Cellulose (2020) 27:4189–4208
https://doi.org/10.1007/s10570-020-03058-0

ORIGINAL RESEARCH

Cellulose nanofiber diameter distributions from microscopy image analysis: effect of measurement statistics and operator

Shaun Ang · Judith Raj Narayanan · Wriju Kargupta · Victoria Haritos · Warren Batchelor

Received: 6 October 2019 / Accepted: 12 February 2020 / Published online: 28 February 2020

Abstract In the field of cellulose nanofibers (CNF), the determination of their nano-scale fibre diameters is crucial for characterising the material and helping end-users compare different products available on the market. Yet, measuring CNF diameters from microscopy images is often a tedious task due to the limited field of view at high magnification, high fibre counts and complex entanglement of fibre networks. In this work, the effect of measuring all fibres in an entire image, 100 representative fibres in an image and all fibres in a reduced area of an image were investigated. Two CNF samples were analysed: a coarser more heterogeneous fibre sample and a finer more homogeneous fibre sample. Four high magnification images were analysed per sample. Different people (operators) were given the same task of measuring the fibres to study the effect of operator bias. All fibres from all images were first measured by Operator 1 (main operator) to determine the baseline distribution. Operators 2 and 3 were tasked with selecting and measuring 100 representative fibres from each image. The results showed that when the operators were asked to select fibres, they showed consistent bias towards larger fibres thus rendering the results unreliable. A more promising approach was found when Operators 2 and 3 analysed all fibres within a reduced area within an image. This showed a maximum deviation from the baseline fibre diameter distribution of 9 nm for the coarser more heterogeneous sample and 2 nm for the finer more homogeneous sample. Statistical analysis showed that approximately 150 fibres needed to be measured in order to obtain representative results, provided all fibres within the selected area were measured. Kruskal–Wallis statistical testing also showed that the technique of measuring all fibres in a reduced area of an SEM image yielded more representative results to the baseline distribution compared to the technique of measuring 100 representative fibres. The results here provide the foundation for future accurate measurements of CNF diameter distributions.

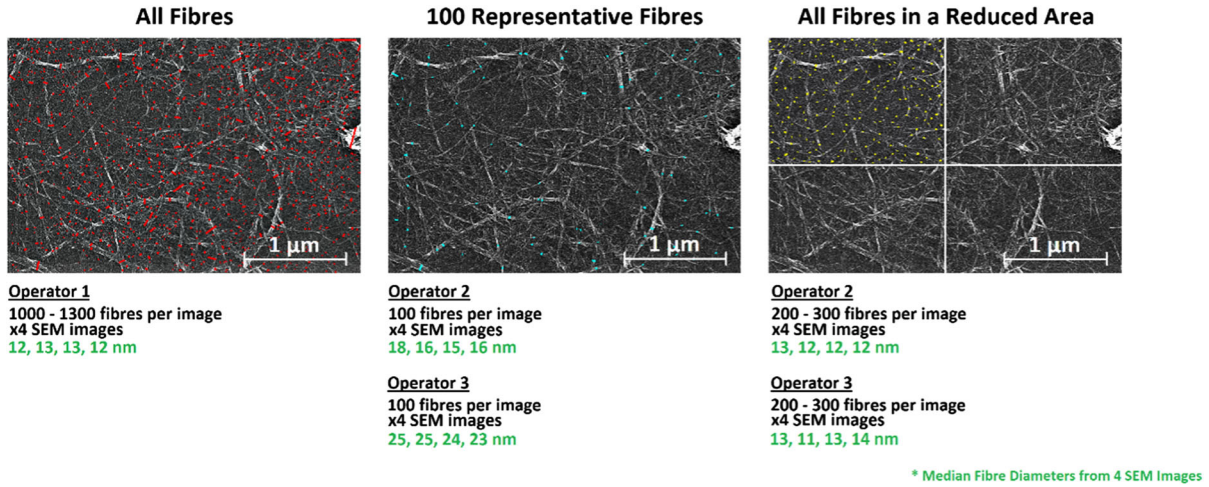
Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10570-020-03058-0) contains supplementary material, which is available to authorized users.

S. Ang · J. R. Narayanan · W. Kargupta · V. Haritos · W. Batchelor (✉)
Department of Chemical Engineering, Bioresource Processing Research Institute of Australia, Monash University, Melbourne, VIC 3800, Australia
e-mail: warren.batchelor@monash.edu

S. Ang
e-mail: shaun.ang@monash.edu

V. Haritos
e-mail: victoria.haritos@monash.edu

Graphic abstract



Keywords Cellulose nanofiber · SEM · Fibre diameter · Histogram · Statistics

Introduction

Comprehensive understanding of cellulose nanofiber (CNF) dimensions is necessary for proper characterisation of the material and understanding differences among the varieties available on the market and those produced in labs. This is necessary to track CNF development with processing to ensure that the production of CNF is of a consistent quality (Osong et al. 2016). The variation in fibre sizes, fibre morphology and formation of aggregates greatly affect performance in applications such as papermaking or as barriers (Osong et al. 2016). Common measures of fibre quality include fibre diameter (Zhang et al. 2012), fibre aspect ratio (Raj et al. 2016), surface charge (Dimic-Misic et al. 2018), crystallinity (Mandal and Chakrabarty 2011), specific surface area (Spence et al. 2010) and degree of polymerisation (Iwamoto et al. 2007).

One of the most important measures of CNF fibre quality is the fibre diameter distributions. For example, the CNF diameter is a critical parameter for applications such as CNF spray coating systems, to prevent nozzle clogging (Shanmugam et al. 2018). Microscopic imaging is the most widely used method used for CNF diameter evaluation (Kangas et al. 2014). These techniques include scanning electron microscopy (SEM) (Abe and Yano 2009; Chaker et al. 2013; Henriksson et al. 2008), transmission electron microscopy (TEM) (Bhatnagar and Sain 2005; Habibi et al. 2009; Isogai 2013) and atomic force microscopy (AFM) (Baati et al. 2018; Ding et al. 2019; Santucci et al. 2016).

With conventional macroscopic wood pulp fibres, measurements of fibre diameters are typically performed with fibre analyser instruments (Chen et al. 2019). However, these industrial instruments are limited to the micron-scale nature of wood fibres and cannot accurately resolve the nano-scale dimensions of CNF. Gel point analysis techniques have been developed for CNF to estimate their aspect ratio (length/diameter) (Varanasi et al. 2013) but do not provide accurate and detailed fibre diameter distributions. Dynamic light scattering (DLS) experiments have also been investigated to track the changes in CNF fibre dimensions with processing (Errokh et al. 2018). However, criticisms are often raised at the assumed spherical morphology of the CNF necessary for DLS measurements (Errokh et al. 2018), which is contradictory to the well-established thin and highly entangled structure of CNF. As such, powerful microscopic imaging technique such as SEM, TEM and AFM are often utilised to resolve the nano-scale nature of CNF.

Challenges with microscopy images include their limited field of view at very high magnification. This raises questions of how representative smaller images

are of a larger fibre sample (Kangas et al. 2014). Microscopy image analysis is also very time-consuming (Kangas et al. 2014; Osong et al. 2016) which requires delicate sample preparation techniques such as gold (Mandal and Chakrabarty 2011), carbon (Siró et al. 2011) and iridium (Jiang et al. 2017) coating prior to observation. A trade-off often exists with the concentration of fibres used during imaging and the resulting diameter measurements. Too low a fibre concentration will produce samples with too low a fibre count to be representative of the larger distribution. Images taken at too high a sample concentration causes challenges with analysing the highly entangled fibre networks and measuring all fibres.

Computerised image analysis programs such as Diameter J have been developed for the measurement of steel wire nanofibre diameters (Hotaling et al. 2015). However, unlike steel nanofibres which have relatively smooth fibre edges, CNF have the added complexity of exhibiting varying degrees of fibrillation along their fibre edges. This introduces issues with computerised image segmentation due to the varying levels of contrast throughout the image from the overlapping and entwined nature of the CNF. Besides that, program limitations such as the under-representation of larger fibres compared to smaller fibres have also been noted which leads to a bias weighted average measurement (Hotaling et al. 2015). To date there are no effective methods for measuring CNF diameters from microscopy images. This further emphasises that more research is necessary to improve the accuracy and reproducibility of results. From the literature, the determination of CNF diameters from image analysis has been approached and reported using four different methods, with increasing levels of complexity.

At the first and lowest level of complexity, an estimated range of fibre diameters is inferred from observations and basic measurements with microscopy images. Examples of this in the literature include, nanofibrillated cellulose from hardwood kraft pulp with diameters in the range of 15–45 nm (Taheri and Samyn 2016), microfibrillated cellulose diameters from softwood kraft pulps with diameters in the range of 20–90 nm (Taniguchi and Okamura 1998), CNF from wheat straw hulls with diameters in the range of 10–80 nm (Alemdar and Sain 2008) and CNF from soybeans with diameters in the range of 50–100 nm (Wang and Sain 2007). This method lacks a high level of statistical analysis but has the advantage of being simple to execute. It is often used to confirm the presence of CNF and describe features as opposed to thoroughly tracking the change in fibre diameters with progressive treatments.

At the second level of complexity, basic statistics such as mean or average fibre diameter are calculated from measuring a selected number of fibres. Examples of this in the literature include cellulose nanofibres from stone groundwood pulp with an average diameter of 33.7 nm from measuring 18 fibres (Delgado-Aguilar et al. 2016), cellulose nanofibres from thermomechanical pulp with an average diameter of 12.9 nm from measuring 100 fibres (Diop et al. 2017), and cellulose nanofibrils from cotton pulp with an average diameter of 394 nm from measuring 15 fibres (Karande et al. 2011). This method has an added benefit of using statistical analysis to better describe a measure of central tendency. However, there may be selection bias from measuring a limited number of fibres. Problems might also arise from bias towards noticeably smaller or larger fibre diameters to support a particular hypothesis. Further, the use of mean or average fibre diameter is not always a robust measure of central tendency. The heterogeneous nature of many CNF samples with a high number of liberated microfibrils and a lower number of larger fibre bundles (Delgado-Aguilar et al. 2016) often results in a high degree of skewness. This causes an asymmetrical distribution which reduces the utility of the mean fibre diameter alone.

At the third level of complexity, a selected number of fibres in an image are measured and their diameter distributions are plotted as histograms to describe the spread of data. This histogram method has the advantage of being able to express a wide range of data points and frequency peaks, compared to a single measure of mean. One study utilised this technique to track the change in CNF diameter distributions with increasing homogenisation treatments (Raj et al. 2016). The authors measured 50 observable fibres from one SEM image per fibre sample and plotted a frequency versus fibre bin centre histogram distribution (Raj et al. 2016). In another study, 71 and 161 cellulose nanofibres were measured from an SEM and TEM image, respectively (Wang et al. 2012). Histograms were then plotted as a percentage distribution versus the range of fibril diameter bins. However, concerns are still raised at the selection bias of a few fibres to adequately express an entire distribution,

although the use of a histogram is helpful to describe a wider spread of data.

At the fourth level of complexity, all fibres in an image are measured and repeated for multiple images. The CNF distributions are plotted as histograms and statistical analysis is performed. One study produced two CNF samples from a valley beater and a shaker mill (Zhang et al. 2012). Multiple SEM images were taken at different magnifications for both samples. All fibres within each SEM image were measured. The results were then plotted as a histogram distribution with a total of 398 fibres for the valley beater sample and 123 fibres for the shaker mill sample (Zhang et al. 2012). In our previous study, multiple CNF samples were produced from various combinations of refining and homogenisation (Ang et al. 2019). One low magnification SEM image (to capture the coarser fibres) and four high magnification SEM images (to capture the finer CNF fibres) were taken for each sample, with total measurements of up to 4000 fibres for some samples (Ang et al. 2019). Histograms of frequency (counts/m^2) versus bin centre fibre diameter were plotted. Statistical analysis performed resulted in very positive skewness and kurtosis values, indicative of an asymmetrical distribution. This is consistent with the heterogeneous nature of many CNF samples. While this method of analysis avoids bias fibre selection, it was very time-consuming and tedious to perform.

In summary, there are multiple methods for the determination of CNF diameters from image analysis, each with varying degrees of complexity. However, much is still unknown regarding the number of fibres in an image that need to be measured to obtain an accurate distribution. It is also uncertain if measuring a few randomly selected fibres is sufficient to accurately describe a distribution. Other questions include, would measuring all fibres in a reduced area of a single image still result in representative results. Besides that, not much is known regarding the effect of using different operators to analyse similar images on the fibre diameter distribution.

In this study, we investigated the effect of manually measuring all fibres in an image, 100 representative fibres in an image and all fibres in a reduced area of an image by people (operators). The measurements of all fibres in an image were first performed by Operator 1 to obtain a baseline distribution. Two independent operators (Operators 2 and 3) were given instructions to measure 100 representative fibres in an image and all fibres in a reduced area of an image. The results here show a promising method to improve the analysis of fibre diameter distribution for CNF characterisation.

Experimental

CNF sample production

Two samples of CNF were produced from never-dried bleached eucalyptus kraft (BEK) pulp. This pulp was refined in a PFI mill refiner in accordance with the TAPPI T248 sp-00 standard and high-pressure homogenised in a GEA Niro Soavi PANDAPLUS 2000 operating at 1000 bar pressure at 1 wt% consistency. The first sample was processed for 10,000 PFI mill revolutions and 1 high-pressure homogenisation pass, and is referred to as the 10K1P sample. The second sample was processed for 50,000 PFI mill revolutions and 3 high-pressure homogenisation passes, and is referred to as the 50K3P sample. This allowed for the comparison of two different types of CNF. A coarser more heterogeneous 10K1P fibre sample and a finer more homogeneous 50K3P fibre sample.

It is worth noting that within the context of this research, the term "coarser" is in reference to the nano-sized CNF with the less processed 10K1P sample. This term is not in reference to the largely unbroken micron-sized fibres in the sample which can be viewed under an optical microscope and are significantly fewer in number compared to these nano-sized CNF.

Scanning electron microscopy (SEM) imaging

The 10K1P and 50K3P fibre samples were diluted to a 0.001 wt% suspension concentration. A single drop of each suspension was cast on a silicon chip and left to air dry. The dried samples were then coated with iridium prior to imaging. Both samples were observed under an FEI Magellan 400 XHR FEGSEM fitted with a monochromator and operating at a 3.0 kV spot size and 6.3 pA current.

For each sample, four high magnification SEM images were taken. The 8 images across both samples were labelled 10K1P-A, 10K1P-B, 10K1P-C, 10K1P D, 50K3P-A, 50K3P-B, 50K3P-C and 50K3P-D. To

allow for a direct comparison, all 8 images were taken at the same 60,000 magnification, had the same horizontal field width of 3.45 μm, with a fixed 1 μm length scale bar and an image area of 7.395×10^{-12} m^2.

Diameter distribution from SEM image analysis

All fibre diameters were manually measured as image pixels using Image J processing software then correlated to a pre-determined pixel-to-length scale to obtain measurements in nm. Each image was annotated to mark the measured fibres. Three different people (operators) were involved with this measurement process.

Operator 1 (main operator), counted and measured all observable fibres from all SEM images for each sample. In total, 2946 fibres were measured across all four images of the 10K1P sample and 4723 fibres were measured across all four images of the 50K3P sample. For each sample, the fibre diameter distribution was plotted as a histogram of relative frequency (%) versus bin centre for fibre diameter. For consistency and easy comparison between the fibre samples, all histograms plotted had the same 5 nm bin sizes on the x-axis and relative frequency scale on the *y*-axis.

The variation in fibre diameter distribution between the separate A, B, C and D images for each sample was then investigated. Separate histograms were plotted for each of the fibre distributions measured in an SEM image. This allowed for the comparison between the different 10K1P images of A (721 fibres), B (888 fibres), C (605 fibres) and D (732 fibres); and 50K3P images of A (981 fibres), B (1161 fibres), C (1229 fibres) and D (1352 fibres).

In the next study, two independent operators (Operator 2 and Operator 3) were given the instructions to pick and measure 100 representative fibres in each SEM image for all 8 images. This also allowed for the investigation of operator bias when different operators tasked with the same objective. Results here were plotted into histograms with the same axes as the previous section.

Finally, the fibres in the top-left quadrant of each SEM image were measured. The isolated quadrant had an image area of 2.15×10^{-12} m^2, consistent for all 8 images. Operators 2 and 3 were only given the instructions to count and measure every fibre in the isolated quadrant for each image. The aim was to investigate the effect of measuring all fibres in a reduced area of an image on the diameter distribution, while avoiding bias in fibre selection. As with the previous section, the results were plotted into histograms with the same axes to allow for the comparison between different operators.

Figure 1 shows an example of annotated images from measuring all fibres in an image (Fig. 1a), 100 representative fibres in an image (Fig. 1b) and all fibres in the top-left quadrant in an image (Fig. 1c). The annotated images from measuring all fibres in an image were published in our previous study on CNF sheet strength (Ang et al. 2019). All annotated images from measuring 100 representative fibres and all fibres in the top-left quadrant can be found in the supplementary material.

In this study, the manual measurement of all fibres in an SEM image required about 1.5–2 h on average per SEM image. The manual measurement of 100 representative fibres in an SEM image required about 10–15 min on average per SEM image. The manual measurement of all fibres in the top-left quadrant (reduced area) of an SEM image required about 20–25 min on average per SEM image.

Measures of central tendency and moving median analysis

Descriptive statistics were performed for all samples to determine their measures of central tendency. This includes the measures of minimum, maximum, median, arithmetic mean, standard deviation, skewness and kurtosis for each case. Then, a moving median analysis was conducted by plotting the median fibre diameter as a function of fibre counts for all cases. This was done with the goal of determining the minimum fibre counts necessary for the median to converge to a value.

Kruskal–Wallis-1-way-ANOVA on ranks H test

Statistical testing was performed using the Kruskal–Wallis-1-way-ANOVA on ranks H test (referred to as the Kruskal–Wallis test). The Kruskal–Wallis test is a non-parametric test (does not assume a normal distribution) which makes it suitable for the fibre distributions in this study. Briefly, the Kruskal–Wallis test (McDonald 2009) ranks the data sets within a group and between groups to determine if different groups

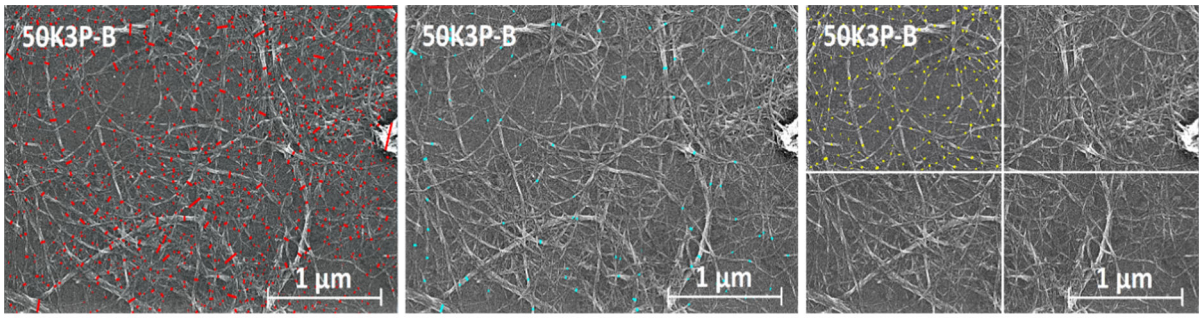


Fig. 1 Annotated SEM images of 50K3P-B from measuring **a** all fibres in an image **b** 100 representative fibres in an image and **c** all fibres in the top-left quadrant in an image

have similar distributions. This allows for the comparison between two or more groups, regardless of the sample size.

The Kruskal–Wallis test provides a H-statistic value for every comparative test performed. This H-statistic value is compared to a standard χ^2 value (determined at an alpha level, α of 0.01 and is a function of the number of degrees of freedom, d.f.). A H-statistic value χ^2 value means that the null hypothesis is rejected, and all groups are not similarly distributed. A H-statistic value $< \chi^2$ value means that the null hypothesis cannot be rejected, and all groups are similarly distributed. The magnitude of the H-statistic value is used to compare between distributions at the same degree of freedom. A full detailed explanation of this test can be found at (McDonald 2009).

In this study, a Kruskal–Wallis test was first performed to compare between the baseline (measurement of all fibres from all images) distribution of the 10K1P sample and the 50K3P sample by Operator 1. This test was referred to as T1, the overall test.

Then, 14 Kruskal–Wallis tests were performed to compare between the individual images and groups of images from measuring all fibres in an image by Operator 1 for the 10K1P sample (tests T2–T8) and 50K3P sample (tests T9–T15), respectively. Detailed information regarding the description of each test can be found in Tables 4 and 5 in the "Results and discussion" section.

Lastly, 32 Kruskal–Wallis tests were performed comparing within and between the groups of operators (Operators 2 and 3), images (A, B, C and D) and methods (measure all fibres in an image, measure 100 representative fibres in an image, measure all fibres in the top-left quadrant of a reduced area of an image) versus the baseline distribution for the 10K1P sample (tests T16–T31) and 50K3P sample (tests T32–T47), respectively. Detailed information regarding the description of each test can be found in Table 6 and 7 in the "Results and discussion" section.

Results and discussion

SEM Imaging

Figures 2 and 3 shows the original four high magnification SEM images taken for the 10K1P and 50K3P sample, respectively. As shown in Fig. 2, 10K1P treatments resulted in a coarser and more heterogeneous mixture of both large and small CNF. The entanglement of both larger and smaller fibres is prevalent in all images shown. From Fig. 3, the 50K3P treatments resulted in a more homogeneous mixture of finer CNF. The high number of finer fibres also resulted in a highly entangled structure.

Measures of central tendency for all fibre diameter distribution

Tables 1 and 2 shows the results from statistical analysis of fibre diameter distribution for the 10K1P and 50K3P sample, respectively. The histograms for all cases are shown in Figs. 4, 5, 6 and 7, and are discussed in the later sections of the discussion.

As can be seen, all distributions in Tables 1 and 2 had very positive skewness values. Skewness is measure of how asymmetrical a distribution is compared to a normal distribution which has a skewness of

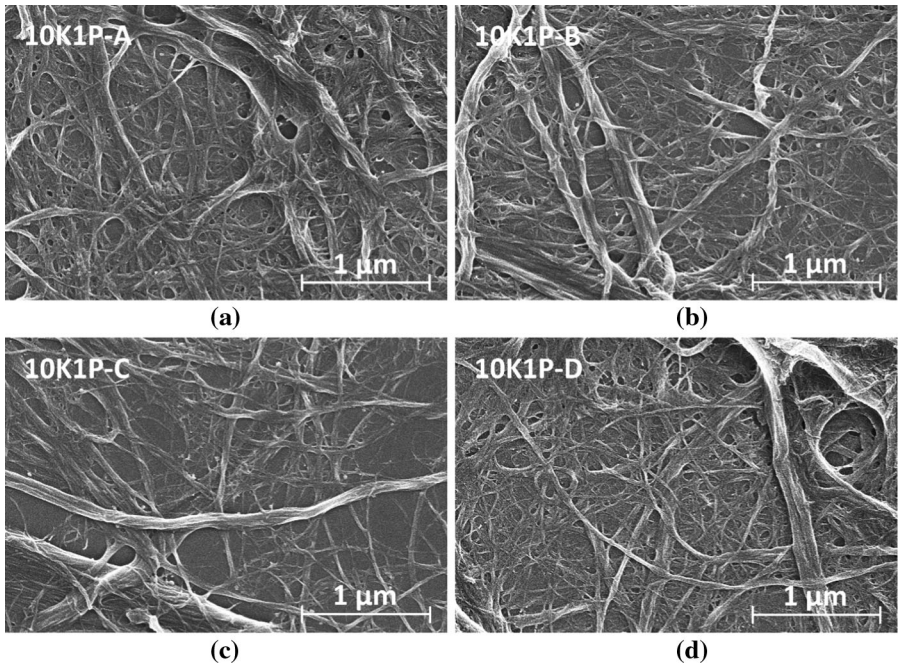


Fig. 2 Four main SEM images of 10K1P CNF fibres

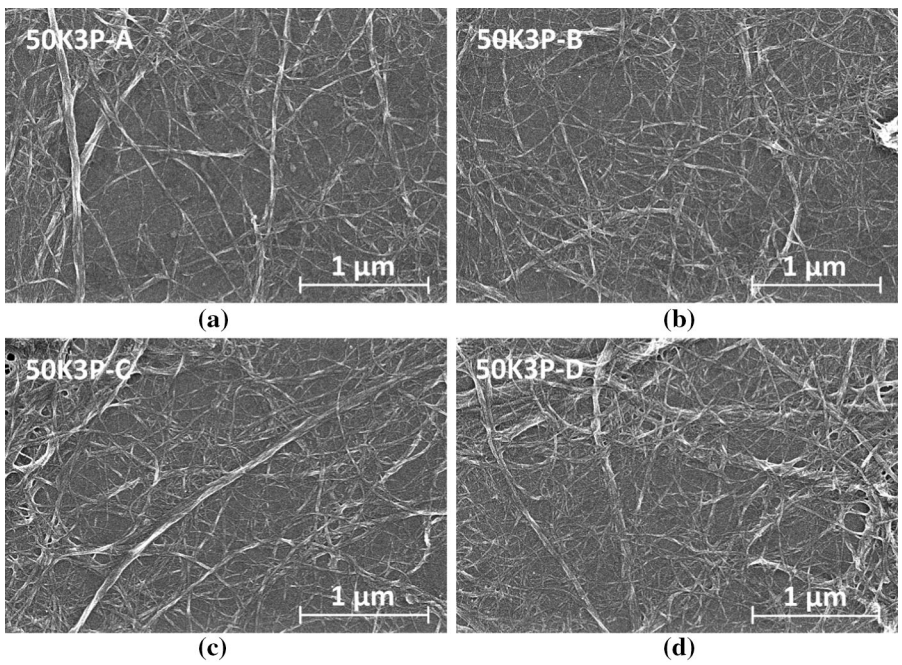


Fig. 3 Four main SEM images of 50K3P CNF fibres

zero. Here, a very positive skewness value indicates that the distribution has a long right-tail end (shown in Figs. 4, 5 and 6). This means that there is a much higher number of smaller-sized fibres, compared to the relatively lower number of bigger-sized fibres.

All distributions also had large positive kurtosis. Kurtosis is a measure of how much the tail-ends in a

Table 1 Statistical analysis of 10K1P fibre diameter distribution

Operator	Image (s)	Median diameter (nm)	Skewness	Kurtosis	Min diameter (nm)	Max diameter (nm)	Arithmetic mean diameter (nm)	SD (nm)	Sample size (counts)
All fibres in each SEM image									
1	A	20	4	21	6	267	28	26	721
	B	15	5	36	4	291	22	24	888
	C	17	6	54	4	392	26	31	605
	D	18	3	11	6	213	28	26	732
All fibres in all four SEM images (baseline)									
1	A,B,C,D	17	5	34	4	392	26	27	2946
100 Representative fibres in each SEM image									
2	A	20	1	1	8	69	25	14	100
	B	21	5	36	8	226	28	26	100
	C	18	4	27	7	179	25	21	100
	D	19	3	13	5	149	25	21	100
3	A	27	1	1	7	96	31	18	100
	B	32	1	2	18	105	38	18	100
	C	32	2	6	14	145	39	24	100
	D	31	2	4	12	164	42	29	100
All fibres in top-left quadrant in each SEM image									
2	A	16	3	12	6	147	23	18	193
	B	18	3	9	7	95	22	14	186
	C	15	4	26	6	114	18	11	224
	D	17	3	11	6	133	23	17	231
3	A	26	2	5	6	125	31	21	219
	B	20	2	4	6	122	27	19	209
	C	21	2	4	5	122	29	22	205
	D	24	2	7	5	152	31	22	208

distribution differ to the tail-ends of a normal distribution. A large positive kurtosis value indicates a sharp peak in the data set, which is referred to as a leptokurtic distribution. This is also evidence of a notably higher number of smaller-sized fibres and lower number of bigger-sized fibres.

These skewness and kurtosis values are consistent with an asymmetrical distribution. This is also expected by the larger standard deviation values compared to arithmetic mean diameter values for many cases. Accordingly, the median diameter was instead chosen as a more representative measure of central tendency as it is robust in the presence of outliers. The median diameter also forms the basis of comparison between the separate studies and was discussed in the following sections of this paper.

Distribution from measuring all fibres in all SEM images

Figure 4a, b shows the diameter distribution from measuring all fibres in the four SEM images for the 10K1P and 50K3P sample, respectively. The 10K1P sample had a mode of 10–15 nm, while the 50K3P had a higher proportion of smaller fibres with a mode of 5–10 nm. In both samples, 60% to 75% of fibres had diameters in the 5–20 nm range. The relative frequency of fibres greater than 45 nm was higher with the 10K1P sample compared to the 50K3P sample, as 13% of the 10K1P sample had fibre diameters greater than 45 nm. These results support the higher degree of heterogeneity with the 10K1P sample, shown in Fig. 2.

Table 2 Statistical analysis of 50K3P fibre diameter distribution

Operator	Image (s)	Median diameter (nm)	Skew-ness	Kurt-osis	Min diameter (nm)	Max diameter (nm)	Arithmetic mean diameter (nm)	SD (nm)	Sample size (counts)
All fibres in each SEM image									
1	A	12	5	33	3	158	15	12	981
	B	13	6	61	3	228	17	15	1161
	C	13	5	39	3	203	17	15	1229
	D	12	6	66	2	298	17	18	1352
All fibres in all four SEM images (baseline)									
1	A,B,C,D	13	6	60	2	298	17	15	4723
100 Representative fibres in each SEM image									
2	A	18	3	13	8	79	20	10	100
	B	16	3	9	5	68	19	10	100
	C	15	2	10	8	55	16	7	100
	D	16	1	3	6	45	17	7	100
3	A	25	2	3	10	76	28	11	100
	B	25	2	10	10	86	26	10	100
	C	24	7	59	11	207	28	21	100
	D	23	3	9	9	101	26	14	100
All fibres in top-left quadrant in each SEM image									
2	A	13	3	12	5	66	15	8	207
	B	12	1	2	5	29	13	4	220
	C	12	2	8	5	49	14	7	301
	D	12	4	29	5	83	14	7	302
3	A	13	2	5	2	53	14	7	207
	B	11	1	2	2	33	13	5	209
	C	13	9	102	3	203	16	16	211
	D	14	3	11	3	94	17	12	212

As given in Tables 1 and 2, the median fibre diameters of the 10K1P and 50K3P sample were 17 nm and 13 nm, respectively. For this paper, these values are labelled the baseline medians, to which the other studies were compared.

While measuring all fibres from multiple SEM images eliminates the bias in fibre selection, this technique is very time-consuming. Hence, two alternative methods of measuring 100 representative fibres and all fibres in a reduced area of the image were explored, to find a more efficient way to determine a representative fibre diameter distribution.

Distribution from measuring 100 representative fibres in each SEM image

Figure 5 shows the diameter distributions from measuring 100 representative fibres in an image by two different operators (Operators 2 and 3). The results varied greatly between the different operators for a given fibre sample. For the 10K1P sample in Fig. 5a, a mode of 10–15 nm and 15–20 nm was obtained by Operator 2 depending on which of the four images were analysed. However, Operator 3's analysis of the 10K1P sample resulted in very inconsistent distributions between the different images as shown in Fig. 5c. For the more homogeneous 50K3P sample, Operators 2 and 3 obtained a much more consistent spread of data points between the four images

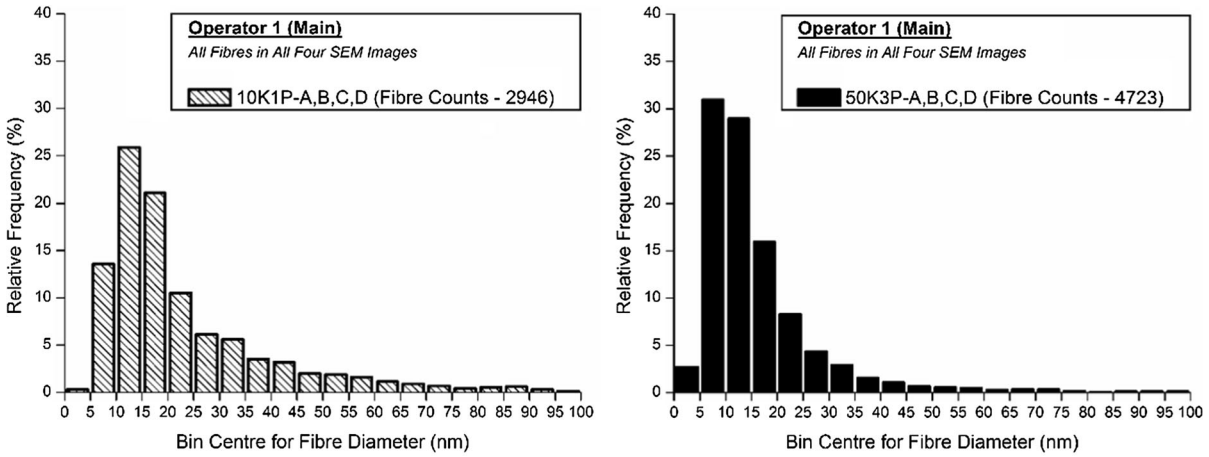


Fig. 4 Diameter distribution from measuring all fibres in all four SEM images for **a** 10K1P and **b** 50K3P fibres by Operator 1

compared to the 10K1P sample. However, their modes still differed from one another as shown in Fig. 5b, d; with Operator 3 consistently measuring larger fibres compared to Operator 2.

From Table 1, measuring 100 representative fibres for the four 10K1P images by Operator 2 resulted in median diameters of 20 nm, 21 nm, 18 nm and 19 nm, compared with median diameters of 27 nm, 32 nm, 32 nm and 31 nm for the four 10K1P images from Operator 3. Here, a maximum error of 4 nm and 15 nm was obtained for Operators 2 and 3 respectively, compared to the baseline median of 17 nm. For the 50K3P sample in Table 2, the median diameters from the four images were determined to be 18 nm, 16 nm, 15 nm and 16 nm from Operator 2's measurements and 25 nm, 25 nm, 24 nm and 23 nm from Operator 3's measurements. Operator 2 and Operator 3 had a maximum error of 5 nm and 12 nm, respectively compared to the baseline median of 13 nm.

Overall, it was observed that when tasked with measuring 100 representative fibres, both operators were consistently biased to measuring the larger fibres. Operator 3's measured diameters were also noted to be much larger than Operator 2's measured diameters regardless of the heterogeneity of the sample. Here, the technique of measuring 100 randomly selected representative fibres was shown to be very prone to operator bias and resulted in inconsistent and non-representative fibre diameter distributions.

Distribution from measuring all fibres in top-left quadrant in each SEM image

The diameter distributions from measuring all fibres in the top-left quadrant in each SEM image are shown in Fig. 6. When comparing Figs. 6c to Fig. 5c in the previous section, it becomes evident that Operator 3's distributions were more consistent when measuring all fibres in a reduced area compared to measuring 100 representative fibres. Also, the distributions for Operator 3 (Fig. 6c) much more closely resembled those for Operator 2 (Fig. 6a), unlike the previous case of Fig. 5a, c. For the more homogeneous 50K3P sample, both operators had very similar modes as shown in Fig. 6b, d from measuring all fibres in the reduced area.

Statistical analysis from measuring all fibres in the top-left quadrant in Table 1 and 2 also produced results that were closer to the baseline distribution, than the 100 fibres approach. For the 10K1P sample in Table 1, median fibre diameters of 16 nm, 18 nm, 15 nm and 17 nm were obtained for Operator 2 and 26 nm, 20 nm, 21 nm and 24 nm were obtained for Operator 3, for the four SEM images. These values here were much closer to the baseline median of 17 nm, than when 100 representative fibres were measured. For the more homogeneous 50K3P sample Table 2, median fibre diameters of 13 nm, 12 nm, 12 nm and 12 nm were obtained for Operator 2 and 13 nm, 11 nm, 13 nm and 14 nm were obtained from Operator 3, for the four SEM images. These median

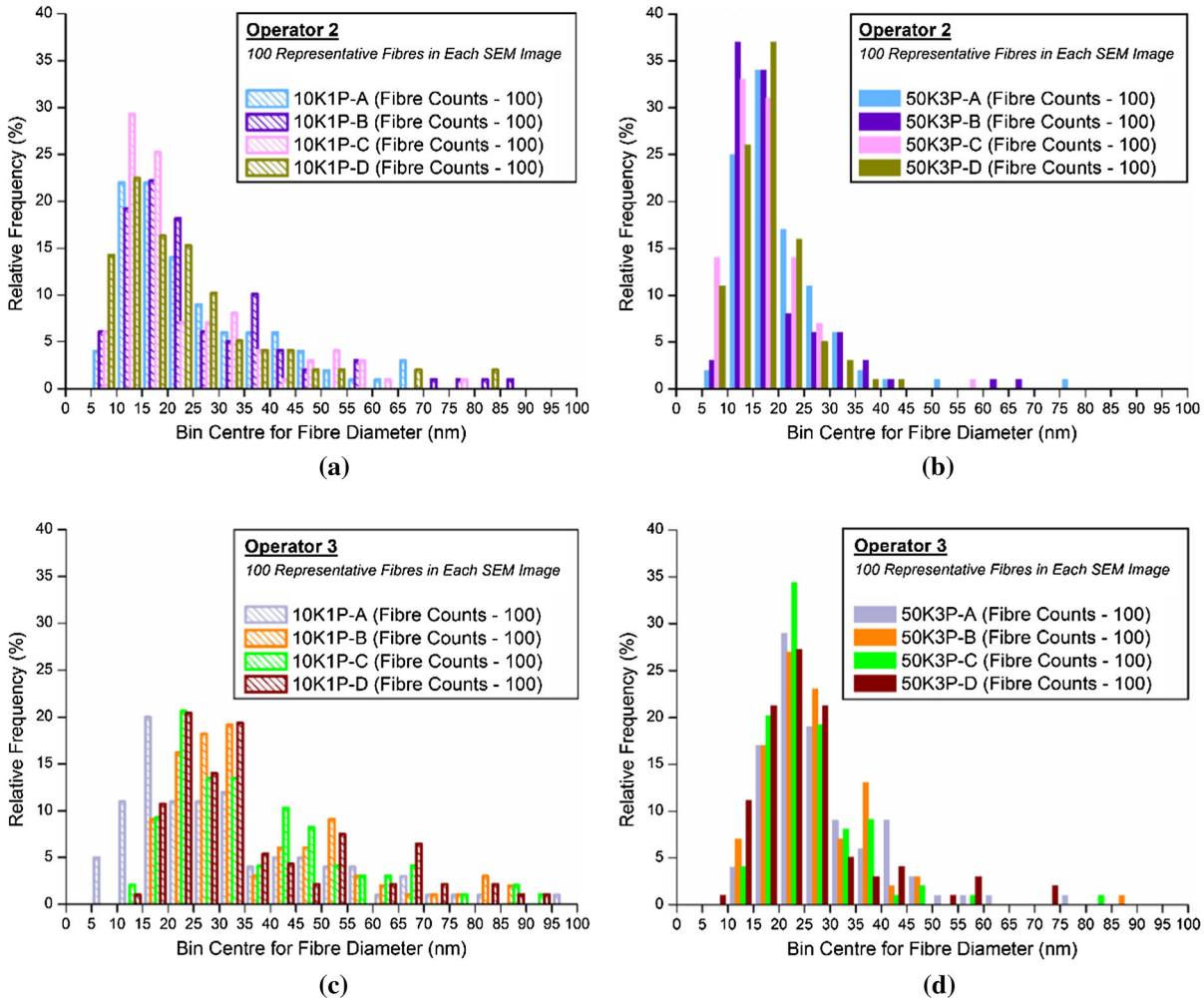


Fig. 5 Diameter distribution from measuring 100 representative fibres in each SEM image for **a** 10K1P and **b** 50K3P fibres by Operator 2 and **c** 10K1P and **d** 50K3P fibres by Operator 3

values were very close to the baseline median of 13 nm.

Overall, a maximum error of 9 nm and 2 nm was obtained for the 10K1P and 50K3P sample, respectively using this technique of measuring fibres in a reduced area. This was a notable improvement compared to the maximum error of 15 nm and 12 nm for the 10K1P and 50K3P sample when measuring 100 representative fibres.

Clearly, analysing a set area of a SEM image yields more accurate results compared to analysing 100 representative fibres. Measuring all fibres within the set area also eliminates operator bias.

Distribution from measuring all fibres in each SEM image

For comparison, the baseline distributions in Fig. 4 were replotted as separate distributions from the individual images and shown in Fig. 7a, b for the 10K1P and 50K3P sample, respectively. Median diameters for the 10K1P images were found to be 20 nm, 15 nm, 17 nm and 18 nm. Median diameters for the 50K3P images were found to be 12 nm, 13 nm, 13 nm and 12 nm. At most, these median values varied by 17% and 7% from the baseline median values of the 10K1P fibres and 50K3P fibres, respectively. The technique also required the measurement

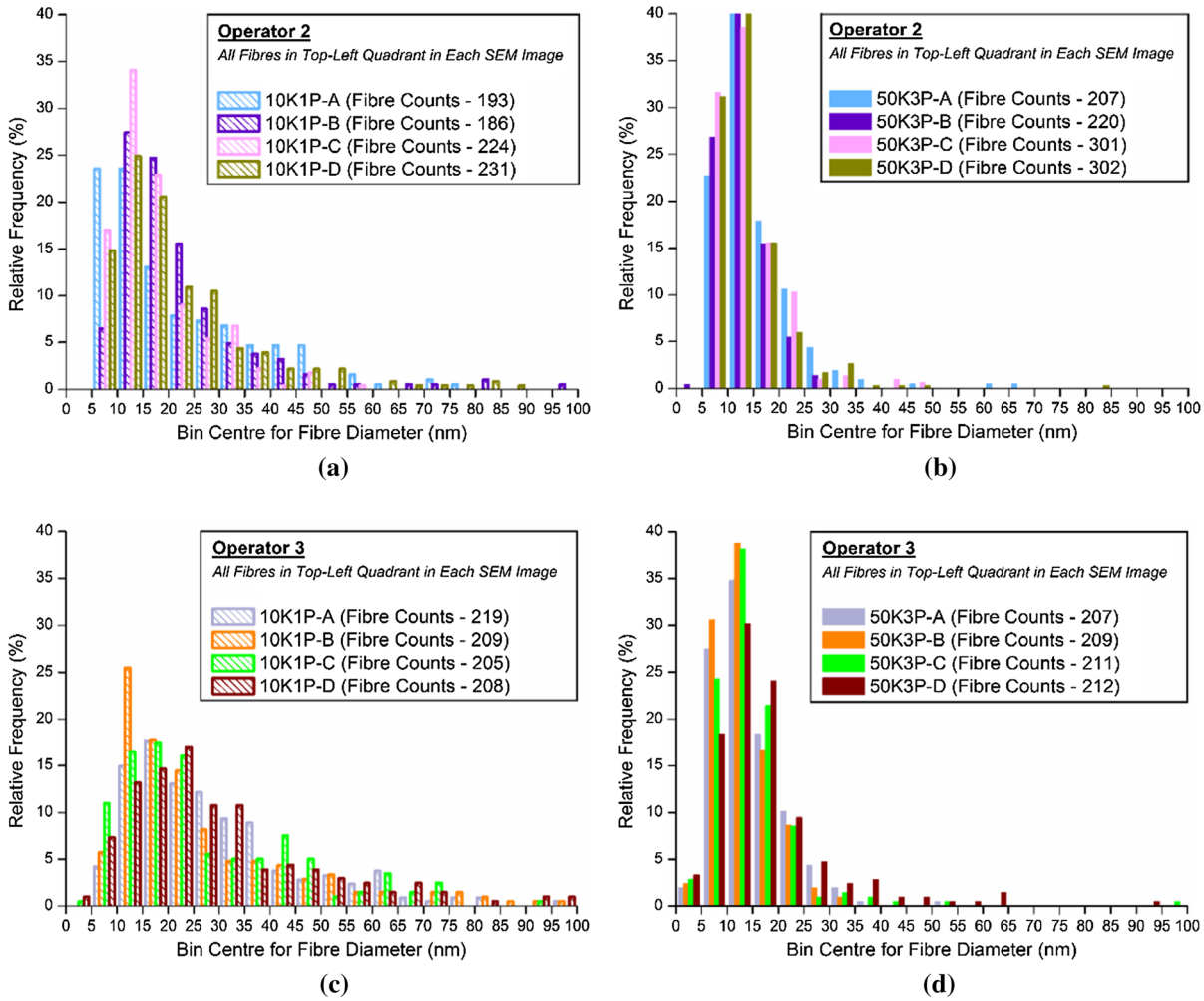


Fig. 6 Diameter distribution from measuring all fibres in top-left quadrant in each SEM image for **a** 10K1P and **b** 50K3P fibres by Operator 2 and **c** 10K1P and **d** 50K3P fibres by Operator 3

of all fibres in an entire image which was a very time-consuming exercise.

Moving median analysis

Figure 8a, b shows the median fibre diameter as a function of fibre counts from measuring all fibres within an image for the 10K1P and 50K3P sample, respectively. From Fig. 8a, the median fibre diameters were shown to converge after approximately 500 counts for most cases. A similar observation was also made with the 50K3P sample in Fig. 8b. Interestingly, the results from both figures also support the previous claim that measuring in excess of 3000–4000 fibres from multiple images did not significantly improve the accuracy of results.

Figure 9a, b shows the median fibre diameter as a function of fibre counts from measuring 100 representative fibres in an image for the 10K1P and 50K3P sample, respectively. For Operator 2, the median was shown to converge after approximately 60 fibre counts with the 10K1P sample in Fig. 9a. In comparison for Operator 3, the median fibre diameter was shown to fluctuate more and only converged after about 80 fibre counts with the 10K1P sample in Fig. 9a. Both Operator 2 and Operator 3's converged median values were also notably different from one another. For the more homogeneous 50K3P fibres, the median fibre diameter was shown to converge after approximately

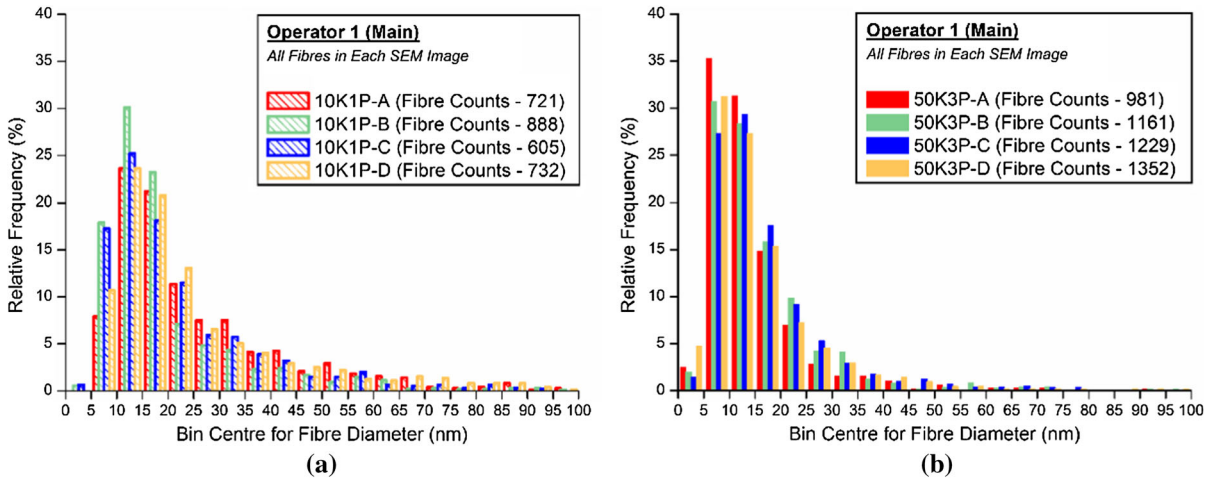


Fig. 7 Diameter distribution from measuring all fibres in each SEM image for **a** 10K1P and **b** 50K3P fibres by Operator 1

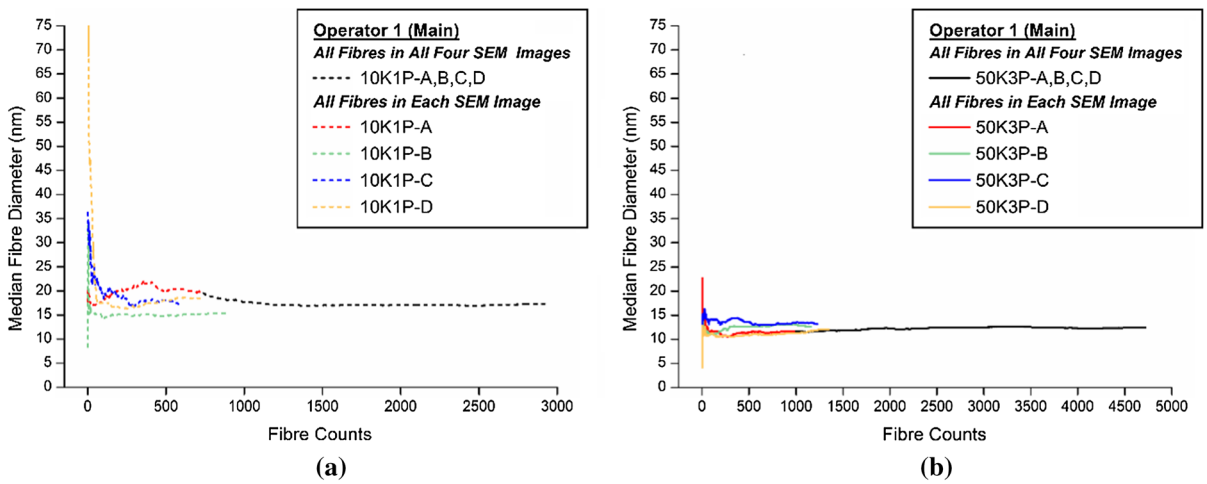


Fig. 8 Median fibre diameter as a function of fibre counts for **a** 10K1P fibres and **b** 50K3P fibres from measuring all fibres in all four SEM images and in each SEM image

50 fibre measurements for all cases as shown in Fig. 9b.

Nonetheless, the method of measuring 100 representative fibres did not produce accurate results as the converged median values were larger than those from the baseline distribution for all images, independent of the operator.

Figure 10a, b depicts the median fibre diameter as a function of fibre counts from measuring all fibres in the top-left quadrant in each image for the 10K1P sample and 50K3P sample, respectively. From Fig. 10a, the median fibre diameters of the coarser 10K1P fibres were shown to converge after about 150 fibre measurements for both operators. The difference between the converged median values of Operator 2 and 3 in Fig. 10a was also smaller compared to those in Fig. 9a.

Interestingly for the 50K3P fibres, both operators obtained very similar results as indicated by the multiple overlapping plots with similar medians in Fig. 10b. Here, the median was also shown to converge after approximately 150 fibre measurements.

The results show that the measurement of approximately 150 fibres in a reduced area of an image would yield representative results, provided all fibres were measured. This method was noted to be much less tedious to perform compared to measuring all fibres in the image.

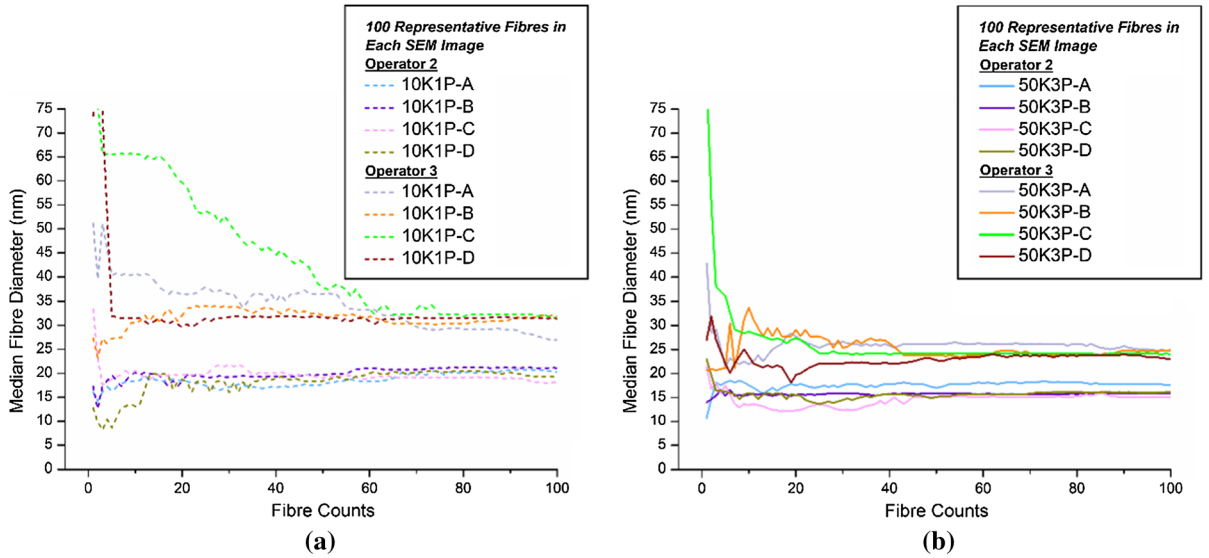


Fig. 9 Median fibre diameter as a function of fibre counts for **a** 10K1P fibres and **b** 50K3P fibres from measuring 100 representative fibres in each SEM image

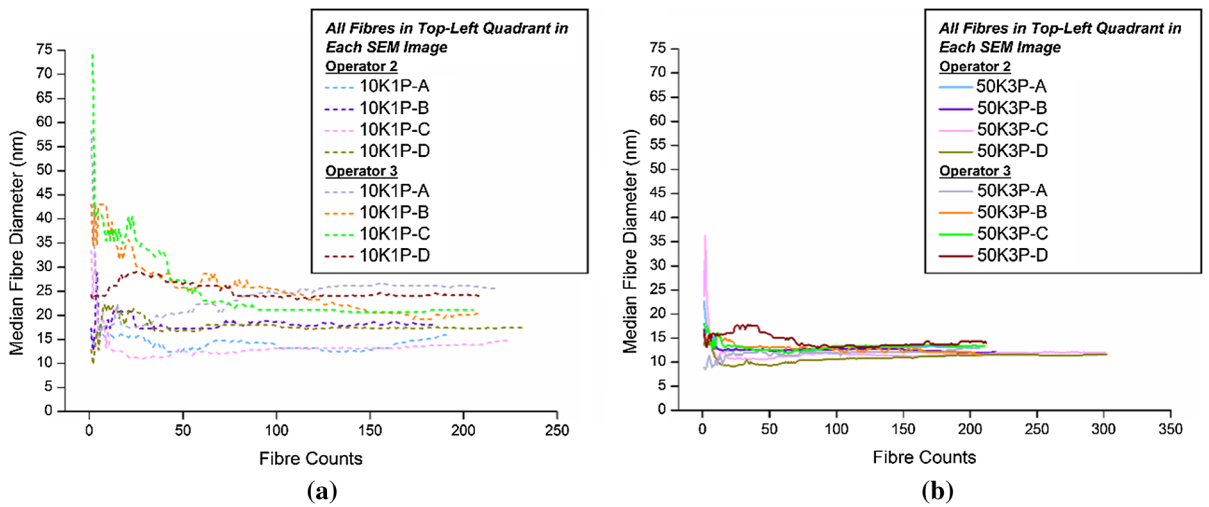


Fig. 10 Median fibre diameter as a function of fibre counts for **a** 10K1P fibres and **b** 50K3P fibres from measuring all fibres in top-left qudrant in each SEM image

Kruskal–Wallis-1-way-ANOVA on ranks H test

The results from the Kruskal–Wallis test for the 10K1P baseline versus the 50K3P baseline are shown in Table 3 (referred to as T1, the overall test). The results from the Kruskal–Wallis tests between the individual images and groups of images from measuring all fibres in an image by Operator 1 for the 10K1P sample (T2–T8) and 50K3P sample (T9–T15) are shown in Tables 4 and 5, respectively.

The results from the Kruskal–Wallis test within and between groups of operators, images and methods for the 10K1P distributions (T16–T31) and the 50K3P distributions (T32–T47) are shown in Tables 6 and 7, respectively. As previously mentioned in the "Methods" section, a H-statistic value greater than the χ^2 value means that the groups do not have similar distributions while a H-statistic value less than the χ^2 value means that the groups have similar distributions.

Table 3 Kruskal–Wallis test results of the 10K1P baseline distribution versus the 50K3P baseline distribution

Test	Description of Kruskal–Wallis test (overall)	N (no. of observations across all groups)	H-statistic	χ^2 (α = 0.01, d.f. = 1)
T1	10K1P baseline versus 50K3P baseline	7669	614	6.63

Table 4 Kruskal–Wallis test results of between individual and groups of images for Operator 1 for the 10K1P distributions

Test	Description of Kruskal–Wallis test (10K1P)	N (no. of observations across all groups)	H-statistic	χ^2 (α = 0.01)
Operator 1 (OP1): all fibres in an SEM image (d.f. = 3)				
T2	OP1 all fibres in image A versus image B versus image C versus image D	2946	84.4	11.3
Operator 1 (OP1): all fibres in an SEM image (d.f. = 1)				
T3	OP1 all fibres in images ABC versus images ABD	4555	3.47	6.63
T4	OP1 all fibres in images ABC versus images ACD	4272	19.5	6.63
T5	OP1 all fibres in images ABC versus images BCD	4439	0.73	6.63
T6	OP1 all fibres in images ABD versus images ACD	4399	6.99	6.63
T7	OP1 all fibres in images ABC versus images BCD	4566	7.34	6.63
T8	OP1 all fibres in images ACD versus images BCD	4283	26.9	6.63

Table 5 Kruskal–Wallis test results of between individual and groups of images for Operator 1 for the 50K3P distributions

Test	Description of Kruskal–Wallis test (50K3P)	N (no. of observations across all groups)	H-statistic	χ^2 (α = 0.01)
Operator 1 (OP1): all fibres in an SEM image (d.f. = 3)				
T9	OP1 all fibres in image A versus. image B versus image C versus image D	4723	34.5	11.3
Operator 1 (OP1): all fibres in an SEM image (d.f. = 1)				
T10	OP1 all fibres in images ABC versus images ABD	6865	5.51	6.63
T11	OP1 all fibres in images ABC versus images ACD	6933	0.87	6.63
T12	OP1 all fibres in images ABC versus images BCD	7113	0.59	6.63
T13	OP1 all fibres in images ABD versus images ACD	7056	2.07	6.63
T14	OP1 all fibres in images ABC versus images BCD	7236	9.72	6.63
T15	OP1 all fibres in images ACD versus images BCD	7304	2.89	6.63

The overall test for the 10K1P baseline distribution versus the 50K3P baseline distribution in Table 3 resulted in a H-statistic value of 614. This H-statistic value is significantly larger than the χ^2 value of 6.63. This establishes the overall basis that the 10K1P and 50K3P samples have statistically different distributions. These results are consistent with the more heterogeneous nature of the 10K1P sample and the more homogeneous nature of the 50K3P sample as shown in Figs. 2 and 3, respectively.

The test results from comparing images A versus B versus C and versus D for the 10K1P sample (T2) in Table 4 resulted in a H-statistic value of 84.4 which is significantly larger than the χ^2 value of 11.3. A similar

Table 6 Kruskal–Wallis test results within and between groups of operators, images and methods for the 10K1P distributions

Test	Description of Kruskal–Wallis test (10K1P)	N (no. of observations across all groups)	H-statistic	χ^2 (α = 0.01, d.f. = 1)
Operator 2 (OP2): 100 representative fibres in an SEM image				
T16	OP2 100 rep. fibres in image A versus 10K1P baseline	3046	7.66	6.63
T17	OP2 100 Rep. fibres in image B versus 10K1P baseline	3046	7.06	6.63
T18	OP2 100 rep. fibres in image C versus 10K1P baseline	3046	0.92	6.63
T19	OP2 100 rep. fibres in image D versus 10K1P baseline	3046	0.22	6.63
Operator 3 (OP3): 100 representative fibres in an SEM image				
T20	OP3 100 rep. fibres in image A versus 10K1P baseline	3046	25.8	6.63
T21	OP3 100 rep. fibres in image B versus 10K1P baseline	3046	91.3	6.63
T22	OP3 100 rep. fibres in image C versus 10K1P baseline	3046	79.8	6.63
T23	OP3 100 rep. fibres in image D versus 10K1P baseline	3046	82.8	6.63
Operator 2 (OP2): all fibres in top-left quadrant (reduced area) in each SEM image				
T24	OP2 all fibres in reduced area image A versus 10K1P baseline	3139	4.15	6.63
T25	OP2 all fibres in reduced area image B versus 10K1P baseline	3132	0.36	6.63
T26	OP2 all fibres in reduced area image C versus 10K1P baseline	3170	22.8	6.63
T27	OP2 all fibres in reduced area image D versus 10K1P baseline	3177	0.22	6.63
Operator 3 (OP3): all fibres in top-left quadrant (reduced area) in each SEM image				
T28	OP3 all fibres in reduced area image A versus 10K1P baseline	3165	44.6	6.63
T29	OP3 all fibres in reduced area image B versus 10K1P baseline	3155	9.52	6.63
T30	OP3 all fibres in reduced area image C versus 10K1P baseline	3151	13.4	6.63
T31	OP3 all fibres in reduced area image D versus 10K1P baseline	3154	34.9	6.63

observation was also made in the case of comparing images A versus B versus C versus D for the 50K3P sample (T9) in Table 5. Here, a H-statistic value of 34.5 was obtained which is much larger than the χ^2 value of 11.3. In both 10K1P and 50K3P cases, the results indicate that the individual images (A, B, C and D) did not have similar distributions to one another. This is an interesting observation given the relatively similar medians obtained for the 10K1P sample (20 nm, 15 nm, 17 nm and 18 nm) in Table 1 and the 50K3P sample (12 nm, 13 nm, 13 nm and 12 nm) in Table 2, respectively for Operator 1 by counting all fibres in an image. These results support the notion that measuring fibres from more than one image is necessary to adequately describe the true population distribution of a sample.

The test results from comparing distributions of groups of images against other groups of images for the 10K1P sample by Operator 1 in Table 4 resulted in H-statistic values of 3.5, 19, 0.7, 6.9, 7.3 and 26 for tests T3–T8, respectively. Here, two tests had statistically similar distributions (T3 and T5), two tests were very close to being statistically similar (T6 and T7) with H-statistic values of 6.9 and 7.3 slightly larger than the χ^2 value of 6.63, and 2 tests did not have statistically similar distributions (T4 and T8). The test results from comparing distributions of groups of images against other groups of images for the 50K3P sample by Operator 1 in Table 5 resulted in H-statistic values of 5.5, 0.9, 0.6, 2.1, 9.7 and 2.9 for tests T10–T15, respectively. Here, five out of the six tests resulted in statistically similar distributions (T10, T11, T12, T13 and T15) with one test very close to being statistically similar (T14) with a H-statistic value of 9.7 slightly larger than the χ^2 value of 6.63.

Overall, the combination of images resulted in a large majority of tests with statistically similar distributions. This as opposed to considering the distributions in each image separately (T2 and T9) which was shown to have different distributions, across all 4 individual images for both 10K1P and 50K3P samples. These results indicate that by combining the distributions of all the images together, more statistically similar and representative results could be

Table 7 Kruskal–Wallis test results within and between groups of operators, images and methods for the 50K3P distributions

Test	Description of Kruskal–Wallis test (50K3P)	N (no. of observations across all groups)	H-statistic	χ^2 (α = 0.01, d.f. = 1)
Operator 2 (OP2): 100 representative fibres in an SEM image				
T32	OP2 100 rep. fibres in image A versus 50K3P baseline	4823	48.1	6.63
T33	OP2 100 rep. fibres in image B versus 50K3P baseline	4823	27.5	6.63
T34	OP2 100 rep. fibres in image C versus 50K3P baseline	4823	11.0	6.63
T35	OP2 100 rep. fibres in image D versus 50K3P baseline	4823	22.2	6.63
Operator 3 (OP3): 100 representative fibres in an SEM image				
T36	OP3 100 rep. fibres in image A versus 50K3P baseline	4823	135	6.63
T37	OP3 100 rep. fibres in image B versus 50K3P baseline	4823	122	6.63
T38	OP3 100 rep. fibres in image C versus 50K3P baseline	4823	127	6.63
T39	OP3 100 rep. fibres in image D versus 50K3P baseline	4823	107	6.63
Operator 2 (OP2): all fibres in top-left quadrant (reduced area) in each SEM image				
T40	OP2 all fibres in reduced area image A versus 50K3P baseline	4930	2.92	6.63
T41	OP2 all fibres in reduced area image B versus 50K3P baseline	4943	2.76	6.63
T42	OP2 all fibres in reduced area image C versus 50K3P baseline	5024	1.44	6.63
T43	OP2 all fibres in reduced area image D versus 50K3P baseline	5025	2.11	6.63
Operator 3 (OP3): all fibres in top-left quadrant (reduced area) in each SEM image				
T44	OP3 all fibres in reduced area image A versus 50K3P baseline	4930	0.01	6.63
T45	OP3 all fibres in reduced area image B versus 50K3P baseline	4932	4.50	6.63
T46	OP3 all fibres in reduced area image C versus 50K3P baseline	4934	0.72	6.63
T47	OP3 all fibres in reduced area image D versus 50K3P baseline	4935	7.12	6.63

obtained to adequately describe the entire distribution of a sample population as a whole. As such, all testing of Operator 2 and Operator 3 results were compared against the baseline distributions (combining all images together) for the 10K1P and 50K3P sample and is shown in Tables 6 and 7, respectively.

The test results from measuring 100 representative fibres in an SEM image versus the 10K1P baseline distributions in Table 6 resulted in H-statistic values of 7.6, 7.1, 0.9 and 0.2 for Operator 2 (T16–T19) and 26, 91, 80 and 83 for Operator 3 (T20–T23) for Images A, B, C and D, respectively. The results here indicate that Operator 2 had two statistically similar distributions to the baseline distribution (T18 and T19) while Operator 3 had no statistically similar distributions to the baseline distribution. The results from measuring all fibres in a reduced area in an SEM image versus the 10K1P baseline distributions in Table 6 resulted in H-statistic values of 4.2, 0.4, 23 and 0.2 for Operator 2 (T24–T27) and 45, 10, 13 and 35 for Operator 3 (T28–T31) for Images A, B, C and D, respectively. The results here indicate that Operator 2 had three statistically similar distributions to the baseline distribution (T24, T25 and T27) while Operator 3 had no statistically similar distributions to the baseline distribution.

Overall, the method of measuring all fibres in a reduced area resulted in improvements compared to the method of measuring 100 representative fibres. Operator 2 had three statistically similar distributions to the baseline using the reduced area method compared to two statistically similar distributions to the baseline using the 100 representative fibres method. In the case of Operator 3, both methods resulted in statistically different distributions to the baseline distribution across all images. However, it is worth noting that the average H-statistic values for the reduced area method at 25.6 were lower and an improvement compared to the 100 representative fibres method at 69.9.

The test results from measuring 100 representative fibres in an SEM image versus the 50K3P baseline distributions in Table 7 resulted in H-statistic values of 48, 28, 11 and 22 for Operator 2 (T32–T35) and 135, 122, 127 and 107 for Operator 3 (T36–T39) for Images A, B, C and D, respectively. The results here indicate

that the method of measuring 100 representative fibres resulted in distributions which were statistically very different to the baseline distribution, regardless of the operator or image for a homogeneous CNF sample. These results are consistent with previous observations in "Distribution from measuring 100 representative fibres in each SEM image" section which note that when operators were tasked to measure 100 representative fibres, they consistently exhibited bias in measuring the larger fibres. The results from measuring all fibres in a reduced area in an SEM image versus the 50K3P baseline distributions in Table 7 resulted in H-statistic values of 2.9, 2.7, 1.4 and 2.1 for Operator 2 (T40–T43) and 0.01, 4.5, 0.7 and 7.1 for Operator 3 (T44–T47) for Images A, B, C and D, respectively. The results here indicate that all four of Operator 2's distributions (T40, T41, T42 and T43) were statistically similar to the baseline distribution. Also, three of Operator 3's distributions (T44, T45 and T46) were statistically similar to the baseline distribution, with one distribution (T47) coming very close with a H-value of 7.12, compared to the χ^2 value of 6.63.

Overall, the method of measuring all fibres in a reduced area resulted in significant improvements to obtain a representative distribution of the 50K3P sample compared to the method of measuring 100 representative fibres.

In summary, the technique of measuring all fibres in a reduced area was shown to yield more consistent and representative results with the baseline distribution for both 10K1P and 50K3P cases. With the more heterogeneous 10K1P sample, Operator 2 had a higher number of distributions similar to the baseline using the method of measuring all fibres in a reduced area. Operator 3 had lower H-statistic values closer to the χ^2 value using the reduced area method. The results here form the foundation of future accurate measurement techniques for heterogeneous CNF samples. In the case of the more homogeneous 50K3P sample, the method of measuring 100 representative fibres consistently resulted in bias towards counting larger fibres for both operators, which renders the results inaccurate. The method of measuring all fibres in a reduced area was shown to be a significant improvement for the 50K3P sample. All four images for Operator 2 were statistically similar to the baseline distribution using the reduced area method. Three out of the four images for Operator 3 were statistically similar to the baseline distribution using the reduced area method, with the last image for Operator 3 having a H-statistic value only slightly larger than the χ^2 value. The method of measuring all fibres in a reduced area was shown to be very promising for the measurement of CNF diameter distributions and is worth researching further.

Conclusion

The measurement of CNF diameters from microscopy images is an arduous task due to the high number of fibre counts and complex entanglement of fibre networks. In this study, the effect of measuring all fibres in an image, 100 representative fibres in an image and all fibres in a reduced area of an image on the diameter distribution were investigated for two samples. One sample consisted of coarser more heterogeneous fibres while the other consisted of finer more homogeneous fibres. The results showed that when two independent operators were tasked with measuring 100 representative fibres, there was a bias in the selection towards larger fibres, which led to inaccurate results. An alternative method of measuring all fibres in a reduced area of an image provided a more representative distribution. This method resulted in a maximum error of 8 nm for the more heterogeneous sample and 1.5 nm for the more homogeneous sample, which was much lower compared to the 100 representative fibre approach.

Statistical analysis also showed that measuring about 150 fibres in a reduced image area was sufficient to result in a converged median value, provided all the fibres were unbiasedly measured. This converged median value was also very close to the baseline median obtained from measuring all fibres in multiple images. Kruskal–Wallis statistical testing also showed that the technique of measuring all fibres in a reduced area of an SEM image yielded more representative results to the baseline distribution compared to the technique of measuring 100 representative fibres.

The results here introduce a less time-consuming method for the measurement of CNF diameter distribution and is worth researching further. While the methods in this study were performed on SEM images, future work would be to investigate if this method would translate well to TEM images which also struggle from similar issues of having a large number of fibrillated and entangled fibres.

Acknowledgments The authors wish to thank Australian Paper Maryvale for their ongoing support with the project. The authors wish to acknowledge the Monash Centre of Electron Microscopy (MCEM) for use of their SEM facilities. Shaun Ang would like to thank the Australian Government for a Research Training Program scholarship.

References

Abe K, Yano H (2009) Comparison of the characteristics of cellulose microfibril aggregates of wood, rice straw and potato tuber. Cellulose 16:1017–1023. https://doi.org/10.1007/s10570-009-9334-9

Alemdar A, Sain M (2008) Biocomposites from wheat straw nanofibers: Morphology, thermal and mechanical properties. Compos Sci Technol 68:557–565. https://doi.org/10.1016/j.compscitech.2007.05.044

Ang S, Haritos V, Batchelor W (2019) Effect of refining and homogenization on nanocellulose fiber development, sheet strength and energy consumption. Cellulose 26:4767–4786. https://doi.org/10.1007/s10570-019-02400-5

Baati R, Mabrouk AB, Magnin A, Boufi S (2018) CNFs from twin screw extrusion and high pressure homogenization: a comparative study. Carbohydr Polym 195:321–328. https://doi.org/10.1016/j.carbpol.2018.04.104

Bhatnagar A, Sain M (2005) Processing of cellulose nanofiber-reinforced composites. J Reinf Plast Compos 24:1259–1268. https://doi.org/10.1177/0731684405049864

Chaker A, Alila S, Mutjé P, Vilar MR, Boufi S (2013) Key role of the hemicellulose content and the cell morphology on the nanofibrillation effectiveness of cellulose pulps. Cellulose 20:2863–2875. https://doi.org/10.1007/s10570-013-0036-y

Chen JH, Liu JG, Su YQ, Xu ZH, Li MC, Ying RF, Wu JQ (2019) Preparation and properties of microfibrillated cellulose with different carboxyethyl content. Carbohydr Polym 206:616–624. https://doi.org/10.1016/j.carbpol.2018.11.024

Delgado-Aguilar M, González I, Tarrés Q, Pèlach MT, Alcalà M, Mutjé P (2016) The key role of lignin in the production of low-cost lignocellulosic nanofibres for papermaking applications. Ind Crops Prod 86:295–300. https://doi.org/10.1016/j.indcrop.2016.04.010

Dimic-Misic K, Maloney T, Gane P (2018) Effect of fibril length, aspect ratio and surface charge on ultralow shear-induced structuring in micro and nanofibrillated cellulose aqueous suspensions. Cellulose 25:117–136. https://doi.org/10.1007/s10570-017-1584-3

Ding Q, Zeng J, Wang B, Tang D, Chen K, Gao W (2019) Effect of nanocellulose fiber hornification on water fraction characteristics and hydroxyl accessibility during dehydration. Carbohydr Polym 207:44–51. https://doi.org/10.1016/j.carbpol.2018.11.075

Diop CIK, Tajvidi M, Bilodeau MA, Bousfield DW, Hunt JF (2017) Isolation of lignocellulose nanofibrils (LCNF) and application as adhesive replacement in wood composites: example of fiberboard. Cellulose 24:3037–3050. https://doi.org/10.1007/s10570-017-1320-z

Errokh A, Magnin A, Putaux JL, Boufi S (2018) Morphology of the nanocellulose produced by periodate oxidation and reductive treatment of cellulose fibers. Cellulose 25:3899–3911. https://doi.org/10.1007/s10570-018-1871-7

Habibi Y, Mahrouz M, Vignon MR (2009) Microfibrillated cellulose from the peel of prickly pear fruits. Food Chem 115:423–429. https://doi.org/10.1016/j.foodchem.2008.12.034

Henriksson M, Berglund LA, Isaksson P, Lindström T, Nishino T (2008) Cellulose nanopaper structures of high toughness. Biomacromolecules 9:1579–1585. https://doi.org/10.1021/bm800038n

Hotaling NA, Bharti K, Kriel H, Simon CG Jr (2015) DiameterJ: a validated open source nanofiber diameter measurement tool. Biomaterials 61:327–338

Isogai A (2013) Wood nanocelluloses: fundamentals and applications as new bio-based nanomaterials. J Wood Sci 59:449–459. https://doi.org/10.1007/s10086-013-1365-z

Iwamoto S, Nakagaito AN, Yano H (2007) Nano-fibrillation of pulp fibers for the processing of transparent nanocomposites. Appl Phys A Mater Sci Process 89:461–466. https://doi.org/10.1007/s00339-007-4175-6

Jiang E et al (2017) Cellulose nanofibers as rheology modifiers and enhancers of carbonization efficiency in polyacrylonitrile. ACS Sustain Chem Eng 5:3296–3304. https://doi.org/10.1021/acssuschemeng.6b03144

Kangas H, Lahtinen P, Sneck A, Saariaho A-M, Laitinen O, Hellen E (2014) Characterization of fibrillated celluloses. A short review and evaluation of characteristics with a combination of methods. Nord Pulp Paper Res J 29:129–143

Karande VS, Bharimalla AK, Hadge GB, Mhaske ST, Vigneshwaran N (2011) Nanofibrillation of cotton fibers by disc refiner and its characterization. Fibers Polym 12:399–404. https://doi.org/10.1007/s12221-011-0399-3

Mandal A, Chakrabarty D (2011) Isolation of nanocellulose from waste sugarcane bagasse (SCB) and its characterization. Carbohydr Polym 86:1291–1299. https://doi.org/10.1016/j.carbpol.2011.06.030

McDonald JH (2009) Handbook of biological statistics, vol 2. Sparky house publishing, Baltimore

Osong SH, Norgren S, Engstrand P (2016) Processing of wood-based microfibrillated cellulose and nanofibrillated cellulose, and applications relating to papermaking: a review. Cellulose 23:93–123. https://doi.org/10.1007/s10570-015-0798-5

Raj P, Mayahi A, Lahtinen P, Varanasi S, Garnier G, Martin D, Batchelor W (2016) Gel point as a measure of cellulose nanofibre quality and feedstock development with mechanical energy. Cellulose 23:3051–3064. https://doi.org/10.1007/s10570-016-1039-2

Santucci BS, Bras J, Belgacem MN, Curvelo AADS, Pimenta MTB (2016) Evaluation of the effects of chemical composition and refining treatments on the properties of nanofibrillated cellulose films from sugarcane bagasse. Ind Crops Prod 91:238–248. https://doi.org/10.1016/j.indcrop.2016.07.017

Shanmugam K, Doosthosseini H, Varanasi S, Garnier G, Batchelor W (2018) Flexible spray coating process for smooth nanocellulose film production. Cellulose

25:1725–1741. https://doi.org/10.1007/s10570-018-1677-7

Siró I, Plackett D, Hedenqvist M, Ankerfors M, Lindström T (2011) Highly transparent films from carboxymethylated microfibrillated cellulose: the effect of multiple homogenization steps on key properties. J Appl Polym Sci 119:2652–2660. https://doi.org/10.1002/app.32831

Spence KL, Venditti RA, Habibi Y, Rojas OJ, Pawlak JJ (2010) The effect of chemical composition on microfibrillar cellulose films from wood pulps: mechanical processing and physical properties. Bioresour Technol 101:5961–5968. https://doi.org/10.1016/j.biortech.2010.02.104

Taheri H, Samyn P (2016) Effect of homogenization (microfluidization) process parameters in mechanical production of micro- and nanofibrillated cellulose on its rheological and morphological properties. Cellulose 23:1221–1238. https://doi.org/10.1007/s10570-016-0866-5

Taniguchi T, Okamura K (1998) New films produced from microfibrillated natural fibres. Polym Int 47:291–294. https://doi.org/10.1002/(SICI)1097-0126(199811)47:3%3c291:AID-PI11%3e3.0.CO;2-1

Varanasi S, He R, Batchelor W (2013) Estimation of cellulose nanofibre aspect ratio from measurements of fibre suspension gel point. Cellulose 20:1885–1896. https://doi.org/10.1007/s10570-013-9972-9

Wang B, Sain M (2007) Isolation of nanofibers from soybean source and their reinforcing capability on synthetic polymers. Compos Sci Technol 67:2521–2527. https://doi.org/10.1016/j.compscitech.2006.12.015

Wang QQ, Zhu JY, Gleisner R, Kuster TA, Baxa U, McNeil SE (2012) Morphological development of cellulose fibrils of a bleached eucalyptus pulp by mechanical fibrillation. Cellulose 19:1631–1643. https://doi.org/10.1007/s10570-012-9745-x

Zhang L, Batchelor W, Varanasi S, Tsuzuki T, Wang X (2012) Effect of cellulose nanofiber dimensions on sheet forming through filtration. Cellulose 19:561–574. https://doi.org/10.1007/s10570-011-9641-9