I&EC research
Industrial & Engineering Chemistry Research

pubs.acs.org/IECR
Article

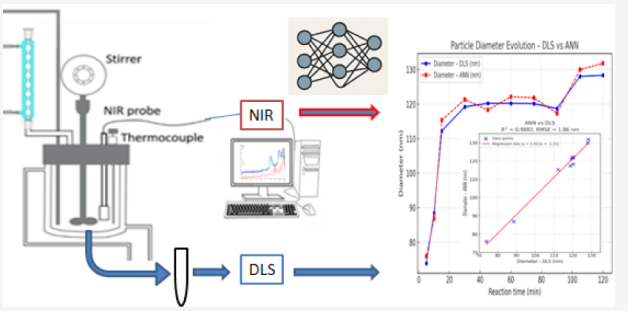
Relationship between NIR Spectrum and Polymer Particle Size for Monitoring Emulsion Polymerization Processes

Published as part of Industrial & Engineering Chemistry Research special issue "Celebrating the Legacy of Prof. Jose M. Asua: Emulsion Polymerization and Polymer Reaction Engineering".

Douglas Rozendo da Silva and Reinaldo Giudici*

Cite This: *Ind. Eng. Chem. Res.* 2025, 64, 19045−19063

ABSTRACT: Emulsion polymerization is widely used in industrial applications, such as paints and resins, due to its versatility and ability to create materials with varied properties. To enhance these processes, in-line and online monitoring techniques are crucial, particularly for tracking the polymer particle size during reactions. Near-infrared (NIR) spectroscopy combined with fiber optic sensors has demonstrated potential for this purpose; however, direct measurement is not feasible, which requires the development of calibration models to correlate spectral data with desired properties. The present study includes experimental data from the synthesis of 11 homopolymers and 3 copolymers of methyl methacrylate (MMA) and styrene (St) and analyzes their NIR spectra. The spectral region between 9000 and 14,000 cm^{-1} showed high sensitivity to particle size changes during emulsion polymerization. Calibration models using Principal Component Regression (PCR), Partial Least Squares regression (PLS), and Artificial Neural Networks (ANN) with spectral ranges of 14,000−4000 cm^{-1} and 14,000−9000 cm^{-1} were developed. Results indicated that the ANN model with the 14,000−4000 cm^{-1} range performed best, demonstrating that a robust calibration model for monitoring particle size across various formulations and conditions was achievable.



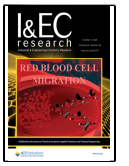
1. INTRODUCTION

Monitoring polymer properties is often carried out through sampling and offline measurement of quality parameters on a laboratory scale. Due to the extended time required for offline analysis, such data are generally inadequate for control purposes. To achieve the production of a polymer resin with specified end-use properties, the availability of efficient control techniques is crucial. A gap between the polymerization process and the control technique must be filled with precise and robust instrumentation for online and in-line monitoring and calculation.[1] Thus, the development of real-time measurement techniques for polymerization process variables has been an area of intense research over the last decades.[2−5]

Various approaches are available for monitoring polymerization processes, including techniques based on ultrasound, calorimetry, near-infrared (NIR) spectroscopy, and Raman spectroscopy. Many different techniques used offline can be adapted to appropriate sampling and preparation/dilution devices for use in online applications.[6,7] Regarding the particle size and particle size distribution, many works have reviewed the available measurement techniques (based on microscopic, on separative procedures, and on light scattering) with their pros and cons.[8−13] More recently new techniques (e.g., photon density wave spectroscopy) have been arisen for use in online monitoring of particle size in polymerization processes.[14−17]

Among the different techniques available, near-infrared (NIR) spectroscopy—a type of vibrational spectroscopy with wavenumbers ranging from 4000 to 12,000 cm^{-1} (a physical quantity inversely proportional to wavelength)—has been widely applied as an effective tool for monitoring and detecting polymer characteristics, owing to its rapid spectral recording, nondestructive measurement, and real-time in-line analysis. Using optical fibers and probes, NIR spectra of polymers can be recorded in situ in reactors, especially for colloidal polymer particles in an aqueous dispersion system.[18] Furthermore, this technique is noninvasive, requires minimal sample preparation (the probe can be directly inserted in the reaction medium), and is able to obtain both chemical and physical data

Received: June 29, 2025
Revised: September 8, 2025
Accepted: September 11, 2025
Published: September 17, 2025



https://doi.org/10.1021/acs.iecr.5c02647
Ind. Eng. Chem. Res. 2025, 64, 19045−19063

acquisition in only one instrument, making it applicable in various processes, including monitoring chemical reactions such as polymerization.

However, the information obtained by this technique is not a direct measurement of the variables of interest. To estimate properties and characterize the reaction medium using spectroscopic methods, it is necessary to develop a calibration model that correlates spectral data with the properties of interest measured by offline reference techniques. For particle size monitoring, reference measurements are usually performed using dynamic light scattering techniques, which require sampling and dilution, or through direct observation using transmission electron microscopy, which involves more complex and time-consuming sample preparation and analysis. This calibration model development stage is crucial for the success of the technique, generally requiring not only samples collected during polymerization trials but also samples synthetically prepared to cover a broader range of variables and thus attempt to represent all possible variations expected in the process.

Despite the variety of calibration methods in use, the main issues lie in the complexity of the spectrum's nature in the NIR region, where peaks of interest are almost certainly overlapped by one or more peaks, creating interferences, broad and scattered bands.[19] Additional difficulties arise in the case of polymerization in heterogeneous media, such as emulsion polymerization processes, where the heterogeneous nature of the medium, i.e., the presence of dispersed polymer particles in the aqueous medium, affects the spectrum. On the other hand, the effect of particle presence on spectra can be used to extract information about particle size, which is crucial for product quality.

Some studies have been developed in this direction and reported the effective possibility of monitoring particle size in-line from near-infrared spectra.[20−26] However, the calibration models found are valid exclusively for the conditions under which the calibration was performed. Any alteration in the system, even if minor, requires a new calibration, which means spending significant time on this stage again and also involving financial costs, thus limiting the broader application of this monitoring technique in the chemical industry. This limitation is generally circumvented by constructing robust models that attempt to cover the maximum expected variations of the system, but a model valid for different reaction systems has not yet been found, despite the monitored variable being a physical measure, in this case, particle size, mainly due to the overlap of bands in the spectrum.

Previous studies suggest an opportunity for investigation in this direction: many studies indicated that NIR spectra can be sensitive to changes of chemical composition (due to absorption) and particle size (due to light scattering), and the spectral region at shorter wavelengths (higher wavenumbers, say, beyond 9000 cm^{-1}) shows high sensitivity to the evolution of particle size during emulsion polymerization, with minimal interference from other components present in the medium, thus presenting the potential for predicting particle size evolution.[19−26] However, the behavior of the NIR spectrum in this region and its relation to particle size is quite complex and nonlinear: in some experiments conducted by Silva et al.,[19] it was observed that the spectrum intensity increased with the increase in particle size, but in other situations, the opposite was observed. Thus, a detailed study aimed at clarifying the root cause of the NIR spectrum behavior in the spectral region of higher wavenumbers could enable the future proposition of a single robust calibration model capable of monitoring the particle size in emulsion polymerizations using different formulations (monomers) and process conditions, as this region appears to be free of absorbance from other components typically present in these systems.

This work aims to study more systematically the relationship between the NIR spectrum and the particle size. For this purpose, the relationship between this spectroscopic technique and the average particle diameter (d) will be presented through NIR spectral analysis and a calibration model.

2. METHODOLOGY

Distilled water, methyl methacrylate (MMA) 99 wt % (Sigma-Aldrich), styrene (St) 99 wt % (Sigma-Aldrich), sodium lauryl sulfate (SLS) (Quimica Moderna), potassium persulfate (KPS) >99 wt % (Sigma-Aldrich), hydroquinone >99 wt % (Sigma-Aldrich), and acetone >99.5 wt % (Sigma-Aldrich) were used without further purification.

The experimental setup used for the emulsion polymerization reactions comprised a 1 L jacketed reactor with four ports, a condenser, a nitrogen sparger, a thermocouple, a mechanical stirrer, an optical fiber, and a near-infrared (NIR) probe. Additionally, a cold bath connected to the condenser and a hot bath for temperature control and system heating were employed. Figure 1 illustrates the apparatus used for the reaction system in the experimental assays.

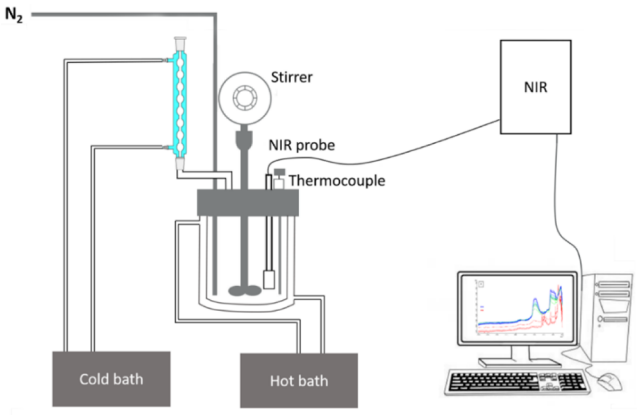


Figure 1. Apparatus for emulsion polymerization reactions.

To study the relationship between the near-infrared (NIR) spectrum and the size and concentration of polymer particles in the emulsion and to analyze some aspects involved in this relationship, an experimental plan was implemented with the following steps:

Step 1: A series of ten experiments was conducted using methyl methacrylate (MMA), varying the concentrations of emulsifier and initiator to produce samples with different particle sizes, maintaining a solid content of approximately 30 wt %.

Step 2: No new syntheses were performed in this stage. Instead, the synthesized polymers were diluted to adjust the particle concentrations (30, 25, 20, 15, and 10 wt %), while keeping the particle size constant, to study the effect of particle concentration on the NIR spectra.

Step 3: No new syntheses were conducted in this stage either. A new set of measurements was carried out by mixing

https://doi.org/10.1021/acs.iecr.5c02647
Ind. Eng. Chem. Res. 2025, 64, 19045−19063

previously produced particles of different sizes (d1 and d2) in various ratios (100:0, 85:15, 50:50, 25:75, and 20:80). The NIR spectra were recorded to investigate how the nonuniformity in particle sizes affects the spectra.

Step 4: Three copolymerization experiments were conducted under the same conditions as the previous syntheses but with varying MMA and St ratios in the formulation (20:80 and 80:20) to obtain copolymer particles with different compositions along with the copolymer synthesized with an equimolar ratio of the two monomers (50:50). The spectra of these samples were compared to assess the effect of particle composition on the spectra.

The emulsion polymerization reactions for the production of the particles used in this study were conducted using both batch and batch methods with intermittent addition processes.

Three MMA homopolymerization reactions and one styrene (St) homopolymerization reaction were carried out using batch processes, following the procedure initially based on the methodology used by Fontenot and Schork.[27] The reactor agitation was set to 400 rpm, and the temperature was maintained at 60 ± 10 °C throughout the reaction for MMA homopolymers and at 70 ± 10 °C for the St homopolymer. Initially, the reactor was charged with water, emulsifier, and monomer (MMA or St), which was sparged with nitrogen for 30 min to remove oxygen from the reaction medium and prevent inhibition of the polymerization reaction. The mixture was then heated using water from a hot bath circulating through the reactor jacket until reaching the desired temperature of 60 or 70 °C, at which point the initiator was added, and polymerization then started. After the reaction began, nitrogen was maintained at a reduced flow rate. The MMA homopolymerization reactions lasted for 120 min, while the St homopolymerization reaction lasted for 240 min. By varying the concentrations of the emulsifier and initiator appropriately, polymer particles of defined sizes were produced, covering the range of interest for the study. Table 1 shows the formulations for the MMA1, MMA2, MMA3, and ST1 experiments.

Table 1. Formulations of the Batch Homopolymerization Experiments

Experiment	H_2O/g	SLS/g	KPS/g	MMA/g	St/g
MMA1	406.2065	1.0962	0.5555	177.0483	-
MMA2	406.2040	3.9033	1.8308	177.0003	-
MMA3	405.4798	0.8804	0.5103	176.9400	-
ST1	406.1750	3.9033	1.8304	-	184.1287

In the batch homopolymerization experiments of MMA with intermittent addition, previously produced polymer particles were used as seeds to obtain polymers with larger particle diameters. These procedures were also based on the methodology used by Fontenot and Schork[27] regarding operational conditions, while the formulations were initially based on the methodology of Silva et al.[19] The total reaction time was 120 min. The main difference in these experiments, besides using previous batches as seeds, was the addition of monomer at four distinct times: at 0, 20, 40, and 60 min. Each intermittent addition was performed quickly with the aid of a funnel and lasted only a few seconds. Table 2 presents the formulations used in these batch homopolymerization experiments with intermittent addition.

In addition, three copolymerization experiments were conducted, varying the molar ratio of MMA to St in the formulations (20:80, 50:50, and 80:20) to obtain copolymer particles with different compositions. The reactor agitation was set to 600 rpm, and the temperature was maintained at 70 ± 10 °C throughout the reactions. Initially, the reactor was charged with water, emulsifier, and monomers (MMA and St), which was sparged with nitrogen for 30 min. The reaction mixture was then heated using water from a hot bath circulating through the reactor jacket until it reached 70 °C, at which point the initiator was added and copolymerization commenced. The reactions lasted approximately 240 min, and, as in the previous experiments, nitrogen was maintained at a reduced flow rate after the reaction began. Table 3 shows the formulations used in the MMA and St copolymerization experiments.

Table 3. Formulations of the MMA and St Copolymerization Experiments

Experiment	H_2O/g	SLS/g	KPS/g	St/g	MMA/g
STMMA1	406.2098	3.9032	1.8304	92.5928	88.4967
STMMA2	406.2654	3.9017	1.8335	147.3107	35.4167
STMMA3	406.2078	3.9028	1.8307	36.7908	141.5727

Offline NIR analyses were conducted through sampling to collect spectra used for monitoring the polymerization parameters of the present study. Additionally, in-line analyses were performed with the same equipment to monitor the synthesis of certain polymers. Samples from all conducted experiments were analyzed using NIR techniques, employing an immersion infrared sensor (Hellma model 661-622-NIR with a transflection probe and an optical path length of 1 mm) connected to a Bruker IFS 28/N spectrometer.

The spectra obtained from the offline analysis had resolution 2 cm^{-1} and a number of scans suitable for a short acquisition time, approximately 20 to 30 s per spectrum. In contrast, in-line tests were collected every 30 s, with each spectrum generated using 4 scans. For the offline tests, after completing

Table 2. Formulations of the Batch Experiments with Intermittent Addition

					MMA/g − Intermittent addition			
Experiment	H_2O/g	SLS/g	KPS/g	Seed/g	0 min	20 min	40 min	60 min
MMA4	200.59	0.438	0.510	100.08 (MMA3)	50.17	50.05	50.31	50.06
MMA5	240.72	0.409	0.513	100.04 (MMA1)	35.04	35.01	35.01	35.06
MMA6	400.00	0.440	0.512	100.03 (MMA3)	50.02	50.04	50.04	50.20
MMA7	270.26	0.460	0.551	98.55 (MMA2)	30.03	30.03	30.07	30.05
MMA8	270.00	0.461	0.512	100.46 (MMA1)	35.02	35.06	35.06	35.01
MMA9	400.15	0.440	0.510	100.00 (MMA18)	50.04	50.08	50.05	50.03
MMA10	200.00	0.321	0.499	100.23 (MMA18)	50.22	50.01	-	-

https://doi.org/10.1021/acs.iecr.5c02647
Ind. Eng. Chem. Res. 2025, 64, 19045−19063

the experimental trials, the synthesized polymers were diluted to alter the particle concentration in terms of solid content (30%, 25%, 20%, 15%, and 10%), while maintaining the same average diameter. This approach aimed to dissociate the evolution of particle concentration from the evolution of diameter, enabling a systematic study of the effect of NIR on particle concentration and size as well as generating data for the development of the calibration model for this study.

The conversion of monomer to polymer was measured offline by gravimetric technique.[28] The average particle size of the emulsion polymer particles was determined by the Dynamic Light Scattering (DLS) technique, also known as Photon Correlation Spectroscopy, a widely used technique to determine the average size of emulsions, particles, and molecules in suspension with diameters ranging from 0.3 nm to 6 μm.[29] Size measurements by light scattering calculate the equivalent diameter of the particles through an intensity-weighted average of the scattered light.

3. RESULTS AND DISCUSSION

3.1. Experimental Results. The volatilization gravimetry technique was used to quantify the overall experimental conversion of the reactions. Satisfactory results were obtained concerning the overall conversion, with values exceeding 96%. Figure 2 shows the conversions for the MMA batch homopolymerization reactions, and Figure 3 presents the conversions for the MMA batch homopolymerization experiments with intermittent addition. Finally, Figure 4 shows the conversions for the styrene homopolymerization and styrene-MMA copolymerizations.

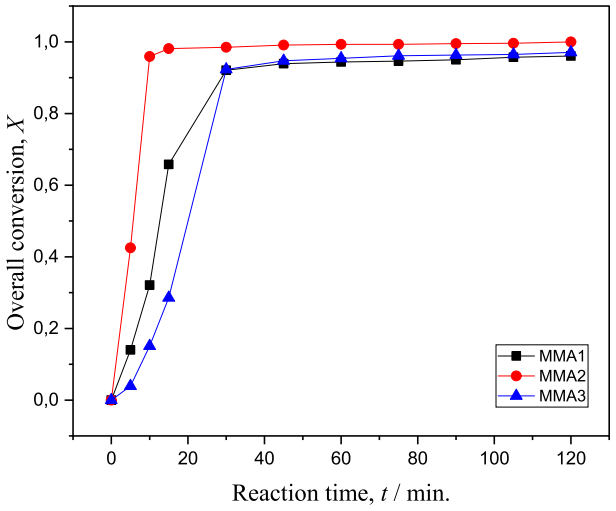


Figure 2. Conversions of MMA batch homopolymerization reactions.

The expected overall conversions in MMA emulsion polymerization syntheses can vary depending on the specific reaction conditions and process objectives. However, high overall conversions are generally anticipated to obtain solid polymers with the desired properties. Czajka and Armes,[30] in their study on X-ray scattering during the polymerization of methyl methacrylate, achieved overall conversions of 93% to 95%.

It can be observed from the conversion graphs in Figures 2, 3, and 4 that the polymerization of styrene generally proceeded more slowly than that of MMA. The fact that PMMA is more hydrophilic than polystyrene allows a greater surface area to be stabilized by the same amount of surfactant, leading to a higher number of particles. Polystyrene produced more coagulum and achieved lower overall conversions.

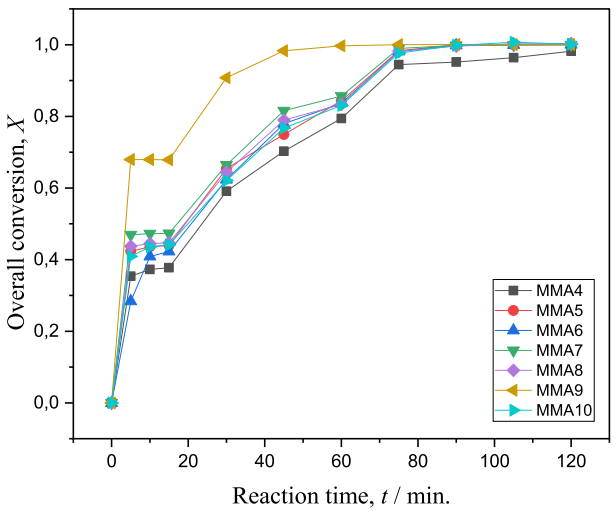


Figure 3. Conversions of MMA batch homopolymerizations with intermittent addition.

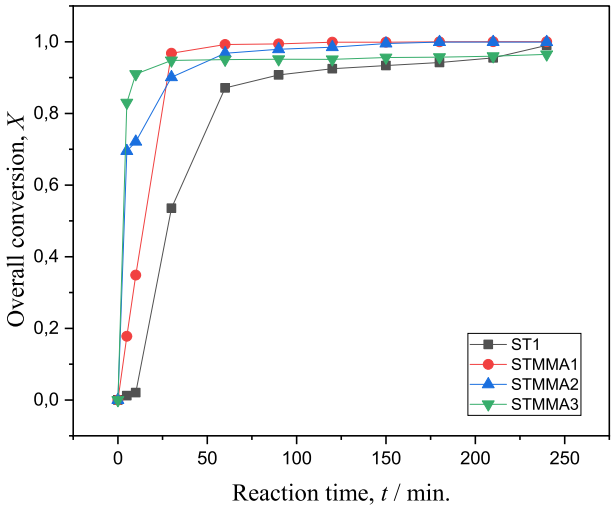


Figure 4. Conversions of the styrene (St) homopolymerization and styrene (St) and MMA copolymerizations.

The high monomer conversion rate in emulsion polymerization reactions of both MMA and styrene can be attributed to several factors, including particle size, emulsion stabilization, the presence of suitable emulsifiers and initiators, temperature control, and the removal of the generated heat, among others.

Figure 5 shows the variation in average diameters for the 10 MMA homopolymerization reactions conducted. These results indicate that a good range of sizes was obtained, enabling the study of this variable with particles ranging from 100 to 465 nm.

Figure 6 presents the average diameters obtained during the styrene homopolymerization reaction and styrene-MMA copolymerizations. In this case, it is observed that smaller average particle diameters were obtained compared to those of the products from MMA homopolymerizations. The conditions and formulations used contributed to this result, as did the chemical structure of the monomers. Styrene is a monomer that contains a benzene ring, making it more bulky and rigid.

https://doi.org/10.1021/acs.iecr.5c02647
Ind. Eng. Chem. Res. 2025, 64, 19045−19063

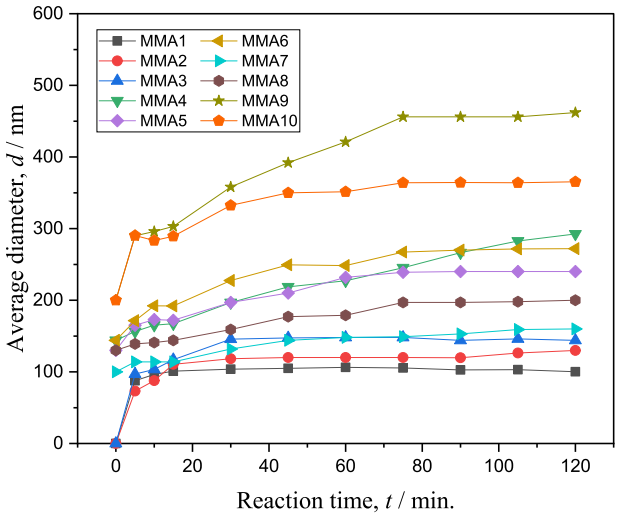


Figure 5. Average diameters of the particles obtained through MMA homopolymerization.

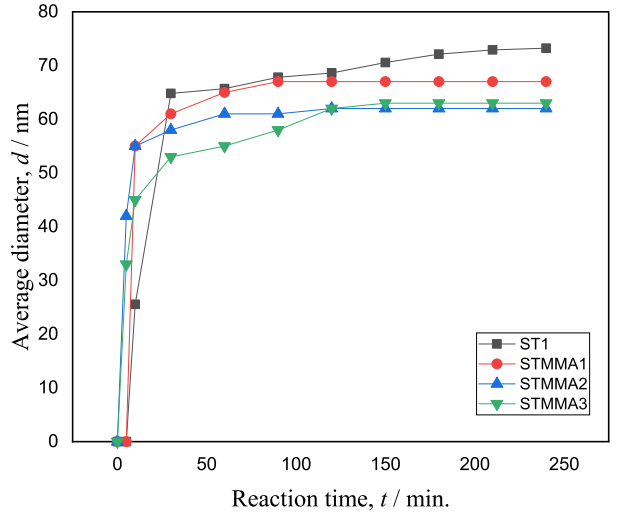


Figure 6. Average diameters of the particles obtained through styrene homopolymerization and styrene-MMA copolymerizations.

During polymerization, styrene molecules tend to organize in a more compact manner, resulting in smaller and denser polymer chains. On the other hand, methyl methacrylate has a more flexible and less bulky structure, and during polymerization, polyMMA molecules can arrange themselves in a more spaced-out manner, leading to the formation of larger and less dense chain arrangements compared to that for polystyrene.

Similar to the average diameter, the particle size distribution obtained from DLS analysis is weighted by the intensity of light scattered by the particles, and PSD affects the emulsion rheology, adhesion, optical properties, mechanical strength, and latex stability.[31] Figure 7 shows the size distribution as a graph of the relative intensity of scattered light versus size classes on a logarithmic scale for the following samples: MMA1, MMA2, MMA6, MMA7, ST1, and STMMA1.

The polydispersity index (PDI) of the sample is provided by the equipment and is a measure of the range of particle size distribution. According to Horgan and Vincent,[32] this index provides an indication of the micelle/particle size distribution. Values above 0.2 are considered indicative of broad distributions, while narrow and monodisperse samples typically have PDI values below 0.1. Thus, it can be observed from the graphs in Figure 7 that the analyzed samples can be considered monodisperse, as PDI values were obtained between 0.038 and 0.090.

3.2. Qualitative Analysis of the Influence of Average Particle Size, Solids Content, and Copolymer Composition on the NIR Spectra. The NIR spectra collected during the polymerization reaction experiments display the behavior shown in Figure 8, in this case during the MMA3 experiment, where the evolution of the reaction can be observed in the in-line NIR spectra.

At time t = 0, during the initial collection, the reaction mixture consisted only of monomer (MMA), water, emulsifier, and initiator. Thus, in the absence of polymeric particles at the initial moments, the region from 14,000 to 9000 cm^{-1} shows virtually zero absorbance. This region does not exhibit significant absorption bands related to chemical components, whereas in the region from 6500 to 5400 cm^{-1}, peaks related to the double bond present in the monomer ($C{=}CH_2$) can be observed. This double bond disappears as the monomer is converted into a polymer during the polymerization. Additionally, the spectrum also shows the highest concentration of C−H bonds of MMA in the form of ($-CH_2$) in the region between 5600 and 6200 cm^{-1}.[19]

As the reaction progresses, the baseline of the spectrum shows a change in the slope. This slope is related to the formation of particles and the resulting light scattering. Thus, the spectral regions associated with the chemical part of the process (consumption of the double bond and formation of the polymer chain) are correlated to the physical part (formation and growth of polymeric particles).

In NIR spectroscopy, dispersed polymer particles influence the spectra mainly through light scattering. As particle size increases, the scattering intensity changes the baseline slope and overall signal intensity, especially in regions without strong chemical absorption bands. This effect occurs because scattering reduces the transmitted light reaching the detector, altering the apparent absorbance in a size-dependent manner.

The effect of particle concentration on the NIR spectra was studied through dilutions, which allowed different concentrations in terms of solid content (SC) to be obtained while maintaining the same particle diameter (d).

The spectra were compared with different preprocessing methods: (a) the original spectrum with only a noise filter (smoothing) over 25 points; (b) the first derivative; and (c) the second derivative. Figures 9, 10, and 11 show the NIR spectra corresponding to the MMA2 experiment.

The spectra in Figures 9, 10, and 11 exhibit notable sensitivity in both the original spectra and the first and second derivatives, across the three observed regions. This observation aligns with the established understanding that variations in the solid content can significantly affect NIR spectra. Studies have demonstrated that increases in solid content typically result in enhanced absorbance in specific spectral regions, which can be attributed to changes in the interaction between NIR light and the sample matrix.[33] The consistent patterns observed, as indicated by the arrows, reflect the systematic relationship between the solid content and the spectral response. The use of first and second derivative treatments enhances the detection of these patterns by minimizing baseline shifts and highlighting subtle changes in the spectra.[34,35]

For the study of the effect of the average particle diameter on NIR spectra, a region of the near-infrared spectrum, located

https://doi.org/10.1021/acs.iecr.5c02647
Ind. Eng. Chem. Res. 2025, 64, 19045−19063

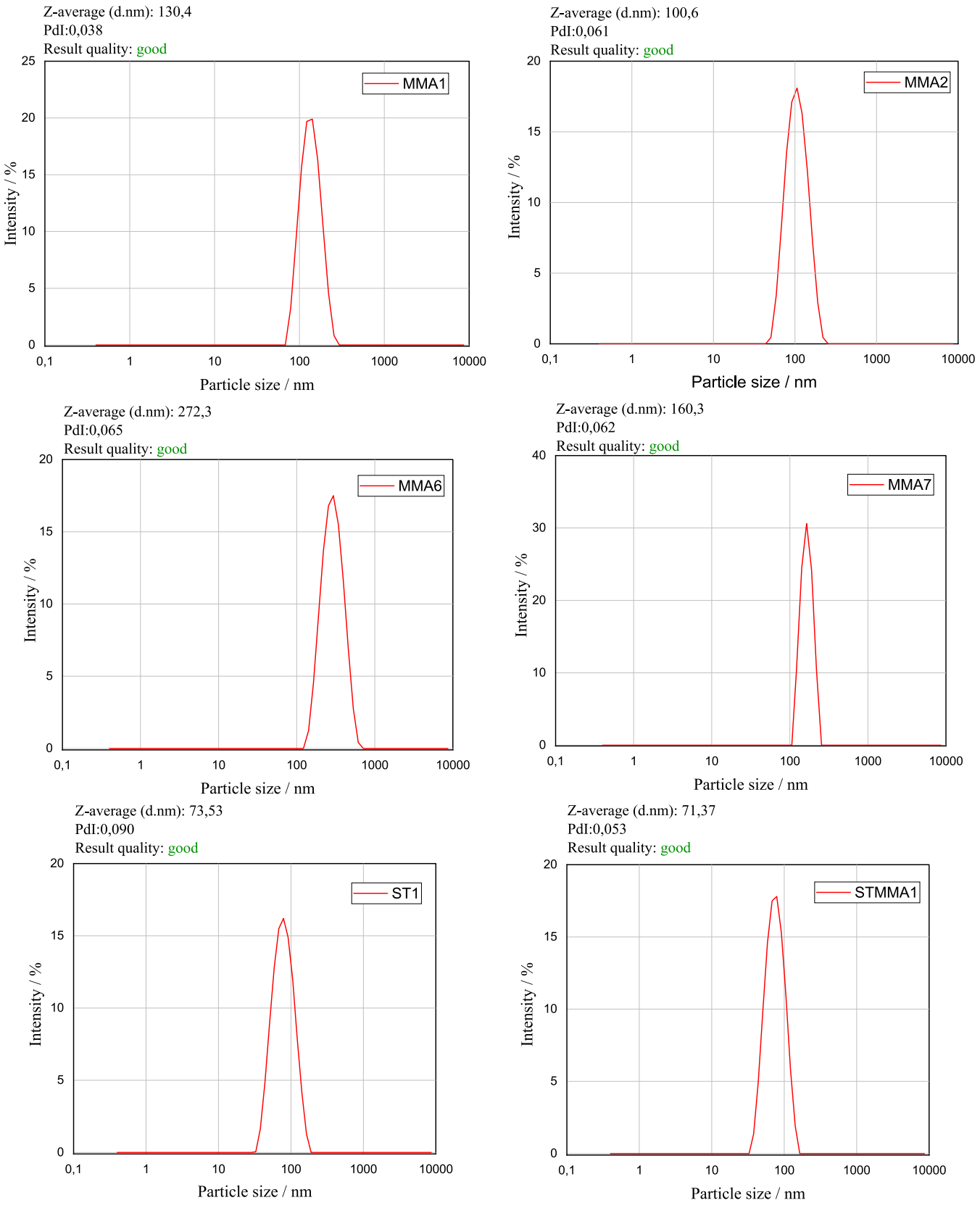


Figure 7. Particle size distribution of MMA and St homopolymerization samples and St-MMA.

between 10,475 and 13,000 cm^{-1}, was observed. According to previous studies, this region exhibits high sensitivity to changes in particle size.[19−22] It is also important to highlight that this study is crucial for evaluating the potential to develop a robust calibration model capable of monitoring particle diameter in emulsion polymerization reactions.

The NIR spectra of samples with different particle diameters at fixed solid contents are presented in Figure 12.

https://doi.org/10.1021/acs.iecr.5c02647
Ind. Eng. Chem. Res. 2025, 64, 19045−19063

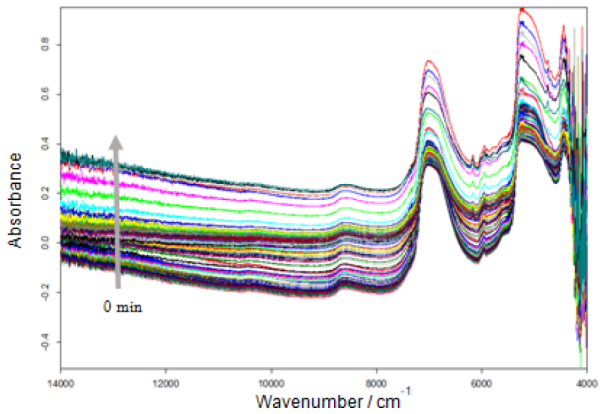


Figure 8. In-line NIR spectra collected during the MMA3 experiment.

Figure 12 reveals a discernible pattern in the measured spectra. The signal intensity increases with larger particle diameters, but there is also a noticeable angular variation, as indicated by the arrows in the presented graphs. Additionally, spectra from larger particles (200, 240, 272, 295, 365, and 462 nm) exhibit more pronounced bands and additional vibrations compared with those from smaller particles (100, 130, 144, and 160 nm). This observation aligns with findings from studies that show how the particle size affects the scattering and absorption characteristics in NIR spectra. Larger particles tend to interact more significantly with infrared light, resulting in more distinct spectral features. This pattern supports the feasibility of developing a robust calibration model for particle size monitoring.

In the NIR region between 9000 and 14,000 cm^{-1}, significant changes are observed, as noted by Torraga and Giudici.[36] This region, often termed the "physical" region of the spectrum, lacks prominent absorption bands related to chemical functionalities but is sensitive to physical changes such as particle size. Studies by Horgan and Vincent[37] also confirm that this spectral range is crucial for monitoring physical properties due to its sensitivity to variations in particle size and distribution. Their research underscores the importance of this region in distinguishing between different sizes of polymer particles, providing a reliable basis for calibration models used in the real-time monitoring of polymerization processes. These findings suggest that with appropriate calibration, the NIR spectral data from this region could be effectively utilized to develop predictive models for

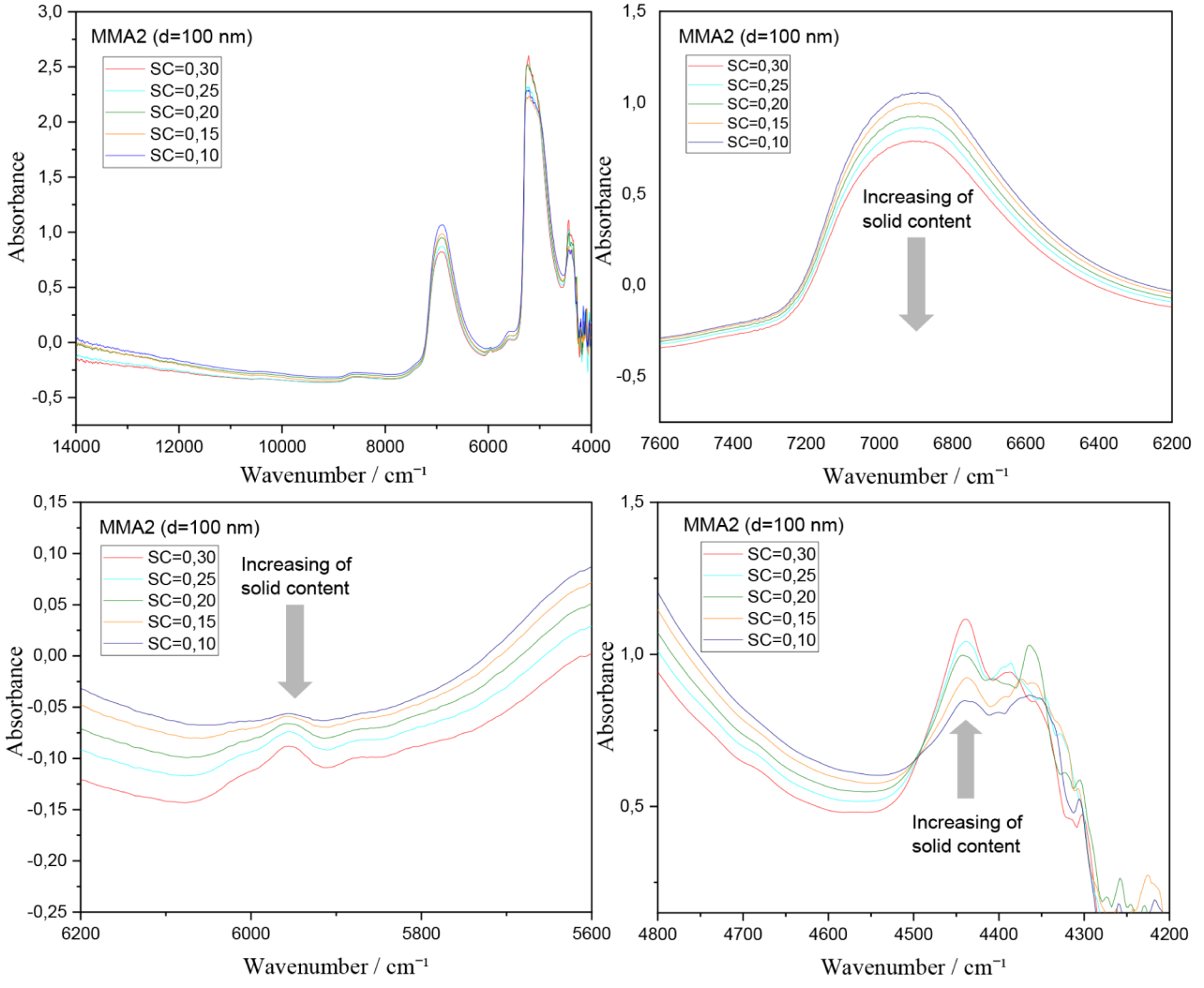


Figure 9. NIR spectra of the same emulsion at different dilutions (different solid content, SC, same particle diameter) from the MMA2 experiment.

https://doi.org/10.1021/acs.iecr.5c02647
Ind. Eng. Chem. Res. 2025, 64, 19045−19063

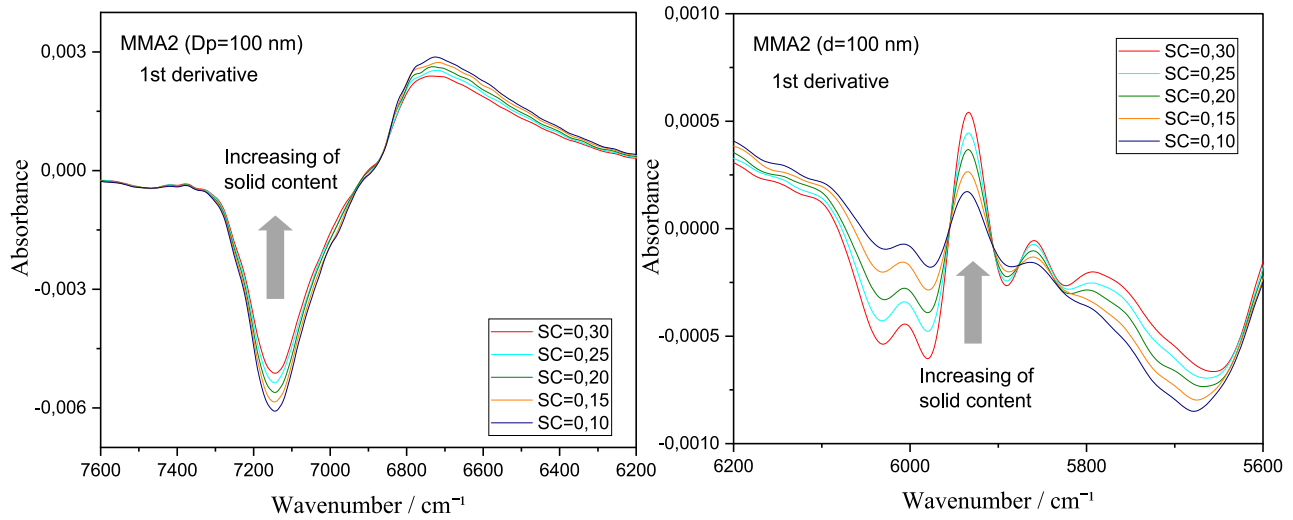


Figure 10. NIR spectra after first derivative treatment of the dilutions from the MMA2 experiment.

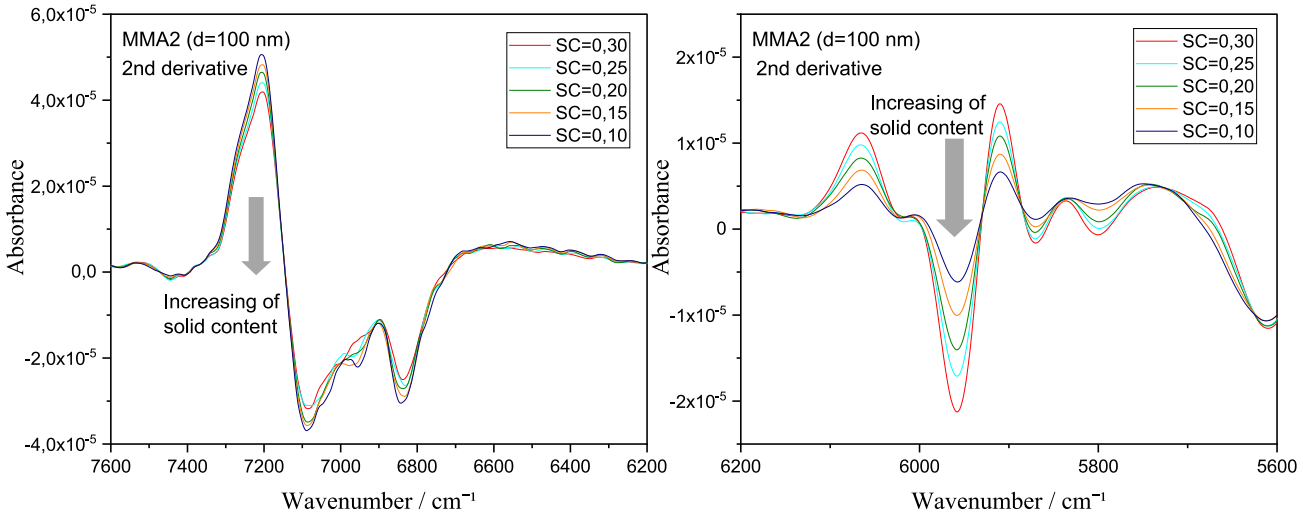


Figure 11. NIR spectra after the second derivative treatment of the dilutions from the MMA2 experiment.

particle size, enhancing the accuracy and efficiency of polymerization monitoring and control.

After mixing particles of different sizes ($d1$ and $d2$) in different mass ratios (100:0, 75:25, 50:50, 25:75, 0:100) with the same solid content, NIR spectra were measured (Figure 13) to assess how the nonuniformity of particle sizes might affect the spectra.

Figure 13 shows that the spectra change according to the indicated arrows with increasing intensity and shifting angular positions as the mass ratio of the larger particle rises. Additionally, in mixtures of smaller particles (<200 nm) with larger particles, the spectra at a 50:50 mass ratio exhibit bands that are more characteristic of the larger particles, with these bands becoming more pronounced.

The study of the effect of particle composition on the NIR spectrum was conducted by comparing the spectra of particles obtained from three experiments varying the molar ratio of MMA and St. These experiments are STMMA1 (50:50 MMA:St), STMMA2 (20:80 MMA:St), and STMMA3 (80:20 MMA:St). The residual concentrations of MMA and St in the copolymerization experiments STMMA1, STMMA2, and STMMA3 were estimated by using gas chromatography (GC). Thus, the composition of each copolymer was estimated according to the data in Table 4.

Consequently, it can be analyzed that the compositions of the copolymers were closed to the molar ratio of the experiments. This was expected since the monomer conversion was high. Moreover, during the copolymerization, the composition drift is expected to be small because the monomers MMA and St have very similar reactivity rates of 0.46 and 0.52, respectively. Furthermore, the GC analysis also confirmed the conversions previously obtained by gravimetry, which were approximately 99% in all three experiments. To evaluate the effect of these different compositions on the NIR spectrum, the spectra of the three copolymer experiments were measured, and the spectra presented in Figure 14 were obtained. It can be observed in the spectra of Figure 14 that despite varying the composition of the copolymers, the difference between the spectra is quite small. Since the diameters of these particles are very similar, this supports the notion that NIR demonstrates good sensitivity for monitoring this physical variable, which could enable the creation of a robust model for monitoring the particle size obtained in emulsion polymerization reactions with different formulations and reaction parameters.

https://doi.org/10.1021/acs.iecr.5c02647
Ind. Eng. Chem. Res. 2025, 64, 19045−19063

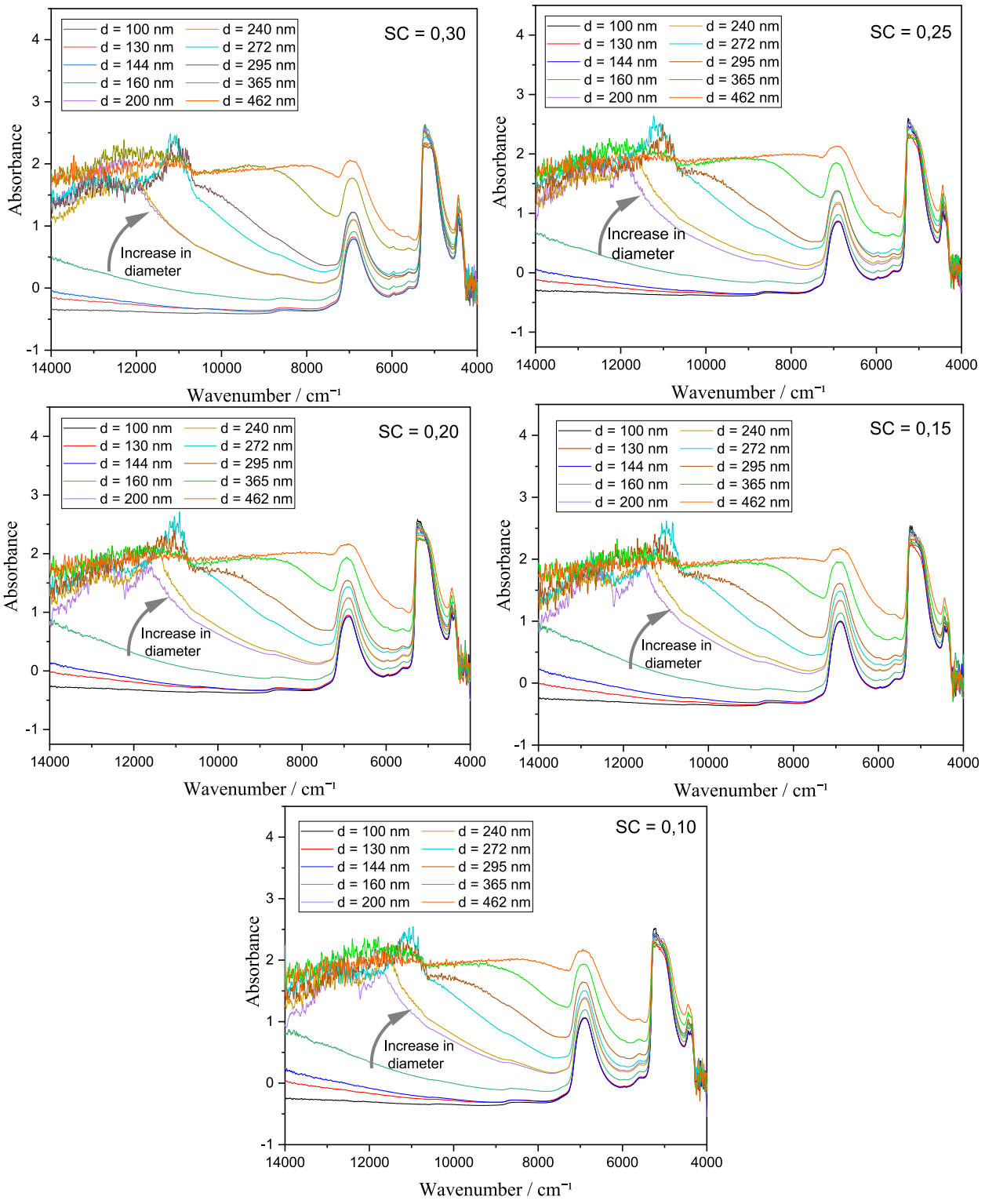


Figure 12. Effect of the particle diameter on NIR at different solid contents.

3.3. NIR Calibration Models. The selection of the test set for calibration and external validation of the particle diameter monitoring model was made by using in-house samples collected during the emulsion polymerization reactions, both from the final samples and from the dilutions. According to Mevik and Wehrens,[38] the process of selecting a diverse calibration set is crucial for ensuring that the model captures the full variability of the system being studied, improving the robustness of the predictions. Moreover, the selection of calibration samples that encompassed a wide range of particle diameters was prioritized, as suggested by Esbensen et al.,[39] who emphasize that a wide distribution of sample sizes is essential for minimizing bias in multivariate calibration models. Samples with different solid contents were also used, which can help account for matrix effects that could influence the particle diameter measurements.

https://doi.org/10.1021/acs.iecr.5c02647
Ind. Eng. Chem. Res. 2025, 64, 19045−19063

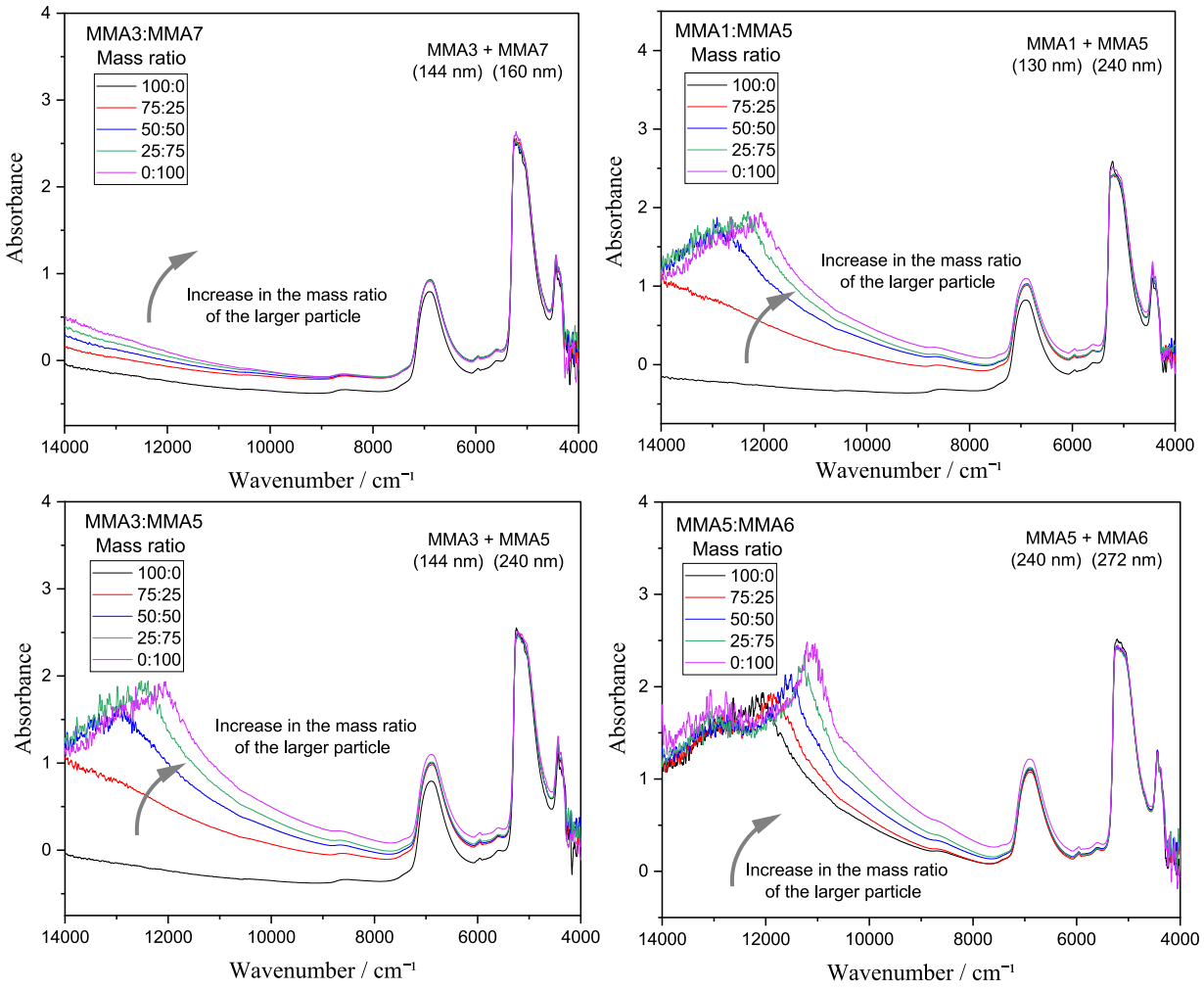


Figure 13. NIR spectra of mixtures of particles of different sizes.

Table 4. Compositions of the Copolymers Obtained by GC

Experiment	MMA Composition	St Composition
STMMA1	0.5008	0.4992
STMMA2	0.1985	0.8015
STMMA3	0.7917	0.2083

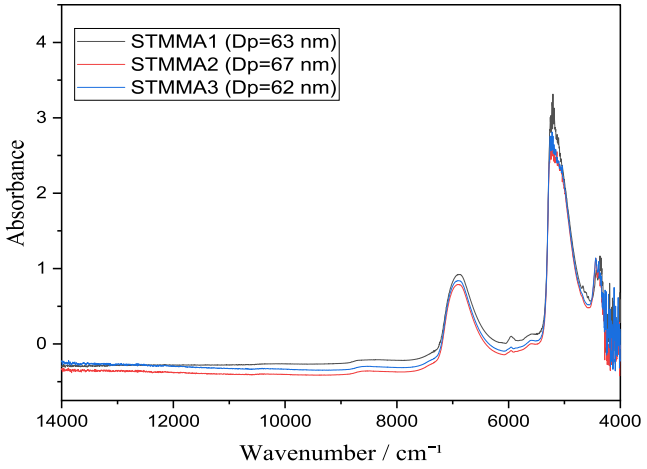


Figure 14. NIR spectra of STMMA1, STMMA2, and STMMA3 with TS of 30%.

A total of 54 data points (DLS-measured diameters and the corresponding NIR spectra) were employed for the model calibration, covering the range of particle size 62−462 nm and solids content 10−30% (see Figure 15) and including data of homopolymerizations (MMA, ST) and copolymerization (STMMA). The final latexes of the polymerization reactions at 30% solids content were diluted to alter the particle concentration (to solids content 25%, 20%, 15%, and 10%), while keeping the same average diameter. The separation of data for calibration and validation was carried out through

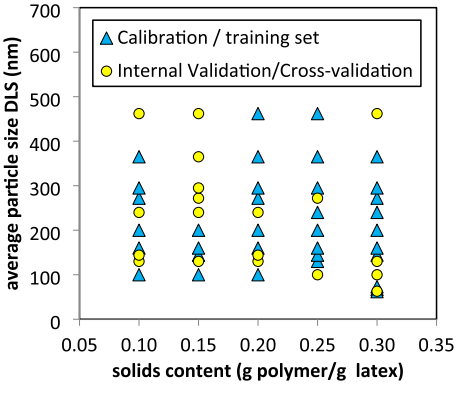


Figure 15. Samples used for calibration and validation.

https://doi.org/10.1021/acs.iecr.5c02647
Ind. Eng. Chem. Res. 2025, 64, 19045−19063

random partitioning in RStudio, where 65% of the data (35 data points) were used for calibration and 35% for validation (19 data points). This approach follows the recommendations of Kuhn and Johnson,[40] who advocate for random data partitioning to reduce potential bias and improve generalizability. Additionally, during the development of the models, all data underwent min-max normalization. Normalization is a key step in machine learning preprocessing, as it rescales each data point between a minimum and maximum value (typically 0 and 1), ensuring that features are on the same scale, thereby preventing features with larger ranges from dominating the model.[41] This technique is widely used for models that rely on distance-based metrics or gradient-based optimization.

In terms of the bands chosen for calibration and validation of the model, two distinct databases (spectral ranges) were used. The first one utilized only the absorbances between the wavenumbers of 14,000 to 9000 cm^{-1}, as both the literature and our previously presented results indicate that this is a region of interest regarding particle diameter. The second database covered the entire spectrum, from 14,000 to 4000 cm^{-1}. Both databases were used in the development of calibration models using Principal Component Regression (PCR), Partial Least Squares regression (PLS), and Artificial Neural Networks (ANN) in RStudio.

These three chemometric approaches (PCR, PLS, and ANN) were selected to represent a range of commonly applied calibration strategies in NIR spectroscopy, from conventional linear methods (PCR, PLS) to a more advanced nonlinear approach (ANN), allowing both baseline and state-of-the-art performance to be assessed under the same data set.

The main idea of PCR and PLS is the reduction of the number and collinearity of the independent variables of the regression (the absorbances measured in the NIR spectra) by the techniques of projection in reduced spaces. Strong collinearity exists in the spectra because each chemical group corresponds to a number of different peaks/bands; in addition, the broadbands of NIR spectra imply that several points that form a given band will also change in a similar way. After the projection in the reduced space, the new independent variables, called principal components (PCs) in PCR, or latent variables (LVs) in PLS, are actually linear combinations of the original ones, grouping the collinear information into a small group of new noncollinear variables that contain virtually the same information as the full spectrum. The PCs are determined from the analysis of the independent variables, while the determination of LVs also accounts for the variability of the corresponding dependent variable. Artificial Neural Networks (ANNs) were also tested because they can account for nonlinearities in the relationship between the independent and the dependent variables that may exist in the case under study (due to multiple scattering, concentrated medium, etc.). In order to reduce the number of independent variables in ANNs, we employed the reduced number of PCs as independent variables, instead of using the full number of absorbances measured in each spectrum (2500 absorbances for the spectral range 14,000−9000 cm^{-1}, or 5000 absorbances for the spectral range 14,000−4000 cm^{-1}). This is particularly convenient for two reasons: (a) reducing the number of inlet variables strongly reduces the number of weights of the ANN model; (b) using a reduced number of PCs instead of the absorbances avoids the collinearity of the independent variables, making the ANN more robust and the estimation of the model parameters (weights and biases) more reliable.

Additionally, the criteria used for evaluating and selecting the models were the coefficient of determination (R^2), which measures the proportion of data variability explained by the model, and the root-mean-square error of cross-validation (RMSECV), which assesses the model's accuracy. The use of R^2 as a primary metric is common in regression analysis, as it indicates how well the model explains the variance in the dependent variable.[42] A high R^2 value, typically close to 1, is indicative of a strong model fit, meaning the model is able to explain a large portion of the data's variability (although some care must be exercised when interpreting R^2 values).[43]

In addition to R^2, the RMSECV is a critical measure for evaluating the accuracy of predictive models during cross-validation. As pointed out by Chai and Draxler,[44] RMSECV quantifies the discrepancy between predicted and observed values, with lower values indicating better predictive performance. It provides a direct measure of the model error during cross-validation, ensuring that the model not only fits the training data but also generalizes well to data not used in the model training.[45] The combination of these two metrics, R^2 and RMSECV, ensures a comprehensive evaluation of both the model fit and predictive accuracy, aligning with established best practices in model validation.

3.3.1. Principal Component Regression (PCR) Models. In the development of the PCR calibration model, the smallest possible number of principal components (PCs) was analyzed, so that the model still achieved a reasonable R^2 value and the lowest RMSECV value. The careful selection of principal components is critical in PCR, as it helps to balance model complexity and predictive accuracy. According to Wold et al.,[46] the number of PCs used should be minimized to avoid overfitting while still capturing the essential variance in the data. This approach aligns with the principles outlined by Massy,[47] who emphasized the importance of retaining only those components that contribute significantly to the explanation of variability, thereby enhancing the interpretability of the model.

The PCR approach was first applied considering the spectral range from 14,000 to 9000 cm^{-1} that does not contain bands corresponding to chemical bonds ("chemical information"). In Figure 16a, the root-mean-square errors of cross-validation (RMSECV) are plotted as a function of the number of PCs. This visualization is essential for identifying the optimal number of components, as it allows for the assessment of the RMSECV in relation to the complexity of the model. As noted by Martens and Naes,[48] monitoring the RMSECV as the number of components increases helps identify the point at which additional components no longer provide significant improvement in model performance, effectively guiding the model selection process. Thus, the analysis of latent variables is fundamental to developing robust and efficient PCR models. It can be observed that RMSECV decreases as the number of PCs increases up to 4 components, and then the RMSECV value does not change significantly, indicating that the choice of 4 components (with RMSECV = 0.0762) is sufficient. Figure 16b presents the observed and predicted values of average diameter (R^2 = 0.9572) using only the data set for calibration (65% of the database), and Figure 16c shows the comparison of the predictions for the data set for testing (35% of the database), for which R^2 = 0.9499 and RMSECV = 0.0869 were obtained.

A second attempt was the application of the PCR for the full spectral range measured (14,000 to 4000 cm^{-1}) that contains

https://doi.org/10.1021/acs.iecr.5c02647
Ind. Eng. Chem. Res. 2025, 64, 19045−19063

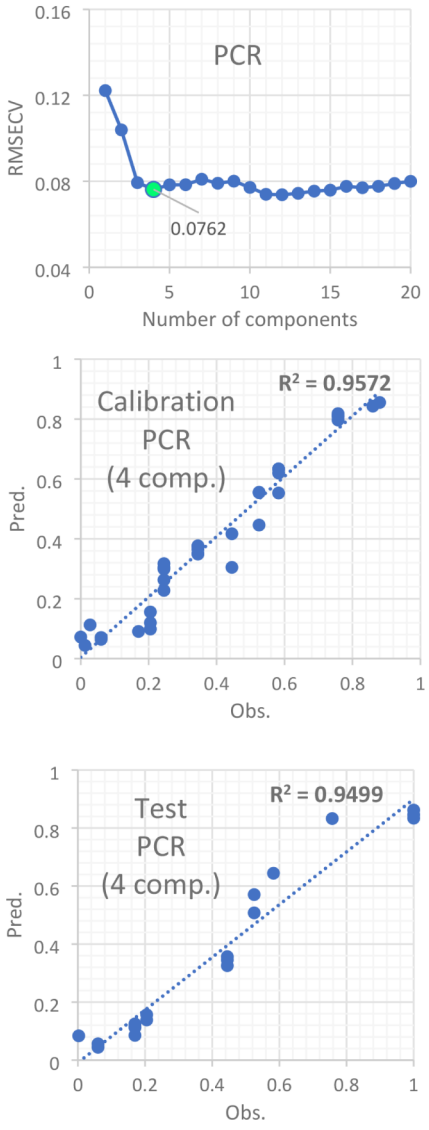


Figure 16. PCR model (spectral range 14,000 to 9000 cm^{-1}) with 4 PCs (principal components).

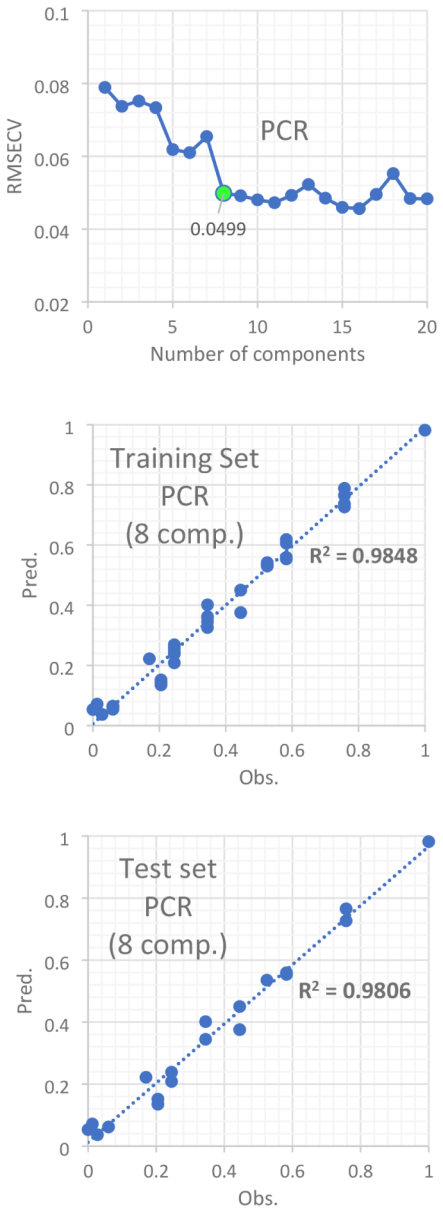


Figure 17. PCR model (spectral range 14,000 to 4000 cm^{-1}) with 8 PCs (principal components).

also the bands related to the chemical information. In Figure 17a, the graph of the root-mean-square errors of cross-validation by the number of PCs can be observed for the model that used the complete database with the full range of absorbances measured in the NIR spectra. It is noted that the trend of RMSECV does not change significantly for more than 8 principal components. Therefore, in this case, the choice was made for 8 components in this model, with RMSECV = 0.0499. Figure 17b shows the fitting of the model for the calibration set (R^2 = 0.9848) and Figure 17c for the testing set (R^2 = 0.9806 and RMSECV = 0.0494). Comparing the two spectral ranges tested, the PCR model for the full range required a large number of PCs (8) and, in turn, gave better predictions for the particle diameter.

3.3.2. Partial Least Squares (PLS) Models. In the development of the PLS calibration models, the goal was also to use the smallest possible number of latent variables while still achieving a reasonable R^2 value and the lowest RMSECV. The effective selection of latent variables in PLS is crucial for enhancing model interpretability and performance. Recent studies emphasize that PLS is particularly advantageous for analyzing data with multicollinearity, as it can extract relevant latent structures from complex data sets.[49,50]

In Figure 18a, the graph of the root-mean-square errors from cross-validation by the number of PLS components is presented, again starting by the use of the reduced spectral range of 14,000 to 9000 cm^{-1}. This visualization is vital for determining the optimal number of components. As noted by Zhang et al.,[51] monitoring RMSECV across varying component numbers helps to ensure the model achieves a balance between fitting the training data and maintaining predictive accuracy on data not used in model training. Furthermore, employing robust validation techniques, including cross-validation and independent test sets, as highlighted by Ruckstuhl et al.,[52] is essential for confirming the reliability and generalizability of the PLS model's predictions. The graph in Figure 18a illustrates that, beginning with 6 components, the RMSECV value starts to rise gradually, and after reaching 10 components, the error increases significantly. Consequently,

https://doi.org/10.1021/acs.iecr.5c02647
Ind. Eng. Chem. Res. 2025, 64, 19045−19063

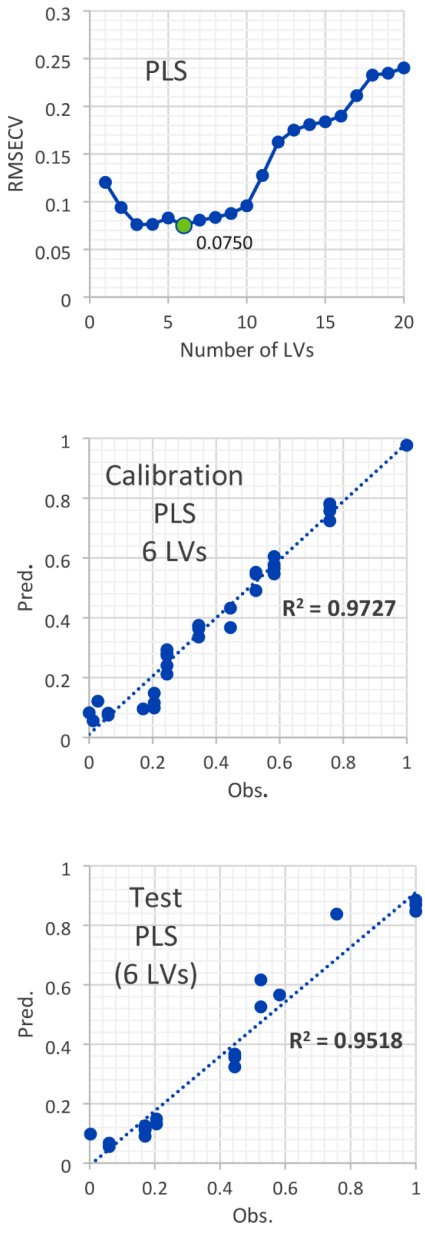


Figure 18. PLS model (spectral range 14,000 to 9000 cm^{-1}) with 6 LVs (latent variables).

the decision was made to select 6 components for this model, with RMSECV = 0.0750. Figure 18b presents the curve generated by the calibration model for the observed and predicted values for the data used for calibration (65% of the database) with the previously determined number of components, yielding a R^2 = 0.9727. Figure 18c shows the comparison of the predictions for the separated database for model validation with R^2 = 0.9518 and RMSECV = 0.0827.

The same approach of the PLS calibration model was performed using the full spectral range (14,000 to 4000 cm^{-1}). Figure 19a shows that the choice of 6 latent variables for this model is appropriate, with RMSECV = 0.0479. This PLS model fitted the data well for the calibration set (R^2 = 0.9921, Figure 19) and also for the test set (R^2 = 0.9796, Figure 19).

Comparing Figures 18 and 19, we see that the use of the full spectral range provides a better calibration model for the same number of latent variables.

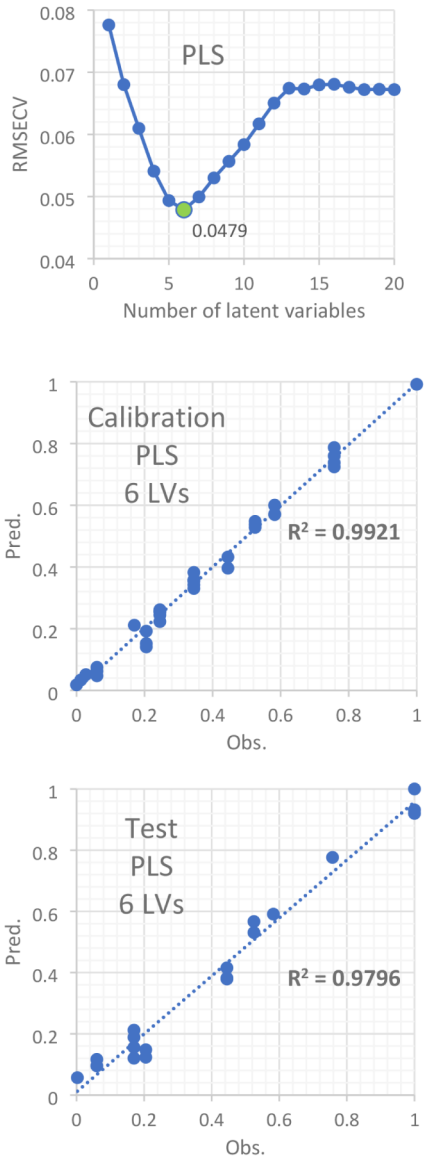


Figure 19. PLS model (spectral range 14,000 to 4000 cm^{-1}) with 6 LVs (latent variables).

3.3.3. Artificial Neural Network (ANN) Models. 3.3.3.1. Principal Component Analysis (PCA). In an Artificial Neural Network model for obtaining the particle size as a function of the NIR spectrum, the input (independent) variables are, in principle, the full set of absorbances measured at several different wavenumbers. As already discussed, such a large set of independent variables is highly correlated to each other (highly collinear), thus requiring a previous treatment to reduce these variables to a smaller number of truly independent variables. Thus, in the creation of a calibration model using Artificial Neural Networks, the execution of a Principal Component Analysis (PCA) of the spectral data helps to reduce the dimensionality of the data by removing redundant information and highlighting key patterns. This dimensionality reduction is crucial as it can accelerate the training of the network and reduce the risk of overfitting. Overfitting occurs when the model is overly complex, capturing noise and random variations in the training data,

https://doi.org/10.1021/acs.iecr.5c02647
Ind. Eng. Chem. Res. 2025, 64, 19045−19063

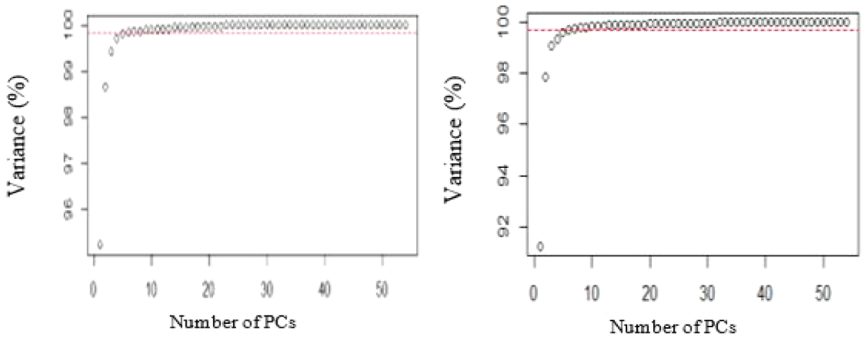


Figure 20. Variance of the spectra as a function of the number of PCs (a) for the spectral database from 14,000 to 9000 cm^{-1} and (b) for the spectral database from 14,000 to 4000 cm^{-1}.

which impairs its ability to generalize well to new data not previously used in model training.[53,54]

In this case, to obtain the diameter calibration models using an ANN, a PCA was initially performed, resulting in new reduced sets of variables, called principal components (in the context of PCR) or latent variables (in the context of PLS). These components effectively represent the independent information contained in the spectrum, allowing for a more robust calibration model with the monitored variables.[55,56] Figure 20a,b presents the variance of the data as a function of the number of principal components (PCs) for the NIR database covering the absorbance ranges of 14,000 to 9000 cm^{-1} and 140,00 to 4000 cm^{-1}, respectively. The graphs illustrate how the selected components account for the variance in the data, supporting the selection of an optimal number of PCs for subsequent analysis.[57,58] For the database from 14,000 to 9000 cm^{-1}, approximately 99.7% of the total variance of the data set is captured with only 5 PCs. In contrast, for the database from 14,000 to 4000 cm^{-1}, this variance of about 99.7% was achieved with 6 PCs, thus defining the number of PCs for both cases.

After the Principal Component Analysis was performed, a new database was generated, consisting solely of the principal component data and the average particle diameters (d). The scores of the principal components were used as the input for the neural network, while the average particle diameters represented the output. Normalization was performed to adjust the values of the numerical columns in the data set to a common scale, which is crucial for ensuring that all features contribute equally to the distance calculations during training.[59,60] This process helps to preserve the relationships among different variables while minimizing the impact of differing ranges.

Additionally, feedforward networks were developed using the backpropagation algorithm, a widely used method for training neural networks that optimizes the weights by minimizing the error through iterative updates.[53,61] The ANN models used are of one hidden layer, using a sigmoidal function as the nonlinear activation function of each neuron.

For model development, only the data set aside for calibration was used, which constituted 65% of the data, as previously mentioned. In configuring the structure of the neural network, it was necessary to vary the number of neurons in the hidden layer (NH) until the minimum RMSECV was achieved. The "for" command in RStudio was utilized, which efficiently repeats commands controlled by a variable, to execute the ANN function while varying NH from 1 to 20, with 5 repetitions for each value to train the network.[62,63] This approach ensured that the error values for each NH were systematically obtained for the training data set, as shown in Figure 21. From this figure, it was found that the minimum

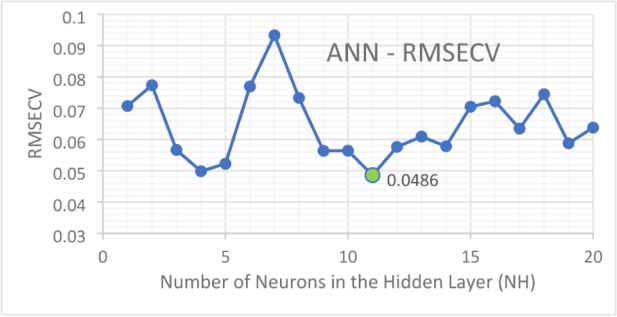


Figure 21. RMSECV of the ANN model as a function of the NH in the hidden layer for training data (database from 14,000 to 9000 cm^{-1}).

RMSECV for the test data is around 0.0486 when the ANN comprises 11 neurons in the hidden layer, establishing this number as the optimal quantity of neurons for that layer.

After training, the ANN model with 11 neurons in the hidden layer, obtained for the database from 14,000 to 9000 cm^{-1}, using 5 PCs as input variables is illustrated in Figure 22. This model fitted the training data very well (R^2 = 0.9992, RMSECV = 0.00751) and also the testing data very satisfactorily (R^2 = 0.9658, RMSECV = 0.0744), as shown in Figure 23.

The same procedure was employed for fitting an ANN model for the broader database from 14,000 to 4000 cm^{-1} and the results are presented in Figure 24 for the determination of the number of neurons in the hidden layer, which was 6 in this case, Figure 25 for the resulting structure and parameters of the ANN model, and Figure 26 for the performance of the fitting of this ANN model, for training data R^2 = 0.9996 and RMSECV = 0.0056, and for the testing data R^2 = 0.9802 and RMSECV = 0.0381.

3.4. Comparisons of the Developed Diameter Calibration Models. Tables 5 and 6 present the comparison of the developed particle diameter models using different methods for selecting bands from the NIR spectrum between

https://doi.org/10.1021/acs.iecr.5c02647
Ind. Eng. Chem. Res. 2025, 64, 19045−19063

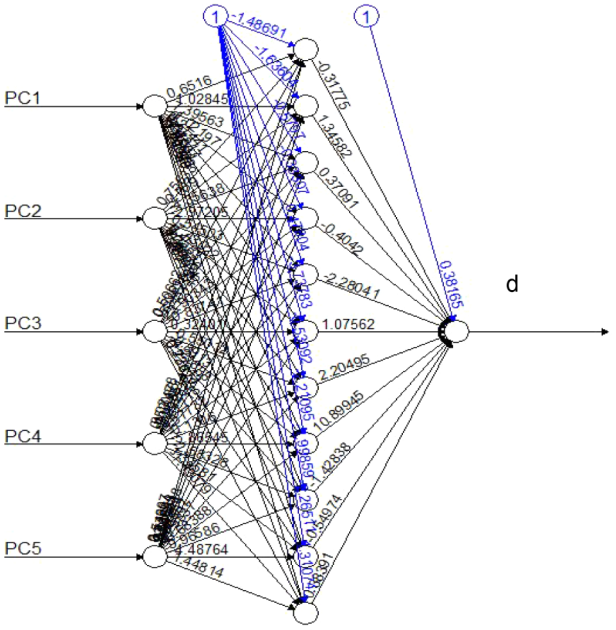


Figure 22. Structure and parameters of the Artificial Neural Network model (database from 14,000 to 9000 cm^{-1}).

14,000 to 9000 cm^{-1} and 14,000 to 4000 cm^{-1} (the entire spectrum), respectively.

Overall, it can be observed that the selection of a reduced band does not yield significant improvements in the model results (in terms of R^2 and RMSECV), indicating that slightly better calibration models were obtained without the need for spectral band selection. This finding aligns with studies suggesting that band selection may inadvertently remove important information from the spectrum, leading to suboptimal model performance.[64,65] The "physical part of the spectrum" (higher wavenumbers) refers to the light scattered by the polymer particles, and this can be related to the particle size and particle concentration (solids content). The "chemical part of the spectrum" (lower wavenumbers) contains information about the amount of water. As the main components of the emulsion are water and polymer particles, more water means less polymer (and vice versa). So, it seems that the use of the full spectral range may help the calibration model to better account for the effects of particle size and particle concentration. This is probably the reason why the calibration models using the full spectral range (Table 6) performed a little better than the corresponding calibration models that used the reduced spectral range (Table 5).

Moreover, it is noted that in general, the ANN model achieved a slightly better predictive accuracy compared to the other two methods. Recent research supports the notion that ANN can capture complex nonlinear relationships within the

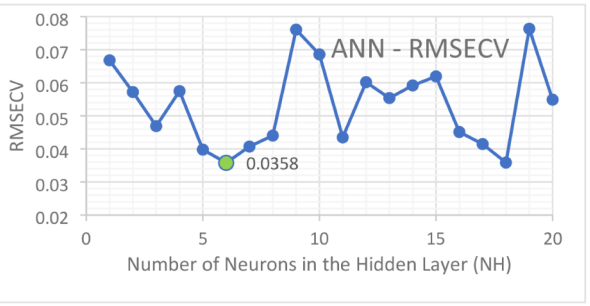


Figure 24. RMSECV of the ANN model as a function of the NH in the hidden layer for training data (database from 14,000 to 4000 cm^{-1}).

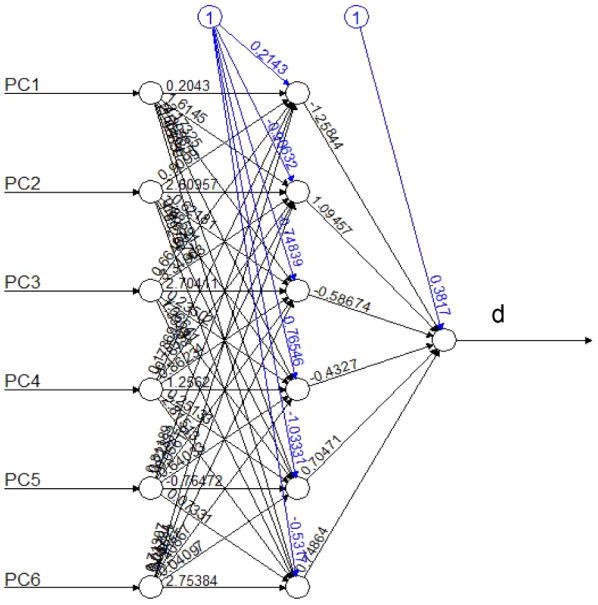


Figure 25. Structure and parameters of the Artificial Neural Network model (database from 14,000 to 4000 cm^{-1}).

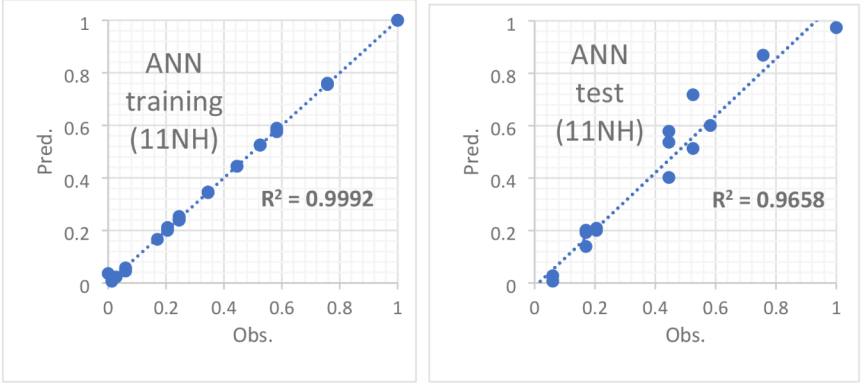


Figure 23. Fitting performance of the ANN calibration model (database from 14,000 to 9000 cm^{-1}).

https://doi.org/10.1021/acs.iecr.5c02647
Ind. Eng. Chem. Res. 2025, 64, 19045−19063

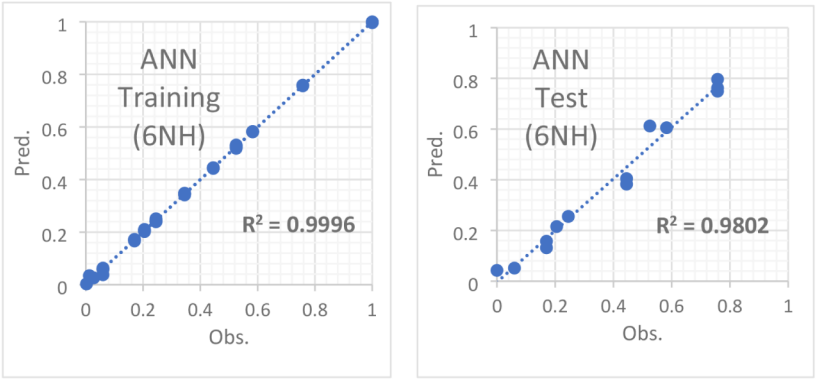


Figure 26. Fitting performance of the ANN calibration model (database from 14,000 to 4000 cm^{-1}).

Table 5. Comparison of the Calibration Model Results for Particle Diameter (Database from 14,000 to 9000 cm^{-1})

Method	Components	RMSECV	R^2
PCR	4	0.0762	0.9572
PLS	6	0.0750	0.9727
ANN	5	0.0486	0.9992

Table 6. Comparison of the Calibration Model Results for Particle Diameter (Database from 14,000 to 4000 cm^{-1})

Method	Components	RMSECV	R^2
PCR	8	0.0499	0.9848
PLS	6	0.0479	0.9921
ANN	6	0.0358	0.9996

data more effectively than traditional methods, resulting in enhanced predictive accuracy.[66,67] Additionally, it can be seen that the PLS models provided more satisfactory results compared to those developed using PCR, as expected, which is consistent with findings that highlight PLS's robustness in dealing with multicollinearity and its ability to extract meaningful latent structures.[68,69]

Artificial Neural Networks (ANN) offer several advantages over traditional calibration models like Principal Component Regression (PCR) and Partial Least Squares (PLS) due to their ability to handle nonlinear relationships and capture complex patterns in the data. ANNs can automatically learn from the data, adapting to intricate features that may be missed by linear models like PCR and PLS, making them particularly useful in applications where data structures are complex.[70,71] It is known that chemical information represented by the absorbance is linearly related to the concentrations of the chemical groups. On the other hand, particle size is related to the scattering, diffraction, reflection, and refraction of light in a much more complex way; moreover, for in-line application in an undiluted medium (high particle concentration), multiple scattering certainly occurs, making the relationship between the near-infrared spectrum and particle size more complex and nonlinear. These conditions indicate that flexible, nonlinear models, such as ANN models, may be more likely to establish a reliable relationship between particle size and NIR spectra

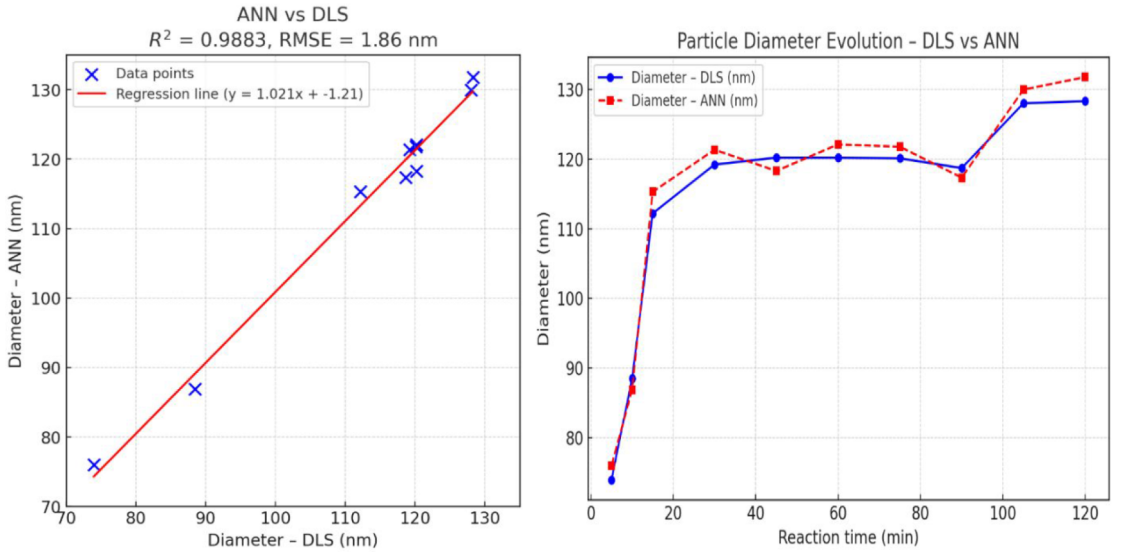


Figure 27. (a) ANN−NIR predicted vs DLS-measured particle diameters for the independent MMA11 reaction. (b) Time evolution of the particle diameter during the MMA11 reaction measured by DLS and predicted by the ANN−NIR model.

https://doi.org/10.1021/acs.iecr.5c02647
Ind. Eng. Chem. Res. 2025, 64, 19045−19063

Moreover, ANNs are better equipped to manage large data sets and interrelated variables, enabling them to model intricate relationships effectively.[72,73] However, the training of ANNs often requires more extensive data sets for effective learning and hyperparameter tuning, which can make their implementation more complex. Additionally, they can be more challenging to interpret than linear models, posing a significant barrier in fields where model transparency is crucial.[74,75]

Ultimately, the choice between these methods depends on the specific needs of the problem at hand as well as the availability of data and resources. While ANNs can provide superior performance in certain contexts, simpler models like PCR and PLS may be more appropriate for cases where interpretability and ease of use are prioritized.[76,77]

In the case of ANN models, it is important to emphasize that the number of neurons in the hidden layer has a significant impact on the performance and learning complexity. If the hidden layer has too few neurons, the network may not be able to capture the complexity of the patterns in the data, resulting in low learning capacity and underfitting.[78] On the other hand, if the hidden layer has too many neurons, the network may become excessively complex and struggle to generalize to new data, leading to overfitting.[79]

Near-infrared spectroscopy has proven to be a simple, rapid, nondestructive, low-cost, and environmentally friendly technique for monitoring emulsion polymerization reactions. Based on the results obtained, it can be said that the models developed, particularly the one using neural networks with the entire spectrum, are capable of monitoring the particle diameter across different systems. It is important to emphasize, however, that in-line monitoring with NIR spectroscopy is subject to interferences that may affect the calibration model. Such interferences include the presence of air bubbles in the optical path of the probe, sample nonuniformities, elevated concentrations in the samples, differences in the optical path of the probe, and variations in the intensity of the NIR source.

3.5. External Validation of the ANN−NIR Model Using Data Collected in an Independent Emulsion Polymerization Reaction. To verify the predictive performance of the developed ANN−NIR calibration model, an independent emulsion polymerization reaction (MMA11) was monitored. This was a batch reaction carried out at 60 °C with 400 g of water, 175.4 g of monomer MMA, 1.2862 g of emulsifier SLS, and 0.5675 g of initiator KPS. Samples were collected at time intervals over a total of 120 min of reaction time; particle diameters were measured by Dynamic Light Scattering (DLS) and NIR spectra were collected and used to predict the particle diameter using the trained ANN−NIR model.

Figure 27a shows the parity plot between the ANN−NIR predicted and DLS-measured diameters for this experiment MMA11. The results exhibit a strong linear relationship ($y = 1.021x - 1.21$), with $R^2 = 0.9883$ and RMSE = 1.86 nm, confirming the model's ability to reproduce DLS results for new conditions not previously used in the model calibration.

Figure 27b presents the time evolution of the particle size for the same reaction. Both ANN−NIR and DLS results follow the same trend: a rapid increase in particle diameter during the initial 15 min, a plateau between approximately 30 and 90 min, and a slight increase toward the end of the polymerization. Minor deviations between the curves remain within expected measurement uncertainties, confirming the robustness of the ANN−NIR approach for real-time monitoring.

4. CONCLUSION

The present study involved successful experiments focused on the formation of emulsion polymer particles with varying average diameters, enabling the proposed investigation of the relationship between NIR spectra and average particle size in the emulsion polymerization process. All experiments demonstrated high conversion rates and resulted in monodisperse particles. Initially, a visual analysis was conducted to assess the effects of particle concentration, particle diameter, size uniformity, and the composition and uniformity of the particles on the NIR spectra.

Following spectral analysis, several regions of good sensitivity concerning the studied properties were identified, with the most significant being in the range of 14,000 to 9000 cm^{-1}, highlighting a correlation between concentration and mean particle diameter with NIR spectroscopy.

Calibration models were developed by using different spectral regions and various multivariate calibration methods. However, it was observed that restricting the region to 14,000 to 9000 cm^{-1} did not result in improved model performance. Furthermore, the models created using Artificial Neural Networks exhibited slightly better performance and accuracy.

Thus, the study reinforces the feasibility of creating a robust calibration model capable of monitoring homopolymerizations and copolymerizations in emulsions, utilizing different formulations (monomers) and process conditions. Future work could amplify the number of formulations of different copolymerization systems for broader tests of the calibration models here studied.

The models developed for determining particle diameter, in comparison to common measurement techniques that require sample preparation/dilution, such as dynamic light scattering (DLS), allow for results to be obtained much more rapidly. This speed represents a significant advantage for quality control in the industry, enabling quicker responses and informed decision-making based on the relevant data.

AUTHOR INFORMATION

Corresponding Author

Reinaldo Giudici − *Universidade de São Paulo, Escola Politécnica, Department of Chemical Engineering, São Paulo 05508-010, Brazil;* orcid.org/0000-0001-6544-7098; Email: rgiudici@usp.br

Author

Douglas Rozendo da Silva − *Universidade de São Paulo, Escola Politécnica, Department of Chemical Engineering, São Paulo 05508-010, Brazil*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.iecr.5c02647

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding

The Article Processing Charge for the publication of this research was funded by the Coordenacao de Aperfeicoamento de Pessoal de Nivel Superior (CAPES), Brazil (ROR identifier: 00x0ma614).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work has been supported by the São Paulo Research Foundation (FAPESP, grant number 2020/00332-7), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, grant numbers 130919/2020-9 and 310125/2021-9), and CAPES (financial code 001).

■ NOMENCLATURE AND ACRONYMOUS

ANN, Artificial Neural Networks; *d*, Average Particle Diameter (nm); DLS, Dynamic Light Scattering; GC, Gas Chromatography; H_2O, Water; KPS, Potassium Persulfate; MMA, Methyl Methacrylate; NH, Number of Neurons in the Hidden Layer; NIR, Near-Infrared; PC, Principal Component; PCA, Principal Component Analysis; PCR, Principal Component Regression; PDI, Polydispersity Index; PLS, Partial Least Squares Regression; R^2, Coefficient of Determination; RMSECV, Root Mean Square Error of Cross-Validation; SC, Solid Content; SLS, Sodium Lauryl Sulfate; St, Styrene; *t*, Reaction time (min); wt %, Weight Percent; *X*, Overall Conversion (%)

■ REFERENCES

(1) de Faria, J. R.; Machado, F.; Lima, E. L.; Pinto, J. C. Monitoring of Vinyl Chloride Suspension Polymerization Using NIRS. 1. Prediction of Morphological Properties. *Comput.-Aided Chem. Eng.* **2009**, *27*, 327−332.

(2) Chien, D. C. H.; Penlidis, A. On-line sensors for polymerization reactors. *J. Macromol. Sci., Rev. Macromol. Chem. Phys.* **1990**, *30* (1), 1−42.

(3) Kammona, O.; Chatzi, E. G.; Kiparissides, C. Recent developments in hardware sensors for the on-line monitoring of polymerization reactions. *J. Macromol. Sci., Rev. Macromol. Chem. Phys.* **1999**, *39* (1), 57−134.

(4) Richards, J. R.; Congalidis, J. P. Measurement and control of polymerization reactors. *Comput. Chem. Eng.* **2006**, *30* (10−12), 1447−1463.

(5) Fonseca, G. E.; Dubé, M. A.; Penlidis, A. A critical overview of sensors for monitoring polymerizations. *Macromol. React. Eng.* **2009**, *3* (7), 327−373.

(6) Alb, A. M.; Reed, W. F. Fundamental measurements in online polymerization reaction monitoring and control with a focus on ACOMP. *Macromol. React. Eng.* **2010**, *4* (8), 470−485.

(7) Alb, A. M.; Drenski, M. F.; Reed, W. F. Automatic continuous online monitoring of polymerization reactions (ACOMP). *Polym. Int.* **2008**, *57* (3), 390−396.

(8) Li, H.; Li, J.; Bodycomb, J.; Patience, G. S. Experimental methods in chemical engineering: Particle size distribution by laser diffraction—PSD. *Can. J. Chem. Eng.* **2019**, *97* (7), 1974−1981.

(9) Pei, Y.; Hinchliffe, B. A.; Minelli, C. Measurement of the size distribution of multimodal colloidal systems by laser diffraction. *ACS Omega* **2021**, *6* (22), 14049−14058.

(10) Zhang, S.; Zhang, Q.; Shang, J.; Mao, Z. S.; Yang, C. Measurement methods of particle size distribution in emulsion polymerization. *Chin. J. Chem. Eng.* **2021**, *39*, 1−15.

(11) Elizalde, O.; Leal, G. P.; Leiza, J. R. Particle size distribution measurements of polymeric dispersions: A comparative study. *Part. Part. Syst. Charact.* **2020**, *17* (5−6), 236−243.

(12) Nijhu, R. S.; Khatun, A.; Hossen, M. F. A comprehensive review of particle size analysis techniques. *Int. J. Pharm. Res. Dev.* **2024**, *6* (1), 01−05.

(13) Coones, R. T.; Kestens, V.; Minelli, C. A comparison of hydrodynamic diameter results from MADLS and DLS measurements for nanoparticle reference materials. *J. Nanopart. Res.* **2025**, *27* (7), 170.

(14) Schlappa, S.; Brenker, L. J.; Bressel, L.; Hass, R.; Münzberg, M. Process characterization of polyvinyl acetate emulsions applying inline photon density wave spectroscopy at high solid contents. *Polymers* **2021**, *13* (4), 669.

(15) Schlappa, S.; Bressel, L.; Reich, O.; Münzberg, M. Advanced particle size analysis in high-solid-content polymer dispersions using photon density wave spectroscopy. *Polymers* **2023**, *15* (15), 3181.

(16) Aspiazu, U. O.; Paulis, M.; Leiza, J. R. Photon Density Wave spectroscopy to monitor the particle size in seeded semibatch emulsion copolymerization reactions. *Chem. Eng. J.* **2024**, *483*, 149292.

(17) Aspiazu, U. O.; Gómez, S.; Paulis, M.; Leiza, J. R. Real-Time Monitoring of Particle Size in Emulsion Polymerization: Simultaneous Turbidity and Photon Density Wave Spectroscopy. *Macromol. Rapid Commun.* **2024**, *45* (20), 2400374.

(18) Yi, M.; Qiu, T.; Okubo, M.; Li, X.; Guo, L. Innovative on-line near-infrared (NIR) spectroscopy to estimate content of each phase in composite polymer particles prepared by seeded emulsion polymerization. *Vib. Spectrosc.* **2018**, *95*, 23−31.

(19) Silva, W. K.; Chicoma, D. L.; Giudici, R. In-situ real-time monitoring of particle size, polymer, and monomer contents in emulsion polymerization of methyl methacrylate by near infrared spectroscopy. *Polym. Eng. Sci.* **2011**, *51* (10), 2024−2034.

(20) Gossen, P. D.; MacGregor, J. F.; Pelton, R. H. Composition and particle diameter for styrene-methyl methacrylate copolymer latex using UV and NIR spectroscopy. *Appl. Spectrosc.* **1993**, *47*, 1852−1870.

(21) Santos, A. F.; Silva, F. M.; Lenzi, M. K.; Pinto, J. C. Monitoring and control of polymerization reactors using NIR spectroscopy. *Polym. -Plast. Technol. Eng.* **2005**, *44* (1), 1−61.

(22) Reis, M. M.; Araújo, P. H. H.; Sayer, C.; Giudici, R. Correlation between Polymer Particle Size and in-situ NIR Spectra. *Macromol. Rapid Commun.* **2003**, *24*, 620−624.

(23) Reis, M. M.; Araújo, P. H. H.; Sayer, C.; Giudici, R. Comparing near infrared and Raman spectroscopy for on-line monitoring of emulsion copolymerization reactions. *Macromol. Symp.* **2004**, *206*, 165−178.

(24) Vieira, R. A. M.; Sayer, C.; Lima, H. L.; Pinto, J. C. Detection of monomer droplets in a polymer latex by near-infrared spectroscopy. *Polymer* **2001**, *42*, 8901−8906.

(25) Vieira, R. A. M.; Sayer, C.; Lima, H. L.; Pinto, J. C. In-line and in situ monitoring of semi-batch emulsion copolymerizations using near-infrared spectroscopy. *J. Appl. Polym. Sci.* **2002**, *84*, 2670−2682.

(26) Chicoma, D. L.; Sayer, C.; Giudici, R. In-Line Monitoring of Particle Size during Emulsion Polymerization under Different Operational Conditions using NIR Spectroscopy. *Macromol. React. Eng.* **2011**, *5*, 150−162.

(27) Fontenot, K.; Schork, F. J. Sensitivities of droplet size and stability in monomeric emulsions. *Ind. Eng. Chem. Res.* **1993**, *32*, 373−385.

(28) Santos, M. P.; Torraga, M. G.; Barbosa, N. B.; Tavares, D. T.; Giudici, R. Estimation of Reactivity Ratios in the Copolymerization of Styrene and VeoVa-10. *Macromol. React. Eng.* **2020**, *14* (3), 2000001.

(29) Dialetachi, E. L. G. *Espalhamento dinâmico de luz em sistemas coloidais diluídos. Dissertação (Mestrado em Física)*; Universidade de São Paulo, 2017.

(30) Czajka, A.; Armes, S. P. Time-Resolved Small-Angle X-ray Scattering Studies during Aqueous Emulsion Polymerization. *J. Am. Chem. Soc.* **2021**, *143* (3), 1474−1484.

(31) Edouard, D.; Sheibat-Othman, N.; Hammouri, H. Observer design for particle size distribution in emulsion polymerization. *AIChE J.* **2005**, *51*, 3167−3185.

(32) Horgan, A.; Vincent, B. Polystyrene nanoparticles based on poly(butyl methacrylate-g-methoxypoly(ethylene glycol)) and poly(methyl methacrylate-g-methoxypoly(ethylene glycol)) graft copolymers. *J. Colloid Interface Sci.* **2003**, *262*, 536−547.

(33) Dawson, J. L.; Reineccius, G. A. Influence of solid content on near-infrared spectra of food matrices. *J. Sci. Food Agric.* **1988**, *47*, 207−213.

(34) Smith, B. C. *Fundamentals of Fourier Transform Infrared Spectroscopy.*; CRC Press, 1999.

https://doi.org/10.1021/acs.iecr.5c02647
Ind. Eng. Chem. Res. 2025, 64, 19045−19063

(35) Workman, J. J.; Weyer, L. R. *Practical Guide to Interpretive Near-Infrared Spectroscopy.*; CRC Press, 2012.
(36) Torraga, M. G. F.; Giudici, R. Stabilizer-free dispersion copolymerization monitoring by in-line NIR spectroscopy. *Ind. Eng. Chem. Res.* **2020**, *59* (35), 15497−15505.
(37) Reis, M. M.; Araújo, P. H.; Sayer, C.; Giudici, R. In situ near-infrared spectroscopy for simultaneous monitoring of multiple process variables in emulsion copolymerization. *Ind. Eng. Chem. Res.* **2004**, *43* (23), 7243−7250.
(38) Mevik, B. H.; Wehrens, R. The pls Package: Principal Component and Partial Least Squares Regression in R. *J. Stat. Software.* **2007**, *18* (2), 1−23.
(39) Esbensen, K. H.; Guyot, D.; Westad, F.; Houmoller, L. P. *Multivariate Data Analysis: In Practice.*; CAMO Software, 2012.
(40) Kuhn, M.; Johnson, K. *Applied Predictive Modeling.*; Springer, 2013.
(41) Patro, S. G. K.; Sahu, K. K. Normalization: A Preprocessing Stage. *Int. Adv. Res. J. Sci. Eng. Technol.* **2015**, *29*, 414−425.
(42) Draper, N. R.; Smith, H. *Applied Regression Analysis.*; Wiley, 1998.
(43) Frost, J. How to interpret R-squared in regression analysis. *Statistics By Jim.* 2017, *16*, https://statisticsbyjim.com/regression/interpret-r-squared-regression/
(44) Chai, T.; Draxler, R. R. Root mean square error (RMSE) or mean absolute error (MAE)? Geoscientific Model Development. *Geosci. Model Dev.* **2014**, *7* (3), 1247−1265.
(45) Hawkins, D. M. The Problem of Overfitting. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1−12.
(46) Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. J. The Collinearity Problem in Linear Regression: The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM J. Sci. Stat. Comput.* **1983**, *5*, 735−743.
(47) Massy, W. F. Principal Components Regression in Exploratory Statistical Research. *J. Am. Stat. Assoc.* **1965**, *60*, 234−241.
(48) Martens, H.; Naes, T. *Multivariate Calibration.*; Wiley, 1989.
(49) Bohlin, J.; Sjoberg, E.; Hage, H. A Comprehensive Review of Partial Least Squares Regression: Methodological Developments and Applications. *Chemom. Intell. Lab. Syst.* **2020**, *203*, 104056.
(50) Shankar, A.; Devi, M.; Choudhury, S. Advances in Partial Least Squares Regression: A Comprehensive Review. *Anal. Chim. Acta* **2021**, *1155*, 338−356.
(51) Zhang, Y.; Zhang, L.; Wang, Y. Model Selection and Evaluation in Partial Least Squares Regression: A Practical Approach. *J. Chemom.* **2022**, *36*, No. e3258.
(52) Ruckstuhl, H.; Brechbühler, C.; Gander, M. Validation Strategies for Chemometric Models: Addressing Overfitting and Model Robustness. *Food Chem.* **2023**, *401*, 134018.
(53) Bishop, C. M. *Pattern Recognition and Machine Learning.*; Springer, 2006.
(54) Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning.*; MIT Press, 2016.
(55) Liu, Y.; Zhang, C.; Chen, L. Application of Principal Component Analysis for the Dimensionality Reduction in ANN Model of Spectroscopic Data. *Talanta* **2019**, *204*, 668−674.
(56) Kwan, S. C.; Xiong, Z.; Tan, T. Using PCA for Enhanced ANN Performance in Spectroscopic Data Analysis. *J. Chemom.* **2020**, *34*, No. e3211.
(57) Ghosh, S.; Kumar, M.; Kumar, V. Principal Component Analysis for Enhanced Modeling of Spectroscopic Data. *Anal. Chim. Acta* **2021**, *1147*, 12−21.
(58) Yang, M.; Jiang, Z.; Wu, Y. Optimizing the Use of Principal Components in Machine Learning Applications for Spectroscopy. *Chemom. Intell. Lab. Syst.* **2023**, *240*, 104548.
(59) Han, J.; Kamber, M.; Pei, J. *Data Mining: Concepts and Techniques.*; Morgan Kaufmann, 2019.
(60) O'Sullivan, D.; McCarthy, U.; O'Donoghue, A. Normalization Techniques in Neural Network Applications. *J. Comput. Sci.* **2021**, *50*, 101176.
(61) Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. Learning Representations by Back-Propagating Errors. *Nature* **1986**, *323*, 533−536.
(62) Tzeng, J. E.; Wu, Y. C.; Wang, J. Y. Optimizing Neural Network Architecture: A Case Study Using R. *J. Comput. Theor. Nanosci.* **2020**, *17*, 4549−4554.
(63) Velez, M.; Zuniga, M.; Rodriguez, L. Implementation of Neural Networks with R for Predictive Modeling in Engineering. *Eng. Appl. Artif. Intell.* **2021**, *97*, 104451.
(64) Kumar, V.; Yadav, R.; Singh, A. Impact of Variable Selection Techniques on the Performance of Regression Models in Spectroscopy. *J. Chem.* **2021**, *35*, No. e3210.
(65) Van der Voet, H.; Wesselingh, J. A.; Huurman, M. The Effects of Wavelength Selection on the Predictive Performance of Spectroscopic Models. *Anal. Chim. Acta* **2022**, *1200*, 338−346.
(66) Gonzalez, A.; Martinez, J. L.; Torres, L. Artificial Neural Networks for Predictive Modeling: A Comparative Study of Performance with Spectroscopic Data. *Chemom. Intell. Lab. Syst.* **2020**, *200*, 104012.
(67) Mirzaei, M.; Mohammadi, A.; Rahimi, H. Comparison of ANN and PLS Techniques for Spectroscopic Data Analysis: Insights from Recent Applications. *Talanta* **2023**, *265*, 124496.
(68) Batzias, F. A.; Tzempelikos, N.; Katsoulis, K. A Comparative Study of PLS and PCR in Spectroscopic Applications: Benefits and Limitations. *Food Chem.* **2020**, *328*, 127104.
(69) Liu, Y.; Zhang, C.; Chen, L. Recent Advances in Partial Least Squares Regression: Methodological Developments and Applications. *Chemom. Intell. Lab. Syst.* **2022**, *227*, 104588.
(70) Zhang, Y.; Wang, Y.; Xu, C. Nonlinear Data Modeling Using Artificial Neural Networks: A Review. *Artif. Intell. Rev.* **2020**, *53*, 2375−2395.
(71) Pannu, S. S.; Dutta, S.; Gupta, S. Artificial Neural Networks in Chemical Engineering: An Overview and Applications. *Chem. Eng. Sci.* **2021**, *228*, 115947.
(72) Fang, Y.; Jiang, Z.; Yu, Y. A Comparative Study of ANN and PLS for Multivariate Calibration: Methodological Insights and Practical Applications. *Chemom. Intell. Lab. Syst.* **2019**, *195*, 103889.
(73) Bhat, S.; Kothari, S. L.; Gupta, D. Utilizing Deep Learning Approaches in Spectroscopic Data Analysis: A Review. *Anal. Chim. Acta* **2022**, *1181*, 338951.
(74) Doshi-Velez, F.; Kim, B. Towards a rigorous science of interpretable machine learning. *Proceedings Of The 34th International Conference On Machine Learning*; International Conference On Machine Learning, 2017, *70*, 1−20.
(75) Casado, F. J.; Martinez, M. A.; Ortiz, J. Explaining Machine Learning Models in Chemometrics: A Case Study on Model Interpretation Strategies. *Anal. Chim. Acta* **2022**, *1202*, 339089.
(76) Sharma, S.; Sharma, M.; Gupta, R. Choosing the Right Model: An Evaluation of Machine Learning Techniques for Predictive Analytics in Chemical Processes. *Chem. Eng. Res. Des.* **2023**, *194*, 1−16.
(77) Liu, Y.; Zhang, C. Recent Advances in Machine Learning for Spectroscopy: Comparative Insights and Future Directions. *TrAC, Trends Anal. Chem.* **2024**, *156*, 116756.
(78) Shibata, K.; Ikeda, Y.Effect of number of hidden neurons on learning in large-scale layered neural networks. *ICCAS-SICE*. IEEE: Fukuoka, Japan, 2009, 5008−5013
(79) Sheela, K. G.; Deepa, S. N. Review on methods to fix number of hidden neurons in neural networks. *Math. Probl. Eng.* **2013**, 425740.

https://doi.org/10.1021/acs.iecr.5c02647
Ind. Eng. Chem. Res. 2025, 64, 19045−19063