

CIVIL & ENVIRONMENTAL ENGINEERING | RESEARCH ARTICLE

Evaluation of image texture recognition techniques in application to wastewater coagulation

Nataliia Sivchenko, Knut Kvaal and Harsha Ratnaweera

Cogent Engineering (2016), 3: 1206679



Received: 31 March 2016
Accepted: 23 June 2016
First Published: 29 June 2016

*Corresponding author: Nataliia Sivchenko, Department of Mathematical Sciences and Technology, Norwegian University of Life Sciences, P.O. Box 5003-IMT, Aas 1432, Norway
E-mails: nataliia.sivchenko@gmail.com, nataliia.sivchenko@nmbu.no

Reviewing editor:
Yiu Fai Tsang, The Hong Kong Institute of Education, Hong Kong

Additional information is available at the end of the article

CIVIL & ENVIRONMENTAL ENGINEERING | RESEARCH ARTICLE

Evaluation of image texture recognition techniques in application to wastewater coagulation

Nataliia Sivchenko^{1*}, Knut Kvaal¹ and Harsha Ratnaweera¹

Abstract: Flocs formation and growth are important characteristics in wastewater coagulation process. The shape and size of flocs highly affect further separation processes, therefore resulting treatment efficiency of wastewater after coagulation. Observed images of flocs tend to show strong relations to coagulation parameters: dose and coagulation time. In this article, three texture recognition techniques were evaluated for the ability to mathematically describe the relationship between the images of flocs and coagulant dosages. The easily computable texture analysis methods were found to be potential techniques for the characterization of the particles images. Ten out of eleven co-occurrence matrix-based grey level co-occurrence matrix (GLCM) texture features were found to be significant for the dosage prediction by a principal component regression model with only one principal component. Two features (Inverse difference moment and Variance) were selected for the multiple linear regression model. Test set prediction accuracy varied from 83 to 96% depending on texture analysis method and multivariate model. Best dosage prediction and image classification results were achieved by GLCM and angle measure technique. The results of image texture analysis coupled with multivariate modelling techniques indicate that it is possible to characterize and relate flocs images, captured during coagulation, with different coagulant dosages, as well as predict those dosages.

ABOUT THE AUTHORS

The authors are members of Water, Environment, Sanitation and Health (WESH) research group at the Norwegian University of Life Sciences (NMBU), Department of Mathematical Sciences and Technology (IMT). WESH group focuses on water and wastewater related issues and is heavily involved in teaching and supervision of MSc and PhD programmes in water engineering and technology. The group also has one of the largest externally funded research and educational project portfolios at IMT, and collaborates with partners from EU, North America, Eurasia, Asia and Africa. It also has a number of Research, Development and Innovation projects with Norwegian partners. The main research and development areas of the WESH group are: process control and optimization of coagulation and biological treatment processes; membrane fouling and filtration processes; microbial water quality and risk assessment; decentralized wastewater systems; modelling of sewer systems.

PUBLIC INTEREST STATEMENT

The performance of coagulation process is crucial in wastewater treatment. By observing changes in flocs' aggregation, it is potentially possible to predict treatment efficiency and optimize the coagulation process – on-line dosage control. In this article, we present the new approach in flocs images description. It is based on the characterization of the whole image (texture image analysis) instead of characterizing each floc separately. Texture image analysis methods applied to images of flocs and coupled with multivariate modelling seem to be a potential tool for on-line prediction of the optimal coagulant dosage and resulting treatment efficiency.

Subjects: Image Processing; Intelligent Systems; Systems & Control Engineering; Water Engineering; Water Science

Keywords: coagulation; model wastewater; images of flocs; image texture analysis; multivariate data analysis

1. Introduction

Coagulation is a well known and essential water purification process for drinking water, wastewater, and industrial water treatment. Added to the colloidal suspension, organic or inorganic coagulant destabilizes system which leads to particles' aggregation and subsequent separation. Intelligent dosage control systems are required to optimize and control the removal of particles and phosphates, minimize the use of a coagulant, achieve better sludge management (by reduction of the sludge volumes) and enhance the plant availability of phosphates. Nowadays optimization and process control of coagulation are mainly done by simple flow-proportional dosing concept or based on data from jar tests (Ratnaweera & Fettig, 2015). Nevertheless, with the recent development of a big range of on-line sensors the different coagulation control systems appear (Annadurai, Sung, & Lee, 2004; Juntunen, Liukkonen, Lehtola, & Hiltunen, 2013; Ratnaweera, Lei, & Lindholm, 2002). However, the initial costs of such control systems are usually quite high. Furthermore, the control systems based on outlet water parameters have a perceptible time lag between dosing point and effluent, varying from 30 min to several hours depending on the separation method. This peculiarity makes such systems inappropriate for wastewater coagulation dosing control, where influent water parameters and its flow rate changes dramatically within a short time. Hence, water and wastewater treatment industry are seeking for simple, cheap, robust and precise real-time dosing control systems.

It has been relatively long since researchers started to model the mechanisms of flocculation (Vold, 1963). The developments in this area are summarized in the review by Thomas, Judd, and Fawcett (1999). Visualization, modelling, and understanding of flocs formation mechanisms reveals the ability to determine relationships between particles' properties and coagulation efficiency thus optimize the coagulation process. Many researchers have made their contribution to the understanding of the particles aggregation, breakage and regrowth mechanisms. Jarvis, Jefferson, Gregory, and Parsons (2005) summarized the development of floc strength and breakage mechanisms in their review. Influence of coagulation mixing conditions (e.g. share rates, impeller construction) into flocs formation, breakage and regrowth were also studied (Cao et al., 2011; He, Nan, Li, & S., 2012). However, due to complicated nature of flocs and the huge number of parameters influencing the coagulation process there is still no comprehensive or universally accepted mathematical description of floc formation mechanisms.

The modelling of drinking water treatment process, which included coagulation stage, had been recently performed by Juntunen, Liukkonen, Pelo, Lehtola, and Hiltunen (2012). The authors investigated both laboratory and process data. They found aluminium dose to be one of the most important factors affecting the treated water quality parameters (turbidity and residual aluminium).

Recently, computer vision image processing techniques started to be employed more frequently in order to study and characterize complex size, shape and features of particles. Photographic image acquisition techniques make it possible to capture and analyse flocs' images *in situ* (Chakraborti, Gardner, Atkinson, & Van Benschoten, 2003). Our hypothesis is that it should be possible to correlate certain floc properties with coagulant dosages and treatment efficiencies using images of flocs. The distinct effect of coagulant dosage on aggregates' features was detected by Lin, Huang, Chin, and Pan (2008) in kaolin water suspension. Authors used wet scanning electron microscopy in order to obtain morphology of flocs. Jin (2005) used the high-resolution digital camera to study the influence of temperature on flocs properties under different coagulant dosages for river water coagulation. The relation between projected area of particles and coagulant dosages were found. Wang et al. (2011) investigated the changes in flocs' characteristics due to different dosages and coagulation pH

in humic acid suspension, obtaining images by digital CCD camera. Our approach in this article is to prove that images of model wastewater flocs captured *in situ* by the nonintrusive photographic method are descriptive in terms of coagulant dosage prediction. Thus, the technique could potentially be used for on-line coagulation dosage control. Advanced dosage control then will be possible in means of determining lowest coagulant dosage which leads to required solid, liquid and phosphorous separation.

Image analysis of particles is a fairly old (Tambo & Watanabe, 1979b) and known technique to be used in different water treatment applications, for instance: activated sludge processing (Alves et al., 2000; Amaral & Ferreira, 2005; da Motta, Pons, Roche, & Vivier, 2001; Dagot et al., 2001; Gorczyca & Ganczarczyk, 2002; Jenné, Banadda, Smets, Deurinck, & Van Impe, 2007; Jenné et al., 2006; Mesquita, Amaral, & Ferreira, 2011), aggregates settling velocity measurements (Gorczyca & Ganczarczyk, 1996; Vahedi & Gorczyca, 2012, 2014), membrane fouling (Mendret, Guigui, Schmitz, Cabassud, & Duru, 2007), natural organic matter removal (Xiao, Lam, Li, Zhong, & Zhang, 2011). Image analysis applied to coagulation and aggregation of particles was also widely studied in different waters: clay (kaolinite) suspension (He et al., 2012; Lin et al., 2008; Tambo & Hozumi, 1979; Tambo & Watanabe, 1979a; Xiao et al., 2011), mineral suspensions (Gorczyca & Ganczarczyk, 1996), latex particles suspension (Chakraborti et al., 2003), humic acid solution (Wang et al., 2011), water after lime softening (Vahedi & Gorczyca, 2011, 2012, 2014), lake water (Chakraborti, Atkinson, & Van Benschoten, 2000; Kim, Moon, & Lee, 2001), textile wastewater (Yu, Chen, Cheng, & Chu, 2009; Yu, Cheng, & Huang, 2012), sewer system wastewater (Zheng et al., 2011). However, not many studies concerned the floc-dosage relation in the coagulation process regarding wastewater or model wastewater.

Li and Ganczarczyk (1989) described the spatial structure of particles appearing during water and wastewater treatment processes by fractal theory, i.e. fractal dimension. Since then, it has become the most common way of flocs characterization. In this research, we were not much concerned with the study of the particles' characteristics itself, instead, we were testing the different image analysis techniques to correlate images of flocs with coagulant dosages. The technique presented in this article analyses the whole captured image, not its particular parts or regions of interest, and is known as image texture analysis. The advantage of such method is that the flocs do not have to be segmented from the image, which is quite difficult and challenging in some cases, e.g. in a high turbidity waters and with a high concentration of particles when they are overlapping in the image.

Texture analysis is a well-known method of image recognition and especially classification in numerous computer vision applications. However, as far as we know, it has not been previously reported to be used in application to images of coagulated particles. Texture is a loosely defined term without an accepted or universal quantitative meaning. Texture can be defined as a measure of the image's surface roughness in a meaning of brightness, colour, shape and size variations within some region and its repetitiveness. Texture is a pattern which can be completely distinct or completely random. Furthermore, texture could be isotropic (without any preferred orientation) or anisotropic (has definite pattern structure). Images of flocs presented in this study were described by texture recognition techniques.

There have been many attempts in combining the digital image analysis with coagulation/flocculation process. The clear correlations between floc properties and the flocculation process parameters were found by Juntunen, Liukkonen, Lehtola, and Hiltunen (2014). The authors documented that, for instance, the surface area of floc and the number of floc particles highly correlate with the process data, such as lime feed, pH, turbidity etc. However, in contrast to our research, the authors performed the analysis for drinking water treatment plant with the characterization of flocs by object recognition methods in image analysis. In this paper, we evaluate the texture image analysis methods to be used for coagulant dose prediction by the images of flocs. It seems to be a promising tool for further coagulation–flocculation dosage control studies. Potentially, the texture analysis methods in application to particles separation processes may simplify the image characterization part.

2. Materials and methods

2.1. Raw water

Model wastewater was used to perform all the experiments in this research. In order to have a better representation of the typical domestic wastewater characteristics, the synthetic wastewater was prepared according to previous studies (Fettig, Ratnaweera, & Ødegaard, 1990; Ødegaard, Fettig, & Ratnaweera, 1990; Ratnaweera, 1991). The model wastewater contains both the organic and inorganic components. Dried milk, potato starch, and humic acid represent the organic part of the suspension. Different inorganic salts are used to well define amounts of orthophosphate and ammonium. Bentonite is used to imitate the inorganic particles. For this particular research, model wastewater was prepared with a medium concentration level of salts and particles and represents typical soft Norwegian wastewater. Components and their concentrations were as follows: dried milk (Nestle, Norway)—300 mg/l, potato starch (Hoff, Norway)—60 mg/l, bentonite (Alfa Aesar, USA)—80 mg/l, NaCl (Merck, Germany)—400 mg/l, K_2HPO_4 (Merk, Germany)—50 mg/l, NH_4Cl (Kebo, Germany)—100 mg/l, Humic acid sodium salt (Merck, Germany)—5 mg/l, $NaHCO_3$ (Merck, Germany)—60 mg/l. Ås city tap water was used as a solvent, which is typical soft Norwegian lake water with low concentrations of most components. The model wastewater was prepared for each experimental series and used within 1.5–4 h after preparation. Average model wastewater parameters were next: 8.03 ± 0.13 pH, 187 ± 10.7 mg/l suspended solids, 256 ± 12 FNU turbidity, 11.2 ± 0.24 mg/l total Phosphorus, 9.5 ± 0.07 mg/l Orthophosphates. The initial model wastewater particles had size distribution 0.52–150 μm with two pics at 1,88 and 14.5 μm , determined by Malvern Mastersizer 3000 (Malvern Instruments Ltd, UK).

2.2. Coagulation procedures

All investigations were performed in jar-test scale, Flocculator 2000 (Kemira, Sweden) with programmable mixer units and 1 l beakers. The mixing conditions during coagulation were: 1 min rapid mixing (400 revolutions per min—RPM), 10 min slow mixing (30 RPM) and 20 min sedimentation without mixing. Prepolymerized aluminium chloride coagulant PAX XL61 (Kemira, Sweden) was used in this research. All tests were performed with wastewater temperature 16–17 C and at a constant pH of 7.5 adjusted by required acid or base addition. Coagulant doses varied from 0.29 to 1.08 mmol Al/l, 11 dosages in total. After sedimentation stage, approximately 200 ml of treated water samples were taken from 5 cm depth by a peristaltic pump.

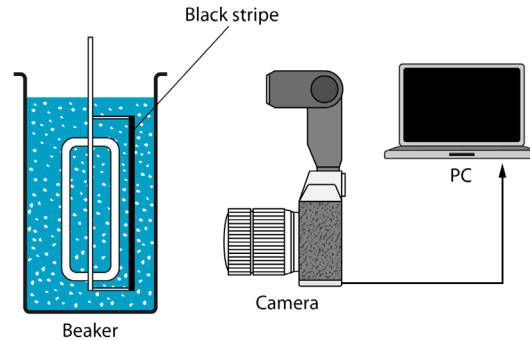
Each sample of treated water taken after sedimentation step was analysed for turbidity, total suspended solids, orthophosphates and total phosphorous concentrations.

2.3. Image acquisition and pre-processing

A nonintrusive optical sampling technique was used to capture images of appearing aggregates. The principal scheme is shown in Figure 1. The images of flocs were obtained during the whole slow mixing period of coagulation with repeatability 1 image per 20 seconds provided by free remote camera control software—DigiCamControl 1.2.0. Image capturing equipment used during the investigations was as follows: digital single lens reflex (DSLR) Nikon D600 camera, 105 mm Nikkor AF-S Micro 1:2.8 G ED lens (Nikkor, China), SpeedLite YN460 flash (Yongnuo, China). The size of the image-capturing zone in the beaker was 4.8×3.3 cm. In order to obtain flocs with the proper depth of field, the mixer unit was modified with an attachment of 4.5 cm width black curved plastic stripe, which also become a background for the images of flocs. The choice of the background colour was based on a fact that the particles, present in model wastewater, are white in colour. Thus, using a contrasting background it is easier to perform the further image analysis.

Images obtained during coagulation have a resolution of 24.3 megapixels each. They were processed in the open source image analysis software *ImageJ* v.1.49 (Rasband, 1997/2016) that bases on plugins and macros. For each image $3,690 \times 3,690$ pixels (3×3 cm) area was cropped by manual investigation of the area. Because of slight changes in lighting conditions during image acquisition

Figure 1. Schematic representation of image acquisition.



for different dosages, all images had been pre-processed in order to have the same brightness intensity. Knut Kvaal wrote *ImageJ* plug-in to mean centre the images' grey-tone values.

In total, coagulation jar tests were performed using 11 coagulant dosages, 10 images from each experiment after 6 min of coagulation were selected as representative samples, resulting in 110 samples. Coagulant dose was set as the response *Y* parameter in multivariate models.

2.4. Fractal dimension of aggregates

Fractal dimension is a very common and widely used parameter to characterize the geometrical features of fractal particles, such as flocs. For a 2-dimensional (2D) projected particle image, the fractal dimension, D_{pf} defines how the projected area of the particles rises with the perimeter (He et al., 2012):

$$A \propto P^{2/D_{pf}} \quad (1)$$

where *A*—projected area; *P*—perimeter of the particles. For the 2D projection of an image, the value of fractal dimension varies from $D_{pf} = 1$ for the circle shape floc to $D_{pf} = 2$ for a chain of particles (a line).

The size of irregular shape particle can be determined as the equivalent diameter, d_p (He et al., 2012):

$$d_p = (4A/\pi)^{1/2}. \quad (2)$$

Flocs' morphological characteristics were obtained using *ImageJ* and plug-in "Analyse particles". First, all images collected during the slow mixing period of coagulation were gathered in a stack. Then 3,690 × 3,690 pixels area was manually estimated and cropped. A stack of images was converted to 8-bit greyscale. Plug-in "Subtract Background" was applied on images in order to equalize background pixel values for easier thresholding. If needed, brightness and contrast were adjusted. Onwards, thresholding was used to obtain binary images, with a threshold value estimated manually or by Otsu method (Otsu, 1979). Flocs within this transformation had greyscale value 0—true black, while the background was set to value 255—true white. Finally, plug-in "Analyse particles" was applied to the binary images of flocs. Many geometrical and statistical parameters might be subtracted from the images of aggregates by this plug-in. For this research, significant parameters of flocs were: number of particles in the image, mean area, and perimeter of flocs. Knowing that 1,230 pixels in the image equal 1 cm, we were able to recalculate flocs' features from pixel values to quantitative values in centimetres.

Mean fractal dimension of flocs was calculated by equation:

$$D_f = 2 \times \log(P_{pix}) / \log(A_{pix}) \quad (3)$$

where A_{pix} —mean area of particles, in pixels; P_{pix} —perimeter, a count of pixel edges; 2 is a constant number (Yu et al., 2009).

2.5. Image analysis by histogram

The histogram is a frequency distribution of the intensity values that occur in an image. The typical 8-bit greyscale image has 256 intensity values (grey shades). Histogram of a greyscale 8-bit image is a one-dimensional (1D) vector of length 256, which contain information about the number of pixels of particular intensity in the image. However, histograms do not contain the spatial information about pixels in the image. The histogram is a fast and simple way to illustrate statistical information of the image—distribution of intensity values, thus, it is a popular tool for real-time image processing. Histogram analysis had also been illustrated as a tool for elimination of poor quality images (Liukkonen, Hiltunen, & Hiltunen, 2015). The information from histograms could be enough for images comparison.

In order to decrease the number of variables (intensity values) the 3 sizes of histogram bins were tried in the further multivariate analysis: 256, 128 and 64 bins. It means that initial histogram was scaled 2 and 4 times using the plug-in in *ImageJ* created by Knut Kvaal “Compute hist stack bins”. The resulting outputs were given as vectors of 256, 128 and 64 intensity values for each image in the stack. Hence, the resulting matrixes were: 110×256 , 110×128 and 110×64 .

2.6. Image analysis by GLCM

GLCM is a typical statistical method of measuring texture. Haralick, Shanmugam, and Dinstein (1973) invented this method, which bases on spatial-dependence GLCM of pixels with estimation of image features using second-order statistics.

ImageJ has a plug-in “GLCM Texture” v.0.4, created by Julio E. Cabrera and further updated to “GLCM Texture Too” v. 0.008 by Toby C. Cornish. Knut Kvaal has done some in-house revision of the plug-in and called the version 0.009, which was used in this analysis. GLCM algorithm was applied considering the distance between the pixel pairs: 1, 2, 3, 5, 10 pixels and the angles were 0° , 45° , 90° and 135° . The resulting output was given as a vector of the next 11 parameters per each image in the stack: angular second moment (ASM), contrast, correlation, inverse difference moment (IDM), entropy, energy, inertia, homogeneity, prominence, variance, and shade. Hence, 20 matrixes were obtained with the size 110×11 each. The detailed description, explanation, and equations for above GLCM texture features are given elsewhere (Conners, Trivedi, & Harlow, 1984; Haralick et al., 1973; Zheng, Sun, & Zheng, 2006).

2.7. Image analysis by AMT

Andrle (1994) introduced angle measure technique (AMT) as a method to characterize the complexity of geomorphic lines. The purpose was to detect changes in coastlines complexity as a function of scale, so firstly it was an approach to analyse 1D data. Later Esbensen, Hjelmén, and Kvaal (1996) have inducted AMT into chemometrics for textural analysis of generic “measurement series” and have shown that this technique is also applicable for the images (2D data). Further successful implementation of AMT (Dahl & Esbensen, 2007; Fongaro & Kvaal, 2013; Fongaro, Lin Ho, Kvaal, Mayer, & Rondinella, 2016; Huang & Esbensen, 2000, 2001; Kucheryavski, 2007; Kvaal, Wold, Indahl, Baardseth, & Næs, 1998) proved that the method could be used for images’ texture description. AMT spectrums can also be used for interpretative purposes. The AMT is implemented as an in-house made *ImageJ* plug-in (Kvaal, 2014).

AMT is a transform-based texture recognition method, which in effect creates a completely new domain – scale domain. The AMT algorithm, regarding image analysis, can be described by five steps.

- (1) Unfold the image (Kvaal et al., 2008; Mortensen & Esbensen, 2005) into a 1D vector of grey level values, also called generic “measurement series”. Row by row or spiral unfold are most commonly used.

Table 1. Results of treated wastewater after sedimentation

Class	Dose (mmol/l)	pH	TSS (mg/l)	Turbidity (FNU)	Ortho-P (mg/l)	Total P (mg/l)	Repl. ¹	Number of images		
								Total	Cal. ²	Val. ³
Raw water		8.03 ± 0.13	187 ± 10.7	256 ± 12.22	9.51 ± 0.07	11.2 ± 0.24	8			
L	0.29	7.53 ± 0.01	26 ± 0.5	14.3 ± 1.90	7.38 ± 0.36	8.84 ± 0.33	2	60	5	5
L	0.3	7.56 ± 0.05	12 ± 1.0	5.98 ± 0.13	7.07 ± 0.08	8.29 ± 0.10	2	96	5	5
L	0.38	7.51 ± 0.02	16 ± 2.0	4.39 ± 0.16	6.29 ± 0.22	6.97 ± 0.27	2	60	5	5
L	0.44	7.54 ± 0.00	12 ± 1.5	2.52 ± 0.28	5.87 ± 0.11	6.12 ± 0.17	2	60	5	5
M	0.51	7.50 ± 0.02	10 ± 1.0	1.42 ± 0.36	4.61 ± 0.09	5.05 ± 0.20	2	54	5	5
M	0.57	7.53 ± 0.01	10 ± 0.5	1.15 ± 0.02	3.71 ± 0.13	3.89 ± 0.18	2	60	5	5
M	0.64	7.52 ± 0.04	8 ± 2.0	1.34 ± 0.18	2.62 ± 0.38	2.76 ± 0.42	2	59	5	5
H	0.7	7.55 ± 0.01	6 ± 2.0	1.48 ± 0.08	1.24 ± 0.19	1.38 ± 0.19	2	60	5	5
H	0.75	7.48 ± 0.01	5 ± 1.0	2.71 ± 0.22	0.63 ± 0.02	0.89 ± 0.09	2	60	5	5
H	0.85	7.52 ± 0.01	7 ± 1.0	2.85 ± 0.58	0.54 ± 0.01	0.65 ± 0.05	2	60	5	5
H	1.08	7.49 ± 0.005	32 ± 2.5	3.86 ± 0.91	0.26 ± 0.06	0.44 ± 0.07	2	60	5	5

Notes: 1—number of dose replicates, 2—number of images in calibration data-set, 3—number of images in validation data-set.

- (2) Generate a set of random initial starting points, the number of these points should be chosen according to the image resolution.
- (3) Develop the AMT complexity measures depending on a current value of the local scale (radius). The average value for a given number of starting points is calculated.
- (4) Increase the local scale by 1 and repeat the last step specified number of times—manually selected value of scale.
- (5) Transform the image into 1D spectrum, plotted as scale (radius) vs. measured mean angle (MA). Thus, the resulting complexity spectrum contains all the angle measures pertaining to all scales of the measurement series (unfolded image).

A detailed description of the AMT algorithm is given elsewhere (Esbensen et al., 1996; Fongaro et al., 2016; Kvaal et al., 2008).

Further, in the analysis the AMT spectrums data were used as the X-input for multivariate statistical analysis. Obtained matrix consisted of 110 samples and 200 variables (scale value).

2.8. Mathematical modelling

Advanced mathematical and statistical tools are often used to explore, evaluate, model, calibrate and predict data from both laboratory-scale and full-scale experiments. Certain process can be optimized with the help of mathematical modelling. Multivariate regression analysis is a powerful statistical tool, which enables the ability to derive relationships between several input parameters and one or several output parameters. It is especially helpful when the initial data are broad, i.e. massive influencing X-parameters is huge and it is not easy to establish the relationships between different parameters by simple linear regression. Furthermore, in many cases, the input parameters are highly correlated.

The Unscrambler® X 10.3 (CAMO Software AS, Norway) was used for the multivariate data analysis of 110 image samples.

Principal component analysis (PCA) was primarily used to explore the data, while principal component regression (PCR) and partial least squares regression (PLSR) were used to model the relations between input X parameters and corresponding Y responses. Models were first validated by

cross-validation and then using a half of the data for calibration and half of the data for test set validation. The number of principal components (PCs)/factors was chosen according to the explained variance. Multiple linear regression (MLR) was used to predict the dosages by some GLCM feature vectors.

Linear discriminant analysis (LDA) with PCs was applied to obtain confusion matrixes of the histogram, GLCM, and AMT spectral data classification. Each data-set was divided into 3 classes according to total *P* efficiency (coded *L*, *M*, and *H*): *L* (low)—coagulant dosages which lead to low (20–50%) treatment efficiency after sedimentation; *M* (medium)—50–85% of total *P* removal; *H* (high)—above 85% of total *P* removal (Table 1). As a result, class *L* consisted of the data for 4 dosages: 0.29, 0.30, 0.38 and 0.44 mmol Al/l with 10 images from each experiment, 40 samples in total. Class *M*—3 dosages: 0.51, 0.57, 0.64 mmol Al/l, 30 samples in total. Class *H*—4 dosages: 0.70, 0.75, 0.85, 1.08 mmol Al/l, 40 samples in total.

3. Results and discussion

3.1. Images of flocs

The presented imaging technique shows the ability to investigate images of flocs during the process of coagulation–flocculation. Capturing images each 20 s gives an opportunity to observe and describe changes in aggregates formation. Figure 2 shows the evolution of flocs during coagulation. Even by human eye, there is a very noticeable difference between the images. At the very first moment after fast mixing, when mixing unit stops to change the rotation speed, first image was taken and it was called “zero seconds” of coagulation. Particles in water are quite small at this moment, initial flocs. However, within the next 20 s these initial particles rapidly form larger aggregates. The growth of flocs is distinctly visible until approximately 3–5 min of coagulation. After that, the flocs visually seem to be quite stable on their geometrical characteristics. However, according to the results of further image analysis, flocs are continuing to grow slightly with time (Table 2).

3.2. Coagulant doses and images of flocs

Figure 3 shows the differences in flocs’ structure due to coagulant dose (given in mmol Al/l). The figure describes relationships between dose and total *P* treatment efficiency. Images on the graph show how flocs looked during coagulation at some given time (400 s after the start of slow mixing). From Figure 3, the increase of coagulant dose results in a decrease of flocs’ size, which is consistent with findings by Wang et al. (2011).

It is noticeable from the images that detection of physical characteristics (thereby fractal dimension) of aggregates in this particular case (image scale) is possible when coagulant dosages are

Figure 2. Eight representative samples of pre-processed images, which illustrate the flocs aggregation process. The number on image shows time of coagulation in seconds. Coagulant dose—0.3 mmol Al/l.

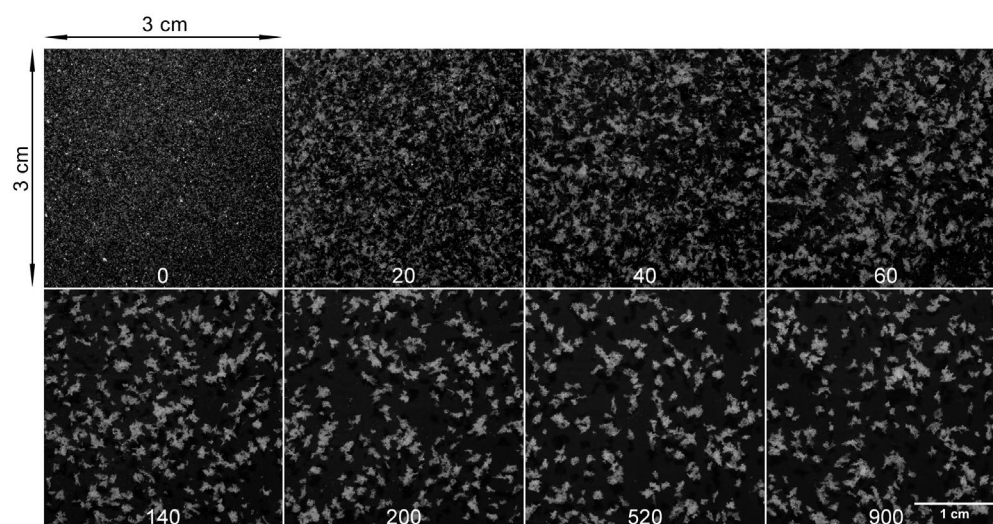


Table 2. Changes in mean fractal dimension and equivalent diameter of flocs with coagulation time

Time (s)	Flocs count	Mean D_f^*	d_p^{**} (mm)
0	18,733	1.87 ± 0.08	0.09 ± 0.03
20	908	1.65 ± 0.07	0.29 ± 0.20
40	587	1.57 ± 0.07	0.36 ± 0.30
60	412	1.55 ± 0.07	0.47 ± 0.39
100	195	1.53 ± 0.07	0.65 ± 0.57
200	186	1.50 ± 0.07	0.72 ± 0.48
400	144	1.48 ± 0.06	0.85 ± 0.54
1,180	127	1.46 ± 0.05	0.90 ± 0.49

*Fractal dimension.

**Mean equivalent diameter.

quite low, but inapplicable for dosages with adequate treatment efficiencies. Closer to optimal coagulant dosages, flocs appear to be smaller, hence, they are overlapping in the image and accurate floc feature detection becomes inapplicable in this scale.

The results of coagulation are presented in Table 1.

3.3. Fractal dimension

Flocs' features as characteristics of particles were detected from the images of flocs for low coagulant dosages.

The results of one time series of experiments, dose—0.29 mmol Al/l, are shown in Table 2. Expectedly, the size (equivalent diameter) of particles is growing with coagulation time. The mean fractal dimension of initial particles is close to 2. As particles aggregate, fractal dimension rapidly decreases during the first minute of coagulation and then continues slowly decreasing. These results are consistent with findings of Chakraborti et al. (2003). Figure 4 shows the examples of 4 flocs (from 23,000 particles analysed for one dose) with different fractal dimensions.

Figure 3. Dependence of total P removal from coagulant dose and its associated image of flocs during coagulation.

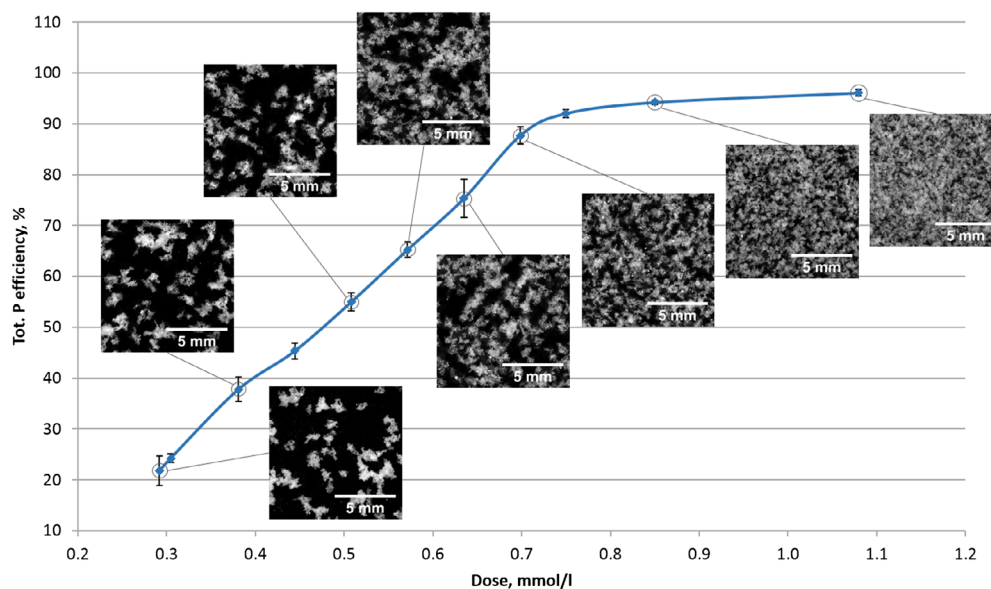
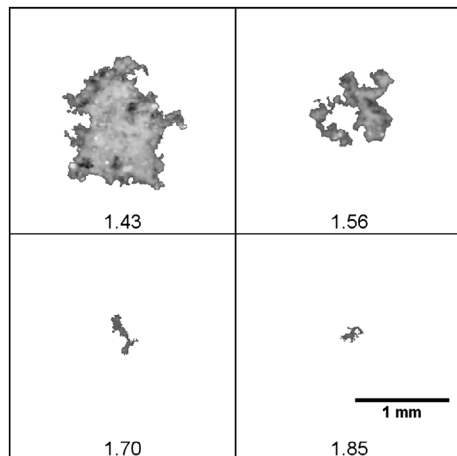


Figure 4. Examples of flocs detected for the coagulant dose 0.29 mmol Al/l, the number on the image equals fractal dimension.



3.4. Histogram analysis

Histograms provide the grey level distribution of pixels in the image. Images of flocs for each coagulant dose had their own typical histograms, so it was possible to distinguish which image's histogram corresponds to which coagulant dose. The results of classification are shown in Table 3.

PCA showed a poor explanation of X variance of 256 grey level histogram data. With 2 PCs 21.59% of calibration X variance was explained and 18.11% of cross-validation. Each next PC increased the explained variance by 1–3%.

The number of grey level values on histogram was decreased by 2 and 4 times, resulting in a better prediction of coagulant dosage. Thus, regression analysis of 64 bins histograms gave the best PCR and PLSR models, listed in Table 4.

3.5. GLCM analysis

PCR analysis was applied to all 20 GLCMs (5 different pixel steps and 4 directions) to find the best combination of pixel step and direction. X was set as 11 GLCM variables for each image, Y —coagulant dose. However, it was found that neither differences in pixel step nor direction give a significant change in coagulant dosage prediction by PCR. Cross-validation R^2 fluctuated between 95 and 96%, meaning that images of flocs are invariant and independent of the difference in pixel step. Hence, all further analyses were done for the GLCM obtained with the pixel step 1 and direction 0°.

The results of PCR model are shown in Figure 5. Coagulant dosages were quite well separated in the score plot with the first two PCs, where the total explained X variance equals 87.69% (PC1 = 73.82%, PC2 = 13.87%) for calibration and 84.17% (PC1 = 70.33%, PC2 = 13.84%) for cross-validation. Validation method was based on systematic block cross-validation, which created 11 submodels where 5 samples (replicates of one dose) had been left out in each of the submodels. The total explained Y variance hence, the accuracy of prediction equals 95.50% (calibration) and 94.98% (cross-validation) with only one PC.

All variables in the correlation loadings plot (Figure 5) fall within the circles, which depict 50–100% of explained variance. The uncertainties in various model parameters were estimated by jack-knifing for the model with 1 PC. The variables found significant are marked with circles in the correlation loadings plot and with stripes pattern in the regression coefficients plot. Ten out of eleven X variables are found to be significant at the 95% confidence interval. Variable “Entropy” has uncertainty limits crossing the zero line, so it is not significant at the 5% level for the model with 1 PC.

For the real on-line applications, it is preferred to have a linear regression model as it is simpler to implement. The selection of variables and estimation of a relevant number of model inputs are important parts of the data analysis. In order to perform the correct variable selection, *R console* with the “NMBU” plug-in (Liland & Sæbø, 2016) was used. Two results of the best subsets for a different amount of variables are shown in Appendix A. For instance, the detected best model with two variables “IDM” and “Variance” has the prediction $R^2 = 0.967$ and $R^2_{adj} = 0.965$. Since R^2 and R^2_{adj} are close to 1, the regression model shows the high agreement with the experimental results. With the increase in number of variables, R^2 and R^2_{adj} tend to slightly increase. However, the loadings plot (Figure 5) shows that X variables are highly correlated and it is worth to keep as low amount of variables as possible. In this particular case, we would suggest choosing a model with two variables. Nevertheless, the best way of choosing the correct number of variables would be to make prediction analysis with the completely new test data-set.

The linear equation for the model with two variables is:

$$Dose = \beta_0 + \beta_1 \times IDM + \beta_2 \times Variance + \epsilon, \tag{4}$$

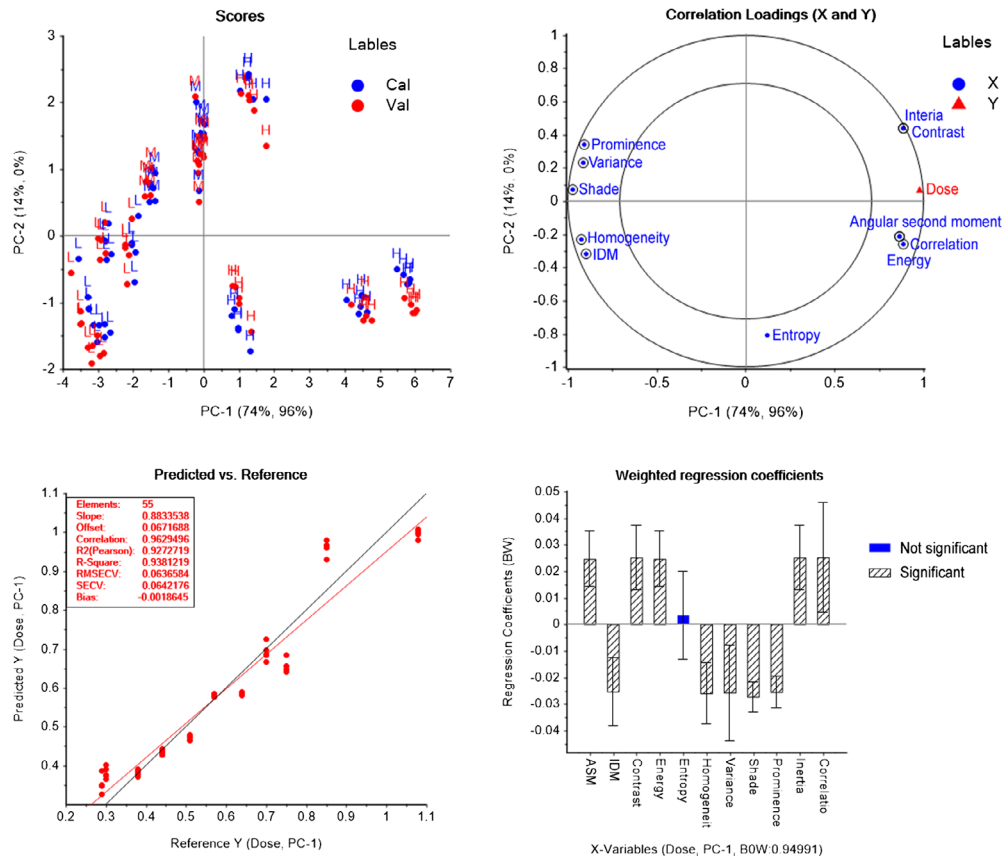
Table 3. Confusion matrixes of PCA-LDA for three image analysis methods

Predicted class	Actual class		
	L	M	H
<i>(a) Histogram, 256 bins, 3 PCs</i>			
L	40	1	0
M	0	29	8
H	0	0	32
<i>(b) GLCM, 11 variables, 3PCs</i>			
L	40	3	0
M	0	27	8
H	0	0	32
<i>(c) AMT, 200 variables, 2 PCs</i>			
L	40	4	0
M	0	26	0
H	0	0	40

Table 4. Results of multiple regression analysis for 3 different image analysis techniques

Regression method	PCR		PLSR	
Number of components/factors	PC-2		Factor-2	
Explained Y variance	R^2 cal. (%)	R^2 val. (%)	R^2 cal. (%)	R^2 val. (%)
Histogram, 256 bins	87.39	83.53	96.35	88.84
Histogram, 128 bins	90.28	88.53	95.07	93.03
Histogram, 64 bins	91.55	89.92	95.51	93.80
GLCM, 4 variables	87.54	87.95	92.89	92.42
GLCM, 6 variables	92.76	92.83	95.58	95.22
GLCM, 11 variables	95.91	95.59	96.89	96.09
AMT, 200 variables	95.84	95.01	96.90	96.19

Figure 5. Results of PCR of 11 GLCM textural features obtained from the images of flocs for different coagulant dosages.



where β_0 —intercept, β_1, β_2 —coefficients which determine the contribution of each independent variable to the estimation of the dependent variable (dose), ε random error. The coefficients were found by MLR with $R = 0.983$, $R^2(\text{cal}) = 0.967$ and $R^2(\text{cross-val}) = 0.947$, $p < 0.001$ and the final model equation was as follow:

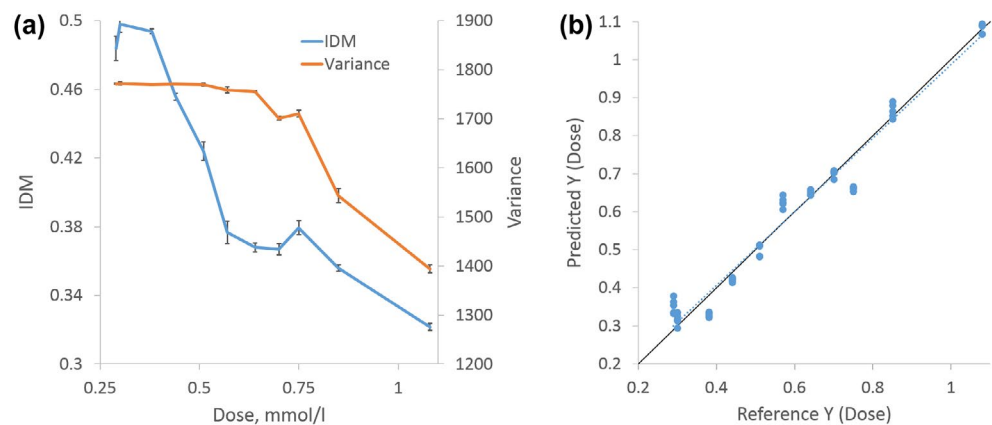
$$\widehat{\text{Dose}} = 3.09 - 2.44 \times \text{IDM} - 9 \times 10^{-4} \times \text{Variance}, \tag{5}$$

It should be noted that the big difference in the values of parameters $\hat{\beta}_1$ and $\hat{\beta}_2$ is caused by the difference in scale of variables “IDM” and “Variance” (Figure 6(a)). The comparison and equal significance of both variables can be seen from the plot “weighted regression coefficients” (Figure 5).

It is not easy to interpret the physical sense of this linear equation because the GLCM feature vectors are the second-order statistical parameters of the original images. When $\text{IDM} = \text{Variance} = 0$, the Dose is 3.09 mmol Al/l. However, the value $\hat{\beta}_0$ does not have a practical interpretation since under the normal conditions images of flocs have grey-tone pixel variations, thus, IDM and Variance differ from 0. When the IDM increases by 1, while Variance remains constant, the estimated mean change of Dose decreases by 2.44. When the Variance increases by 1,000, while IDM remains constant, the estimated mean change of Dose decreases by 0.9. The typical images with high IDM and Variance refer to low dosages (Figure 6(a)), which can be visually seen from Figure 3.

The results of MLR prediction are shown in Figure 6(b). The total explained Y variance equals 96.66% (calibration) and 94.74% (cross-validation) with 2 GLCM feature vectors—“IDM” and “Variance”. It should be emphasized, that this research aims at showing that the texture analysis methods could be applied for determining the coagulant dose. The good repeatability of the results from the images of dynamic coagulation process should also be noted. Despite the high prediction

Figure 6. (a) mean values of variables IDM and Variance vs. dose; (b) scatter plot of predicted vs. observed coagulant dosages for MLR model with two GLCM feature vectors.



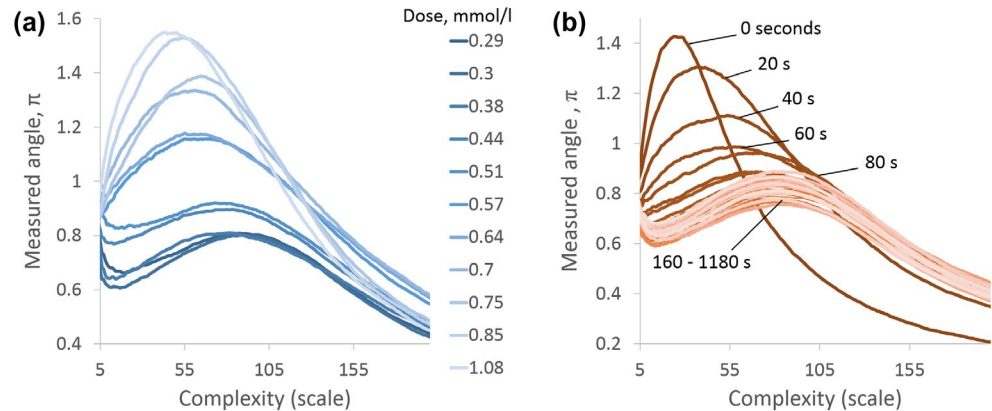
results by PCR and MLR, the models have to be further verified and tested on the bigger data-sets, best—on the data from a wastewater treatment plant.

3.6. AMT

AMT is a potential tool for observing and describing images of flocs both for high and low coagulant dosages. It describes the texture complexity of an image, converting information of grey level values to descriptive spectrums. The spectrum shows the dependence of the MA (calculated for given number of random sample points, in our research it was set to 10,000 samples) in an unfolded image from the scale (radius). Figure 7(a) shows the AMT spectra (complexity) results of flocs images captured at the same coagulation time (400 s) for eleven coagulant dosages from 0.29 to 1.08 mmol Al/l. In a range of low coagulant dosages, the flocs tend to be bigger and detached from each other (see example images in Figure 3). AMT spectrums are not always easy to interpret. However, in application to coagulation, there are visible trends in images' complexities. For the flocs images related to low coagulant dosages the picks of AMT spectrums are lying within larger scale values and have lower measured angles (Figure 7(a)). In contrast, images of flocs associated with higher coagulant dosages have AMT spectrums picks lying in lower scale range and having higher measured angles. This trend—spectrums' vertexes shift to the left, will be easier to explain if we look into the graphs of images' unfold. Figure 8(a) illustrates the unfolding of the first row of an image related to low coagulant dose (0.29 mmol/l). The neighbourhood pixel grey level variations are not dramatic and the pictures of grey levels (which associate with the flocs on the image) are not very frequent. The biggest grey level pixel variations, thus high measured angles, between the neighbouring areas within the whole image occurs in the distance (scale) of approximately 90 pixels, which is visible from Figure 7(a). In contrast, the unfolding of the image associated with higher coagulant dose shows huge and frequent neighbourhood grey-tone variations (Figure 8(b)). Thus, the biggest grey level changes on the image occur in a distance (scale) of approximately 50 pixels, resulting in highest measured angles.

Figure 7(b) shows the AMT spectra results for a time series of flocs images for one certain coagulant dose—0.3 mmol Al/l. For the first seconds of coagulation, AMT spectrums of flocs images are quite sharp, with the maximum angle values within low scales. This can be explained by significant grey level changes in neighbourhood pixel areas in the images due to the high amount of small particles. With time, while aggregates are merging and growing the differences in grey-tone pixel values of close areas are decreasing, resulting in a shift of AMT spectra's vertex to the right, thus to higher scales (bigger distances). From some time, changes in flocs images' textures are becoming almost invisible. The aggregates and their spatial locations change from image to image, because of continuing coagulation, but overall textural characteristics, in other words—images' complexities, remain quite stable, what is also visible from AMT spectrums. Spectrums related to coagulation time 160–1,180 s are overlapping, indicating the stability of images' textures, thus certain stability of flocs' geometrical characteristics. The tendencies described above were observed for all coagulant doses.

Figure 7. AMT spectrums for: (a) different doses at the same time—400 s after the start of coagulation; (b) the time series images of flocs, dose—0.3 mmol Al/l.



Summarizing all above, the AMT spectrums regarding images of flocs can be interpreted as follows: the bigger the flocs and the better they are separated from each other, the more gentle is AMT slope of the spectra; the smaller the flocs and the higher their amount on the image (overlapping), the sharper the resulting AMT spectra, which situates in the lower scales.

The spectrums' data from time series images were processed in *The Unscrambler*[®] by PCA. The results of PCA showed a good separation of “stable” images from those taken during the first seconds of coagulation. Three principle components resulted in 97.5% of explained calibration variance and 96.3% of cross-validation variance.

Figure 9 shows the scores on PC1 of AMT spectrum data for time series of flocs images. Results indicate that after approximately 200 s of coagulation images of flocs become quite stable.

3.7. Mathematical modelling

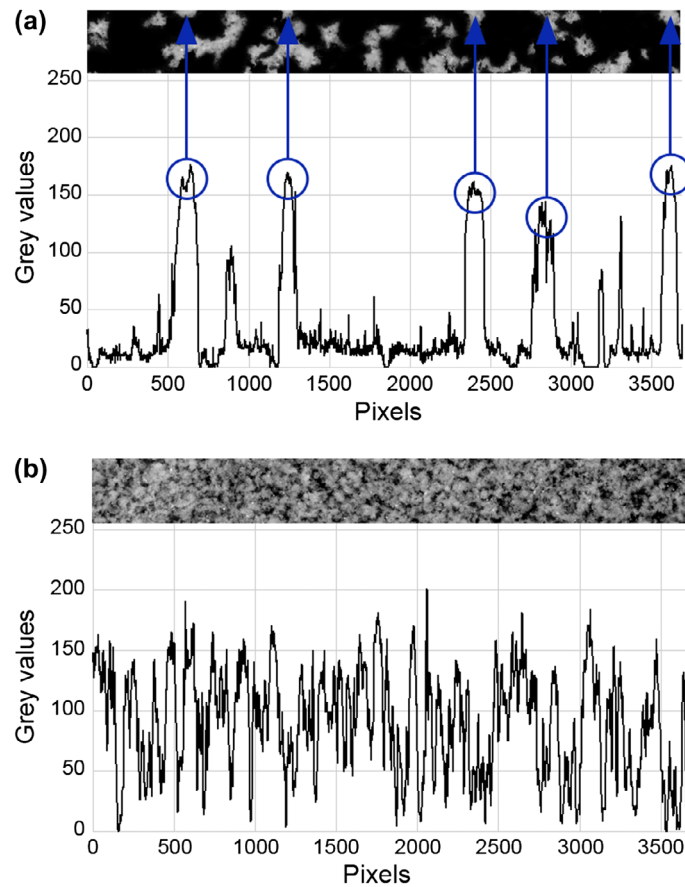
LDA has shown that all three methods of texture image processing give quite good results of data classification. The information about the coagulant dosages and treatment efficiencies was not revealed during LDA, only information from the images was used for classification. LDA was based on score vectors of PCA for classification with estimation of prior probabilities. Confusion matrix results for three image analysis techniques are shown in Table 3.

Data based on histogram analysis of flocs images were classified with 91.82% accuracy (Table 3(a)) using three PCs. One out of ten samples of dose 0.51 mmol Al/l was classified as a low tot. *P* efficiency (class *L*) instead of medium efficiency (class *M*). Also, 8 samples of dose 0.70 mmol Al/l that belongs to class *H* (high tot. *P* efficiency) were classified to *M*.

Data from GLCM image analysis gave the classification accuracy of 90.00% using 3 PCs (Table 3(b)). The classification results are very much similar to the results described above for histogram analysis. However, two more sample images of dose 0.51 mmol Al/l were misclassified.

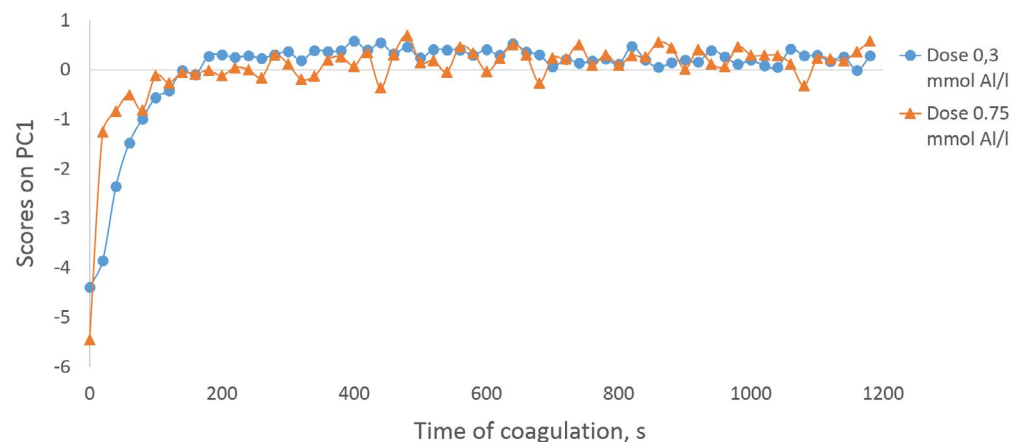
The highest PCA-LDA classification accuracy of 96.36% was determined for the AMT spectrums data of images of flocs using 2 PCs (Table 3(c)). Only 4 samples of dose 0.51 mmol Al/l from class *M* were misclassified to *L*. All 40 samples of high treatment efficiency (class *H*) were classified correctly. This is a very good result, showing that AMT is a perspective tool for image analysis of flocs—AMT spectrums contain necessary data for modelling and precise classification. AMT has the potential to be used for further image analysis of flocs and development of coagulation dosage control system.

Figure 8. Unfold of the first photograph's line of pixels for the coagulant dosages: (a) 0.29 mmol/l; (b) 1.08 mmol/l.



The results of multivariate regression analysis are shown in Table 4. PCR and PLSR were used for mathematical modelling. Response Y is the coagulant dosages in mmol Al/l, X_s are matrixes of histogram data, GLCM feature vectors, and AMT spectrums. Data-sets were divided into two equal parts for calibration and validation of the model. Hence, calibration and validation sets consisted of 11 values of coagulant dosages and 5 replicates of flocs' images per each dose resulting in a total of 55 samples. The obtained matrices of histogram image analysis consisted of 256, 128 and 64 variables (bins). GLCM data was tested with 4 (ASM, contrast, correlation, and entropy), 6 (ASM, contrast, variance, correlation, IDM, entropy) and all 11 variables. AMT data consisted of spectrums with 200

Figure 9. Scores on PC1 of AMT spectrums, calculated from images of flocs for 2 doses—0.3 and 0.75 mmol Al/l.



points, which were set as variables. R^2 of calibration and validation data-sets was used to compare the obtained regression models.

PLSR model based on AMT spectrums showed the best result—96.19% variance explained by 2 factors for validation data. PLSR of GLCM data with 11 variables gave the same result—96.09% Y variance explained. Slightly lower validation R^2 (93.80%) was calculated for PLSR based on 64 bins histogram of flocs' images. All PCR models showed from 1 to 10% lower R^2 compared to PLSR models. This can be explained by the model calculation procedure since PLSR technique models relations between X and Y considering Y during the modelling process, while PCR bases only on X matrix.

All validated models showed quite good results with a minimum of 83% variance explained. It means that images of flocs correlate with coagulant dosages and that the dosages could be predicted by multiple statistical methods. Images of flocs for particular dosage are unique, thus, potentially could be used for dosage control. However, it should be further tested and proved through bigger data with the wider range of coagulant dosages and more sample points. Further analysis should be also performed for other initial parameters of model wastewater and/or industrial wastewater. Among the limitations of the texture recognition methods, the dependence on the lighting conditions should be stated. The described texture analysis methods are dependent on the grey-tone pixel values in the images, thus, to be able to compare the images it is important to have stable lighting conditions or perform mean centring of the images during pre-processing stage. Despite lighting conditions which might vary during laboratory scale experiments, with the full-scale installation it is most likely all conditions are set to a steady state, so this problem should not arise.

4. Conclusions

Changes in flocs' structure depending on the coagulant dosage and treatment efficiency were observed.

According to the treatment efficiency, the 11 coagulant dosages were categorized into three classes for further PCA-LDA. Image analysis by AMT showed the best classification results.

The image analysis method which bases on fractal dimension of flocs has been widely studied. We have observed that there are some difficulties with the application of this method for highly contaminated water—such as wastewater.

Various texture image analysis methods, applied in the other scientific fields were tested for wastewater coagulation. GLCM and AMT were found to be the most promising methods of flocs images identification and correlation with coagulant dosages.

Ten out of eleven co-occurrence matrix-based texture features were found to be significant for dosage prediction. The linear regression model was built based on two of these feature vectors. IDM and Variance were found to adequately describe the flocs' images and their respective dosages.

Texture image analysis methods are easily computable and potentially could be used together or instead of conventional object-recognition image analysis techniques in order to simplify the image analysis methodology, especially in cases when the particles detection is complicated by the initial environment and could lead to under- or overestimation of the flocs' parameters.

Concluding, texture image analysis methods and regression models are the way to correlate images of flocs with dosages and treatment efficiencies, which enables optimal coagulant dosage control.

Funding

This research was supported by the IMPREC project funded by the Polish–Norwegian Research Fund and the Water Harmony project funded by SIU and Ministry of foreign affairs of Norway.

Author details

Nataliia Sivchenko¹

E-mail: nataliia.sivchenko@gmail.com

ORCID ID: <http://orcid.org/0000-0002-1054-0794>

Knut Kvaal¹

E-mail: knut.kvaal@nmbu.no

ORCID ID: <http://orcid.org/0000-0002-0579-7448>

Harsha Ratnaweera¹

E-mail: harsha.ratnaweera@nmbu.no

ORCID ID: <http://orcid.org/0000-0003-1456-2541>

¹ Department of Mathematical Sciences and Technology, Norwegian University of Life Sciences, P.O. Box 5003-IMT, Aas 1432, Norway.

Cover image

Source: Authors.

Citation information

Cite this article as: Evaluation of image texture recognition techniques in application to wastewater coagulation, Nataliia Sivchenko, Knut Kvaal & Harsha Ratnaweera, *Cogent Engineering* (2016), 3: 1206679.

References

- Alves, M., Cavaleiro, A. J., Ferreira, E. C., Amaral, A. L., Mota, M., da Motta, M., & Pons, M. N. (2000). Characterisation by image analysis of anaerobic sludge under shock conditions. *Water Science and Technology*, 41, 207–214.
- Amaral, A. L., & Ferreira, E. C. (2005). Activated sludge monitoring of a wastewater treatment plant using image analysis and partial least squares regression. *Analytica Chimica Acta*, 544, 246–253. doi:10.1016/j.aca.2004.12.061
- Andrle, R. (1994). The angle measure technique: A new method for characterizing the complexity of geomorphic lines. *Mathematical Geology*, 26, 83–97. doi:10.1007/BF02065877
- Annadurai, G., Sung, S. S., & Lee, D. J. (2004). Simultaneous removal of turbidity and humic acid from high turbidity stormwater. *Advances in Environmental Research*, 8, 713–725. [http://dx.doi.org/10.1016/S1093-0191\(03\)00043-1](http://dx.doi.org/10.1016/S1093-0191(03)00043-1)
- Cao, B., Gao, B., Liu, X., Wang, M., Yang, Z., & Yue, Q. (2011). The impact of pH on floc structure characteristic of polyferric chloride in a low DOC and high alkalinity surface water treatment. *Water Research*, 45, 6181–6188. doi:10.1016/j.watres.2011.09.019
- Chakraborti, R. K., Atkinson, J. F., & Van Benschoten, J. E. (2000). Characterization of alum floc by image analysis. *Environmental Science & Technology*, 34, 3969–3976. <http://doi.org/10.1021/es990818o>
- Chakraborti, R. K., Gardner, K. H., Atkinson, J. F., & Van Benschoten, J. E. (2003). Changes in fractal dimension during aggregation. *Water Research*, 37, 873–883. [http://dx.doi.org/10.1016/S0043-1354\(02\)00379-2](http://dx.doi.org/10.1016/S0043-1354(02)00379-2)
- Connors, R. W., Trivedi, M. M., & Harlow, C. A. (1984). Segmentation of a high-resolution urban scene using texture operators. *Computer Vision, Graphics, and Image Processing*, 25, 273–310. doi:10.1016/0734-189X(84)90197-X
- da Motta, M., Pons, M.-N., Roche, N., & Vivier, H. (2001). Characterisation of activated sludge by automated image analysis. *Biochemical Engineering Journal*, 9, 165–173. doi:10.1016/S1369-703X(01)00138-3
- Dagot, C., Pons, M., Casellas, M., Guibaud, G., Dollet, P., & Baudu, M. (2001). Use of image analysis and rheological studies for the control of settleability of filamentous bacteria: Application in SBR reactor. *Water Science and Technology*, 43, 27–33.
- Dahl, C. K., & Esbensen, K. H. (2007). Image analytical determination of particle size distribution characteristics of natural and industrial bulk aggregates. *Chemometrics and Intelligent Laboratory Systems*, 89, 9–25. doi:10.1016/j.chemolab.2007.05.005
- Esbensen, K. H., Hjelmen, K. H., & Kvaal, K. (1996). The AMT approach in chemometrics—first forays. *Journal of Chemometrics*, 10, 569–590. [http://dx.doi.org/10.1002/\(ISSN\)1099-128X](http://dx.doi.org/10.1002/(ISSN)1099-128X)
- Fettig, J., Ratnaweera, H., & Ødegaard, H. (1990, October 1–3). Simultaneous phosphate precipitation and particle destabilization using aluminium coagulants of different basicity. *Chemical Water and Wastewater Treatment: Proceedings of the 4th Gothenburg Symposium Madrid, Spain*, 221–242. doi:10.1007/978-3-642-76093-8
- Fongaro, L., & Kvaal, K. (2013). Surface texture characterization of an Italian pasta by means of univariate and multivariate feature extraction from their texture images. *Food Research International*, 51, 693–705. doi:10.1016/j.foodres.2013.01.044
- Fongaro, L., Lin Ho, D. M., Kvaal, K., Mayer, K., & Rondinella, V. V. (2016). Application of the angle measure technique as image texture analysis method for the identification of uranium ore concentrate samples: New perspective in nuclear forensics. *Talanta*, 152, 463–474. doi:10.1016/j.talanta.2016.02.027
- Gorczyca, B., & Ganczarzyk, J. (1996). Image Analysis of Alum Coagulated Mineral Suspensions. *Environmental Technology*, 17, 1361–1369. doi:10.1080/09593331708616505
- Gorczyca, B., & Ganczarzyk, J. (2002). Flow rates through alum coagulation and activated sludge flocs. *Water Quality Research Journal of Canada*, 37, 389–398.
- Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3, 610–621. doi:10.1109/TSMC.1973.4309314
- He, W., Nan, J., Li, H., & Li, S. (2012). Characteristic analysis on temporal evolution of floc size and structure in low-shear flow. *Water Research*, 46, 509–520. doi:10.1016/j.watres.2011.11.040
- Huang, J., & Esbensen, K. H. (2000). Applications of angle measure technique (AMT) in image analysis: Part I. A new methodology for *in situ* powder characterization. *Chemometrics and Intelligent Laboratory Systems*, 54, 1–19. doi:10.1016/S0169-7439(00)00100-3
- Huang, J., & Esbensen, K. H. (2001). Applications of AMT (angle measure technique) in image analysis: Part II: Prediction of powder functional properties and mixing components using Multivariate AMT Regression (MAR). *Chemometrics and Intelligent Laboratory Systems*, 57, 37–56. doi:10.1016/S0169-7439(01)00120-4
- Jarvis, P., Jefferson, B., Gregory, J., & Parsons, S. A. (2005). A review of floc strength and breakage. *Water Research*, 39, 3121–3137. <http://dx.doi.org/10.1016/j.watres.2005.05.022>
- Jenné, R., Banadda, E., Gins, G., Deurinck, J., Smets, I. Y., Geeraerd, A. H., & Van Impe, J. F. (2006). Use of image analysis for sludge characterisation: Studying the relation between floc shape and sludge settleability. *Water Science & Technology*, 54, 167–174.
- Jenné, R., Banadda, E. N., Smets, I., Deurinck, J., & Van Impe, J. (2007). Detection of filamentous bulking problems: Developing an image analysis system for sludge composition monitoring. *Microscopy and Microanalysis*, 13, 36–41. doi:10.1017/S1431927607070092
- Jin, Y. (2005). *Use of a high resolution photographic technique for studying coagulation/flocculation in water treatment*. Saskatoon: University of Saskatchewan.

- Juntunen, P., Liukkonen, M., Lehtola, M., & Hiltunen, Y. (2013). Dynamic soft sensors for detecting factors affecting turbidity in drinking water. *Journal of Hydroinformatics*, 15, 416–426. <http://dx.doi.org/10.2166/hydro.2012.052>
- Juntunen, P., Liukkonen, M., Lehtola, M., & Hiltunen, Y. (2014). Characterization of alum floc in water treatment by image analysis and modeling. *Cogent Engineering*, 1, 1–13. <http://doi.org/10.1080/23311916.2014.944767>
- Juntunen, P., Liukkonen, M., Pelo, M., Lehtola, M. J., & Hiltunen, Y. (2012). Modelling of water quality: An application to a water treatment process. *Applied Computational Intelligence and Soft Computing*, 2012, 1–9. doi:10.1155/2012/846321
- Kim, S.-H., Moon, B.-H., & Lee, H.-I. (2001). Effects of pH and dosage on pollutant removal and floc structure during coagulation. *Microchemical Journal*, 68, 197–203. doi:10.1016/S0026-265X(00)00146-6
- Kucheryavski, S. (2007). Using hard and soft models for classification of medical images. *Chemometrics and Intelligent Laboratory Systems*, 88, 100–106. doi:10.1016/j.chemolab.2006.08.012
- Kvaal, K. (2014). *AMT plugin for ImageJ*. Retrieved from http://arken.nmbu.no/~kkvaal/eamtexplorer/imagej_plugins.html
- Kvaal, K., Kucheryavski, S. V., Halstensen, M., Kvaal, S., Flø, A. S., Minkinen, P., & Esbensen, K. H. (2008). eAMTExplorer: A software package for texture and signal characterization using angle measure technique. *Journal of Chemometrics*, 22, 717–721. doi:10.1002/cem.1160
- Kvaal, K., Wold, J. P., Indahl, U. G., Baardseth, P., & Næs, T. (1998). Multivariate feature extraction from textural images of bread. *Chemometrics and Intelligent Laboratory Systems*, 42, 141–158. doi:10.1016/S0169-7439(98)00017-3
- Li, D. H., & Ganczarczyk, J. (1989). Fractal geometry of particle aggregates generated in water and wastewater treatment processes. *Environmental Science & Technology*, 23, 1385–1389. <http://doi.org/10.1021/es00069a009>
- Liland, K. H., & Sæbø, S. (2016). *R Commander plug-in for university level applied statistics*. Retrieved from <https://cran.r-project.org/web/packages/RcmdrPlugin.NMBU/index.html>
- Lin, J. L., Huang, C., Chin, C. J. M., & Pan, J. R. (2008). Coagulation dynamics of fractal flocs induced by enmeshment and electrostatic patch mechanisms. *Water Research*, 42, 4457–4466. <http://dx.doi.org/10.1016/j.watres.2008.07.043>
- Liukkonen, M., Hiltunen, T., & Hiltunen, Y. (2015). Evaluation of fluidized bed condition by image analysis and modelling. *IFAC Proceedings Volumes (IFAC-PapersOnline)*, 48, 892–897. doi:10.1016/j.ifacol.2015.05.191
- Mendret, J., Guigui, C., Schmitz, P., Cabassud, C., & Duru, P. (2007). An optical method for *in situ* characterization of fouling during filtration. *AIChE Journal*, 53, 2265–2274. doi:10.1002/aic.11257 [http://dx.doi.org/10.1002/\(ISSN\)1547-5905](http://dx.doi.org/10.1002/(ISSN)1547-5905)
- Mesquita, D. P., Amaral, A. L., & Ferreira, E. C. (2011). Characterization of activated sludge abnormalities by image analysis and chemometric techniques. *Analytica Chimica Acta*, 705, 235–242. doi:10.1016/j.aca.2011.05.050
- Mortensen, P. P., & Esbensen, K. H. (2005). Optimization of the angle measure technique for image analytical sampling of particulate matter. *Chemometrics and Intelligent Laboratory Systems*, 75, 219–229. doi:10.1016/j.chemolab.2004.08.004
- Ødegaard, H., Fettig, J., & Ratnaweera, H. (1990, October 1–3). Coagulation with prepolymerized metal salts. *Chemical water and wastewater treatment: Proceedings of the 4th Gothenburg symposium Madrid, Spain*, 304, 189–220. <http://dx.doi.org/10.1007/978-3-642-76093-8>
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Trans. Systems Man Cybernet.*, 9, 62–66.
- Rasband, W. S. (1997/2016). *ImageJ*, US National Institutes of Health, Bethesda, MD, <http://imagej.nih.gov/ij/>
- Ratnaweera, H. (1991). *Influence of the degree of coagulant prepolymerization on wastewater coagulation mechanisms*. Trondheim: Norwegian Institute of Technology.
- Ratnaweera, H., & Fettig, J. (2015). State of the art of online monitoring and control of the coagulation process. *Water*, 7, 6574–6597. doi:10.3390/w7116574
- Ratnaweera, H., Lei, L., & Lindholm, O. (2002). Simulation program for wastewater coagulation. *Water Science & Technology*, 46, 27–33.
- Tambo, N., & Hozumi, H. (1979). Physical characteristics of flocs—II: Strength of floc. *Water Research*, 13, 421–427. doi:10.1016/0043-1354(79)90034-4
- Tambo, N., & Watanabe, Y. (1979a). Physical aspect of flocculation process—I: Fundamental treatise. *Water Research*, 13, 429–439. doi:10.1016/0043-1354(79)90035-6
- Tambo, N., & Watanabe, Y. (1979b). Physical characteristics of flocs—I: The floc density function and aluminium floc. *Water Research*, 13, 409–419. doi:10.1016/0043-1354(79)90033-2
- Thomas, D. N., Judd, S. J., & Fawcett, N. (1999). Flocculation modelling: A review. *Water Research*, 33, 1579–1592. [http://dx.doi.org/10.1016/S0043-1354\(98\)00392-3](http://dx.doi.org/10.1016/S0043-1354(98)00392-3)
- Vahedi, A., & Gorczyca, B. (2011). Application of fractal dimensions to study the structure of flocs formed in lime softening process. *Water Research*, 45, 545–556. doi:10.1016/j.watres.2010.09.014
- Vahedi, A., & Gorczyca, B. (2012). Predicting the settling velocity of flocs formed in water treatment using multiple fractal dimensions. *Water Research*, 46, 4188–4194. doi:10.1016/j.watres.2012.04.031
- Vahedi, A., & Gorczyca, B. (2014). Settling velocities of multifractal flocs formed in chemical coagulation process. *Water Research*, 53, 322–328. doi:10.1016/j.watres.2014.01.008
- Vold, M. J. (1963). Computer simulation of floc formation in a colloidal suspension. *Journal of Colloid Science*, 18, 684–695. doi:10.1016/0095-8522(63)90061-8
- Wang, Y. L., Feng, J., Dentel, S. K., Lu, J., Shi, B. Y., & Wang, D. S. (2011). Effect of polyferric chloride (PFC) doses and pH on the fractal characteristics of PFC–HA flocs. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 379, 51–61. doi:10.1016/j.colsurfa.2010.11.044
- Xiao, F., Lam, K. M., Li, X. Y., Zhong, R. S., & Zhang, X. H. (2011). PIV characterisation of flocculation dynamics and floc structure in water treatment. *Colloids and surfaces A: Physicochemical and engineering aspects*, 379, 27–35. doi:10.1016/j.colsurfa.2010.11.053
- Yu, R.-F., Chen, H.-W., Cheng, W.-P., & Chu, M.-L. (2009). Simultaneously monitoring the particle size distribution, morphology and suspended solids concentration in wastewater applying digital image analysis (DIA). *Environmental Monitoring and Assessment*, 148, 19–26. doi:10.1007/s10661-007-0135-z
- Yu, R.-F., Cheng, W.-P., & Huang, H.-D. (2012, June 18–20). On-line assessment of the particle separation in chemical flocculent suspension by image analysis. In M. Jekel (Ed.), *International conference on particle separation* (pp. 53–57). Berlin: International Water Association.
- Zheng, C., Sun, D.-W., & Zheng, L. (2006). Recent applications of image texture for evaluation of food qualities—A review. *Trends in Food Science & Technology*, 17, 113–128. <http://doi.org/10.1016/j.tifs.2005.11.006>
- Zheng, H., Zhu, G., Jiang, S., Tshukudu, T., Xiang, X., Zhang, P., & He, Q. (2011). Investigations of coagulation–flocculation process by performance optimization, model prediction and fractal structure of flocs. *Desalination*, 269, 148–156. <http://doi.org/10.1016/j.desal.2010.10.054>

Appendix A

The output from R: results of the variable selection for MLR model by best subsets.

>best subsets (RegModel.3, nbest = 2, nvmax = 11, force.in = NULL)

Reordering variables and trying again:

		ASM	Cont	Corr	Ener	Entr	Homog	IDM	Iner	Prom	Shade	Var	RSS	R ²	R ² _{adj.}	Cp
1	(1)										*		0.2097	0.9296	0.9282	439.5
1	(2)						*						0.3339	0.8878	0.8857	730.1
2	(1)							*				*	0.0994	0.9666	0.9653	183.6
2	(2)			*				*					0.1029	0.9654	0.9641	191.6
3	(1)			*				*			*		0.0574	0.9807	0.9796	87.3
3	(2)							*			*	*	0.0605	0.9797	0.9785	94.6
4	(1)			*			*	*			*		0.0432	0.9855	0.9843	56.1
4	(2)						*	*			*	*	0.0469	0.9843	0.9829	64.6
5	(1)			*		*		*	*		*		0.0283	0.9905	0.9895	23.2
5	(2)		*	*		*		*			*		0.0283	0.9905	0.9895	23.2
6	(1)			*		*	*		*		*	*	0.0209	0.9929	0.9921	7.9
6	(2)		*	*		*	*				*	*	0.0209	0.9929	0.9921	7.9
7	(1)			*		*	*		*	*	*	*	0.0187	0.9937	0.9928	4.7
7	(2)		*	*		*	*			*	*	*	0.0187	0.9937	0.9928	4.7
8	(1)		*	*		*	*	*		*	*	*	0.0185	0.9938	0.9927	6.2
8	(2)			*		*	*	*	*	*	*	*	0.0185	0.9938	0.9927	6.2
9	(1)	*		*		*	*	*	*	*	*	*	0.0184	0.9938	0.9926	8.0
9	(2)	*	*	*		*	*	*		*	*	*	0.0184	0.9938	0.9926	8.0

Notes: ASM—angular second moment, Cont—contrast, Corr—correlation, Ener—energy, Entr—entropy, Homog—homogeneity, IDM—inverse difference moment, Iner—inertia, Prom—prominence, Var—variance.



© 2016 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.

You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

