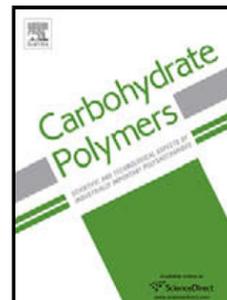


Journal Pre-proof

In-depth characterization of the aggregation state of cellulose nanocrystals through analysis of transmission electron microscopy images

Cristina Campano (Conceptualization) (Methodology) (Validation) (Formal analysis) (Investigation) (Data curation) (Writing - original draft) (Visualization), Patricio Lopez-Exposito (Conceptualization) (Methodology) (Software) (Validation) (Formal analysis) (Investigation) (Data curation) (Writing - review and editing) (Visualization), Laura Gonzalez-Aguilera (Methodology) (Validation) (Investigation) (Data curation) (Visualization), Ángeles Blanco (Resources) (Writing - review and editing) (Supervision) (Project administration) (Funding acquisition), Carlos Negro (Resources) (Writing - review and editing) (Supervision) (Project administration) (Funding acquisition)



PII: S0144-8617(20)31444-2
DOI: <https://doi.org/10.1016/j.carbpol.2020.117271>
Reference: CARP 117271
To appear in: *Carbohydrate Polymers*
Received Date: 3 August 2020
Revised Date: 14 October 2020
Accepted Date: 15 October 2020

Please cite this article as: { doi: <https://doi.org/>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier.

In-depth characterization of the aggregation state of cellulose nanocrystals through analysis of transmission electron microscopy images

Cristina Campano¹, Patricio Lopez-Exposito², Laura Gonzalez-Aguilera¹, Ángeles Blanco¹ and Carlos Negro^{1,}*

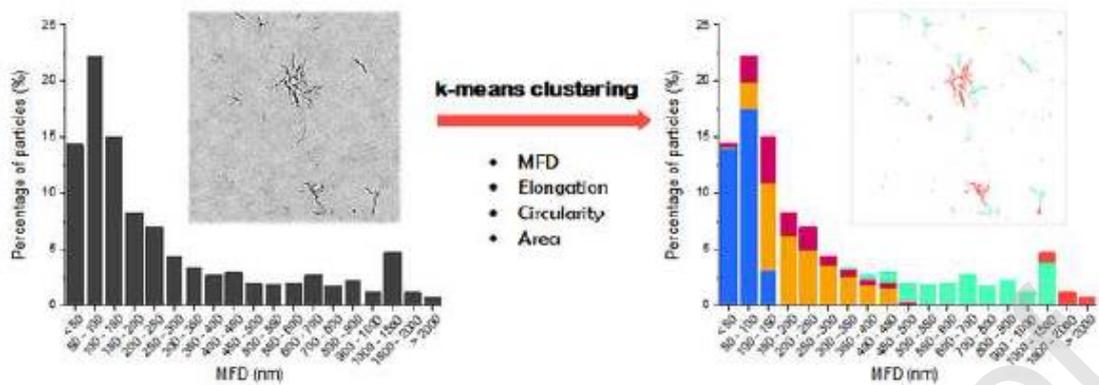
¹Department of Chemical Engineering and Materials, Complutense University of Madrid. Avda. Complutense s/n 28040 Madrid, Spain.

²Departamento de Bioingeniería e Ingeniería Aeroespacial, Universidad Carlos III de Madrid, Leganés, Spain.

e-mail addresses: ccampano@ucm.es (C. Campano), patricioroberto.lopez@uc3m.es (P. Lopez-Exposito), lgonza29@ucm.es (L. Gonzalez-Aguilera), ablanco@ucm.es (A. Blanco), cnegro@ucm.es (C. Negro)

* Corresponding author: Carlos Negro, cnegro@ucm.es. Department of Chemical Engineering and Materials, Complutense University of Madrid. Avda. Complutense s/n 28040 Madrid, Spain. Tel: +34 913944242

Graphical abstract



Highlights

- New characterization of CNC aggregation by TEM and morphological classification
- Models trained with large number of CNC aggregates of diverse morphology and source
- It can detect application-critical high complexity, low compactness CNC aggregates
- Models are persistent and reusable for characterization of other CNC aggregates
- Approach and models usable as standard for CNC characterization with other methods

ABSTRACT

Dispersion of cellulose nanocrystals (CNCs) is of utmost importance to guarantee their reliable application. Nevertheless, there is still no consensual method to characterize CNC aggregation. The hypothesis of this paper is that dispersion could be quantified through the classification of aggregates detected in transmission electron microscopy images. *k*-Means was used to classify image particulate elements of five CNC samples into groups according to their geometric features. Particles were classified into five groups according to their maximum Feret diameter, elongation, circularity and area. Two groups encompassed the most application-critical aggregates: one integrated aggregates of high complexity and low compactness while the other included elongated aggregates. In addition, the characterization of CNC dispersion after different levels of sonication was achieved by assessing the change in the number of elements belonging to each cluster after sonication. This approach could be used as a standard for the characterization of the aggregation state of CNCs.

Keywords: cellulose nanocrystals; nanocellulose; transmission electron microscopy; aggregation state; k-Means clustering; dispersion

1. INTRODUCTION

Cellulose nanocrystals (CNCs) present exceptional properties that make them suitable for many application fields, such as biomedical, wastewater treatment, energy and electronics (Grishkewich, Mohammed, Tang & Tam, 2017). An adequate dispersion of CNCs particles in suspension is crucial to ensure their efficiency (Shojaeiarani, Bajwa & Stark, 2018).

Sulfate ester groups present in sulfuric acid hydrolyzed CNCs confer them a certain stability in water suspension, which commonly declines over time. Such decline is more pronounced at acid pH, resulting in suspensions of particles with very different morphologies encompassing from 20 nm wide compact fibrillar assemblies of single crystallites (2-5 nm each) to clusters formed of the later that measure up to several microns and have a high complexity (Qi, Yu, Zhang & Xu, 2019). Thus, right after the hydrolysis reaction, CNC suspensions are composed of different proportions of both compact assemblies and clusters with different configurations and some of them are unaltered even at neutral pH, if no dispersion method is applied. Although there is a clear acknowledgement of the importance of CNC dispersion (Di Giorgio, Martín, Salgado & Mauri, 2020), to date no consensual method to characterize the aggregation state of CNC suspensions has been agreed. With the term aggregation we are referring to the agglomeration of all types of particles mentioned above, being the state of aggregation the distribution of particles of different morphology in the sample.

Techniques like light diffraction spectroscopy (Rasteiro et al., 2008) or laser reflectance (Campano, Lopez-Exposito, Blanco, Negro & van de Ven, 2019; Lopez-Exposito, Campano, van de Ven, Negro & Blanco, 2019) are useful to monitor CNC flocculation. Nevertheless, their detection limit of is not enough to resolve small CNC aggregates, which have all dimensions in the nano-scale. So far, only some studies have investigated the influence of charge density and ionic strength on the aggregation of CNCs, either by small-angle neutron scattering (SANS) (Cherhal, Cousin & Capron, 2015) or by means of turbidity and

small angle X-ray scattering (SAXS) (Phan-Xuan et al., 2016). Although these techniques are very promising in the detection of slight changes in the CNC aggregation state, the unravelling of their output to provide morphological information of heterogeneous suspensions is rather complex. The lack of a reliable aggregation characterization method implies a large uncertainty in the prediction of CNCs behavior in their application.

In the context of aggregates characterization, transmission electron microscopy (TEM) offers several advantages: it allows the visualization of individual and aggregated particles, its resolution is suited to image nanocelluloses, and it permits the rapid screening of a large population of particles with tunable sampling (Brinkmann et al., 2016; Stinson-Bagby, Roberts & Foster, 2018). However, Chen et al. (2020) indicated that the aggregation detected by microscopy probably reflects a combination of pre-existing aggregates in the initial suspension and clusters formed during the sample deposition process. Despite these limitations associated with the characterization of the bulk morphology of CNCs through TEM, a new reproducible method to describe the CNCs 3D morphology has been recently proposed by this research group (Campano, Balea, Blanco & Negro, 2020). The method involves the previous deposition of a thin layer of Poly-L-Lysine (PLL) on the TEM grids to prevent particles from aggregating during sampling or drying. Although this approach may entail some flattening effects on the particles during drying, especially on large aggregates, we deem it more practicable and accessible than other methods capable of capturing the 3D morphology of particles, such as cryogenic TEM.

Clustering analysis aims to classify a set of elements into groups, considering some parameters measured on the sample so that elements in the same cluster are more similar to each other than to those in other groups (Wu & Chow, 2004). This approach is commonly used in exploratory pattern-analysis, decision-making and machine-learning problems, including data mining, document retrieval, image segmentation, and pattern classification (Jain, Murty & Flynn, 1999).

The hypothesis of this study is that we can describe the aggregation state of CNC suspensions and their evolution through the clustering of the different types of CNC aggregates obtained from TEM images based on their morphological features. To ensure the correct interpretation of the CNC bulk morphology through TEM, we followed the method mentioned above (Campano et al., 2020). Images were binarized and the size and shape of particles was determined. Afterwards, particles were individually skeletonized to obtain the number of nodes and branches in each skeleton. Different combinations of these parameters were used to group akin particles, using an automatic k-Means classification. Finally, the proposed approach was validated by studying the evolution of five types of CNCs with the application of increasing sonication times.

2. MATERIALS AND METHODS

2.1 Materials

Five raw materials were used as CNC source: two bleached kraft pulps of eucalyptus and pine, and three commercial celluloses, namely cotton linters, α cellulose and avicel, supplied by Sigma Aldrich. Chemicals used for CNC production and characterization, H_2SO_4 , Copper(II) ethylenediamine, $K_2Cr_2O_7$ and Ag_2SO_4 , were of analytical reagent grade and also supplied by Sigma Aldrich. PLL used for TEM sample preparation was a 0.1 wt% solution and obtained from Electron Microscopy Sciences.

2.2 Production and characterization of cellulose nanocrystals

Dry eucalyptus and pine pulps were milled with a CT 293 Cyclotec supplied by FOSS A/S (Hillerød, Denmark), filtering through a sieve of 1 mm. Cotton linters, α cellulose and avicel were used in powder form as received. Acid hydrolysis of the five celluloses was performed according to the procedure described by Campano et al. (2020), in which pulp was made to react with 64% H_2SO_4 for 45 min at 45°C at an acid to pulp ratio of 13.5 mL/g.

2.2.1 Hydrolysis yield and dissolved amorphous cellulose

Crystalline yield represents the percentage of dry mass obtained after dialysis related to the initial cellulose content. Amorphous content is the proportion of dissolved cellulose (DAC) during the reaction respect to the initial cellulose content as described by Campano, Miranda, Merayo, Negro & Blanco (2017). The resting proportion to 100% was considered as hydrolysis losses.

2.2.2 Polymerization degree

The CNC suspension was dissolved in a cupri-ethylene-diamine (CED) solution and the intrinsic viscosity (η) of the solution was determined following the standard ISO 5351. Then, polymerization degree (PD) was related to η with the equations 1 and 2, as described by Henriksson, Berglund, Isaksson, Lindstrom & Nishino (2008).

$$\eta = 0.42 \cdot PD \quad PD < 950 \quad (1)$$

$$\eta = 2.28 \cdot PD^{0.76} \quad PD > 950 \quad (2)$$

2.2.3 Crystallinity index

Crystallinity index (Cr.I) was determined from X-ray diffraction (XRD) data, using Segal's method (Segal, Creely, Martin & Conrad, 1959) through the eq. (3).

$$Cr.I(\%) = \frac{I_{002} - I_{am}}{I_{002}} \cdot 100 \quad (3)$$

Where I_{200} is the intensity of the 200 plane at $2\theta = 22.5^\circ$ and I_{am} is the intensity of the amorphous scatter at $2\theta = 18^\circ$. Spectra were obtained with a Philips X'Pert MPD X-Ray diffractometer with an autodivergent slit fitted with a graphite monochromator using Cu- $K\alpha$ radiation operated at 45 kV and 40 mA. The XRD patterns were recorded from 4 to 40° at a scanning speed of $0.6^\circ/\text{min}$.

2.2.4 Zeta Potential

CNC zeta potential was measured at 0.0005%, after pH adjustment to 6, using a NanoBrook 90PlusZeta (Brookhaven Instruments Corporation, NY, USA).

2.2.5 Sulfate ester groups

Total sulfur content of dry CNC samples was obtained from elemental analysis using a LECO CHNS-932 (Michigan, USA). The analysis were carried out by the unit of elemental microanalysis (accredited by ENAC the Spanish Entity of Accreditation).

2.2.6 Hydrodynamic diameter

The hydrodynamic diameter of CNCs was assessed by dynamic light scattering (DLS) using a NanoBrook 90PlusZeta (Brookhaven Instruments Corporation, NY, USA). Scattered light intensity of suspensions at 0.0005% was recorded at 90° and 20 °C.

2.3 Imaging

2.3.1 Image acquisition

Sample micrographs were acquired through a JEOL JEM 1400 plus TEM, operated at 100 kV accelerating voltage, at the Spanish National Centre of Electronic Microscopy (CNME). The CCD camera was an Orius SC200, manufactured by Gatan (Pleasanton, USA), at 2048 x 2048 pixels resolution and pixel size 7.4 microns. 200-mesh copper grids acquired from Electron Microscopy Sciences covered with a continuous layer of 10 nm Formvar and stabilized with 1 nm evaporated carbon film were used in this study. Images were taken through a one-shot acquisition. A minimum of 20 images were taken for each sample, considering different areas of the grid far from each other to have a representation of the whole grid surface.

The procedure used to fix the CNC samples on the grids involved the application of 20 μ L of a 10% solution of PLL on the grids placed on filter paper, leaving them to dry at 60°C.

Then, 10 μL of 0.0005% CNC samples were deposited on the PLL covered grids, also on filter paper, and left to air-dry statically to avoid relocation of particles or their aggregation. The above concentration was selected based on preliminary tests to minimize the effect of dilution on CNCs bulk morphology and to allow an individualization of the particles in the images (Campano et al., 2020).

To achieve different aggregation states CNC suspension samples were submitted to 1.5, 3 and 4.5 min of high-intensity sonication with a sound power of 50 W, using a UP200St digital ultrasonic processor, Heilscher GmbH (Teltow, Germany). Heat released during the sonication treatment was dissipated submerging the samples in an iced bath.

2.3.2 Image analysis

TEM micrographs from the samples were binarized using Fiji, an image processing package, following the procedure described in a previous study (Campano et al., 2020).

Morphological features were measured in all particles not intersecting with the border of the images and having a circularity below 0.8. Particles having greater circularities were identified as impurities. Over 220,000 particles were analyzed. The area, perimeter, maximum Feret diameter (MFD), circularity and length and width of the minimum bounding rectangle (MBR) were directly measured from the particles. Subsequently, particles were skeletonized using the *Skeletonize (2D/3D)* plugin of the Fiji software. The skeletons obtained were analyzed in terms of the number of nodes, i.e. points of union, number of branches and total length of branches, i.e. the sum of the length of all branches. Given the large volume of data, this process was automatized by means of a Python script (attached file).

Other geometrical parameters were obtained from the combination of directly measured ones. Elongation was calculated as the ratio between length and width of the MBR.

Rectangularity was calculated as the ratio of the area of a particle to the area of its MBR. It indicates the normalized discrepancies between the area of the bounding rectangle of the

particle and that of the particle itself (Sinha & Patel, 2014). Eccentricity was estimated as the ratio of particle's width to length subtracted from 1. A shape that is symmetrical across all its axes, e.g. a circle or a square, will have an eccentricity value of 0. Contrarily, as the aspect ratio increases the value tends towards 1 (Mingqiang, Kidiyo & Joseph, 2008).

2.4 Unsupervised machine learning clustering

Automatic clustering was carried out applying the k-Means method, which classifies the vectors of morphology parameters in k different groups. The optimal number of clusters, k , in each case was determined applying the elbow method (Kodinariya & Makwana, 2013). This method assesses the evolution of the model inertia, i.e. the sum of the square distances of each cluster element to their assigned center, as a function of the number of clusters and chooses the number k , i.e. the optimal number of clusters, so that the addition of another cluster does not significantly improve the modeling of the data. This point corresponds to an elbow in the graph of inertias versus number of clusters. For each combination of parameters considered, data were fitted into various k-Means models with values of k from 1 to 10.

k-Means clustering was employed to automatically group particles according to different size and shape features. The k-Means algorithm aims at grouping a set of n data points (particles) belonging to a real d -dimensional space, R^d , around a set of k points (clusters) in R^d , called centroids, so as to minimize the mean squared distance from each data point to its nearest centroid (Kanungo et al., 2002). k-Means clustering analysis was performed using the Scikit-learn Python module.

To aid the visualization of clusters, inter-cluster distance maps were created in which the cluster centroids were embedded in two dimensions by applying a principal components analysis. The distance between the centroids indicates the independence of the clusters.

The analysis was carried out with different input configurations:

- Option A1: MFD, area and perimeter

- Option A2: MFD, area, perimeter and nodes
- Option A3: MFD, area, perimeter and branches
- Option A4: MFD, area, perimeter, nodes and branches
- Option B1: MFD, elongation and circularity
- Option B2: MFD, elongation, circularity and area
- Option B3: MFD, elongation, circularity and nodes
- Option B4: MFD, elongation, circularity, area and nodes

2.5 Characterization of clusters

The average 2D fractal dimension (D_f) of each cluster was estimated through the multi-scale MBR analysis, following the procedure described previously (Wozniak, Onofri, Barbosa, Yon & Mroczka, 2012), in which the following equation (4) is obtained:

$$\ln(A_a) = \frac{D_f}{\alpha} \ln(\sqrt{LW}) + \ln\left(\frac{k_{LW} \cdot A_p^\alpha}{k_\alpha \cdot MFD^{D_f/\alpha}}\right) \quad (4)$$

Where A_a is the projected surface area of an aggregate, A_p the one of the primary particles and k_{LW} , k_α and α are parameters evaluated experimentally. The slope of the representation of $\ln(A_a)$ versus $\ln(\sqrt{LW})$ multiplied by the estimated value of α , 1.1 (Köylü, Faeth, Farias & Carvalho, 1995), would provide the average value of D_f for the particular cluster. The R^2 value of the regression would indicate the degree of similarity among the particles of the group.

For an easier visualization of cluster membership, particles were colored in the images using Fiji. A new black background image was created in which particles were filled with the color corresponding to the cluster they belonged to.

3. RESULTS AND DISCUSSION

3.1 CNC characterization

The results of the characterization of CNCs from the five cellulose sources considered are shown in Table 1. Both eucalyptus and pine yielded a low percentage of CNCs, with values below 10%. The low yield of eucalyptus, around 4,6%, is attributable to the low crystallinity of the raw material, associated to the rapid growth rate of this wood (Xiang et al. 2016). Thus, high proportions of eucalyptus samples were converted into sugars and dissolved in the hydrolysis process (89%). Although the crystallinity of pine was greater than that of eucalyptus, its conversion yield was not proportionally higher, in this case due to a greater material loss during the procedure. The fact that cotton presented a higher crystallinity (90%) resulted in a higher yield (48.5%) and in a relatively low DAC (33.8%). The initial crystallinity of avicel was higher than that of wood, 80%, but lower than in cotton. The yield obtained with avicel was not as high as expected, probably due to the small size of the initial powder, around 50 μm , that caused a greater material loss. The production of CNCs from commercial α -cellulose was also considered on account of its different morphology compared to the other α -cellulose-rich sources, namely cotton and avicel. Commercial α -cellulose presented a Cr.I (61%) similar to that of eucalyptus, as observed before (Carrillo, Mendonça, Ago & Rojas, 2018), which led to a low yield (9.3%) and a high DAC (75.6%).

Table 1. Characterization of cellulose nanocrystals (CNCs) produced from different sources.

	Eucalyptus	Pine	Cotton	Avicel	α -cellulose
Yield (%)	4.6 \pm 0.1	5.8 \pm 0.8	48.5 \pm 0.5	14.8 \pm 0.2	17.5 \pm 0.1
DAC (%)	89.0 \pm 3.7	79.8 \pm 2.8	33.8 \pm 0.5	67.6 \pm 4.1	70.2 \pm 1.3
Zeta potential (mV)	-22.0 \pm 1.6	-22.8 \pm 1.3	-23.9 \pm 0.4	-24.5 \pm 3.5	-23.8 \pm 0.8
Sulfate ester groups (mmol/g)	0.378 \pm 0.011	0.331 \pm 0.015	0.168 \pm 0.009	0.256 \pm 0.012	0.193 \pm 0.008
Length (nm)*	100 \pm 50	105 \pm 50	160 \pm 80	110 \pm 55	125 \pm 60
Width (nm)*	21 \pm 5	21 \pm 7	22 \pm 8	21 \pm 5	21 \pm 11
PD	212 \pm 2	231 \pm 15	241 \pm 21	178 \pm 7	100 \pm 3
Hydrodynamic diameter (nm)	305 \pm 45	430 \pm 60	425 \pm 85	395 \pm 40	360 \pm 27

*DAC: Dissolved amorphous cellulose; RM: Raw material; PD: polymerization degree

As observed in Table 1 and Figure 1, the average length of CNCs was similar in all CNC types, except in the case of cotton, which presented a wider length distribution having some longer crystals than in the other samples. Respect to the width of CNCs, all samples showed a very similar value around 21 nm. Taking into account the preferential reaction of chemicals for amorphous cellulose, one could explain that most of sulfate ester groups remain attached to the edges of the particles. Thus, there should be an indirect relationship between the average length of the CNCs and the proportion of sulfate ester groups, considering a narrow length distribution. Although the presence of sulfate ester groups could contribute to stabilize CNC suspensions, the occurrence of initial aggregations associated to high hydrolysis yields will likely compensate such stabilizing effect. The difference in the values of hydrodynamic diameters shown in Table 1 may be due to the effect of some of their properties influencing the Brownian motion of particles, such as length, stability and the presence of some aggregates.

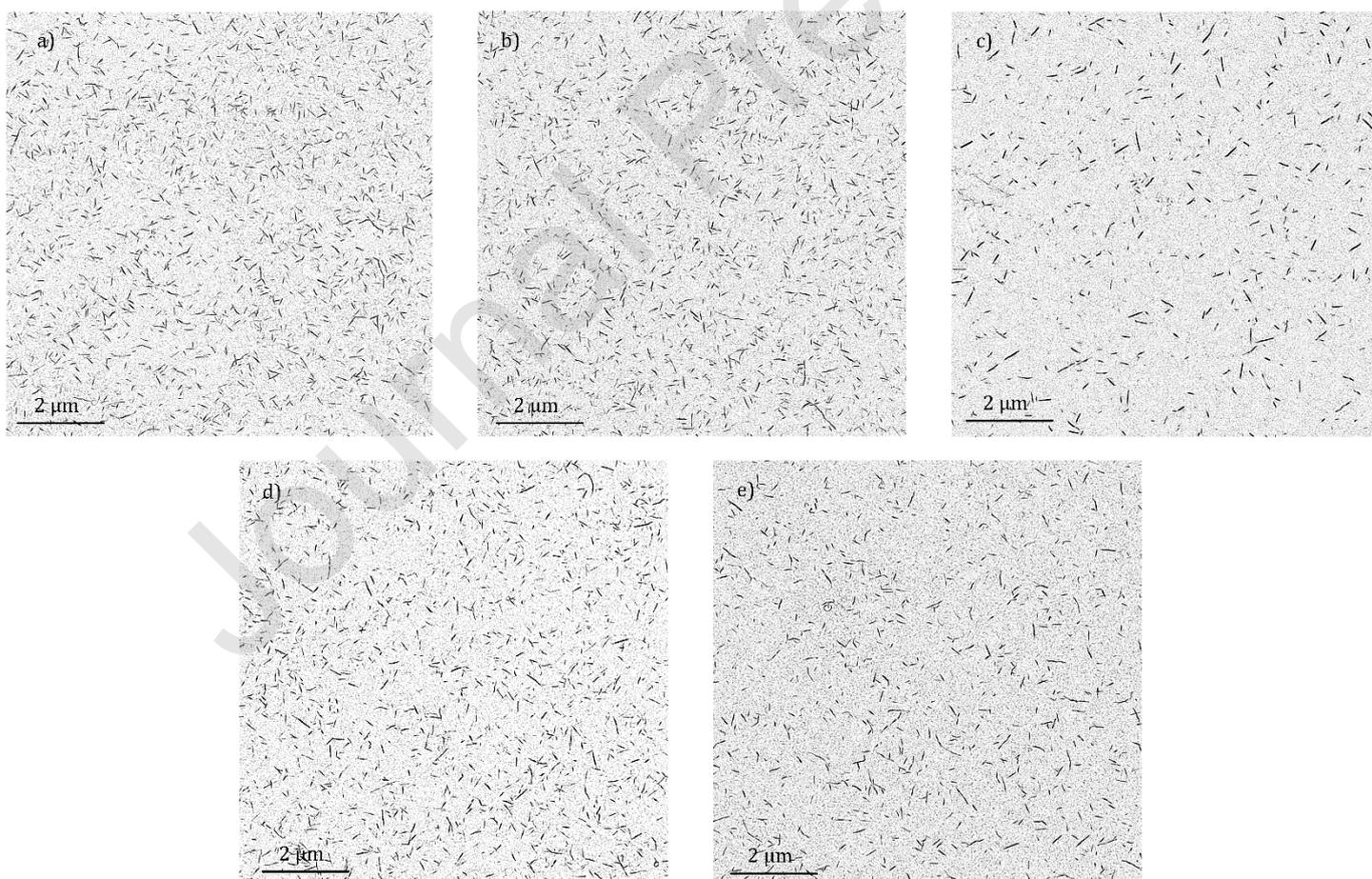


Figure 1. Transmission electron microscopy (TEM) images of fully dispersed cellulose nanocrystals (CNC) samples at 1000x magnification: a) eucalyptus, b) pine, c) cotton, d) avicel and e) α -cellulose.

The most relevant factors influencing the aggregation of CNCs are considered to be size, surface charge, steric impediments and the presence of amorphous regions, as well as medium ionic strength and pH (Cherhal et al., 2015). The last two factors were kept constant in the experiments. We therefore considered the following CNC representative samples: CNCs produced from cotton as the one with longest individuals, CNCs produced from α -cellulose as the most stable suspension (expressed by its zeta potential) and the CNCs obtained from eucalyptus and pine as the ones with the highest amount of sulfate ester groups.

3.2 Image elements clustering

In each clustering approach the corresponding features of particles detected through image analysis (over 220,000) were used to train the particular k-Means model and select the optimal number of classification groups for the set of parameters considered. In all studies, the trained k-Means clustering model was saved so that it can be employed to classify new aggregates outside the scope of the present research. The trained models receive an input in the form to a vector containing the geometric parameters of a given particle (or list of particles) and returns the group to which the particle belongs, as follows:

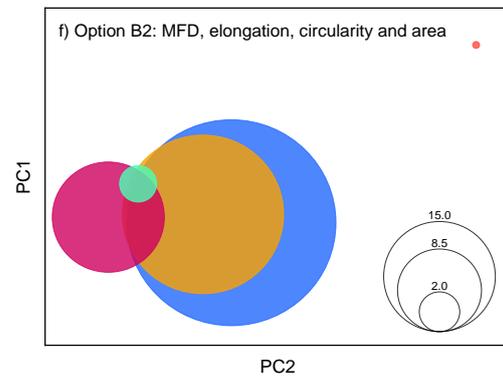
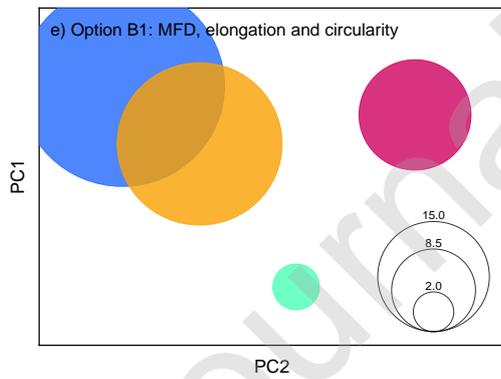
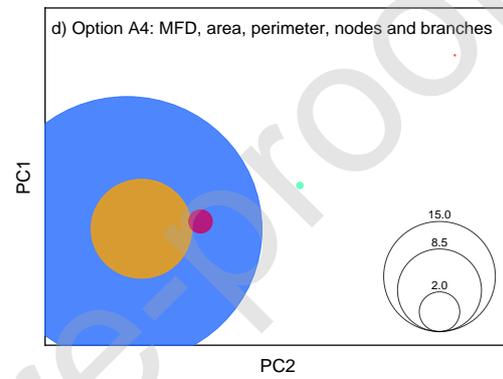
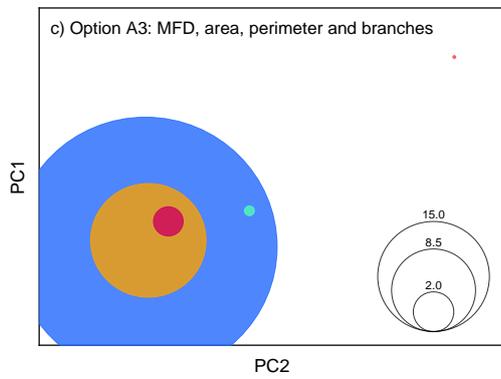
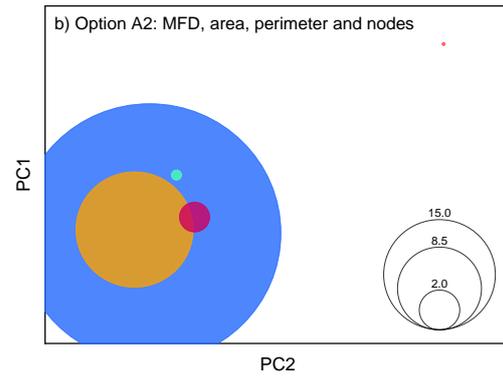
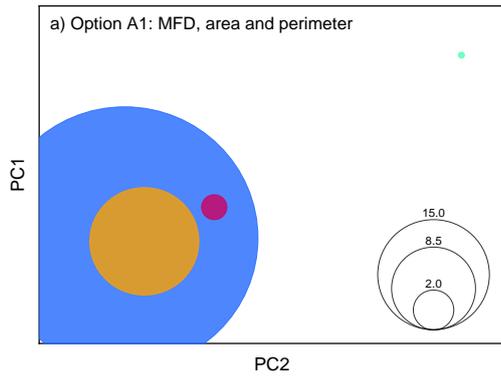
$$f(\text{parameter}_1, \dots, \text{parameter}_n)_{\text{particle}} = \text{group}_i \quad (5)$$

When applying the k-Means clustering method, the selection of the most meaningful variables is crucial to achieve reliable results of both the number of clusters and their membership (Pal, Bezdek & Tsao, 1993). Figure S1 depicts the evolution of inertia as a function of the number of clusters considered. The optimum number of clusters was found to be 4 when considering three parameters, namely options A1 and B1, and it was 5 in the rest of the options.

The inter-clusters distance map of each approach is shown in Figure 2. The clusters are sized by membership. The general goal involves a high inter-clusters distance and a significantly low intra-cluster density (Wu et al., 2004).

Options A1-A4 have three common input factors, namely MFD, perimeter and area, selected for being basic size-related features that convey composite information on the relation between size and shape in CNCs. In these four options, the distribution of members in the groups was characterized for showing a high variability in the number of elements within each cluster, as observed by their areas. One of the clusters (blue) contained over 80% of the particles, while the membership of the second one in size (yellow) was around 15% (Figure 2a-d). The centroids of these two clusters were visibly closer between them compared to the rest of the clusters formed.

The separation between centroids, i.e. the dissimilarity of clusters seems to be enhanced with the inclusion of both nodes and branches (Option A4), as observed in Figure 2d, conveying differentiating information when it comes to group CNC elements.



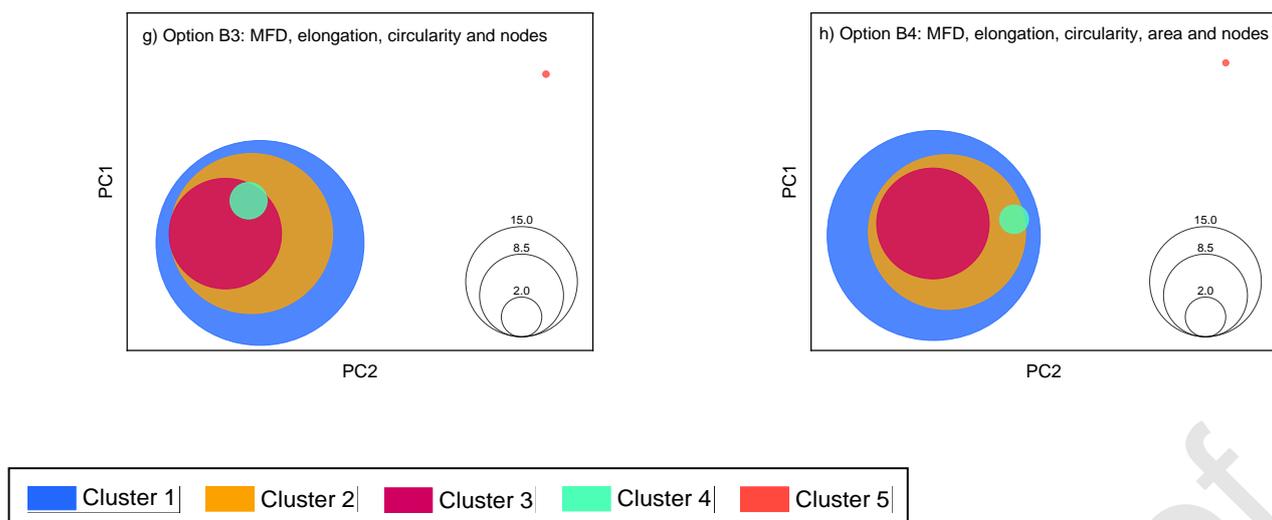


Figure 2. Inter-clusters distance map of the result for the clustering of the cellulose nanocrystals (CNCs) with the input vectors described for the different options. Colors indicate the different clusters (1-5) attending to their membership by decreasing order: blue, yellow, magenta, aquamarine and red coral. Legend indicates the membership scale as area in percentage. Note: overlapping of two clusters in the 2D space does not imply that they overlap in the original feature space.

Input variables of options B1-B4 include not only size-related parameters but also geometrical features. First, a characteristic size: MFD; second, a parameter related to the external geometry: elongation; and third, circularity, a parameter related to the distribution of area over perimeter, which can be considered as an indirect measurement of compactness. Results in options B1-B4 were quite different between them and with respect to those of Options A1-A4.

When considering only MFD, elongation and circularity (Option B1, Figure 2e), a high inter-cluster distance was achieved with a homogenous inter-cluster density, desirable features when defining clusters since they imply a highly meaningful classification. Thus, the clustering of CNCs considering the parameters of Option B1 seemed to be a good starting point. To give relevance to the mass of aggregates, the area was also included

(Option B2). In this case, clusters were also well distanced, although the distribution of particles was more heterogeneous than in Option B1.

In the case of Option B3 and B4 (Figure 2g and 2h), the inter-cluster distance between centroids of clusters 1-3 was quite short, indicating a high similarity among particles belonging to these clusters. Although clusters 4 and 5 were very distant from each other and from the rest, the high similarity between clusters 1-3 would convey little information in the description of the CNC aggregates.

3.3 Description of clusters

As it has shown Options A4, B1 and B2 could be good candidates for the clustering of particles in CNC suspensions to describe their aggregation. To gain a deeper picture on the relationship between size and shape of the CNC aggregates observed in the micrographs, a pairwise analysis of relevant parameters was carried out. Figures 3 and 4 show the representation of these parameters for options A4 and B2. Graphs for the rest of the options are included in the supplementary information file (Figures S2-S7). Clusters were numbered and colored according to the total number of elements belonging to them in decreasing order. Thus, although colors are the same in all options, the elements corresponding to each group are not necessarily the same.

Figure 3 shows that 95.4% of the elements analyzed had a MFD below 300 nm. Generally, when MFD increases, an increase in area was observed. The wide spectrum of areas observed in Figure 3a, indicates that there is a high variety in the shape of the aggregates formed. Regarding elongation, 87% of the particles had a value below 3, but for MFD between 100 and 800 nm, elongation varied in a wide proportion, from 1 to 12. The line patterns observed in the graph were due to an aliasing effect resulting from the resolution of the images being close to the minimum width detected. As expected, the circularity of particles (Figure 3c) followed a marked decreasing trend when MFD increased, given that larger MFD was loosely related with aggregate complexity. However, for particles with

MFD below 300 nm, circularity of particles varies in a wide range from 0.1 to 0.8, which suggests that this parameter plays a significant role in the description of the CNC aggregates. Figure 3d shows that although rectangularity of particles presented a similar trend to that observed in circularity, the distribution of points was more scattered in general.

Journal Pre-proof

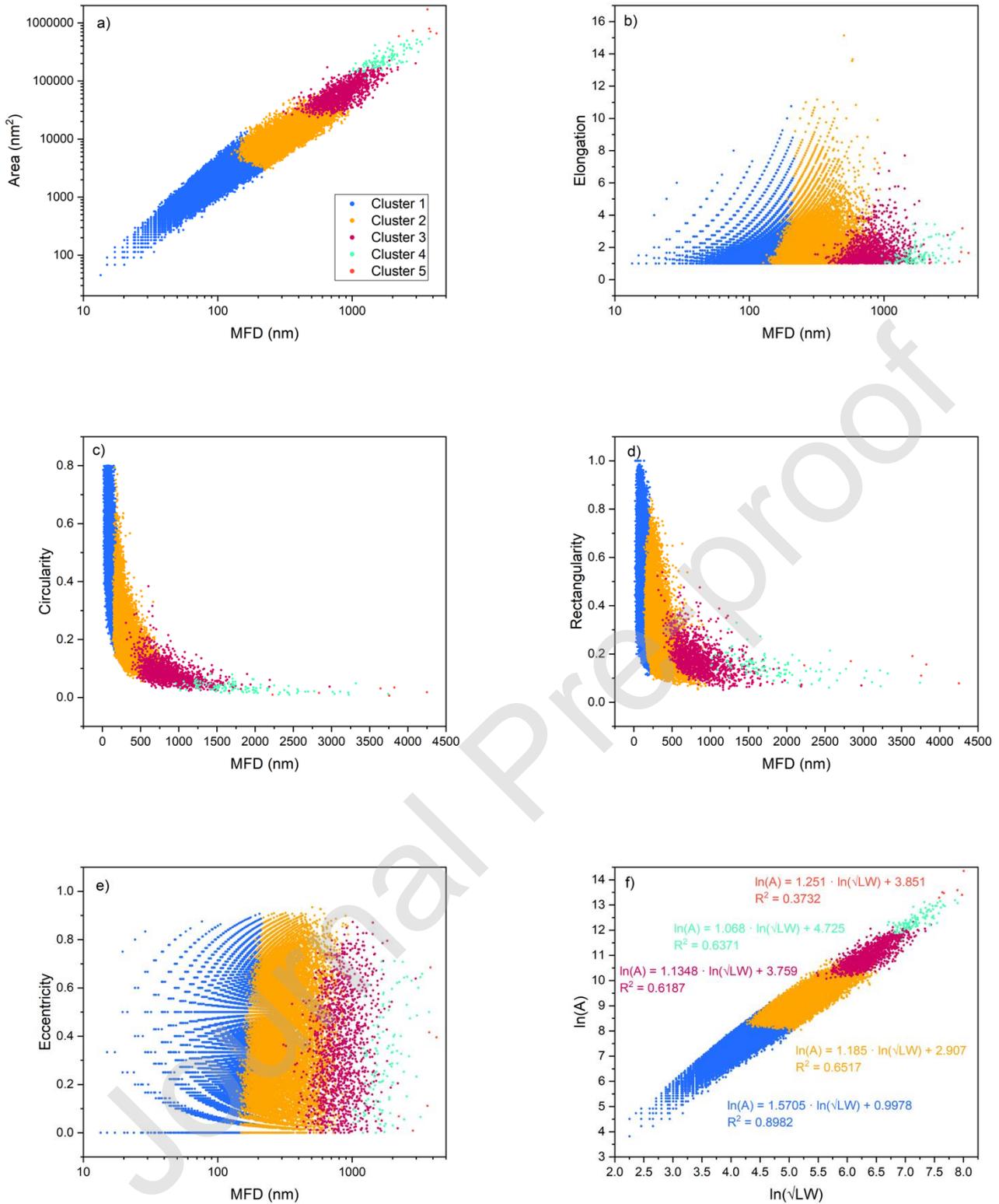


Figure 3. Clustering of particles taking into account features corresponding to Option A4. Evolution of a) area, b) elongation, c) circularity, d) rectangularity and e) eccentricity with maximum feret diameter (MFD), and f) $\ln(A_a)$ versus $\ln(\sqrt{LW})$, where the slope of this curve multiplied by α returns the value of fractal dimension (D_f). Note: legend of a) applies to b), c), d) e) and f).

Eccentricity of particles presented values from 0 to 0.9 with over 57.1% in the range from 0.2 to 0.7. As with elongation, the aliasing effect is visible. Particles present in Figure 3f appear to follow a straight growing tendency, similar to that of Figure 3a. The value of fractal dimension obtained from the slope of the fitting of this curve has a value of 1.68 with R^2 of 0.93. Nevertheless, it is easy to think that some particles or aggregates could have a common fractal dimension in such a way that the aggregates could present multifractality.

The overlapping of clusters observed in Figure 3 expressed the slight improvement in the clustering of particles achieved with the use of the parameters considered in Option A4 compared to that of Options A1-A3, where the split between clusters was almost solely by size (Figures S2-S4). Despite this positive result, particles in most of the clusters seemed to have neither a common external geometry, indicated by elongation and eccentricity, nor a similar distribution of the mass in the space as expressed through circularity, rectangularity and fractal dimension. Therefore, these clustering options do not highlight all geometrical differences that could be needed to assess their individual importance in some CNC applications.

Three of the clusters formed with options B1 and B2 had a MFD below 350 nm, and almost all particles over this size were grouped either in just one cluster (Option B1, Figure S5) or in two clusters (Option B2, Figure 4). In both cases a classification by not just size but also shape was being made, where Cluster 3 grouped particles of an elongation over 3 and an eccentricity over 0.6. Although fractal dimension of this cluster was very similar to that of Cluster 2 (1.70), the fact that they have a quite different external shape makes this classification very promising. Moreover, the good fitting of the different series of particles in Figure 4f and S4f shows that there is a relationship between the particles belonging to each cluster.

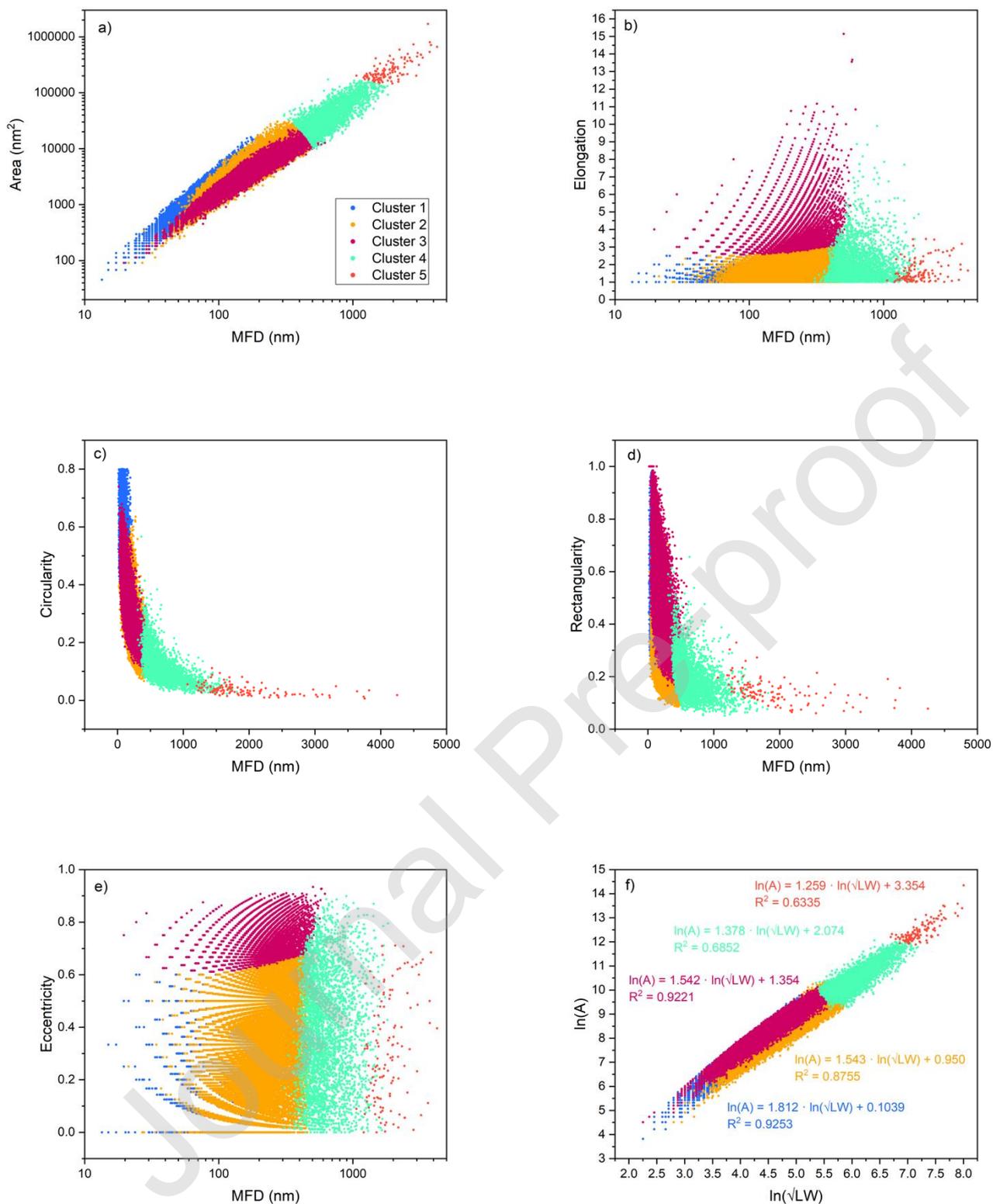
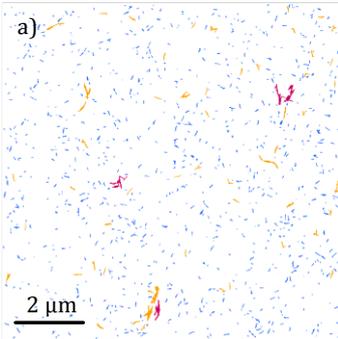


Figure 4. Clustering of particles taking into account features corresponding to Option B2. Evolution of a) area, b) elongation, c) circularity, d) rectangularity and e) eccentricity with maximum feret diameter (MFD), and f) $\ln(A_a)$ versus $\ln(\sqrt{LW})$, where the slope of this curve multiplied by α returns the value of fractal dimension (D_f). Note: legend of a) applies to b), c), d) e) and f).

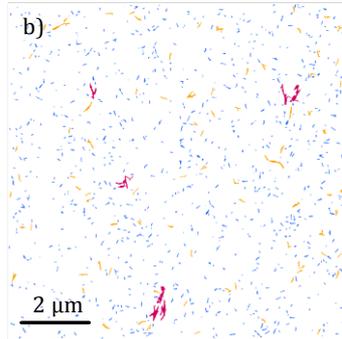
Figure 5 shows two micrographs where particles taking part of each cluster were colored using the same color code than in the graphs, according to the different options considered. Options A1-A4 (Figure 5a-h) corroborated the preliminary conclusion reached before: some aggregates having a similar shape were included in different clusters just by a matter of size. The classifications obtained through options B1-B4 (Figure 5i-p), however, were more similar to those achievable with other techniques that consider the Brownian motion of the particles. In option B1, most of the aggregates were included in Cluster 4 (Figure 5i and 5m). Although this option showed the higher inter-clusters distance, the inclusion of all aggregates in just one cluster may not have sense for the purpose of this study. In this line, the additional consideration of the area of particles in option B2 provided a more descriptive clustering in which large and complex aggregates (Cluster 5) were separated from large and elongated ones (Cluster 4), as observed in Figure 5i-p.

Option A1

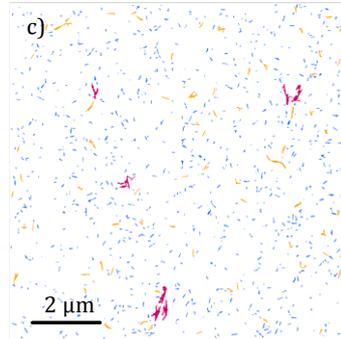
MFD, area and perimeter

**Option A2**

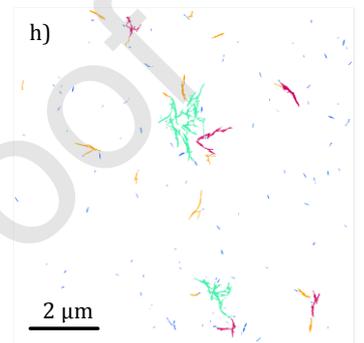
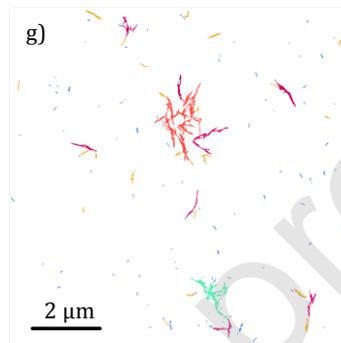
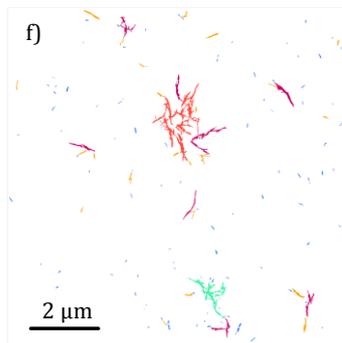
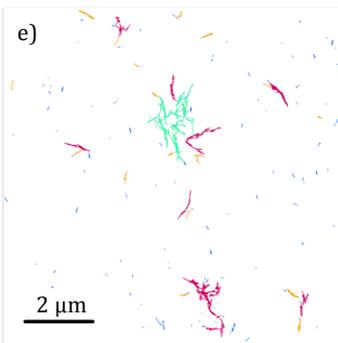
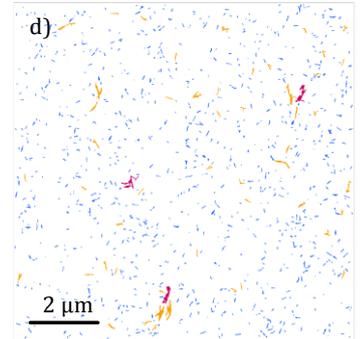
MFD, area, perimeter and nodes

**Option A3**

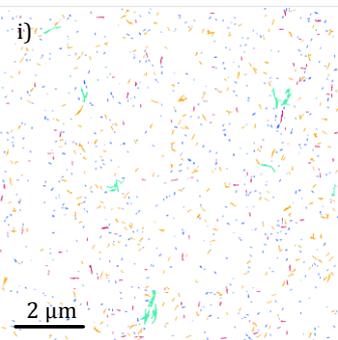
MFD, area, perimeter and branches

**Option A4**

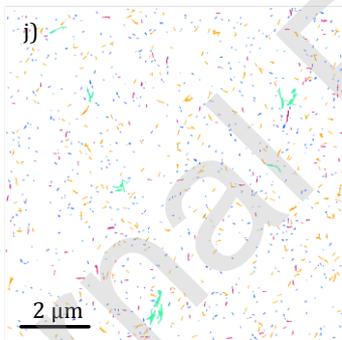
MFD, area, perimeter, nodes and branches

**Option B1**

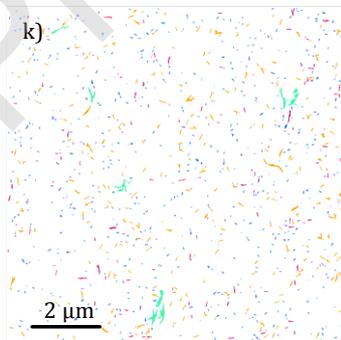
MFD, elongation and circularity

**Option B2**

MFD, elongation, circularity and area

**Option B3**

MFD, elongation, circularity and nodes

**Option B4**

MFD, elongation, circularity, area and nodes

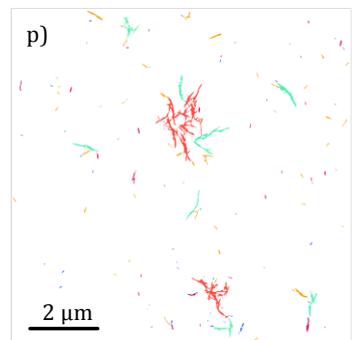
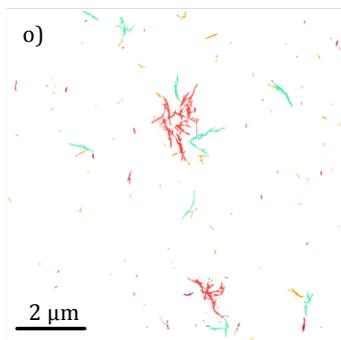
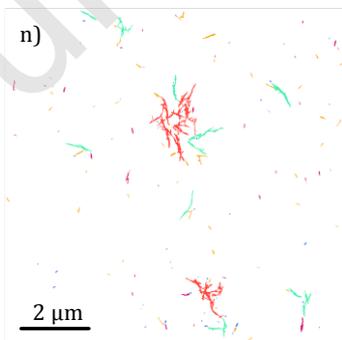
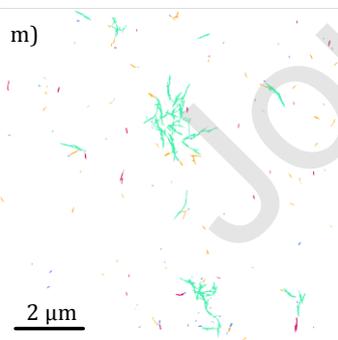
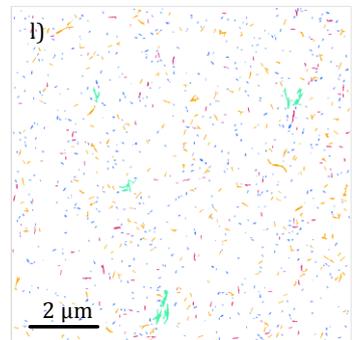


Figure 5. Colored images of an image containing mostly individuals, small fibrillary assemblies and small aggregates (a-d and i-l) and an image containing large aggregates (e-h and m-p) and with the different options considered. Note: colors have been selected based on the number of particles taking part of each cluster: the largest group of each option was colored in dark blue, second in yellow, third in magenta, fourth in aquamarine and the smallest group in red coral.

In the B2 approach, it was possible to identify what could be individual CNCs and small fibrillar assemblies in clusters 1 and 2 as those with 0 nodes and just 1 branch; elongated CNC elements, likely aggregates with high eccentricity in cluster 3; large and elongated aggregates of low circularity, rectangularity and relatively low fractal dimension in cluster 4; and finally, in cluster 5, large, highly complex aggregates with a low fractal dimension. Therefore, this option, which included the MFD, elongation, circularity and area, was selected as the most informative clustering approach to describe aggregation of CNCs.

3.4 Description of the CNC aggregation state

To assess the effectivity of the clustering method chosen to describe the types of CNC aggregates a particle size distribution graph was built for each CNC sample sonicated for different times. Since avicel and pine CNCs represent two cases with different initial aggregation, the size distribution data of these two types of CNC suspensions are shown as an example in Figure 6 and 7. The rest of the samples are included as supplementary material (Figures S8-S10). Most of particles belonging to Cluster 1 (blue) had a MFD below 100 nm. It is highly likely that this group was composed of mostly individual CNCs of the smallest size, conclusion sustained by the increment in the number of particles belonging to this group observed when the sonication treatment time was increased.

It is also noteworthy that Cluster 5 (coral red) only appears in the sample without sonication treatment and that Cluster 4 (aquamarine) particles, initially distributed in a very wide interval of MFD, show a marked decreasing trend in number of elements and

span of distribution when applying sonication. This effect was general for all CNC sources considered although each case presented a particular evolution.

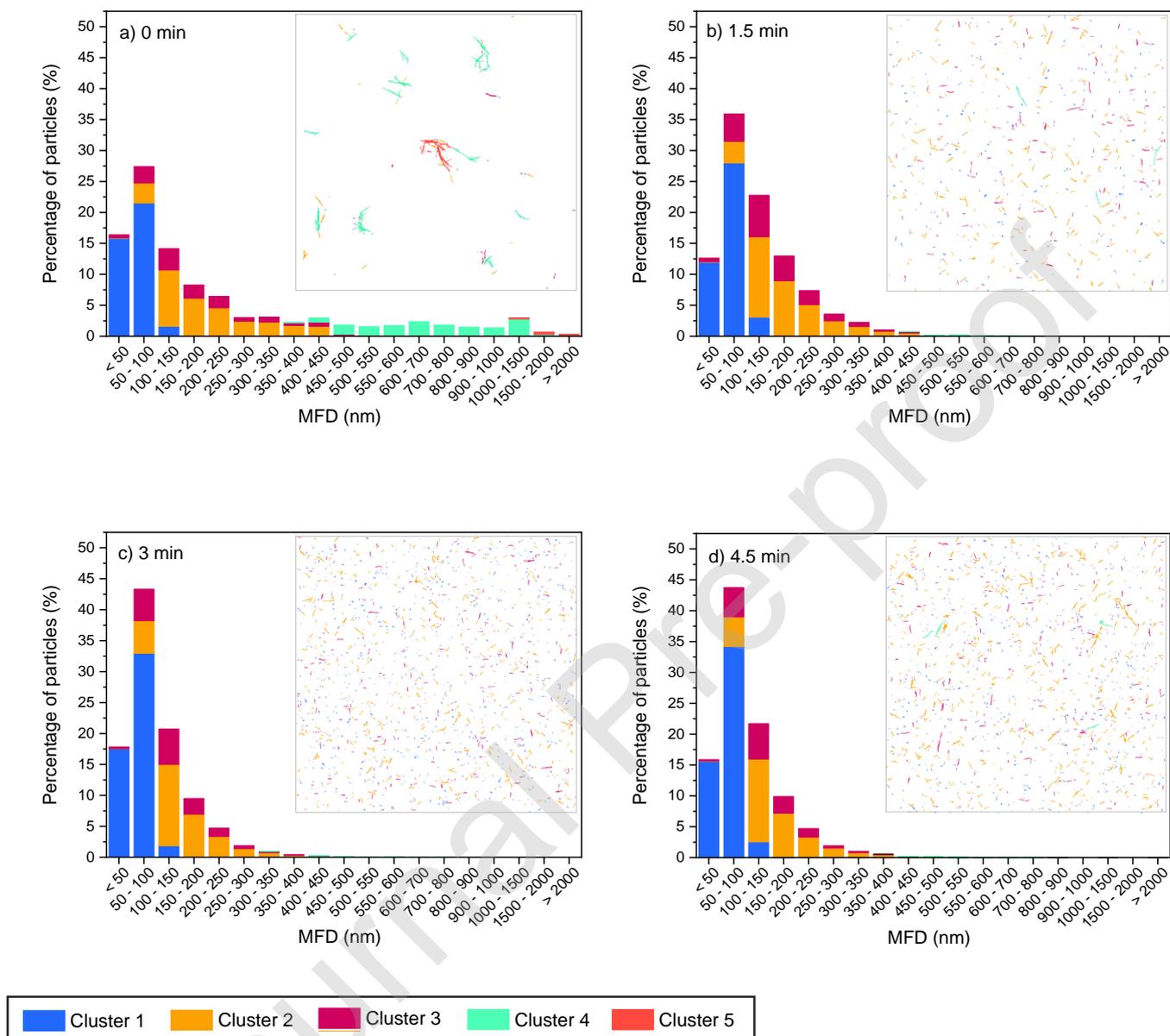


Figure 6. Evolution of the particles size distribution, in terms of maximum feret diameter (MFD), and the visual aspect of the cellulose nanocrystals (CNCs) produced from avicel with the sonication time: a) 0 min, b) 1.5 min, c) 3 min and d) 4.5 min. Colors indicate the different clusters identified: Cluster 1 in blue, Cluster 2 in yellow, Cluster 3 in magenta, Cluster 4 in aquamarine and Cluster 5 in red coral. Note: All particles identified in the whole package of images corresponding to each sample were considered in the graphs.

As for Clusters 2 (yellow) and 3 (magenta), both encompassed particles distributed in a wider MFD interval ranging from 50 to 450 nm in the sample without sonication treatment. The intervals of both clusters were observed to vary in a different manner with the sonication time, but they mostly do in a moderate manner.

In view of the above results, it is possible to state that sonication leads to the disaggregation of particles belonging to Cluster 5 particles due to the disaggregation, which contribute to increase the sum of the memberships of Clusters 1-3 (Figure 6 and 7). The same effect was observed in the aggregates of Cluster 4, but, in this case, a longer sonication time was needed to disrupt the aggregation.

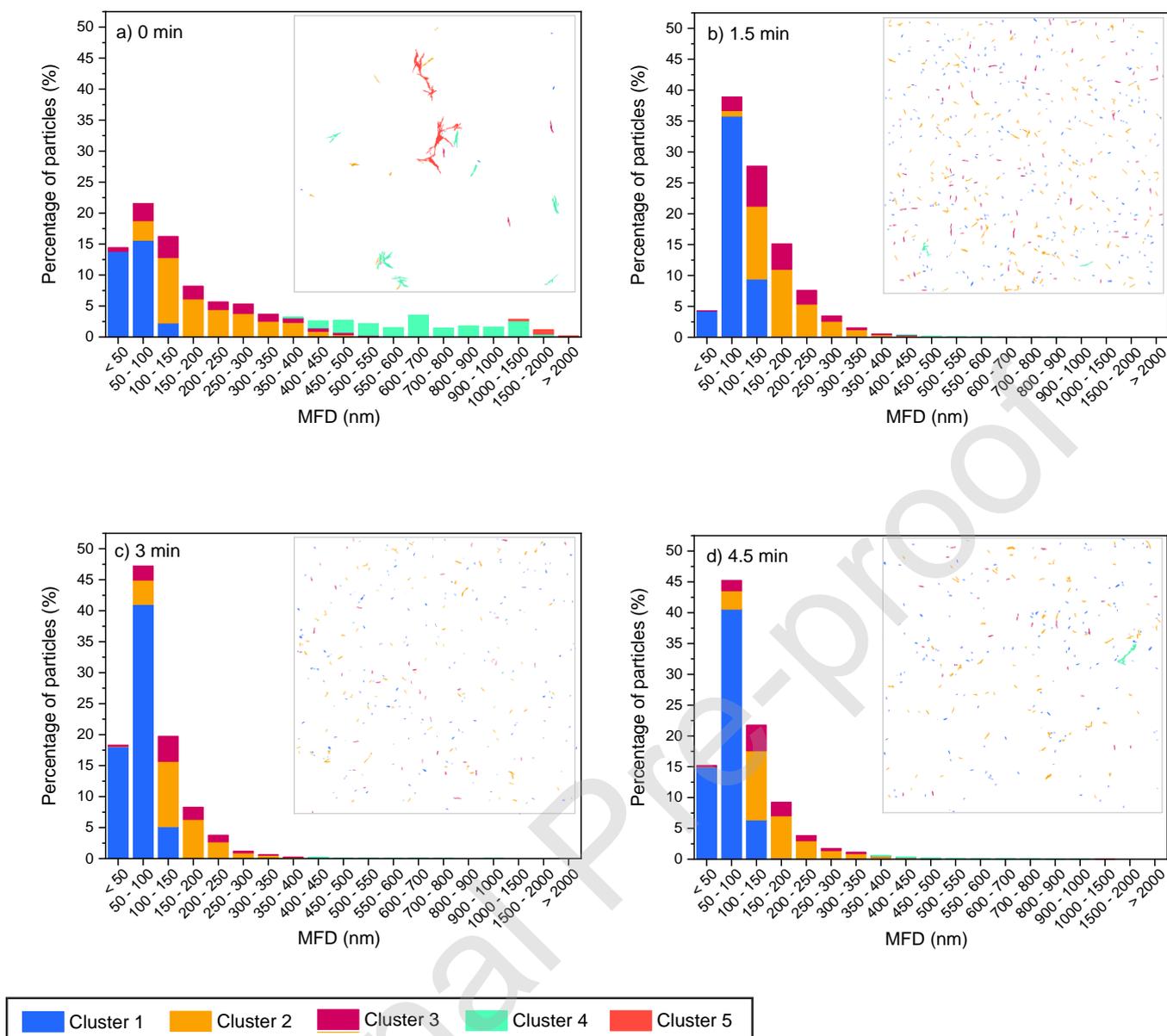


Figure 7. Evolution of the particles size distribution, in terms of maximum feret diameter (MFD), and the visual aspect of the cellulose nanocrystals (CNCs) produced from pine with the sonication time: a) 0 min, b) 1.5 min, c) 3 min and d) 4.5 min. Colors indicate the different clusters identified: Cluster 1 in blue, Cluster 2 in yellow, Cluster 3 in magenta, Cluster 4 in aquamarine and Cluster 5 in red coral. Note: All particles identified in the whole set of images obtained from each sample were considered in the graphs.

Although the profiles of the size distribution graphs (without considering clusters) obtained at different sonication times were very similar between the considered CNC

samples, the TEM images acquired from the suspensions at different sonication conditions tell a very different story, being possible to observe the effect of sonication on particle morphologies with the naked eye. For instance, despite the total percentage of particles between 50 and 100 nm after 3 min of sonication was around 45% in both CNC samples, there is a clear difference between the types of CNC particles observed in each case. The approach proposed in this paper could contribute to solve this disagreement while providing an insight as to the population dynamics of the different types of aggregates resulting from the application of disaggregation treatments.

The description of aggregates could have an enormous significance in some applications. For instance, in the use of CNCs as reinforcing elements, the presence of aggregates with a short elongation, high complexity and low compactness, i.e. those belonging to cluster 5, could entail a loss of efficacy and even a worsening in the macroscopic properties of the material (Shojaeiarani et al., 2018). On the other hand, despite cluster 4 aggregates being also quite complex and porous, their elongated structure makes them not so detrimental for those applications that pretend to improve the strength of the material in just one direction (Liu, Yang, Chang, Wang & Ren, 2020). Moreover, there are applications that leverage the self-assembling of CNCs into chiral nematic liquid crystals as physical templates (Uhlir et al., 2016). In these, a fully dispersed suspension is needed, that can be achieved when only clusters 1 and 2 are identified in the suspension. The method proposed allows, therefore, a detailed characterization of CNC suspensions aggregates. For each particular application, it would be possible to define the adequate treatment to ensure the non-presence of certain aggregates, enabling thus the optimization of nanocellulose use.

4. CONCLUSIONS

A new approach to characterize the aggregation of CNC suspensions based on the analysis of TEM micrographs has been developed. Particles from the images were automatically

classified into groups according to MFD, elongation, circularity and area. This option showed the greatest inter-cluster distance and provided a meaningful description of the CNC aggregates in terms of their eventual application.

With this approach, five clusters were established. Three of them grouped particles of MFD below 400 nm with different morphologies. Cluster 1 was characterized for having non-branched particles with a low length, being probably individuals or small fibrillar assemblies. Clusters 2 and 3 were longer and with a lower fractal dimension than particles of cluster 1. The main difference between clusters 2 and 3 lied in the different elongation and eccentricity, these being of a greater magnitude in cluster 3. Finally, two groups of larger and much more complex aggregates were established as clusters 4 and 5. These two groups of particles can be considered as the most detrimental for homogeneous CNC applications.

In the combination of parameters selected, the approach proved capable to describe the different states of aggregation of CNCs suspensions induced through sonication in terms of the evolution in the number of elements belonging to each cluster. In particular, it was possible to discern how the application of sonication to aggregated CNC suspensions resulted in both, a decrease in the number of aggregates belonging to clusters 4 and 5, an increase in the membership of the cluster 1 and a different evolution in the number of particles of clusters 2 and 3, depending on the sample type. The capability to assess the evolution of aggregation with sonication seems to prove the potential of the method presented, even despite the fact that some inevitable flattening of the particles will occur during sample deposition. Finally, given that the clustering models developed were trained with a very large population of CNC aggregates produced from different cellulose sources and presenting a wide variety of aggregation states, we believe that the approach and models described in this paper can serve as a reference or standard for the characterization of the aggregation state of CNCs and other nanoparticles with similar morphology that tend to aggregate.

5. ACKNOWLEDGEMENTS

Authors want to thank the Spanish Ministry of Economy and Competitiveness for the funding of the project CTQ2017-85654-C2-2-R, as well as the Community of Madrid for funding the RETO-PROSOST-2-CM (P2018/EMT4459). Also thanks to the Spanish National Centre of Electronic Microscopy for the support during image acquisition.

Declarations of interest: none

CRedit author statement

Cristina Campano: Term, conceptualization, methodology, validation, formal analysis, investigation, data curation, writing - original draft, visualization. **Patricio Lopez-Exposito:** Term, conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing - review & editing, visualization. **Laura Gonzalez-Aguilera:** methodology, validation, investigation, data curation, visualization. **Angeles Blanco:** resources, writing - review & editing, supervision, project administration, funding acquisition. **Carlos Negro:** resources, writing - review & editing, supervision, project administration, funding acquisition.

References

- Brinkmann, A., Chen, M. H., Couillard, M., Jakubek, Z. J., Leng, T. Y., & Johnston, L. J. (2016). Correlating Cellulose Nanocrystal Particle Size and Surface Area. *Langmuir*, 32(24), 6105-6114.
- Campano, C., Balea, A., Blanco, Á., & Negro, C. (2020). A reproducible method to characterize the bulk morphology of cellulose nanocrystals and nanofibers by transmission electron microscopy. *Cellulose*.
- Campano, C., Lopez-Exposito, P., Blanco, A., Negro, C., & van de Ven, T. G. (2019). Hairy cationic nanocrystalline cellulose as a novel flocculant of clay. *Journal of Colloid and Interface Science*.
- Campano, C., Miranda, R., Merayo, N., Negro, C., & Blanco, A. (2017). Direct production of cellulose nanocrystals from old newspapers and recycled newsprint. *Carbohydr. Polym.*, 173, 489-496.
- Carrillo, I., Mendonça, R. T., Ago, M., & Rojas, O. J. (2018). Comparative study of cellulosic components isolated from different Eucalyptus species. *Cellulose*, 25(2), 1011-1029.
- Chen, M., Parot, J., Mukherjee, A., Couillard, M., Zou, S., Hackley, V. A., & Johnston, L. J. (2020). Characterization of size and aggregation for cellulose nanocrystal dispersions separated by asymmetrical-flow field-flow fractionation. *Cellulose*, 27(4), 2015-2028.
- Cherhal, F., Cousin, F., & Capron, I. (2015). Influence of charge density and ionic strength on the aggregation process of cellulose nanocrystals in aqueous suspension, as revealed by small-angle neutron scattering. *Langmuir*, 31(20), 5596-5602.
- Di Giorgio, L., Martín, L., Salgado, P. R., & Mauri, A. N. (2020). Synthesis and conservation of cellulose nanocrystals. *Carbohydrate Polymers*, 238, 116187.
- Grishkewich, N., Mohammed, N., Tang, J., & Tam, K. C. (2017). Recent advances in the application of cellulose nanocrystals. *Current Opinion in Colloid & Interface Science*, 29, 32-45.
- Henriksson, M., Berglund, L. A., Isaksson, P., Lindstrom, T., & Nishino, T. (2008). Cellulose nanopaper structures of high toughness. *Biomacromolecules*, 9(6), 1579-1585.
- International Organization for Standardization. (2010). Pulps — Determination of limiting viscosity number in cupri-ethylenediamine (CED) solution (ISO/DIS Standard No. 5351). Retrieved from <https://www.iso.org/standard/51093.html>
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3), 264-323.
- Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*(7), 881-892.
- Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal of Computer Applications*, 1(6), 90-95.
- Köylü, Ü. Ö., Faeth, G., Farias, T. L., & Carvalho, M. d. G. (1995). Fractal and projected structure properties of soot aggregates. *Combustion and Flame*, 100(4), 621-633.
- Liu, X., Yang, K., Chang, M., Wang, X., & Ren, J. (2020). Fabrication of cellulose nanocrystal reinforced nanocomposite hydrogel with self-healing properties. *Carbohydrate Polymers*, 240, 116289.
- Lopez-Exposito, P., Campano, C., van de Ven, T. G., Negro, C., & Blanco, A. (2019). Microalgae harvesting with the novel flocculant hairy cationic nanocrystalline cellulose. *Colloids and Surfaces B: Biointerfaces*.
- Mingqiang, Y., Kidiyo, K., & Joseph, R. (2008). A survey of shape feature extraction techniques. *Pattern recognition*, 15(7), 43-90.
- Pal, N. R., Bezdek, J. C., & Tsao, E.-K. (1993). Generalized clustering networks and Kohonen's self-organizing scheme. *IEEE transactions on Neural Networks*, 4(4), 549-557.

- Phan-Xuan, T., Thuresson, A., Skepö, M., Labrador, A., Bordes, R., & Matic, A. (2016). Aggregation behavior of aqueous cellulose nanocrystals: the effect of inorganic salts. *Cellulose*, 23(6), 3653-3663.
- Rasteiro, M., Garcia, F., Ferreira, P., Blanco, A., Negro, C., & Antunes, E. (2008a). The use of LDS as a tool to evaluate flocculation mechanisms. *Chemical Engineering and Processing: Process Intensification*, 47(8), 1323-1332.
- Qi, W., Yu, J., Zhang, Z., & Xu, H. N. (2019). Effect of pH on the aggregation behavior of cellulose nanocrystals in aqueous medium. *Materials Research Express*, 6(12), 125078.
- Segal, L., Creely, J. J., Martin, A. E., & Conrad, C. M. (1959). An Empirical Method for Estimating the Degree of Crystallinity of Native Cellulose Using the X-Ray Diffractometer. *Textile Research Journal*, 29(10), 786-794.
- Shojaeiarani, J., Bajwa, D. S., & Stark, N. M. (2018). Spin-coating: A new approach for improving dispersion of cellulose nanocrystals and mechanical properties of poly (lactic acid) composites. *Carbohydrate Polymers*, 190, 139-147.
- Sinha, G., & Patel, B. C. (2014). *Medical Image Processing*: PHI Learning Pvt. Ltd.
- Stinson-Bagby, K. L., Roberts, R., & Foster, E. J. (2018). Effective cellulose nanocrystal imaging using transmission electron microscopy. *Carbohydrate Polymers*, 186, 429-438.
- Uhlig, M., Fall, A., Wellert, S., Lehmann, M., Prévost, S., Wågberg, L., von Klitzing, R., & Nyström, G. (2016). Two-dimensional aggregation and semidilute ordering in cellulose nanocrystals. *Langmuir*, 32(2), 442-450.
- Wozniak, M., Onofri, F., Barbosa, S., Yon, J., & Mroczka, J. (2012). Comparison of methods to derive morphological parameters of multi-fractal samples of particle aggregates from TEM images. *Journal of Aerosol Science*, 47, 12-26.
- Wu, S., & Chow, T. W. (2004). Clustering of the self-organizing map using a clustering validity index based on inter-cluster and intra-cluster density. *Pattern recognition*, 37(2), 175-188.