The idea is to reduce every spectrum to a few latent variables, which are a linear combination of the original spectrum intensities (e.g., the principal components that are obtained from a principal component analysis, PCA) and then calculate the regression coefficients between these latent variables and the particle size of the sample using a model. The model is then used to predict particle size in a new material based on the corresponding spectrum. The limitation in NIR spectroscopy arises from the assumption of sample homogeneity that excuses sample characterization by a single spectrum. However, local fluctuations of homogeneity can distort the final model

and eventually restrict its prediction properties. Near-infrared imaging was developed to overcome this limitation and combine the advantages of NIR spectroscopy such as rapid, non-destructive, and remote measurements of solid samples that require no extensive preparation with the potential of hyperspectral imaging. It opened the possibility to evaluate the physical heterogeneity of powder samples (including the size of the particles) through sampling the pixels of an image and performing a statistical analysis of these. Based on the results and constructed models, the particle size can then be inferred and the scattering effect for tested samples can be quantified. NIR-HSI data consist of multiple images that were recorded at subsequent wavelengths within the selected spectral range. Hyperspectral imaging focuses on the region of interest of the sample surface (see Fig. 1). Therefore, compared to basic NIR spectroscopy, the possibilities of HSI are broader, especially when information about material defects or chemical homogeneity is vital. Most HSI applications are analytical and the possibility of evaluating physical properties has been explored less. To the best of our knowledge, HSI has not been extensively used to predict particle size based on the scattering effect. In contrast to DIA, NIR-HSI measurements do not require the particles to be separated, and therefore, the measurement system is less complex and more efficient. Moreover, using the chemometric methods, hyperspectral data can be preprocessed online, and information about the sample is readily available. These features make NIR-HSI attractive for online industrial quality control applications. Unlike the other instrumental techniques described in Section 2.1, HSI is mainly used to characterize the chemical content of a sample. However, based on the reflectance spectra, the chemical and physical properties can be analyzed simultaneously. All of these NIR-HSI features make this technique very promising for examining scattering, relating the scattering with the physical heterogeneity of the sample surface.

3. Experimental section

3.1. Samples preparation

Polypropylene (PP) and polymethyl methacrylate (PMMA) pellets were used as the subjects of the study. Each polymer was ground for two minutes in an FW80 laboratory grinder (Chemland, Stargard Szczecinski, Poland). The obtained powder was sieved manually using nine different sieves with mesh sizes of 1500, 600, 385, 300, 250, 200, 150, 100, and 75 μm , respectively. Eight fractions of the polymer particles were obtained: (1) <1500 and >600, (2) <600 and >385, (3) <385 and >300, (4) <300 and >250, (5) <250 and >200, (6) <200 and >150, (7) <150 and >100, and (8) <100 and >75 μm . Three individual samples were prepared for each fraction. The powdered polymer samples were placed in cylindrical metal cuvettes (8 mm diameter and 2 mm deep). The surface of each sample was flattened and leveled to the upper cuvette rim by gentle tapping.

3.2. Hyperspectral camera and image acquisition settings

Images of the different polymer particle fractions were registered using a Specim FX17e hyperspectral camera (Specim, Spectral Imaging Ltd., Oulu, Finland), which was used in the push-broom mode (640 pixels were captured during each single line scan of an image). For each pixel, the reflectance was measured at 224 equally distributed spectral channels from 935.61 to 1720.2 nm. The region of interest was illuminated by three 50 W tungsten-halogen lamps. The measurement stage with the studied samples was attached to the top of a laboratory scanner, 10 cm below the camera lens and moved steadily during image acquisition. A total of four images were collected. In two subsequent measurements, the image of each set of cuvettes with one type of polymer was collected. The cuvettes were arranged perpendicular to the illumination source to ensure the most homogeneous light illumination. The measurement parameters were adjusted to obtain hyperspectral images with the highest possible signal-to-noise ratio as well as to maintain the appropriate aspect ratio in the resulting image (in x- and y) directions. Therefore, the exposure time, the frame rate and the scanning speed were set to 5 ms, 190 Hz, and 19 $\text{mm}\cdot\text{s}^{-1}$, respectively.

3.3. Light calibration, image masking, and spectral trimming

The light intensity for all of the images was normalized to unity [40] using a Teflon calibration tile and closed camera shutter as a white and dark reference, respectively. The irrelevant image region was trimmed on both sides. The trimmed images were then merged in pairs to collect samples of one type of polymer in a single data cube and facilitate further analysis and interpretation. The dimensions of each data cube were $1000 \times 322 \times 224$ and $1022 \times 322 \times 224$ for the PMMA and PP samples, respectively.

The image foreground and background were detected using adaptive image thresholding with the Gaussian-weighted mean in the neighborhood [40]. The procedure was performed using the MATLAB built-in function 'adaptthresh.m' [41,42]. The input parameters such as the sensitivity factor and the neighborhood size were set to 0.3 and 25, accordingly. In the original hyperspectral image, two redundant components were visible, namely a black rubbery pad and fragments of the metal cuvettes, and therefore these image features were removed. To effectively mask any undesired image components, the image that was used as the input for the 'adaptthresh.m' function was enhanced to expose the polymer powder better. The improved input image for each set of samples was obtained by subtracting two two-dimensional images from the corresponding hyperspectral cube. These images were selected arbitrarily using two wavelengths with different spectral intensities in the polymer spectrum. As a result, the intensity of polymer pixels was increased over the remaining image components because of the image subtraction, which made the further background masking more

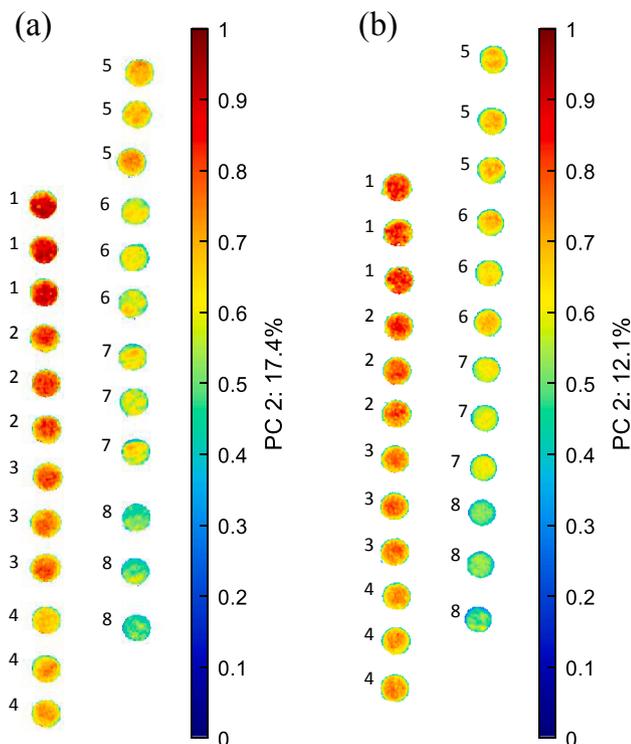


Fig. 1. Image of the score values of the first principal component that were obtained from the original hyperspectral images for fractions of: (a) PMMA and (b) PP.

accurate. For PMMA, the input image was obtained by subtracting the images that were recorded at 1255.9 and 1175.4 nm and for PP, the images that were recorded at 1315.6 and 1189.4 nm. Before any modeling procedure, the data's spectral dimension was trimmed to a 966.7–1688 nm spectral range in order to eliminate noisy wavelengths.

In certain situations, a shiny surface can reflect light stronger than the reference Teflon calibration tile and the intensities of specific spectra are larger than the normalized signal values. Therefore, pixels that corresponded to the spectra with intensity values that exceeded 0–1 were masked. Moreover, the noisiest pixels (the ones with the highest standard deviation were calculated for each spectrum that had previously been subjected to the second derivative) were also eliminated. From each sample (cuvette), one thousand spectra (pixels) with the lowest noise were selected for further analysis.

A more detailed discussion concerning the selection of the region of interest (ROI) and the detection of extreme pixels can be found in Ref. [40], where consecutive steps of image masking correspond to image processing carried out in this study.

3.4. Software used for image collection and data processing

The hyperspectral camera and the laboratory scanner were controlled using Lumo Scanner software (Specim, Spectral Imaging Ltd., Oulu, Finland). All of the image processing steps and analyses were performed in MATLAB R2019a (MathWorks, Natick, MA, USA) using in-house implemented algorithms.

4. Results and discussion

4.1. Initial data exploration

In the first step, the hyperspectral data were explored in an unsupervised manner. A visual evaluation of the chemical inhomogeneity, the level of signal noise and incidental spectroscopic artifacts of various origins and structures (the so-called dead pixels, signal spikes and spectra that exceed the spectral reflection intensity of the reference materials) are essential because these image components may significantly affect the final calibration model. Since the experiment was designed to reduce the impact of the chemical inhomogeneity (pure polymer samples were ground and sieved), we expect that the most considerable influence would correspond to the physical variability of the bulk powder, instrumental noise and spectroscopic artifacts that occasionally appear in images. Because of the lack of chemical variability, the artifact contributions to the overall data variance are substantially greater than in most of the analytical experiments based on hyperspectral imaging that have been conducted to date. On the other hand, incidental artifacts can alter the shape of spectral profiles. Such outlying spectra must be removed before constructing any least-squares model [43]. Therefore, the images were carefully examined before further processing.

Individual PPMA and PP samples, which are visualized as images using the PCA scores, are interesting because they can reveal possible trends between particle size and a reduced spectral representation. The first principal component described most of the data variance (more than 80%), while the second principal component modeled the variability related to the scattering well. This effect became apparent for PPMA and PP samples when the corresponding pseudo images of the score values of the second principal component, which is shown in Fig. 1, was analyzed. The colors of the image pixels changed proportionally to the score values and the size of the particles that constituted the different fractions. For PPMA and PP, the second principal component captured 17.4% and 12.1% of the total data variability. In Fig. 1, the colors of the pixels and their hue changed from navy blue to dark red. It was also readily apparent that the polymer samples in a fraction were more similar to each other in their color intensity and hue than the particles from the other fractions.

Fig. 1 presents the image regions in which the sample's homogeneity fluctuated. They were formed by groups of eye-catching pixels that appeared as spots or rings. Because of the mixing effect, the tiny particles were close to the cuvette rims in all of the fractions regardless of the polymer type (see the characteristic emerald ring around each cuvette in Fig. 1). The possibility to trace sample heterogeneity spatially is the substantial advantage of NIR-HSI over basic "single spot" reflectance NIR spectroscopy.

4.2. Two approaches for modeling the HSI data

Hyperspectral data can be examined using the spectral information from each pixel independently as is presented in Fig. 1. Each region of the sample surface was represented by image pixels of a given color intensity and hue. In this way, the spatial inhomogeneity of a sample can be readily exposed, which is the most valuable advantage of the hyperspectral technique. However, in some cases, the noise and spatial variability distort the NIR spectra and hamper their interpretation. Therefore, an analysis of the averaged spectra that describe a selected image object might be more convenient and sufficient. Moreover, such data representation diminishes the influence of noisy spectra by averaging the spectra that correspond to the pixels that describe each cuvette. In our study, such an approach was considered to mimic the measurements using a traditional NIR spectrometer. On the other hand, the averaged spectrum that is obtained for the selected region is more reliable than the one recorded using a standard NIR spectrometer because of the larger surface sampling.

The trend between the scores of the first principal component and the average polymer particle size are presented in Fig. 2.

There were a few interesting observations. The trend that can be observed in Fig. 1 was less evident when the averaged spectra were used to derive the principal components (see Fig. 2). Generally, the relationship between the different diameters of the fractions and the corresponding score values was maintained. The score plots that are presented in Fig. 2 demonstrate that the fractions that contained the finest particles of PMMA or PP did not follow the overall trend.

For the averaged spectra, the first principal component explained a more significant portion of the total data variance compared to the first principal component that was obtained for the set of the spectra of all of the pixels. For the PMMA and PP samples, it accounted for 92.43% and 95.68%, respectively. Moreover, unlike the pixel representation of the data, the first principal component explained the particle size in the different fractions most accurately for the averaged data representation. This observation was not surprising since the contribution that arises from the physical heterogeneity is minimized by the averaging and the primary source of variance is related to the systematic differences in the particle sizes. However, this did not improve the particle size predictions for the averaged spectra, which indicated that a small part of the data variability supported the modeling. The relationship between the first principal component scores and the particle size was nonlinear, and individual principal components may not carry relevant information to explain the scattering phenomena and related secondary physical features of the powder fractions sufficiently.

In Fig. 2a2 and b2, two loading vectors of the first principal component were plotted for PMMA and PP samples, respectively. Their values can be compared with the corresponding polymer spectra, displayed in Fig. 2a3 and b3. It is worth noting that large absolute values indicate informative wavelength channels. They can be associated with peaks and spectral features in the NIR spectrum that explain the scattering effect related to particle size to the largest extent.

4.3. Modeling particle size using partial least-squares regression

Different partial least-squares models (PLS) that had an increasing number of latent variables have been constructed to study the effect of preprocessing on eliminating the scattering component. They were

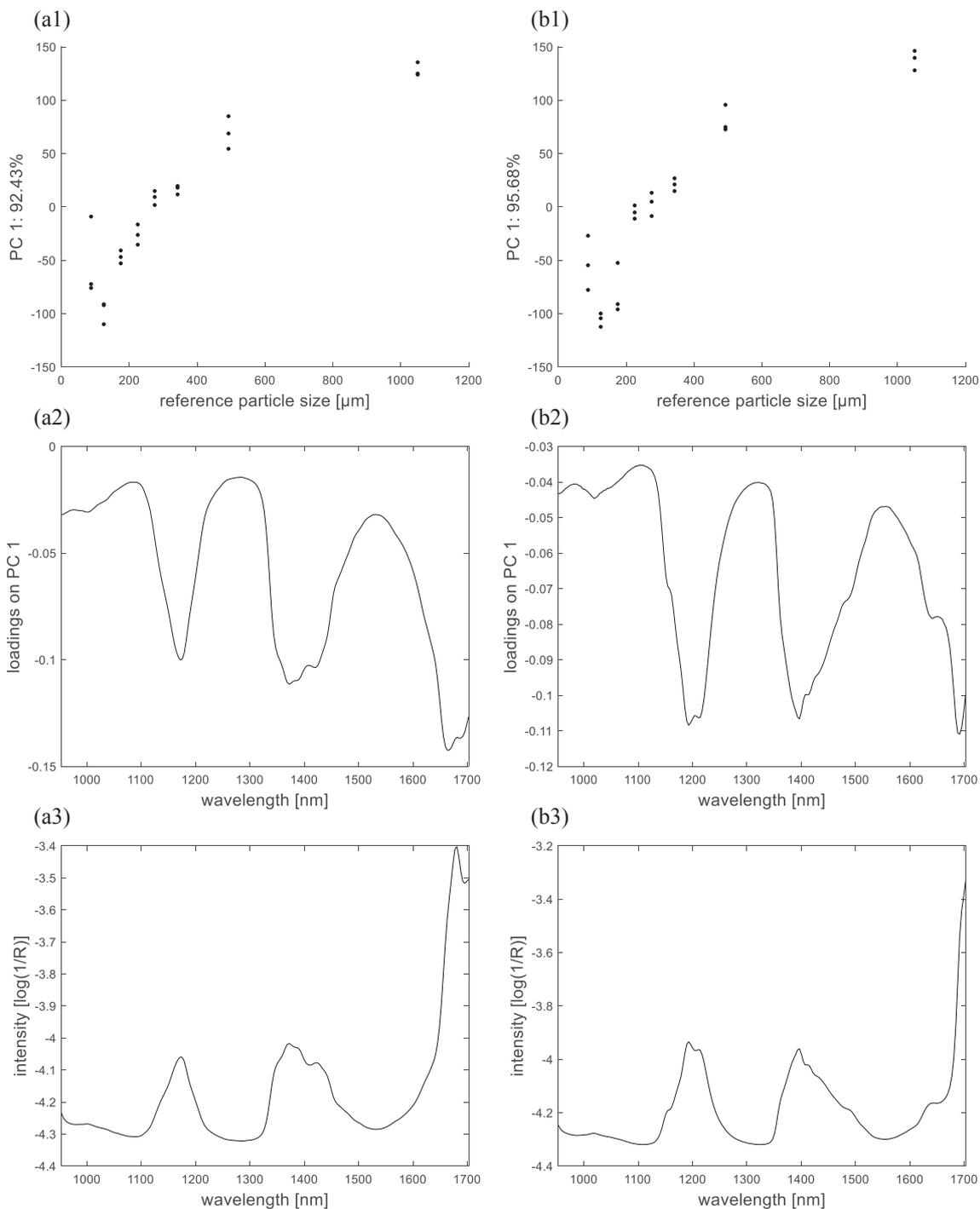


Fig. 2. (a) PMMA samples and (b) PP samples. (1) Trend between the score values of the first principal component for a sample described by the original averaged spectra and the reference mean particle size of a given fraction, (2) loading vector of the first principal component, (3) averaged spectrum of a given polymer sample.

independently built for the PMMA and PP polymer powders that are represented by either individual spectra that are obtained from each image pixel or by the averaged spectra of the pixels that describe the cuvette content. The reference values, mean particle sizes (y_{ref}), corresponded to an arithmetic mean of the upper and lower diameter limit for each fraction. Such a selection of reference values was motivated by the unimodal and 'gaussian-like' distribution of the PCA scores that obtained for the NIR spectra that describe the entire region of a given cuvette.

Initially, the PLS model was built for the averaged spectra and was validated using individual pixel spectra (not considered during model

construction). In Fig. 3, the cuvette content is displayed as images using the predicted response values for each spectrum based on the optimal PLS model for PPMA and PP. The color of each pixel is proportional to the predicted response value.

The white spots in Fig. 3 represent the outlying pixels that correspond to the noisiest and most distorted spectra. Before model construction, they were masked to reduce their harmful effect on the model. The calibration data set consisted of balanced representations of the spectra from each cuvette (1000 spectra). A detailed description of how the region of interest was selected can be found in Section 3.3. The information provided in Fig. 3 should be interpreted with caution. The

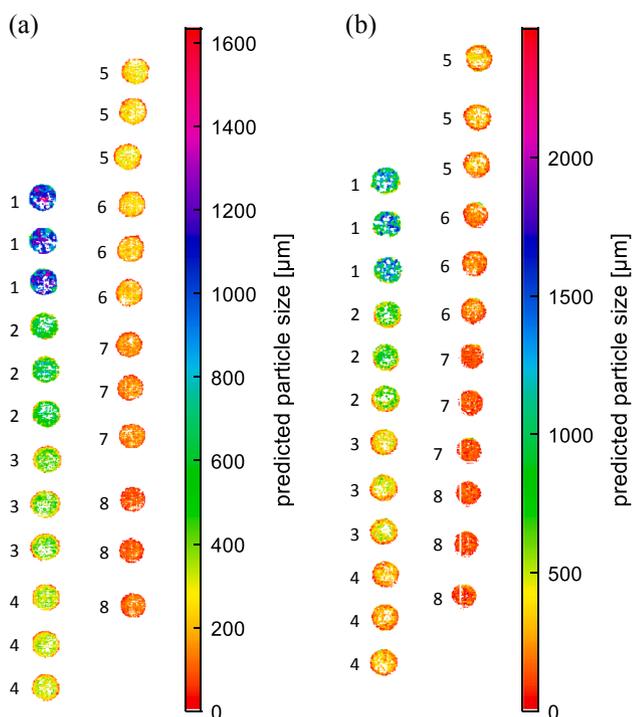


Fig. 3. Image of the predicted size of the particles obtained from the PLS models with six PLS factors for: (a) PMMA and (b) PP original $\log(1/R)$ spectra.

color of a pixel does not resemble the actual size of the particle(s) at a given location. Specific polymer fractions can contain particles with a diameter smaller than the actual pixel size. Therefore, estimating the particle size at a given pixel location represents the average size. Otherwise, the size of the particles from the remaining fractions, for instance, fraction no. 1, are similar or larger than the area of an individual pixel in the image. Even in this situation, samples still contain local inhomogeneities that can be detected on a hyperspectral image, mainly when the edge of a particle or the intersection of a few edges of different particles is found within the area of an individual pixel. Then, the pixel color does not explicitly reflect the actual size but instead informs about the accumulation of local inhomogeneities and therefore, it can be considered for predicting particle size.

The outlying pixels are more frequently observed for fractions with larger particles in which the probability of the occurrence of spatial homogeneity is much higher. This phenomenon is observed when there is a fragment with a smooth surface of larger particles in nearly the entire field of view. Then, the prediction at a specific location, which was estimated by the model, does not correctly represent the studied fraction. This can be seen in the cuvettes in the upper part of the left column in Fig. 3 (fraction no. 1).

The similarities among the samples from the same fraction in Fig. 3 seem to be larger than the ones in Fig. 1. The rings of the smaller particles that gathered near the cuvette rim are visible in many different fractions. In this way, very detailed trends can be detected in the spectral data. Moreover, an analysis confirmed that the predictions of a local inhomogeneity using the PLS model were correct.

4.4. Hyperspectral measurements versus 'single-spot' NIR measurements

It was necessary to adopt the same approach for a quantitative comparison of the different data representations. All of the models that were based on mean spectra were established by averaging 1000 selected spectra from each cuvette and calculating the regression coefficients \mathbf{b}_{pls1} based on the reference mean particle size (\mathbf{y}_{ref1}). In the pixel-based approach, the initial reference values of particle size for each

pixel (\mathbf{y}_{ref2}) were evaluated using the mean spectra PLS model (\mathbf{b}_{pls1}) and an independent set of spectra from each pixel of the hyperspectral image. The pixel-based PLS models were constructed from the evaluated reference particle size values (\mathbf{y}_{ref2}) of individual pixels. Then, the predicted particle size values for the pixels ($\mathbf{y}_{\text{pred2}}$) were averaged and compared with the reference mean values of the actual particle size (\mathbf{y}_{ref1}).

The major steps of the procedure can be summarized as follows:

- (1) calculate the mean reference spectra for each sample;
- (2) construct a vector with reference particle sizes values (\mathbf{y}_{ref1}), corresponding to the mean reference spectra;
- (3) construct the first PLS model that relates mean spectra and the vector with reference particle size values (\mathbf{y}_{ref1});
- (4) use regression coefficient from the first PLS model (\mathbf{b}_{pls1}) for estimation of the initial reference values of particle size for each individual pixel (\mathbf{y}_{ref2});
- (5) construct the pixel-based PLS model for spectra of individual pixels and the estimated value of particle size for each individual pixel (\mathbf{y}_{ref2}).

The 'single spot' NIR measurement was simulated by calculating the arithmetic mean for all of the spectra from each cuvette. The PLS models based on averaged spectra do not require additional evaluation and the second step of calculation to obtain the initial reference values of particle size for each pixel (\mathbf{y}_{ref2}) as in the pixel-based approach. Only the averaged spectra and the vector with initial average reference mean particle size (\mathbf{y}_{ref1}) are required. Thus, average-based spectra modeling represents the most standard regression framework.

Different models were compared in terms of their fit and the prediction power, which are expressed by R^2 and Q^2 values, respectively. The Q^2 values were calculated using the leave-one-fraction-out cross-validation. The values of R^2 and Q^2 are presented in Table 1.

An analysis of the Q^2 values confirmed that it is possible to estimate the average particle size that corresponds to the different polymer powder fractions using the NIR-HSI technique. The most promising PLS models that were constructed for the PMMA and PP powders had Q^2 value above 0.98. The values of R^2 and RMSE indicated that models constructed for the averaged spectra fit the data better than models that had been built for the spectra of image pixels. However, when evaluated for the independent samples, their prediction abilities were expressed by lower Q^2 values than for pixel-based models. Therefore, a NIR hyperspectral camera can provide a set of calibration spectra that support the construction of more reliable models for estimating particle size than the NIR spectrometers that are equipped with a spinner accessory. Furthermore, the logarithmic transformation of a dependent variable on the performance of the PLS models was also examined. The relatively large values of Q^2 indicated that this transformation was beneficial for both types of calibration models (constructed based either on the averaged spectra or the set of spectra that were associated with individual image pixels).

4.5. Pixel-based model performance and its enhancement

The use of derivatives can increase the performance of pixel-based PLS models. The influence of derivatives on the performance of the PLS models estimating particle size was examined by comparing the Q^2 values. The Q^2 values were obtained for different PLS models and calculated in the function of their complexity (see Fig. 4).

For the PMMA and PP polymers, the spectra derivation led to an increase in Q^2 values compared to the Q^2 values of the PLS models that had been constructed for the original spectra. The trend is depicted in Fig. 4a and b. Moreover, regardless of the derivative method that was used (Norris-Williams (NW), Savitzky-Golay (SG), the second or the first derivative), the Q^2 values increased steadily with the number of PLS factors. This observation was somewhat counterintuitive because the

Table 1

Coefficients of the determination (R^2), root mean square error (RMSE), Q^2 values, and the root mean square error of cross-validation (RMSECV), obtained for PLS models with six factors describing the size of particles in PPMA and PP fractions. The PLS models were constructed for differently preprocessed near-infrared spectra using either the pixel-based or mean spectra approaches.

| Approach | Preprocessing | PMMA | | | | PP | | | |
|----------------------|--------------------------|--------|------|--------|--------|--------|------|--------|--------|
| | | R^2 | RMSE | Q^2 | RMSECV | R^2 | RMSE | Q^2 | RMSECV |
| Mean spectra | Original spectra | 0.9904 | 28.6 | 0.7933 | 132.7 | 0.9754 | 45.8 | 0.7038 | 158.9 |
| | log(1/R) | 0.9966 | 17.1 | 0.8372 | 117.8 | 0.9857 | 35.0 | 0.7747 | 138.6 |
| | SG – 7,2,2* | 0.9965 | 17.3 | 0.8513 | 112.6 | 0.9896 | 29.8 | 0.5776 | 189.7 |
| | log(1/R) and SG – 7,2,2* | 0.9992 | 8.2 | 0.9286 | 78.0 | 0.9916 | 26.7 | 0.7288 | 152.0 |
| Pixel-based approach | Original spectra | 0.9677 | 52.5 | 0.8554 | 111.0 | 0.9437 | 69.2 | 0.8671 | 106.4 |
| | log(1/R) | 0.9920 | 26.2 | 0.9420 | 70.3 | 0.9786 | 42.7 | 0.9377 | 72.9 |
| | SG – 7,2,2* | 0.9944 | 21.9 | 0.9872 | 33.0 | 0.9789 | 42.4 | 0.9559 | 61.3 |
| | log(1/R) + SG – 7,2,2* | 0.9977 | 13.9 | 0.9940 | 22.6 | 0.9839 | 37.0 | 0.9837 | 37.3 |

* Window size, polynomial degree, and derivative order.

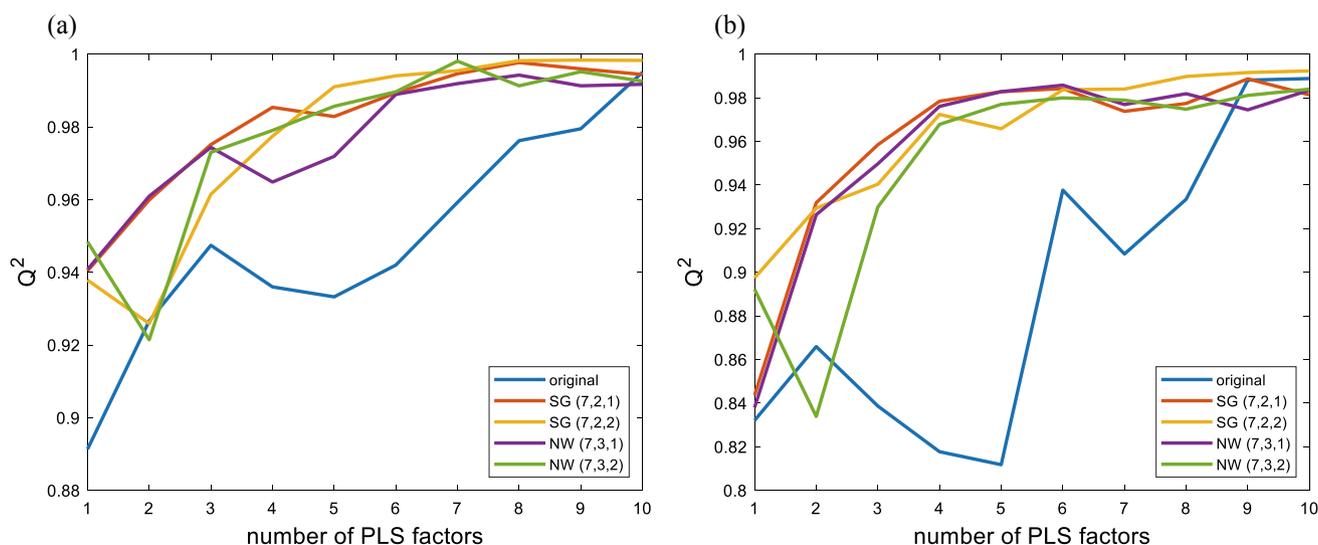


Fig. 4. The Q^2 values as a function of the number of the PLS factors that were obtained for differently preprocessed spectra. The PLS models were constructed using a pixel-based approach based on the log(1/R) spectra that had been extracted from the hyperspectral images that described the PPMA (left column) and PP (right column) samples. The (x, y, z) values in the legend represent the window size (x), polynomial (y) and derivative order (z) for the SG method and the window size (x), gap (y) and derivative order (z) for the NW method.

first derivative was expected to reduce the baseline distortions, and the second derivative was additionally expected to decrease the multiplicative effects. These two spectral features are components of the scattering phenomena that occur in the NIR range. Surprisingly, the predictions of the particle size were more accurate after the spectral derivation.

As was shown in this study, spectra derivatization offered better predictions of the particle size. This beneficial trend was maintained for all of the examined data configurations (averaged spectra approach, pixel-based approach, and spectra with or without prior log(1/R) transformation).

From the perspective of any predictions, the Q^2 values, as a measure of model performance, do not provide comprehensive information about the character of the trend. Therefore, an entire set of mean particle size predicted values were compared with the reference values. Fig. 5 presents the accuracy of the calibration models.

The trends presented in Fig. 5a1 and b1 confirm the strong predictive power of the PLS models for estimating particle size based on hyperspectral images. Moreover, spectra derivative improves the predictions of particle size (see Fig. 5a2 and b2). The derivative emphasizes the spectral features of the signal and does not reduce the scattering. This is consistent with the observation drawn from Fig. 2, where plot loading vectors indicate spectral bands' importance in explaining the scattering phenomena. Derivative preprocessing emphasize the structure of all

spectral features, and thus, it increases model performance.

4.6. Scattering and the spectral regions

As is illustrated in Fig. 6, scattering had a different impact in the different spectral regions of the NIR range. This observation became apparent when the coefficients of determination and the RMSE values were presented for univariate models relating the size of the particles and signal intensity observed at a given wavelength.

Large values of the coefficient of determination point out the spectral regions that explain the studied relationship to the largest extent. For both of the powdered polymers, the most significant spectral regions in terms of scattering prediction coincided with the occurrence of spectral bands. The corresponding models' prediction errors (RMSE) were inversely proportional to the spectral intensities of the peaks and calculated R^2 values.

This demonstrates that in the NIR region, the wavelengths that have a high spectral intensity significantly contribute to predicting the scattering effect. On the other hand, the importance of wavelengths is not linearly proportional to a given band's intensity and the R^2 plot did not resemble the corresponding sample's spectrum. Even small peaks can strongly support the prediction of scattering. Therefore, any new scattering correction methods should include information about the type of spectral pattern if the aim of the method is to reduce the scattering effect

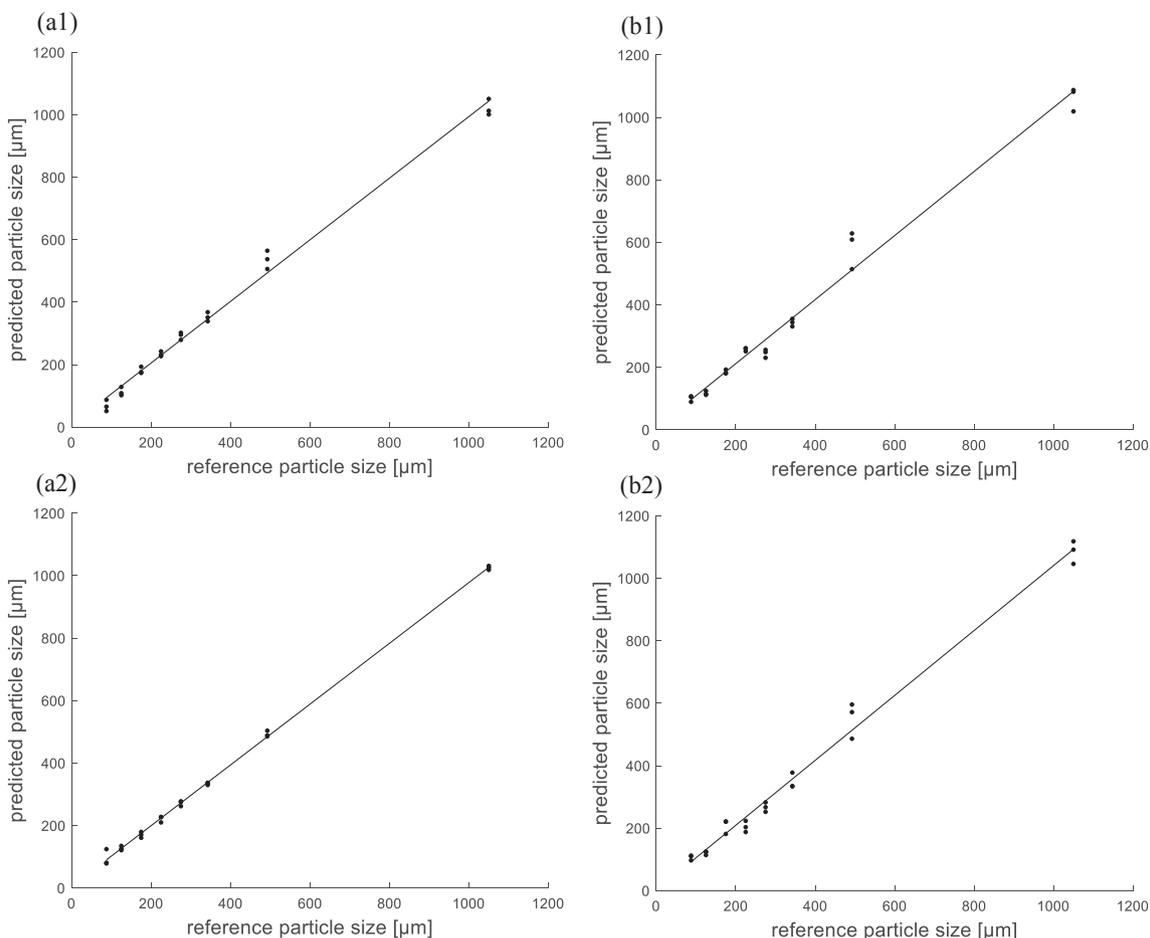


Fig. 5. The relationship between the observed and predicted mean particle diameter for: (a) PMMA and (b) PP samples predicted with the PLS model using six PLS factors and based on (1) the original $\log(1/R)$ spectra, (2) spectra after $\log(1/R)$ and the second derivative with Savitsky-Golay smoothing (the second degree polynomial and window size equal to seven points). The diagonal lines in the figures represent the ideal situation when the particle size that was predicted using the PLS model matched the reference values. The coefficients of the determination for models a1, b1, a2, and b2 are equal to 0.992, 0.979, 0.998, 0.984, respectively.

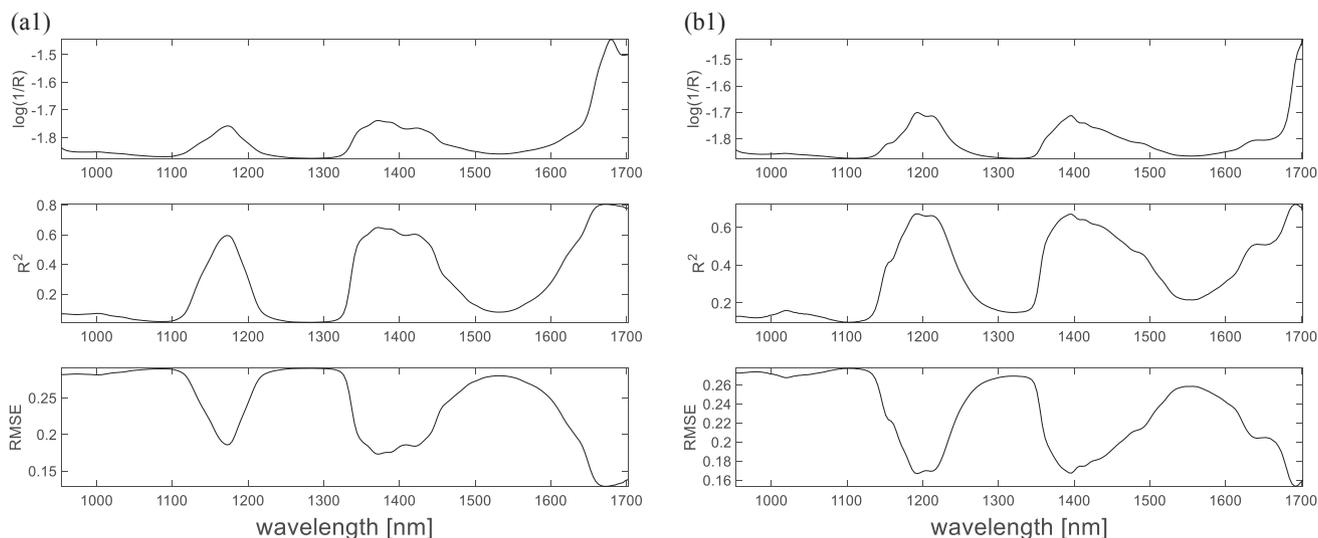


Fig. 6. (a1) $\log(1/R)$ transformed averaged spectrum of the PMMA, coefficients of determination plotted as a function of the selected wavelengths for the construction of a univariate model to predict the average particle size and the corresponding RMSE values; (b1) $\log(1/R)$ transformed averaged spectrum of the PP, coefficients of determination plotted as a function of the selected wavelengths for the construction of a univariate model to predict the average particle size and the corresponding RMSE values.

efficiently.

5. Conclusions

Hyperspectral imaging can reveal the mean particle size of powder samples and the spatial variability and physical heterogeneity of solid samples. Such potential of NIR-HSI supports the construction of calibration models that offer better particle size prediction compared to a single-spot NIR reflectance measurement. This NIR-HSI feature is further enhanced when samples are more heterogeneous. However, to construct a reliable model, a more advanced approach was considered. It included image preprocessing, selecting the region of interest and eliminating outlying pixels. Most of these pretreatment procedures can be automated for a specific sample type and can be performed with little effort. A logarithmic transformation and spectral derivation is beneficial to any NIR data and should be applied to the spectra before any model construction – it does not reduce the scattering phenomena, but it highlights the NIR spectroscopic features. The most remarkable scattering effect was strongly related to the wavelengths that corresponded to the peaks in the NIR spectra of specific samples. Therefore, the development of a novel, efficient scattering correction method should incorporate information about the recorded spectral patterns of solid samples.

CRediT authorship contribution statement

Lukasz Pieszczyk: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Michal Daszykowski:** Conceptualization, Methodology, Formal analysis, Investigation, Resources, Validation, Writing – review & editing, Supervision.

Declaration of Competing Interest

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

Acknowledgements

This research was supported by the National Science Centre, Poland (research grant no. 2018/29/N/ST4/01547).

Data statement

Analyzed HSI data and Matlab algorithms are available upon request from the authors.

References

- Å. Rinnan, F. van den Berg, S.B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, *TrAC, Trends Anal. Chem.* 28 (2009) 1201–1222, <https://doi.org/10.1016/j.trac.2009.07.007>.
- J.-M. Roger, J.-C. Boulet, M. Zeaiter, D.N. Rutledge, 3.01 - Pre-processing methods, in: S. Brown, R. Tauler, B. Walczak (Eds.), *Comprehensive Chemometrics*, second ed., Elsevier, Oxford, 2020, pp. 1–75, <https://doi.org/10.1016/B978-0-12-409547-2.14878-4>.
- H. Martens, S.A. Jensen, P. Geladi, Multivariate linearity transformation for near-infrared reflectance spectrometry, in: *Proceedings of the Nordic Symposium in Applied Statistics*, Stokkand Forlag, Stavanger, 1983, pp. 205–234.
- I.S. Helland, T. Næs, T. Isaksson, Related versions of the multiplicative scatter correction method for preprocessing spectroscopic data, *Chemometr. Intell. Lab. Syst. 29* (1995) 233–241, [https://doi.org/10.1016/0169-7439\(95\)80098-T](https://doi.org/10.1016/0169-7439(95)80098-T).
- N.K. Afseth, A. Kohler, Extended multiplicative signal correction in vibrational spectroscopy, a tutorial, *Chemometr. Intell. Lab. Syst. 117* (2012) 92–99, <https://doi.org/10.1016/j.chemolab.2012.03.004>.
- D.K. Pedersen, H. Martens, J.P. Nielsen, S.B. Engelsen, Near-infrared absorption and scattering separated by extended inverted signal correction (EISC): analysis of near-infrared transmittance spectra of single wheat seeds, *Appl. Spectrosc., AS.* 56 (2002) 1206–1214.
- L. Pieszczyk, H. Czarnik-Matusewicz, M. Daszykowski, Identification of ground meat species using near-infrared spectroscopy and class modeling techniques - aspects of optimization and validation using a one-class classification model, *Meat Sci.* 139 (2018) 15–24, <https://doi.org/10.1016/j.meatsci.2018.01.009>.
- W. Kessler, D. Oelkrug, R. Kessler, Using scattering and absorption spectra as MCR-hard model constraints for diffuse reflectance measurements of tablets, *Anal. Chim. Acta* 642 (2009) 127–134, <https://doi.org/10.1016/j.aca.2009.01.057>.
- P. Mishra, D.N. Rutledge, J.-M. Roger, K. Wali, H.A. Khan, Chemometric pre-processing can negatively affect the performance of near-infrared spectroscopy models for fruit quality prediction, *Talanta* 229 (2021), 122303, <https://doi.org/10.1016/j.talanta.2021.122303>.
- G. Rabatel, F. Marini, B. Walczak, J.-M. Roger, VSN: Variable sorting for normalization, *J. Chemom.* 34 (2020), e3164, <https://doi.org/10.1002/cem.3164>.
- R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra, *Appl. Spectrosc.* 43 (1989) 772–777, <https://doi.org/10.1366/0003702894202201>.
- J.-M. Roger, A. Biancolillo, F. Marini, Sequential preprocessing through ORTHogonalization (SPORT) and its application to near infrared spectroscopy, *Chemometr. Intell. Lab. Syst. 199* (2020), 103975, <https://doi.org/10.1016/j.chemolab.2020.103975>.
- P. Mishra, J.M. Roger, F. Marini, A. Biancolillo, D.N. Rutledge, Parallel pre-processing through orthogonalization (PORTO) and its application to near-infrared spectroscopy, *Chemometr. Intell. Lab. Syst.* 212 (2021), 104190, <https://doi.org/10.1016/j.chemolab.2020.104190>.
- J.L. Ilari, H. Martens, T. Isaksson, Determination of particle size in powders by scatter correction in diffuse near-infrared reflectance, *Appl. Spectrosc., AS.* 42 (1988) 722–728.
- E.W. Ciurczak, R.P. Torlini, M.P. Demkowicz, Determination of particle size of pharmaceutical raw materials using near-infrared reflectance spectroscopy, *Spectroscopy* 1 (1986) 36–39.
- P. Frake, C.N. Luscombe, D.R. Rudd, I. Gill, J. Waterhouse, P. Frake, U. A. Jayasooriya, Near-infrared mass median particle size determination of lactose monohydrate, evaluating several chemometric approaches, *Analyst* 123 (1998) 2043–2046, <https://doi.org/10.1039/a802532k>.
- M. Otsuka, Comparative particle size determination of phenacetin bulk powder by using Kubelka-Munk theory and principal component regression analysis based on near-infrared spectroscopy, *Powder Technol.* 141 (2004) 244–250, <https://doi.org/10.1016/j.powtec.2004.01.025>.
- V. Pauli, Y. Roggo, P. Kleinebudde, M. Krumme, Real-time monitoring of particle size distribution in a continuous granulation and drying process by near infrared spectroscopy, *Eur. J. Pharm. Biopharm.* 141 (2019) 90–99, <https://doi.org/10.1016/j.ejpb.2019.05.007>.
- A. Postelmans, B. Aernouts, W. Saeys, Estimation of particle size distribution from bulk scattering spectra: validation on monomodal suspensions, *Anal. Chem.* 91 (2019) 10040–10048, <https://doi.org/10.1021/acs.analchem.9b01913>.
- A. Postelmans, B. Aernouts, W. Saeys, Estimation of particle size distributions from bulk scattering spectra: sensitivity to distribution type and spectral noise, *Opt. Express* 26 (2018) 15015, <https://doi.org/10.1364/OE.26.015015>.
- B. Aernouts, E. Zamora-Rojas, R.V. Beers, R. Watté, L. Wang, M. Tsuta, J. Lammertyn, W. Saeys, Supercontinuum laser based optical characterization of Intralipid® phantoms in the 500–2250 nm range, *Opt. Express*, OE 21 (2013) 32450–32467, <https://doi.org/10.1364/OE.21.032450>.
- J.T. Alander, V. Bochkov, B. Martinkauppi, S. Saranwong, T. Mantere, A review of optical nondestructive visual and near-infrared methods for food quality and safety, *Int. J. Spectrosc.* 2013 (2013), e341402, <https://doi.org/10.1155/2013/341402>.
- Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond, N. Jent, A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies, *J. Pharm. Biomed. Anal.* 44 (2007) 683–700, <https://doi.org/10.1016/j.jpba.2007.03.023>.
- A.A. Gowen, C.P. O'Donnell, P.J. Cullen, G. Downey, J.M. Frias, Hyperspectral imaging – an emerging process analytical tool for food quality and safety control, *Trends Food Sci. Technol.* 18 (2007) 590–598, <https://doi.org/10.1016/j.tifs.2007.06.001>.
- M. Wang, L. Guo, H. Sun, Manufacture of biomaterials, in: R. Narayan (Ed.), *Encyclopedia of Biomedical Engineering*, Elsevier, Oxford, 2019, pp. 116–134, <https://doi.org/10.1016/B978-0-12-801238-3.11027-X>.
- S.M. Kurtz, Chapter 2 – Synthesis and processing of PEEK for surgical implants, in: S.M. Kurtz (Ed.), *PEEK Biomaterials Handbook* (Second Edition), William Andrew Publishing, 2019, pp. 11–25, <https://doi.org/10.1016/B978-0-12-812524-3.00002-8>.
- R. Xu, Light scattering: a review of particle characterization applications, *Particuology* 18 (2015) 11–21, <https://doi.org/10.1016/j.partic.2014.05.002>.
- J. Stetefeld, S.A. McKenna, T.R. Patel, Dynamic light scattering: a practical guide and applications in biomedical sciences, *Biophys. Rev.* 8 (2016) 409–427, <https://doi.org/10.1007/s12551-016-0218-6>.
- Z. Ma, H.G. Merkus, J.G.A.E. de Smet, C. Heffels, B. Scarlett, New developments in particle characterization by laser diffraction: size and shape, *Powder Technol.* 111 (2000) 66–78, [https://doi.org/10.1016/S0032-5910\(00\)00242-4](https://doi.org/10.1016/S0032-5910(00)00242-4).
- A.F.T. Silva, A. Burggraef, Q. Denon, P. Van der Meeren, N. Sandler, T. Van Den Kerkhof, M. Hellings, C. Vervaeke, J.P. Remon, J.A. Lopes, T. De Beer, Particle sizing measurements in pharmaceutical applications: comparison of in-process methods versus off-line methods, *Eur. J. Pharm. Biopharm.* 85 (2013) 1006–1018, <https://doi.org/10.1016/j.ejpb.2013.03.032>.
- Y. Sun, Z. Cai, J. Fu, Particle morphomics by high-throughput dynamic image analysis, *Sci. Rep.* 9 (2019) 9591, <https://doi.org/10.1038/s41598-019-46062-6>.

- [32] L. Teuber, H. Militz, A. Krause, Dynamic particle analysis for the evaluation of particle degradation during compounding of wood plastic composites, *Compos. A Appl. Sci. Manuf.* 84 (2016) 464–471, <https://doi.org/10.1016/j.compositesa.2016.02.028>.
- [33] L. Li, M. Iskander, Evaluation of dynamic image analysis for characterizing granular soils, *Geotech. Test. J.* 43 (2020) 20190137, <https://doi.org/10.1520/GTJ20190137>.
- [34] T. Gawenda, D. Krawczykowski, A. Krawczykowska, A. Saramak, A. Nad, Application of dynamic analysis methods into assessment of geometric properties of chalcedonite aggregates obtained by means of gravitational upgrading operations, *Minerals* 10 (2020) 180, <https://doi.org/10.3390/min10020180>.
- [35] L. Cervera Gontard, D. Ozkaya, R.E. Dunin-Borkowski, A simple algorithm for measuring particle size distributions on an uneven background from TEM images, *Ultramicroscopy* 111 (2011) 101–106, <https://doi.org/10.1016/j.ultramicro.2010.10.011>.
- [36] D. Buscombe, G. Masselink, Grain-size information from the statistical properties of digital images of sediment, *Sedimentology* 56 (2009) 421–438, <https://doi.org/10.1111/j.1365-3091.2008.00977.x>.
- [37] P. Bosilj, I. Gould, T. Duckett, G. Cielniak, Estimating soil aggregate size distribution from images using pattern spectra, *Biosyst. Eng.* 198 (2020) 63–77, <https://doi.org/10.1016/j.biosystemseng.2020.07.012>.
- [38] M. Budde, S. Leiner, M. Köpke, J. Riesterer, T. Riedel, M. Beigl, FeinPhone: low-cost smartphone camera-based 2D particulate matter sensor, *Sensors* 19 (2019) 749, <https://doi.org/10.3390/s19030749>.
- [39] R. Hussain, M. Alican Noyan, G. Woyessa, R.R. Retamal Marín, P. Antonio Martínez, F.M. Mahdi, V. Finazzi, T.A. Hazlehurst, T.N. Hunter, T. Coll, M. Stintz, F. Muller, G. Chalkias, V. Pruneri, An ultra-compact particle size analyser using a CMOS image sensor and machine learning, *Light Sci. Appl.* 9 (2020) 21, <https://doi.org/10.1038/s41377-020-0255-6>.
- [40] L. Pieszczek, M. Daszykowski, Improvement of recyclable plastic waste detection – a novel strategy for the construction of rigorous classifiers based on the hyperspectral images, *Chemometr. Intell. Lab. Sys.* 187 (2019) 28–40, <https://doi.org/10.1016/j.chemolab.2019.02.009>.
- [41] Adaptthresh, The MathWorks, Inc. (2017). <https://www.mathworks.com/help/images/ref/adaptthresh.html#bu2a9fq-2> (accessed May 8, 2020).
- [42] D. Bradley, G. Roth, Adaptive thresholding using the integral image, *J. Graphics Tools* 12 (2007) 13–21.
- [43] M. Daszykowski, K. Kaczmarek, Y. Vander Heyden, B. Walczak, Robust statistics in data analysis — a review: basic concepts, *Chemometr. Intell. Lab. Sys.* 85 (2007) 203–219, <https://doi.org/10.1016/j.chemolab.2006.06.016>.