An experimental design methodology to evaluate the importance of different parameters on flocculation by polyelectrolytes

Ineide Pinheiro [a,b], Paulo J. Ferreira [a,b], Fernando A. Garcia [b], Marco S. Reis [a,b], Ana C. Pereira [a,b], Christine Wandrey [c], Hamideh Ahmadloo [c], José L. Amaral [d], David Hunkeler [e], Maria G. Rasteiro [a,b,*]

[a] Centre for Chemical Processes Engineering and Forest Products, Coimbra University, Portugal
[b] Chemical Engineering Department, Coimbra University, Pólo II, 3030-790 Coimbra, Portugal
[c] École Polytecnique Fédérale de Lausanne, Switzerland
[d] RAIZ—Institute of Forest and Paper Research, Portugal
[e] Aqua +Tech, Switzerland

ARTICLE INFO

ABSTRACT

Aggregation of particles is fundamental for improving the performance of many solid/liquid processes. Aggregation can be induced by different means, and one of the most common is based on the addition of polymeric additives, namely polyelectrolytes. In this work we have studied the flocculation of precipitated calcium carbonate (PCC), which is used as a filler in papermaking, and is induced by a range of cationic polyacrylamides with different structures, varying in molar mass and degree of branching. The flocculation process was monitored continuously using the LDS technique (laser diffraction scattering), which provides information about the size distribution and structure (fractal dimension) of the aggregates, for each sampling instant and, thus, about the kinetic curve for the flocculation process. This information has then been used to produce a model, based on an experimental design strategy, which allows us to relate flocculation efficiency with the PEL characteristics and concentration. The Partial Least Squares (PLS) regression method was selected to perform the regression, given the significant co linearity among the input variables. The results obtained lead us to conclude that the polymer intrinsic viscosity (related with the PEL molar mass and charge density, and supplying information about the volume of the polymer molecule in the solution) and the degree of branching are the parameters with a stronger influence on flocculation efficiency. The PLS models developed were also successfully validated using an independent data set, which provides confidence on their consistency and prediction accuracy.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Aggregation of particles is fundamental for improving the performance of many solid/liquid processes. Aggregation can be induced by different means, one of the most common being based on the addition of polymeric additives, namely polyelectrolytes (PEL). Flocculation by polyelectrolytes is of importance to many industrial areas, namely in food manufacturing, pharmaceuticals and also in papermaking.

In the case of papermaking, flocculation is the most important phenomena of the wet-end stage since it affects process efficiency (e.g. retention, drainage and runnability) and the quality of the final product (e.g. formation, strength and porosity) [1]. However, to control the flocculation process it is necessary to know and understand how chemical additives act during the whole process.

Various processes occur simultaneously during flocculation: adsorption of polymer molecules at the particle surface; re-arrangement (or re-conformation) of adsorbed polymeric chains; collisions between destabilized particles to form aggregates (flocs); and break-up of flocs. The importance of each process depends on the flocculant characteristics, like structure, molar mass, charge density and concentration; on the characteristics of the suspended particles, like size and charge; on the characteristics of the suspending medium, like pH, conductivity and ionic charge; and, finally, on the contact time and turbulence intensity, among others.

The structure of the polyelectrolyte determines its conformation when adsorbed on the particle surface and, therefore, the predominant flocculation mechanism [2]. In general, if molecular weight is high and charge density is low the polymer adsorbs on the particle surface in such a way that tails and loops are extended far beyond the electric double layer and can interact with the polymer adsorbed on other particles—in this case the flocculation process is dominated by bridging bonds [3,4]. When the charge density is high, the bridging capability is reduced because there is a tendency for the polymer chains to adopt a flatter conformation on the particle surface, which results in the formation of cationic patches that attract the polymer

Table 1
Summary of the flocculants characteristics.

Alpine Floc™	Intrinsic viscosity—IV[a] (ml/g)	Charge density (wt.%)	Number of branches
E1	2308	45.5	0
E1+	1817	48.8	1
E1++	1771	46.6	2
E1++++	1775	42.8	4
E2	1550	47.3	0
E2+	1164	41.4	1
E2++++	977	42.1	4
E3	989	44.6	0
E3++++	594	40.1	4

[a] Schulz-Blaschke in 0.05 M NaCl.

of branching. The particles have been kept constant (PCC particles with a median diameter of 2.05 μm and a surface charge of −32 mV) and all tests were conducted in distilled water. The flocculation process was monitored continuously using the LDS technique (laser diffraction spectroscopy), which supplies information about the size distribution and structure (fractal dimension) of the aggregates [7], for each sampling instant and, thus, about the kinetic curve for the flocculation process. The information obtained is of great importance to understand and predict the flocculation mechanism induced by polyelectrolytes, and thus, to optimize this process.

All this information must be treated adequately, in order to better understand the effect of the different parameters, referred above, on the efficiency of the flocculation process. Flocculation efficiency, having in mind the application under analysis, papermaking, has been defined taking into consideration not only the size of the flocs but also the kinetics of the flocculation process (flocculation rate).

The ultimate objective is to produce a model, based on a sound experimental design strategy, which allows us to relate flocculation efficiency with the PEL characteristics and concentration. For such, Partial Least Squares Regression (PLS) will be employed in order to effectively capture the desired relationship. Unlike Ordinary Least Squares (OLS), PLS effectively handles situations where the X-variables (or input variables) are highly collinear, as happens in the present situation. It basically consists in finding a lower dimensional subspace of the whole X-variable hyperspace, which presents a high covariance with the response (efficiency) and also provides a proper description of the original X data. By first projecting the X-variable observations onto such a predictive subspace, and using the result of such operation as the new predictor variables (also called latent variables), one can effectively handle the correlation structure present, and obtain a stable and robust predictive model.

free surfaces of other particles [4]. On the other hand, the introduction of branches in the polymer chain can alter the PEL conformation on the particle surface and, again, influence the aggregation mechanism. Studies have shown that the degree of branching of the polymer affects the flocculation kinetics and the aggregate properties [5].

The concentration of the flocculant is also a key parameter, since the rate of adsorption depends on the amount of polymer per unit area of the particle surface. Moreover, the flocculant concentration also affects the conformation rate: polymer re-arrangement is relatively fast at low surface concentration but rather slow on crowded surfaces since neighboring molecules interfere with the re-arrangement [3,6].

The methods used to evaluate flocculation are numerous. In this work we have studied the flocculation of precipitated calcium carbonate (PCC), used as filler in papermaking, induced by a range of cationic polyacrylamides with different structures, varying in molar mass and degree

Table 2
Summary of the experimental flocculation efficiency for the different polymers.

| Polymer	Concentration (mg PEL/g PCC)	Efficiency (s⁻¹) Test 1	Test 2
E1	6	0.5859	0.5976
E1	8	0.6140	0.6142
E1	11	0.6017	0.5984
E1	16	0.5348	–
E1+	8	0.5936	0.5929
E1+	10	0.5986	0.5989
E1+	12	0.5839	0.5875
E1++	18	0.6122	0.6091
E1++	20	0.6162	0.6162
E1++	22	0.6181	0.6173
E1++++	8	0.5540	0.5681
E1++++	10	0.6149	0.6156
E1++++	12	0.5227	0.5238
E2	6	0.5119	0.5134
E2	8	0.5103	0.5120
E2	10	0.4860	0.4869
E2	14	0.4773	0.4768
E2+	6	0.5758	0.5745
E2+	10	0.5679	0.5679
E2+	13	0.5523	0.5529
E2++++	6	0.5983	0.6004
E2++++	8	0.5982	0.5978
E2++++	10	0.6011	0.5996
E3	2	0.4857	0.4800
E3	6	0.5045	0.5008
E3	8	0.4964	0.4955
E3	20	0.5178	–
E3++++	6	0.5488	–
E3++++	10	0.5771	0.5782
E3++++	12	0.5815	0.5819
E3++++	14	0.5779	0.5779
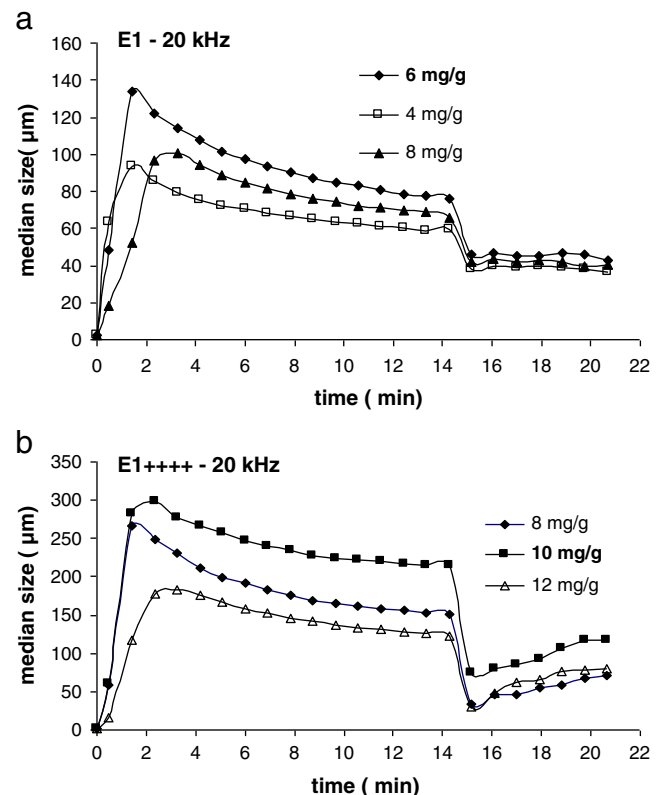
Fig. 1. Flocculation kinetic curves obtained with LDS: flocculation of PCC induced by E1 (a) and E1++++ (b).
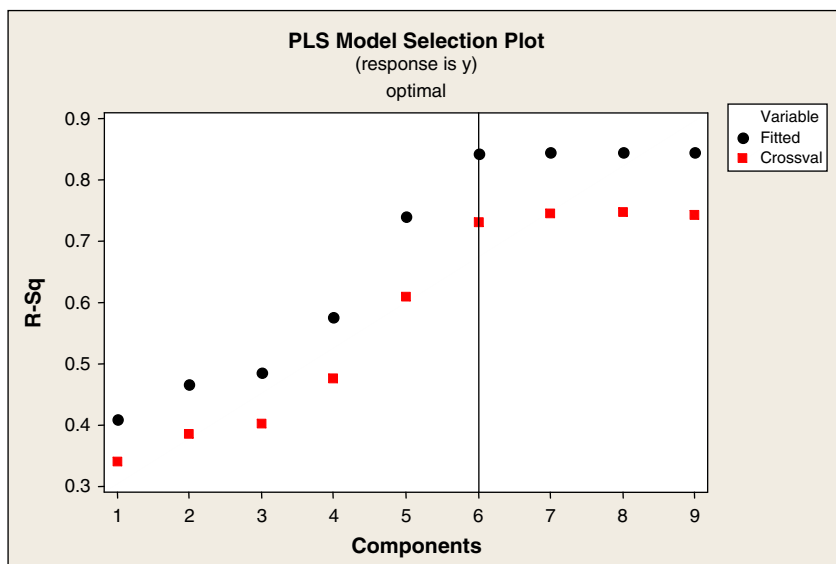
Fig. 2. Evolution of the fitted and cross-validated coefficient of determination, with the number of latent variables considered in the PLS model, when all variables are taken as quantitative.

This paper is organized as follows: in Section 2, we introduce the materials used and describe the experimental procedures followed in this work; then, in Section 3, the data analysis methodologies are briefly reviewed, in order to facilitate the interpretation of the results presented in Section 4; in this latter section, the results obtained are presented in detail and properly discussed, namely regarding the estimation and validation of two PLS models; finally, in Section 4, the main results of this work are reviewed and summarized.

2. Experimental

2.1. Materials

In this work, we have used five cationic polyacrylamide (C-PAM) emulsions of very high molecular weight with a charge density between 40 and 50% weight, copolymers of acrylamide (AM) and acryloyloxyethyltrimethyl ammonium chloride (Q9), developed and supplied by AQUA + TECH [8,9]. The main characteristics of the polyelectrolytes are summarized in Table 1. Flocculant solutions were prepared with distilled water at 0.1% (w/w). The diluted solutions were prepared every day.

The flocculation tests were carried out on a commercial scale nohedral PCC suspension, supplied by OMYA. The PCC suspension was prepared at 1% (w/w) in distilled water and, in order to obtain a good dispersion of the particles, the suspensions were first magnetically stirred for 20 min and then submitted to sonication at 50 kHz during 15 min. After this treatment is completed, the median size of the particles was approximately 2.05 μm (measured by LDS) and the suspension pH, 8.5. The zeta potential of the particles was −32 mV in distilled water.
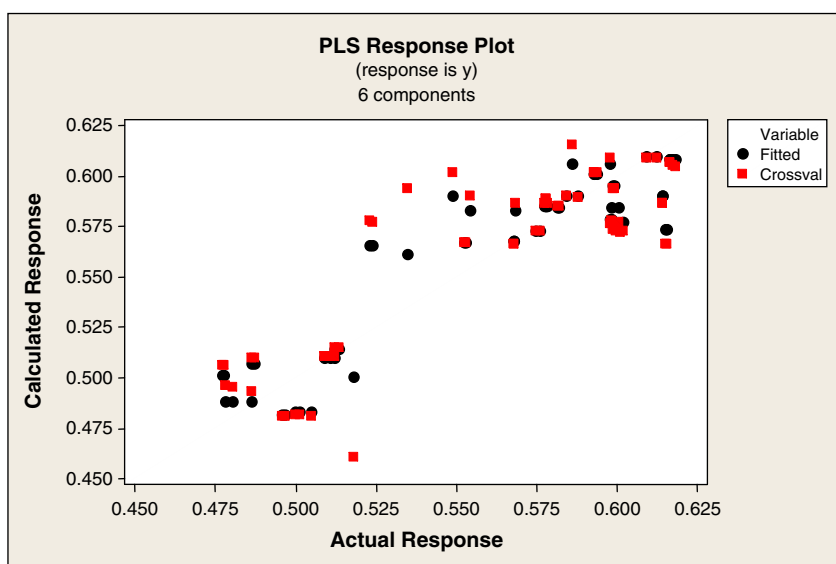

Fig. 3. Fit and cross-validated fitted responses for the PLS response, when all variables are taken as quantitative.

Table 3
Summary of the performance statistics for the PLS model in Case 1.

Components	X variance	Error	R^2	PRESS	R^2 (pred)
1	0.398734	0.0790868	0.409974	0.0883645	0.340759
2	0.645717	0.0716416	0.465520	0.0824299	0.385034
3	0.953899	0.0690158	0.485110	0.0800672	0.402661
4	0.983550	0.0568114	0.576160	0.0702704	0.475750
5	0.989638	0.0347039	0.741092	0.0521775	0.610731
6	0.993841	0.0209809	0.843473	0.0361632	0.730205

Table 4
Coefficients for flocculation efficiency model with 6 components.

	y	y standardized
Constant	0.455156	0.00000
1—x1	0.119750	4.37704
2—x2	0.000511	0.05066
3—x3	0.000015	0.17321
4—x1*x2	−0.000567	−0.23123
5—x1*x3	−0.000013	−0.58982
6—x2*x3	−0.000003	−0.63905
7—x1*x1	−0.019655	−3.00216
8—x2*x2	0.000145	0.35716
9—x3*x3	0.000000	0.86884

2.2. Experimental methodology

The intrinsic viscosity (IV) is a suitable and informative parameter, in particular when the goal is the assessment of the impact of PEL of different chain architectures and structure, on technological processes such as flocculation, where the polymer coil dimensions are important. The prerequisites are known and constant environment/solution conditions, which were contemplated in this study. For a known molar mass, the polymer coil size in solution depends upon the chemical structure, chain architecture (linear, branched), and the solvent quality. IV is dependent on the coil size of the polymer in solution and, thus, can supply information about the conformation adopted by the PEL in solution.
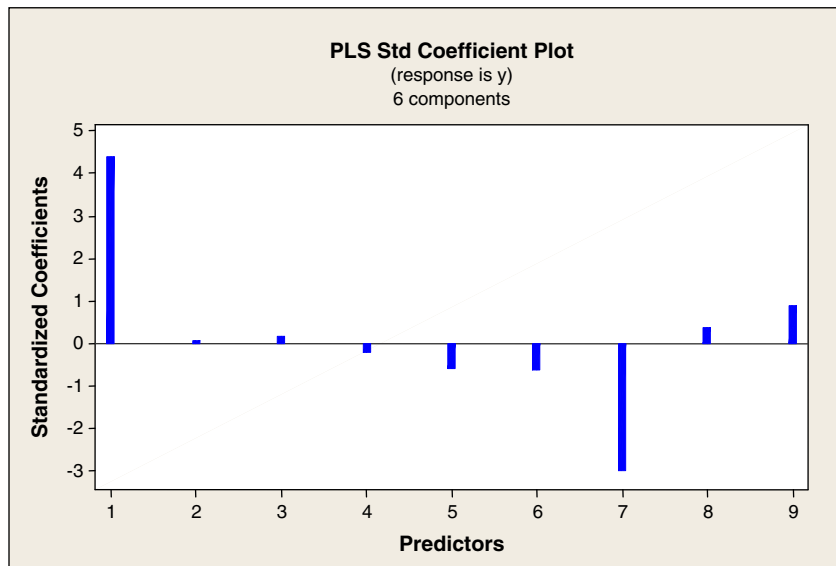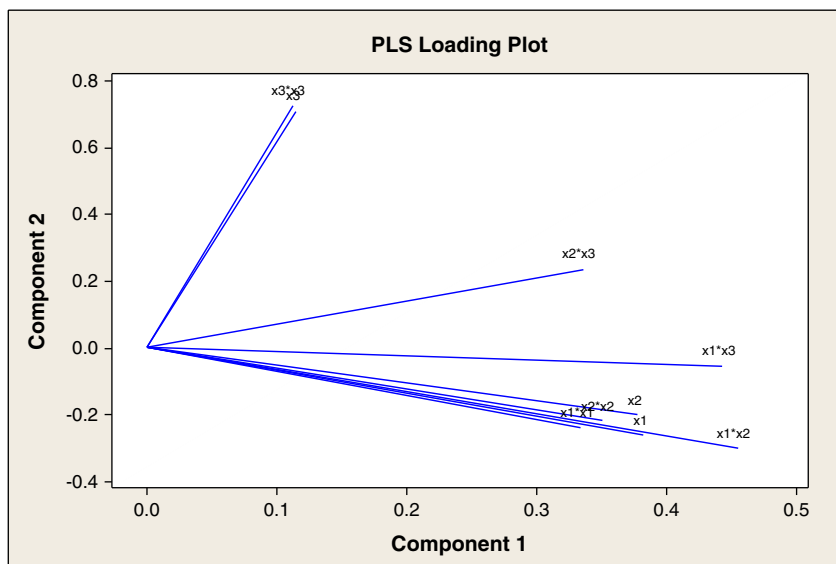

Fig. 4. Standardized PLS coefficients.

Fig. 5. Loading plot for the first two PLS latent variables.
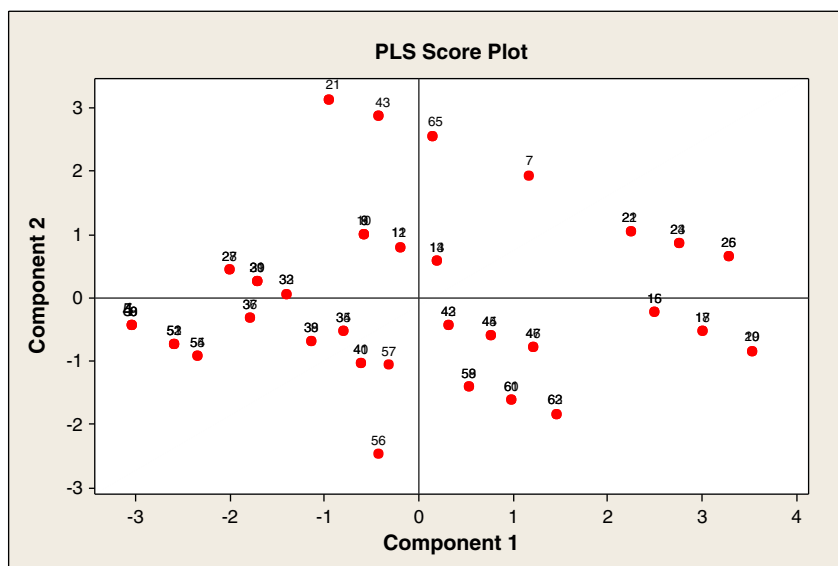
Fig. 6. Score plot for first two PLS component variables.

The IV of the isolated and redissolved copolymers was determined in 0.05 M NaCl aqueous solution at 20 ± 0.1 °C by dilution viscometry, using an automatic capillary viscometer, Viscologic TI1 (Sematech, France), capillary 0.58 mm. The extrapolation to zero concentration was performed according to Huggins' method, as described recently [10].

PCC flocculation was monitored by measuring aggregate size using light diffraction spectroscopy (LDS) in a Malvern Masterziser 2000 (Malvern Instruments). The PCC suspension was added to 700 ml of distilled water in the equipment dispersion unit, until a certain, fixed level of obscuration, was obtained, corresponding to an average PCC concentration around 0.05% (w/w). The tests were carried out with stirring and in turbulent conditions (1400 rpm corresponding to an average shear rate of 312 s^{-1}) as described elsewhere [7]. Obscuration was always kept above 5%, in order to assure a good signal quality [11]. As obscuration decreases pronouncedly during the flocculation test due to floc growth, the tests have to be initiated with a

higher obscuration than usual, to guarantee that, at the end of flocculation, obscuration was always higher than 5%.

Flocculants were tested for a range of concentrations, usually close to the optimum dosage, which is defined as the PEL dosage leading to larger flocs and fastest kinetics. At least one concentration below the optimum (PEL shortage) and one above (excess PEL), were always tested.

To determine the flocculation kinetics curve, a predetermined amount of flocculant was added at once to the suspension and the flocs size distribution was measured every minute during 14 min, i.e., until the flocs size stabilized. The particle size distribution of the PCC was measured before adding the flocculant to the suspension.

Fig. 1 gives two examples of the flocculation kinetic curves obtained for the flocculation of PCC with E1 and E1++++. It is obvious from the analysis of Fig. 1, that the shape of the kinetic curve depends on the polyelectrolyte characteristics. For the linear polymer, the reconformation of the aggregates, after reaching the maximum floc size, is less pronounced
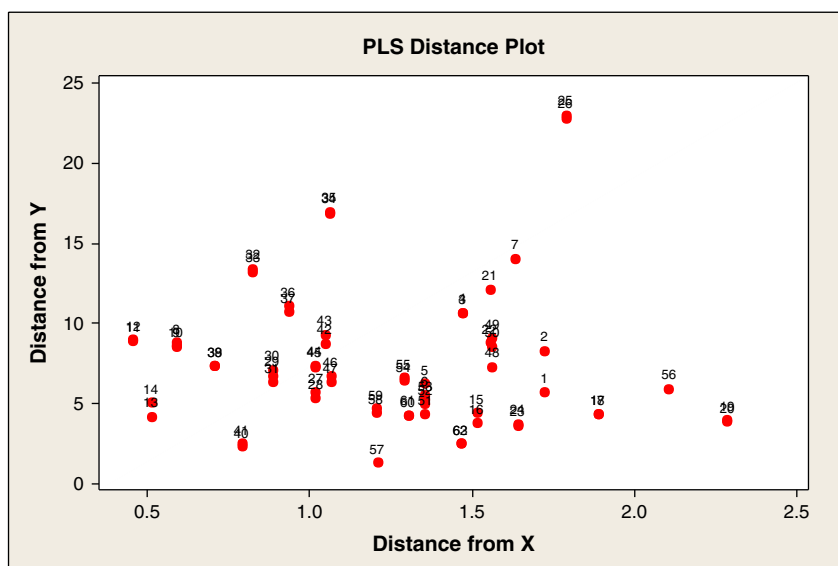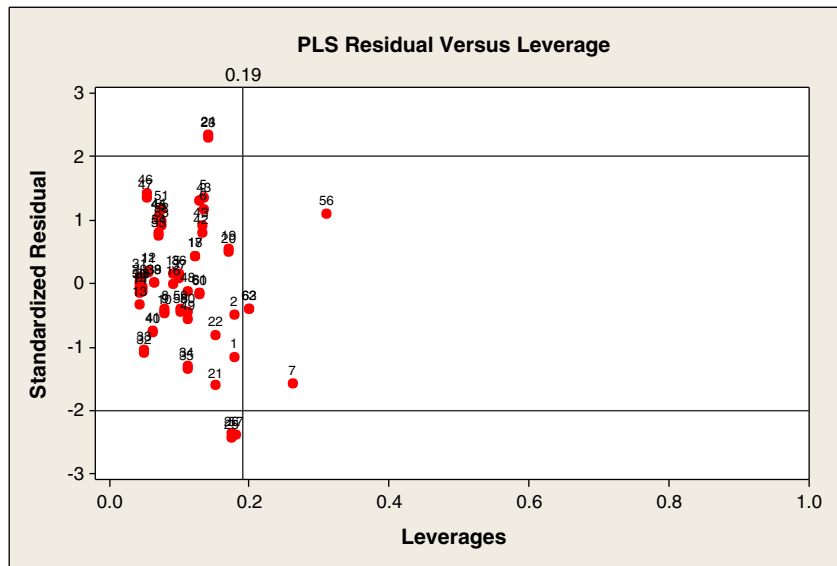

Fig. 7. Distance plot.

Fig. 8. Residual versus leverage.

than with the branched polymer. Moreover, the flocculation rate, during the first stage of flocculation, is faster with the branched polymer. Once the floc size had stabilized (after 14 min) the aggregates were submitted to sonication (20 kHz) which induced breakage of the flocs. After that, partial reflocculation occurred. This can also be observed in Fig. 1.

The reported values of the median particle size (dp$_{50}$) represent an average of at least four replications. The experimental flocculation results corresponding to the range of polymers in Table 1 have been presented in detail elsewhere [12].

3. Methods

3.1. Partial least squares

Partial least squares (PLS) regression is a method for relating two data matrices, the *X*-variables or predictors (or input variables), and the *Y*-variables or responses, through a linear multivariate model.

Unlike Ordinary Least Squares (OLS), PLS is able to cope with highly collinear predictors, as happens in the present work, being furthermore able to handle noisy and even incomplete data in both predictors and responses [13,14].

PLS is a bilinear calibration method that finds those linear combinations of predictors presenting maximal covariance with the response. These linear combinations, called latent variables, are found sequentially, each one spanning a different part of the predictive space, until a point where prediction ability does not improve significantly, or begin getting worse.

PLS-modeling of the relationship between two blocks of variables can be described according to Eqs. 1 and 2.
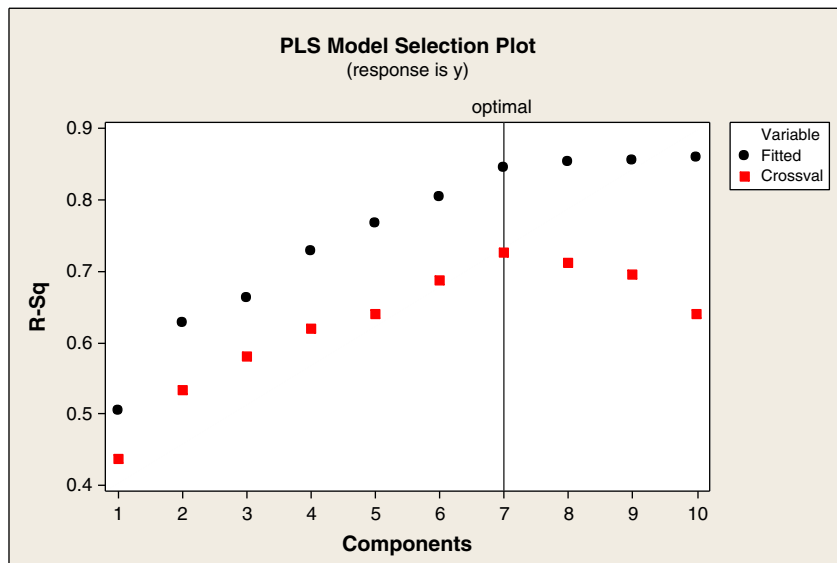
$$X = TP^T + E. \tag{1}$$

$$Y = UC^T + F. \tag{2}$$

Fig. 9. Evolution of the fitted and cross-validated coefficient of determination, with the number of latent variables considered in the PLS model, when branching is taken as a qualitative variable and concentration and intrinsic viscosity as quantitative variables.
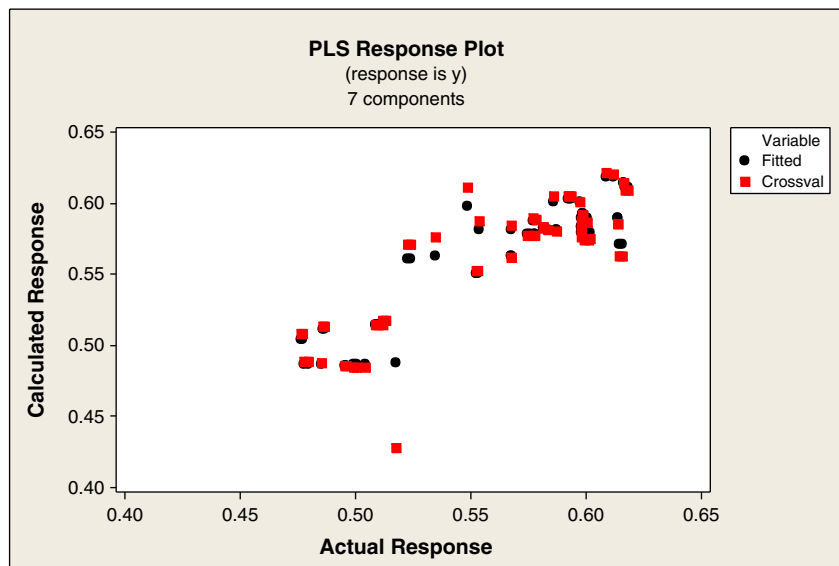
Fig. 10. Fit and cross-validated fitted responses for the PLS response, when branching is taken as a qualitative variable and concentration and intrinsic viscosity as quantitative variables.

Table 5
Summary of the performance statistics for the PLS model in Case 2.

Components	X variance	Error	R^2	PRESS	R^2 (pred)
1	0.275877	0.0663195	0.505225	0.0755221	0.436569
2	0.440029	0.0498522	0.628079	0.0624905	0.533791
3	0.676481	0.0453737	0.661490	0.0563973	0.579250
4	0.782096	0.0363530	0.728790	0.0511681	0.618261
5	0.941290	0.0311904	0.767305	0.0482533	0.640008
6	0.986596	0.0262186	0.804397	0.0419163	0.687284
7	0.991330	0.0206920	0.845628	0.0368681	0.724947

The information about the X and Y observations are stored in the score matrices T and U, respectively. The X-loading matrix P and Y-loading matrix C contain information about the way variables in each block combine with each other, to form the successive latent variables, which can provide insights into the way they are correlated and act together in the predictive model. Matrices E and F are residual matrices, containing unstructured variability not captured by the model [13,15].

The PLS model described by Eqs. 1 and 2, can be re-expressed as a regular regression model, with the same form of one derived from OLS, but with coefficients estimated through the PLS regression algorithm, B_{PLS}, namely:

$$Y = X \cdot B_{PLS} + F. \tag{3}$$

The PLS regression coefficient vector can be expressed in terms of the PLS X-loading matrix, P, and the so-called PLS weights (W), as:

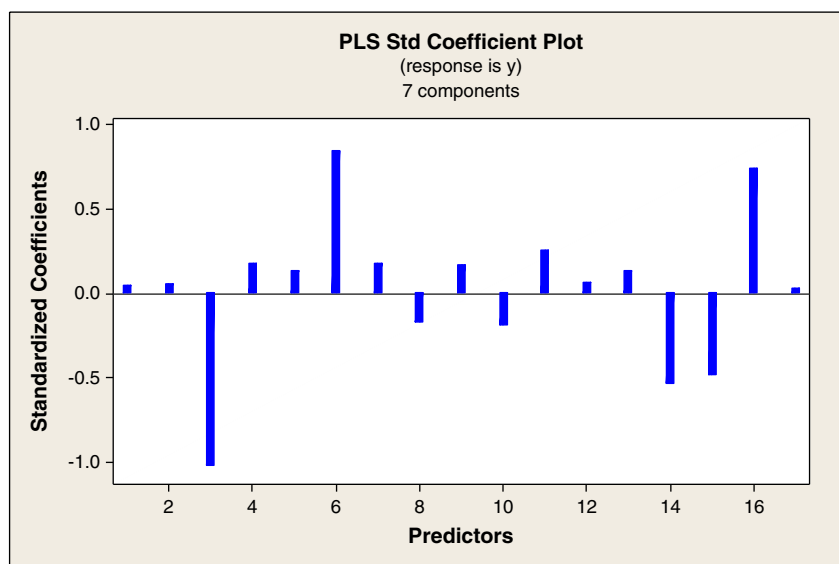$$B_{PLS} = \overbrace{W\left(P^T W\right)^{-1} C^T}^{W^*} = W^* C^T. \tag{4}$$

Fig. 11. Standardized coefficients - Case 2.

Table 6
Coefficients for flocculation efficiency model with 7 components - Case 2.

	y	y standardized
Constant	0.545014	0.00000
1—x2	0.000506	0.05016
2—x3	0.000005	0.05405
3—x1_0	−0.096435	−1.02284
4—x1_1	0.020402	0.17900
5—x1_2	0.020853	0.13271
6—x1_4	0.085191	0.84764
7—x1_0*x2	0.001653	0.17442
8—x1_1*x2	−0.001972	−0.17357
9—x1_2*x2	0.001286	0.16431
10—x1_4*x2	−0.001857	−0.18859
11—x1_0*x3	0.000014	0.25645
12—x1_1*x3	0.000005	0.06290
13—x1_2*x3	0.000012	0.13271
14—x1_4*x3	−0.000043	−0.53333
15—x2*x3	−0.000002	−0.48606
16—x2*x2	0.000000	0.73836
17—x3*x3	0.000010	0.02558

3.2. Data pre-processing

PLS modeling is sensitive to the scale in which variables are expressed. It works better when data are quite symmetrically distributed and have a rather constant error variance. Usually, data are centered and scaled to unit variance before analysis, especially when variables have different units or scales of variability. Scaling all variables to unit variance corresponds to the assumption that all variables are, a priori, equally important in the modeling task [16]. In this work, all variables were centered and scaled to unit variance.

3.3. Cross-validation

In order to validate the model, two types of approaches can be adopted. The best way is to use an independent data set where the model is applied and its prediction ability assessed (external validation). However, this requires an independent data set, also called a "test set", which quite often cannot be provided, as data collected is scarce.

In the absence of such an independent "test set", an alternative consists of using cross-validation, which simulates how well the model predicts new data, but using only the data set available, i.e., the "training"

data set (internal validation) [14,16]. It basically consists in dividing the training set in blocks, and successively leaving one block aside. A model is then estimated with the remaining blocks, which will be employed in the estimation of the response values for the block removed. The prediction errors (or better, the cross-validation prediction errors) are then registered, and the process is repeated for another block, after the one removed is integrated again in the set of blocks to estimate a new model. This process is repeated until all blocks and observations are removed one time, after which the overall cross-validation error can be estimated. This methodology can be used to determine the appropriate number of components or latent variables to retain in the model, i.e., the model dimensionality (number of PLS components) [16].

4. Results

In the present work, three predictor variables were considered: branching (x1), concentration (x2) and intrinsic viscosity (x3). The response variable is the flocculation efficiency (y) which has been defined as:

$$y = \frac{(d_{50\,max} - d_{50PCC})}{d_{50\,max}} \times \frac{1}{t} \tag{5}$$

where, d_{50max} corresponds to the maximum median floc size in the flocculation kinetic curve, for each set of conditions (polymer type and concentration); d_{50PCC} is the median size of the primary precipitated calcium carbonate and t is the time corresponding to the maximum median floc size. Table 2 summarizes the experimental values of y for the different polymers and concentrations.

The main objective of this study was to identify the most important variables for this relationship and to find the combination of factors describing better the flocculation efficiency. The PLS methodology was used for that purpose. In this study we have also compared the models obtained with PLS by considering the polymer branching (variable x1) either as a qualitative or quantitative variable. They differ essentially in the following: the first approach requires the generation of additional variables, that represent the presence/absence of a given level of the variable (the branching degree, in this case); on the other hand, the second approach uses the number of branches directly, as a well-defined quantitative entity.

Every test, for each set of conditions (polymer type and concentration) was repeated at least twice. However, in the PLS regression
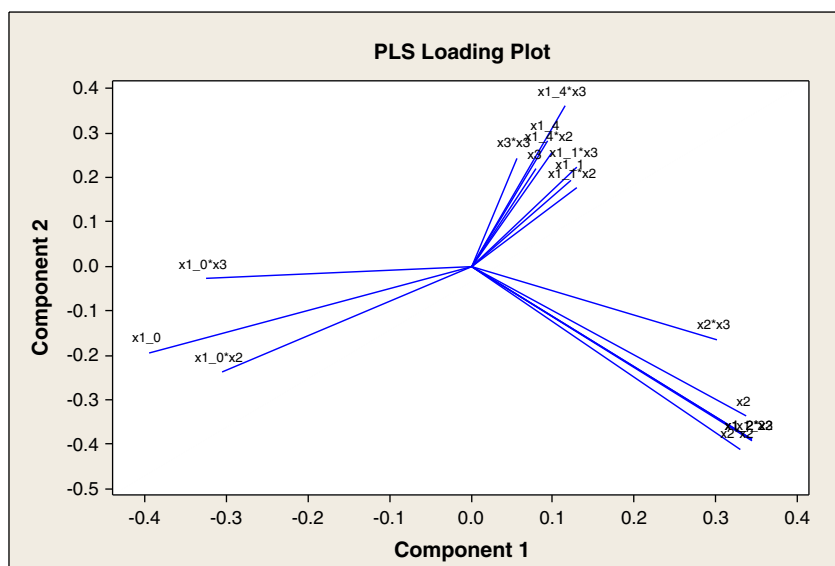

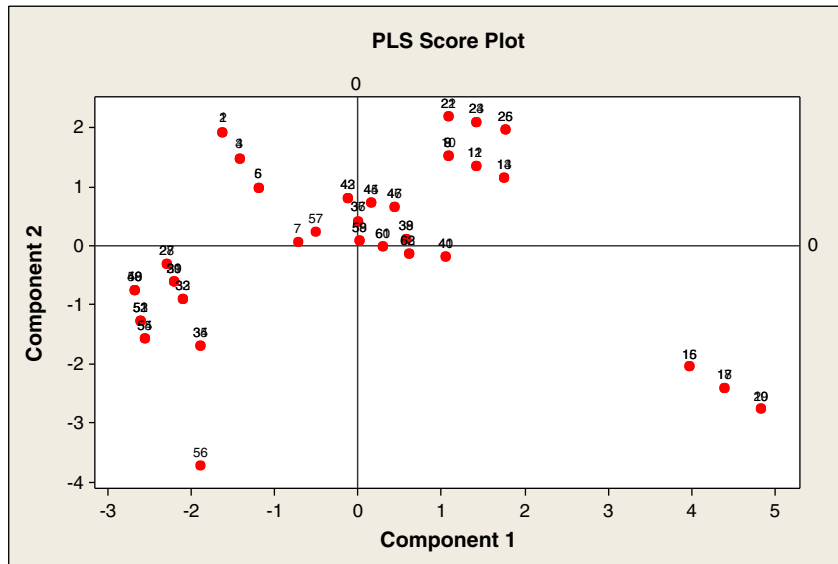**Fig. 12.** Loading plot for the PLS model with 2 components.

Fig. 13. Score plot for the PLS model with 2 components.

modeling, each test was considered individually, i.e. the values in the matrix were not the average values of the repetitions, but the individual values for each sample.

4.1. Case 1: branching, concentration and intrinsic viscosity as quantitative variables

A full quadratic model was estimated by PLS regression, after expanding the original data set with the quadratic and cross-product terms of the quadratic expansion. All variables have been considered as continuous variables.

Fig. 2 represents the evolution of the model's coefficients of determination versus the number of components of the PLS model, and Fig. 3 corresponds to the predicted response versus the actual response in leave-one-out cross-validation (where the cross-validation blocks are formed by a single observation).

The vertical line in Fig. 2 indicates that the optimal model has six components. The response plot (Fig. 3) indicates that the model

predicts the removed observations quite well. Although there are differences between the fitted and cross-validated fitted, none are severe enough to indicate an extreme leverage point.

We selected 6 components for analyzing the model, because they are able to explain 84% of the model response, and not much improvement is achieved in the model with further addition of components.

Table 3 summarizes quantities that characterize the fitting and prediction ability for the models corresponding to the first latent variables (1 to 6), namely X variance explained, error, determination coefficient (R^2), prediction sum squares (PRESS) and the predicted determination coefficient (R^2 (pred)).

The model with 6 components, selected by cross-validation, has a reasonably high R^2 of 84% and R^2 (pred) of 73% (Table 3), providing good indicators for its fitting ability and predictive accuracy.

Fig. 4 and Table 4 display the regression coefficients for the flocculation efficiency (y) model. Obviously, branching and intrinsic viscosity are the most important factors, but the influence of concentration cannot be neglected. The plot of Fig. 4 can be used to interpret the
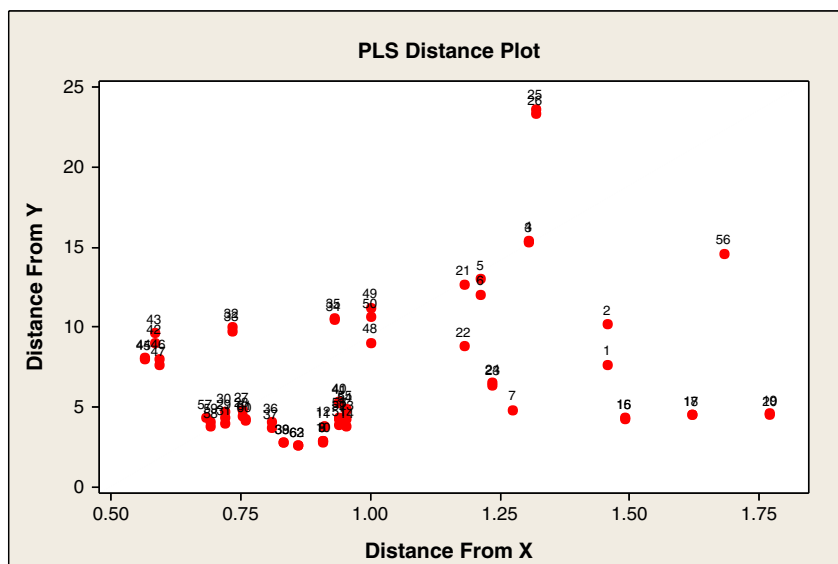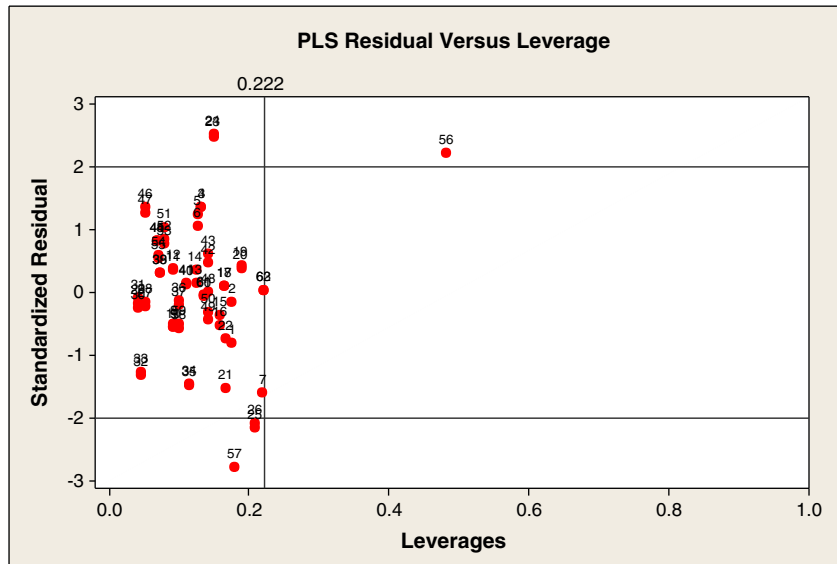

Fig. 14. Distance plot.

Fig. 15. Residual versus leverage.

magnitude and sign of the coefficients. The variables x1, x1∗x1 and x3∗x3 have the largest coefficient and impact on flocculation efficiency. x1, x2∗x2, and x3∗x3 give a positive contribution related to efficiency, while x1∗x3, x2∗x3 and x1∗x1 are negatively related.

Fig. 16. a) Surface curve of flocculation efficiency for Case 1; b) Surface curve of flocculation efficiency for Case 2.

Fig. 5 shows the loading plot for the first two dimensions, whereas Fig. 6 illustrates the score plot also for the first two components.

The loading plot compares the relative influence of the predictors on the response, for each component. The X-loading indicates the importance of the corresponding predictor to the component. Fig. 5 illustrates the relation between component 1 and component 2. In this case, x3 and x3∗x3 have a very short distance in relation to x-axis, indicating that they have low x-loading and are not so much correlated to efficiency. On the other hand, the other variables located further to the right, at longer distances from the center, correspond to higher loadings and thus are more important to the efficiency value. Furthermore, groups of variables lying close to each other in this plot, represent correlated or partially redundant variables. Branching (x1) and concentration (x2) are highly positively correlated because they are quite close to each other, the same applying to their quadratic forms, x1∗x1 and x2∗x2. Variables x3, x3∗x3 and x1 form another group, which are situated in a different quadrant, whereas x1∗x3 and x2∗x3, point in rather specific and distinct directions. (Note that we are just signaling groups of positively related variables, but they can also be correlated with each other in a negative way, if they lie in opposite quadrants, with significant loading coefficients). Furthermore, component 1 seems to be dominated by the influence of x1 and x2, while component 2 is essentially dominated by x3.

The PLS score plots illustrated in Figs. 7 and 8 are useful for understanding the relationship and distribution of the observations. These plots show that there are no obvious outliers. By examining the distance plot (Fig. 7), one can see that all data fall in a cluster, excluding some observations (19, 20, 25, 26 and 56), that are a bit more deviant. The residual versus leverage plot (Fig. 8) shows that some observations (7, 56, 62 and 63) are also potential outliers as they lay outside the horizontal reference lines. Among these, several observations (23, 24, 25, 26 and 57) have also large leverage values, as they are on the right of the vertical reference line.

4.2. Case 2: branching as qualitative variable and concentration and intrinsic viscosity as quantitative variables

In this case, we treated branching as a qualitative variable and unfold such variable in several indicator dummy variables, by using an indexing scheme for the presence/absence of a given branching number. Fig. 9 represents the evolution of the coefficients of determination versus the
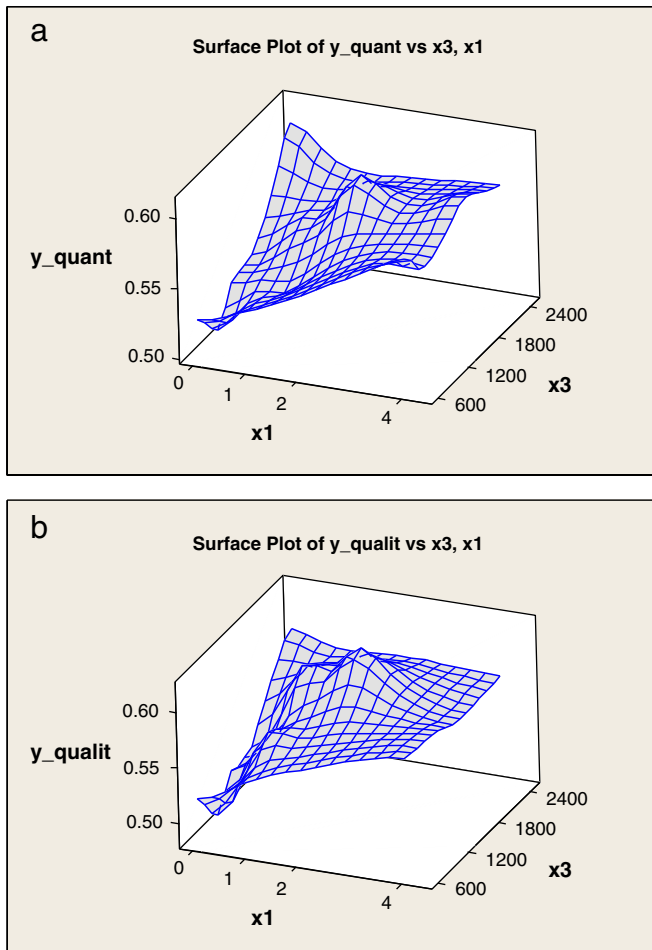
Table 7
Summary of the flocculants characteristics in the test set.

Alpine Floc™	Intrinsic viscosity—IV[a] (ml/g)	Charge density (wt.%)	Number of branches
F1	1399	35.0	0
F1+	1144	29.5	1
F1++	687	37.7	2
F1++++	816	34.4	4

[a] Schulz-Blaschke in 0.05 M NaCl.

number of components, and Fig. 10 corresponds to the calculated response for 7 components versus the actual response.

The optimal component number will be used in the remaining computations (7 components). Table 5 proves that the model is significant with 7 components, which explain 85% of the model.

In Fig. 11 we can see regression coefficients for the flocculation efficiency (y) model obtained from PLS.

In this case, the variables which have greater impact on flocculation efficiency are branching 0 (x1_0), branching 4 (x1_4), interaction between branching (4) and intrinsic viscosity (x1_4∗x3), and the quadratic terms for intrinsic viscosity (x3∗x3) and concentration (x2∗x2). The contribution of the linear term of concentration (x2), is not very significant (see Table 6 and Fig. 11). The variables with the highest positive contribution to efficiency are x1_4 and x3∗x3, while those with the strongest negative contribution to efficiency are x1_0, x1_4∗x3 and x2∗x3.

Fig. 12 shows the 2D loading plot, while Fig. 13 is relative to the score plot for the same dimensionality.

The loading plot (Fig. 12) displays the relationship between all 17 variables at the same time. Variables contributing with similar information are grouped together. Evaluating the results, it is possible to identify three groups of positively related variables (Fig. 12). Analyzing the score plot, it is possible to see that the samples are now more scattered in the 2D space of the first two components, with some observations falling in more distant regions regarding the main cluster.

This behavior can be more accurately checked by analyzing the distance plot (Fig. 14) and the residual versus leverage plot (Fig. 15), which provide information about the distances in the predictors/response spaces, as well as their relative weight in the estimated model (leverage).

From Fig. 14, it is possible to see that a few observations, 19, 20 and 56, have a greater distance value on the y-axis, and observations like 25 and 26, have the longest distance on x-axis. The residual versus leverage plot (Fig. 15) shows that observations 24, 25, 56 and 57 are also potential outliers in model predictions, as they lay outside the horizontal reference lines, some of them having also large leverages associated (high influence in the estimated model).

Fig. 16 shows the surfaces for the flocculation efficiency as a function of x1 and x3 for the two models. It is obvious that the responses of the two models are quite similar, regarding the interaction between the variables most important for the model.

4.3. Validation of the PLS models developed in Case 1 (quantitative variables) and Case 2 (qualitative/quantitative variables)

In order to assess model consistency and prediction accuracy, an independent test set, never used before during the estimation stage, was collected and used for testing the models developed in Case 1 (quantitative) and Case 2 (qualitative/quantitative). This data set constitutes the "F" series of polymers, whose characteristics are summarized in Table 7, comprising a total of 24 new samples.

Introducing the X-variables for the test set in both models developed for Case 1 (quantitative) and Case 2 (qualitative/quantitative), it is possible to obtain their point and interval estimates, which can be compared with the observed values. Of particular importance in this stage is the analysis of the sample-specific prediction intervals (PI), which should contain the observed samples with a given probability, defined a priori, through the specification of the confidence level (in the present case, such probability is of 95%). However, for PLS, there is no analytical closed formula available for computing the sample-specific prediction intervals, as happens for OLS. Therefore, we have implemented a procedure based upon the methodology of "bootstrap residuals", which enables the estimation of the interval limits, in a non-parametric, data-driven way. More details about this technique are available at [17].

Table 8 summarizes the results obtained, where it is possible to verify that, for the model developed in Case 1 only one sample lays outside the prediction intervals, whereas for the model developed in Case 2, three samples are detected in such condition. These results

Table 8
Validation results for the two models developed for Case 1 and Case 2: point estimate and prediction intervals.

Observation	Y (observed)	Case 1: Quantitative			Case 2: Qualitative/quantitative		
		Y (predicted)	PI: lower bond	PI: upper bond	Y (predicted)	PI: lower bond	PI: upper bond
1	0.5461	0.509243	0.469093	0.549394	0.509108	0.468668	0.549547
2	0.5466	0.509243	0.469093	0.549394	0.509108	0.468668	0.549547
3	0.5376	0.501014	0.461285	0.540742	0.505122	0.465336	0.544908
4	0.5377	0.501014	0.461285	0.540742	0.505122	0.465336	0.544908
5	0.5231	0.497434	0.457157	0.53771	0.50147	0.461098	0.541841
6	0.5263	0.497434	0.457157	0.53771	0.50147	0.461098	0.541841
7	0.6045	0.570811	0.530322	0.6113	0.573774	0.533235	0.614314
8	0.6009	0.570811	0.530322	0.6113	0.573774	0.533235	0.614314
9	0.6069	0.567476	0.52741	0.607541	0.56186	0.521559	0.602162
10	0.6118[a]	0.567476	0.52741	0.607541	0.56186[a]	0.521559	0.602162
11	0.5990	0.567729	0.527487	0.607971	0.54242[a]	0.50042	0.58442
12	0.6032	0.567729	0.527487	0.607971	0.54242[a]	0.50042	0.58442
13	0.6214	0.630155	0.585276	0.675034	0.595652	0.552234	0.63907
14	0.6224	0.630155	0.585276	0.675034	0.595652	0.552234	0.63907
15	0.6188	0.642422	0.595283	0.689562	0.597664	0.55331	0.642019
16	0.6197	0.642422	0.595283	0.689562	0.597664	0.55331	0.642019
17	0.6138	0.652053	0.603046	0.70106	0.59911	0.553833	0.644386
18	0.6128	0.652053	0.603046	0.70106	0.59911	0.553833	0.644386
19	0.5825	0.581706	0.541327	0.622085	0.586572	0.546145	0.626999
20	0.586	0.581706	0.541327	0.622085	0.586572	0.546145	0.626999
21	0.6019	0.578394	0.538345	0.618444	0.58042	0.540299	0.620541
22	0.5985	0.578394	0.538345	0.618444	0.58042	0.540299	0.620541
23	0.5904	0.576245	0.535714	0.616776	0.57435	0.533514	0.615186
24	0.5906	0.576245	0.535714	0.616776	0.57435	0.533514	0.615186

[a] Samples whose prediction intervals do not contain the observed values.

confirm the stability and accuracy of the models developed, and provide an additional argument for reassuring the validity of our interpretative considerations, based on their analysis.

5. Conclusions

Flocculation of precipitated calcium carbonate was monitored using light diffraction spectroscopy. Several cationic polyacrylamides with different molar mass and degree of branching, all copolymers of acrylamide (AM) and acryloyloxyethyltrimethyl ammonium chloride (Q9), were tested. LDS supplied information about the kinetic curves for flocs growth. The kinetic curves are influenced by the PEL structure. PELs with different molar mass and degree of branching corresponded to different values for the intrinsic viscosity, indicating differences in the polymer conformation.

A partial least squares (PLS) method was used to identify and correlate the most important variables influencing flocculation efficiency. Two strategies were followed: in the first one, branching was considered as a quantitative variable, while in the second case branching was taken as a qualitative variable. Both models could explain quite well the variability observed in experimental data.

However, the model where branching is considered a continuous variable is less complex and presents a good prediction ability, being therefore particularly interesting. Nevertheless, the predicted responses obtained with the two models are quite similar.

The results obtained, allow us to conclude that the polymer intrinsic viscosity (related with the PEL molar mass and charge density, and supplying information about the volume of the polymer molecule in the solution) and the degree of branching, are the parameters with a stronger influence on flocculation efficiency. The responses of the two models allow us to identify a region of maximum flocculation efficiency corresponding to an intermediate degree of branching and intrinsic viscosity. In fact, branching usually induces a more coiled conformation of the polymer chain in solution, and this can be beneficial for the bridging mechanism. Linear chains of high molar mass polymers tend to interfere with each other in solution, hindering adsorption of the polymer, and rendering the adsorption stage more difficult.

For the range of concentrations tested, all close to the optimum concentration leading to the highest flocculation efficiency for each polymer (larger flocs and fastest kinetics), concentration did not play a very important role on the model, though it could not be neglected.

References

[1] D. Eklund, T. Lindström, Paper Chemistry: An Introduction, DT Paper Science Publ., Finland, 1991.
[2] K.E. Bremmell, G.J. Jameson, S. Biggs, Kinetics polyelectrolyte adsorption at the solid/liquid interface interaction forces and stability, Colloids and Surfaces 139 (1998) 199–211.
[3] S. Biggs, M. Habgood, G.J. Jameson, Yao-de-Yan, Aggregate structures formed via a bridging flocculation mechanism, Chemical Engineering Journal 80 (2000) 13–22.
[4] A. Blanco, E. Fuente, C. Negro, J. Tijero, Flocculation monitoring: focused beam reflectance measurement as a measurement tool, Canadian Journal of Chemical Engineering 80 (2002) 734–740.
[5] E. Antunes, F.A.P. Garcia, P. Ferreira, A. Blanco, C. Negro, M.G. Rasteiro, Effect of water cationic content on flocculation, flocs resistance and reflocculation capacity of PCC induced by polyelectrolytes, Industrial and Engineering Chemistry Research 47 (2008) 6006–6013.
[6] T.G.M. van de Ven, B. Alince, Heteroflocculation by asymmetric polymer bridging, Journal of Colloid and Interface Science 181 (1996) 73–78.
[7] M.G. Rasteiro, F.A.P. Garcia, P. Ferreira, A. Blanco, C. Negro, E. Antunes, The use of LDS as a tool to evaluate flocculation mechanisms, Chemical Engineering and Processing 47 (2008) 1329–1338.
[8] J. Hernandez Barajas, C. Wandrey, D. J. Hunkeler, 2003. U.S. Pat. 6,667,374 B1.
[9] J. Hernandez Barajas, C. Wandrey, D. J. Hunkeler, 2003. U.S. Pat. 6,617,402 B2.
[10] L. Bourdillon, D. Hunkeler, C. Wandrey, The analytical ultracentrifuge for the characterization of polydisperse polyelectrolytes, Progress in Colloid & Polymer Science 131 (2006) 141–149.
[11] M.G. Rasteiro, F.A.P. Garcia, M. del Mar Peréz, Applying LDS to monitor flocculation in papermaking, Particulate Science and Technology 25 (3) (2007) 303–308.
[12] M.G. Rasteiro, I. Pinheiro, F.A.P. Garcia, P. Ferreira, D. Hunkeler, Using light scattering to screen PEL performance in flocculation, Polymers 3 (2010) 915–927.
[13] H. Martens, T. Naes, Multivariate Calibration, Wiley, Chichester, 1989.
[14] S. Wold, M. Sjostrom, L.T. Eriksson, PLS-regression: a basic tool of chemometrics, Chemometrics and Intelligent Laboratory Systems 58 (2001) 109–130.
[15] H.G. Alemea, L.M. Costa, P.J.S. Barbeira, Determination of ethanol and specific gravity in gasoline by distillation curves and multivariate analysis, Talanta 78 (2009) 1422–1428.
[16] S. Wold, M. Sjostrom, L. Eriksson, Partial least squares projections to latent structures (PLS) in chemistry, in: P. von Ragué Schleyer (Ed.), The Encyclopedia of Computational Chemistry, John Wiley and Sons, N.Y., 1998, pp. 2006–2021.
[17] A.C. Pereira, M.S. Reis, P.M. Saraiva, J.C. Marques, Madeira wine ageing prediction based on different analytical techniques: UV–vis, GC–MS, HPLC–DAD, Chemometrics and Intelligent Laboratory Systems 105 (2011) 43–55.