Estimation of *Chlamydomonas reinhardtii* biomass concentration from chord length distribution data

Patricio Lopez-Exposito[1] · Angeles Blanco Suarez[1] · Carlos Negro[1]

Abstract A novel method to estimate the concentration of *Chlamydomonas reinhardtii* biomass was developed. The method employs the chord length distribution information gathered by means of a focused beam reflectance probe immersed in the culture sample and processes the data through a feedforward multilayer perceptron. The multilayer perceptron architecture was systematically optimised through the application of a simulated annealing algorithm. The method developed can predict the concentration of microalgae with acceptable accuracy and, with further development, it could be implemented online to monitor the aggregation status and biomass concentration of microalgal cultures.

Introduction

The production of microalgal biomass has received a great deal of attention in the last years. Several unresolved technical-economical aspects still hinder the transfer of microalgae cultivation from laboratory photobioreactors to industrial scale. Among them, harvesting constitutes a major challenge. Those harvesting methods involving flocculation of microalgal biomass, namely dissolved air flotation, bio-

flocculation, flocculation assisted sedimentation and flocculation assisted filtration, require a good accuracy in what concerns both the degree of aggregation and the concentration of the culture to apply the optimal dosage of flocculant, especially in large-scale facilities. The gravimetric determination of biomass dry weight requires considerable time, and it is not suitable as a monitoring method for continuous harvesting installations. Conventional indirect monitoring techniques are mostly based on associating light scattering or attenuation properties of the medium to a given biomass dry weight through an offline calibration (Reardon et al. 2013). Among these techniques, optical density (OD) measurement is the most popular one. This method, however, often yields inaccurate estimations of dry biomass due to several facts. First, the variable pigment contents that microalgae show during their growth cycle and under different culture conditions may distort the OD measurement (Griffiths et al. 2011). Secondly, the medium itself may undergo changes in turbidity, which will affect the measurement. Finally, the fact that the absorbance measured is a function of cell size, concentration and shape makes that the OD measurements vary according to the cultivation conditions (Chioccioli et al. 2014). While the changes in the quantity of pigments per cell and the intrinsic optical can be overcome choosing a wavelength lying away from the composite one of the culture, the effects of size, shape and concentration factors are hardly avoidable. The literature identifies several techniques, namely 2D fluorometry, IR spectroscopy, multiparameter FC, in situ microscopy and focused beam reflectance measurement probe (FBRM), as a having potential to become the basis of new online monitoring methods (Höpfner et al. 2010; Reardon et al. 2013).

The work presented in this paper focuses on the evaluation of a FBRM as a tool for devising a method to allow the online estimation of biomass concentration in microalgal cultures. The probe was developed to monitor crystallisation (e.g.

✉ Patricio Lopez-Exposito
   plopezex@ucm.es

[1] Chemical Engineering Department, Chemistry Faculty, Complutense University of Madrid, Avda. Complutense s/n, Madrid 28040, Spain

Fig. 1 Schematic microalgal cultures having the same concentration and different chord length distributions
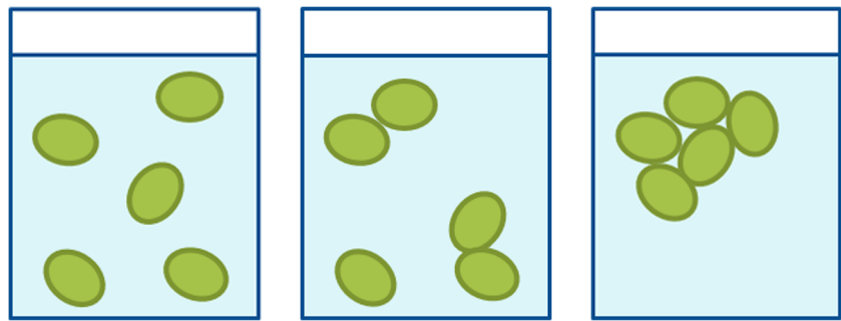
Barrett and Glennon 1999; Li et al. 2014), although it has also been used to monitor flocculation processes (e.g. Blanco et al. 1996, 2002; Jarabo et al. 2013). The FBRM probe is a device that projects a laser beam moving in a circular path into an aqueous medium and registers the backscattered beam produced when the laser path crosses a particle. The number of crossing events is directly related with the density of particles in the medium, and the duration of each event yields the corresponding chord length of the particle. The distribution of chord lengths (CLD) constitutes a representation of the actual particle length distribution (PLD) (Li and Wilkinson 2005).

Given that the FBRM device provides information of the size of the particles in the medium, the probe takes into account the aggregation of cells, thus avoiding the lack of accuracy entailed in the characterisation of suspensions through optical techniques when the concentration of particles is so high that the overlapping effect cannot be compensated or when such suspensions are made up of large particles.

Although the FBRM probe has been occasionally used in flocculation studies of microalgal biomass (Danquah et al. 2010) or to characterise microalgal culture size distributions (Uduman et al. 2011), at the time of writing this article, the literature offers only two references in which the FBRM was employed to estimate biomass concentration, in both cases of plant cell suspensions (McDonald et al. 2001; Kieran et al. 2003). In these studies, starting suspension of plant cells were

Table 1 Codification of network characteristics

	Values	Number of bits
Number of hidden layers	1 or 2	1
Number of neurons in the first hidden layer	3 to 18	4
Number of neurons in the second hidden layer (when applicable)	3 to 18	4
Bias in the first hidden layer	Yes/no	1
Bias in the second hidden layer	Yes/no	1
Bias in the output layer	Yes/no	1
Initial magnitude order of the weights from the input to the first hidden layer	1 0.1 0.01 0.001	2
Initial magnitude order of the biases in the hidden layers (if any)	1 0.1 0.01 0.001	2*2 layers
Initial magnitude order of the weights in the hidden layers	1 0.1 0.01 0.001	2*2 layers
Transfer function of the first hidden layer	Log-sigmoid or hyperbolic tangent sigmoid	1
Transfer function of the second hidden layer	Log-sigmoid or hyperbolic tangent sigmoid	1
Transfer function of the output layer	Log-sigmoid or hyperbolic tangent sigmoid	1
Ratio of the mean square errors and the mean square weights (regularisation)	1 0.5 0.1 0.01	2

diluted several times with the culture medium and characterised with the probe at each dilution point in order to obtain a correlation between the number of particles per second detected and the concentration of biomass. Those works therefore do not take into account the fact that a given biomass concentration can be measured in cultures having a completely different distribution of aggregate sizes. Figure 1 represents three schematic microalgal cultures having the same biomass concentration at different aggregation states.

The use of the FBRM probe to estimate dry biomass concentration presents several drawbacks. First, the size distribution data gathered from microalgal cultures are characterised for having a very wide range of particle chord lengths and rate of particles detected. Second, the geometry of the algal aggregates departs from the spherical shape. This effect has been observed to be even more significant when the number of cells in the aggregate becomes larger as can be seen in the micrographs included in the "Results and discussion" section of this paper (Fig. 6). When aggregation takes place, the culture presents a highly heterogeneous distribution of particle sizes and shapes, and it is possible to find large aggregates of irregular shape, small aggregates and isolated cells of shape close to spherical. Finally, the FBRM probe data only provide a transformed representation of real chord particle length distributions. With these premises, it is not possible to find an analytical or semi-analytical model to estimate algal biomass concentration from CLD data, making it necessary to apply artificial intelligence tools to approach the problem.

The authors chose to model the relationship between CLD data and dry biomass using artificial neural networks (ANN), in particular feedforward multilayer perceptrons, given their proven efficacy in solving problems related to sensing and spectra interpretation in the fields of Biotechnology (Strapasson et al. 2014) and Chemical Engineering (Curteanu and Cartwright 2011; Pirdashti et al. 2013; Ali et al. 2015). ANN have been employed occasionally to estimate biomass concentration of yeasts in fermentation (e.g. Vaněk et al. 2004; Hocalar et al. 2011) based on the concentration or production rates of the chemical species involved in the process. Therefore, the use of ANN in the estimation of microalgal biomass and the use of ANN to process CLD data for estimating biomass concentration represent an important innovation in this field.

The present study investigates the possibility of translating CLD data of *Chlamydomonas reinhardtii* suspensions into a biomass concentration by means of artificial neural networks.

Materials and methods

Chlamydomonas reinhardtii strain from CCAP (CCAP No. 11/32B) was used for this study. It was and cultivated with TAP medium (Gorman and Levine 1965) in shake flasks (115 rpm, 23 °C and 12 h cool while light). After a concentration of biomass around 1 g L^{-1} was obtained, the cultures were transferred to 5.5-L photobioreactors operated at a
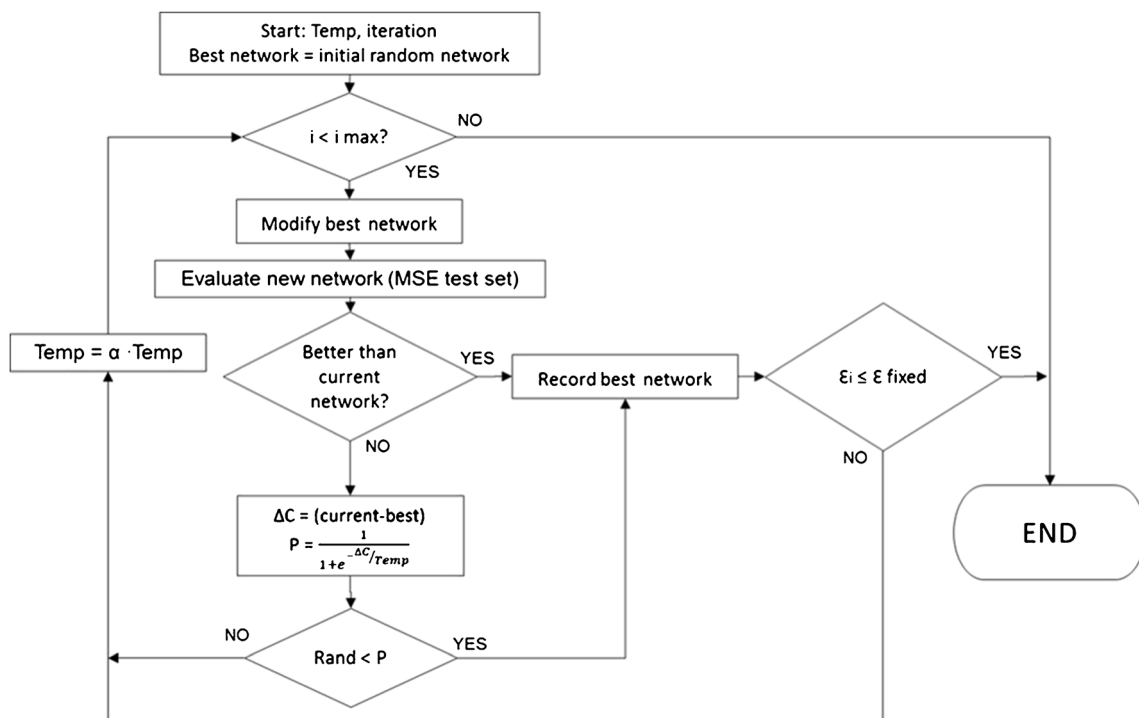
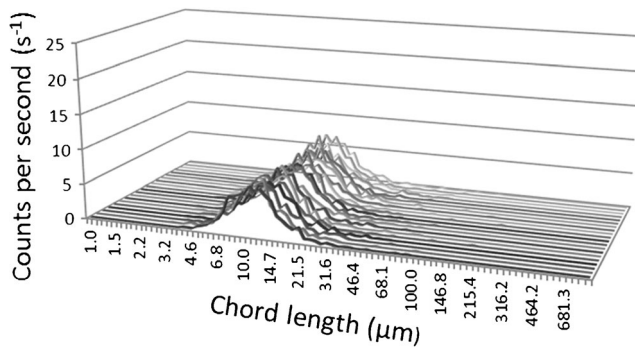**Fig. 2** Algorithm for the selection of the multilayer perceptron architecture

Fig. 3 Chord length distributions of a 0.2 g L^{-1} *C. reinhardtii* culture

temperature between 23 and 25 °C with aeration (2 L min^{-1}) and pH 7.5. pH was controlled through the automatic supply of CO_2 (0.2 L min^{-1}). Light was provided by means of four fluorescent cool while light in a 12-h cycle.

Flocculant

Chitosan from crab shells (Sigma-Aldrich) was used as a flocculant. The flocculant stock solution was prepared by dissolving chitosan in a solution of 1 % glacial acetic acid in ultrapure water subject to mechanical stirring at 400 rpm for 1 h. The solution was left to settle for 1 day before use.

Experimental procedure

Dry biomass determination

The determination of dry biomass was carried in triplicate following the method described in Beckmann et al. (2009). Each sample of 20-mL medium was centrifuged at 8400 rpm for 11 min and washed twice with ultrapure water. The microalgal pellets obtained were transferred to previously dried aluminium dishes and allowed to dry at 101.5 °C for 3 h. The samples were then placed in a desiccator for 45 min and weighted afterwards. The concentration of biomass was calculated as gramme dry biomass per litre of medium.
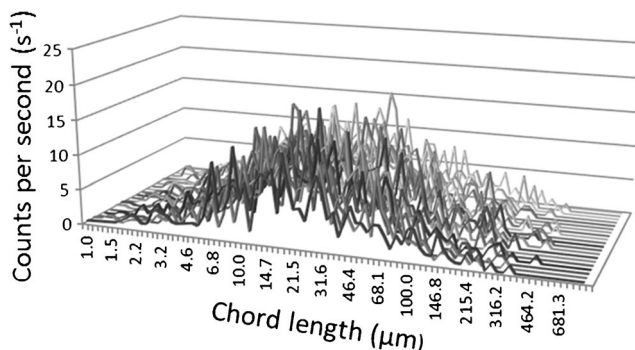

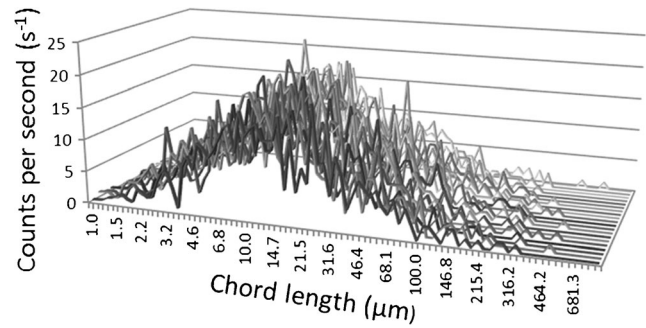**Fig. 4** Chord length distributions of a 1.0 g L^{-1} *C. reinhardtii* culture

Fig. 5 Chord length distributions of a 2.7 g L^{-1} *C. reinhardtii* culture

Characterisation of C. reinhardtii cultures through CLD measurement

An M500L FBRM (Mettler Toledo) was employed to gather data about the distribution of particle chord length in the microalgal cultures. The FBRM is capable of performing the real-time monitoring of the number of particles in a suspension and classifies them in terms of their chord length through a dedicated software system. The software recorded the number of events detected per second and classified them according to their length in one of 90 bins of length intervals organised logarithmically and ranging from 1 to 1000 μm. The cultures were analysed in 10-s sampling intervals over a period of 6 min, which represented 36 data points. In order to represent a wide variety of concentration and aggregation states, the samples were collected at different stages of microalgal growth within the reactor. Fifteen microalgal cultures having different biomass concentrations were considered.

With the purpose of having a wide variety of aggregation states of the cells, each culture was partially flocculated using several doses of 1 % chitosan. Likewise, the original cultures and the flocculated ones were characterised with the FBRM probe taking samples of 200 mL. The samples were placed in 250-mL beakers and mechanically stirred at 200 rpm.
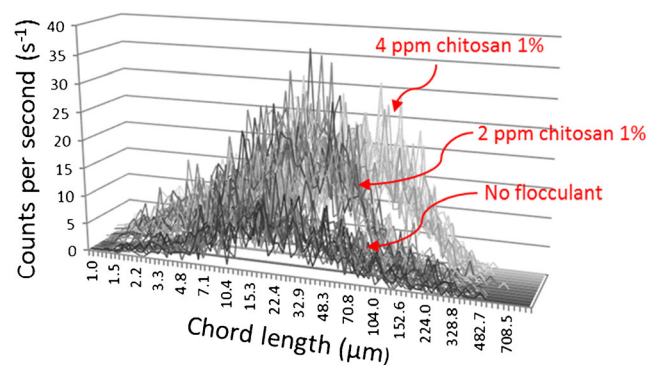

Fig. 6 CLDs of a 1.0 g L^{-1} *C. reinhardtii* sample at different aggregation states
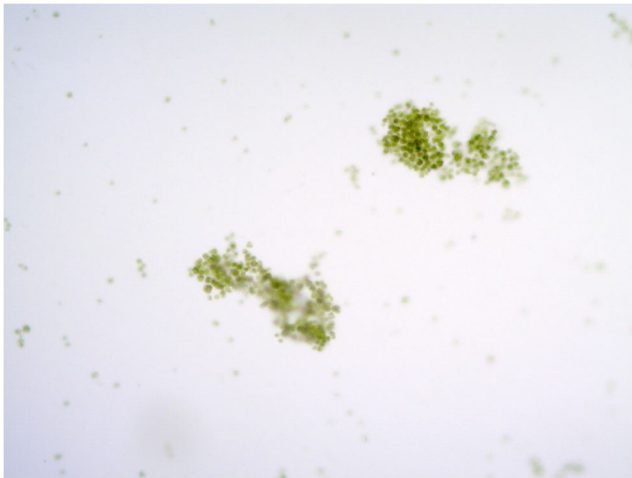
Fig. 7 *Chlamydomonas reinhardtii* flocs

Data processing and network architecture selection

The set of data collected from each algal sample (flocculated and "raw") was exported to a spreadsheet application and converted to a tabular data file. Each data file contained a set of data corresponding to the biomass chord length distribution of the medium sampled, i.e. a collection of vectors of 90 elements each. From the 36 data points gathered for each suspension, the first 8 points and the 8 last ones were discarded, what left a set of 20 consecutive measuring points (vectors) for further processing. From this phase onwards, the processing of data was carried out with the mathematical software package Matlab R2014b. The 20 vectors of each medium sampled were extracted with the mentioned software, and each one was associated to its corresponding dry biomass weight. The final dataset built consisted of a matrix having 1940 vectors of 90 elements (CLD data) and a vector of 1940 elements containing the biomass concentration data. The whole set of data was randomly sorted preserving the correspondence between CLD and biomass concentration.
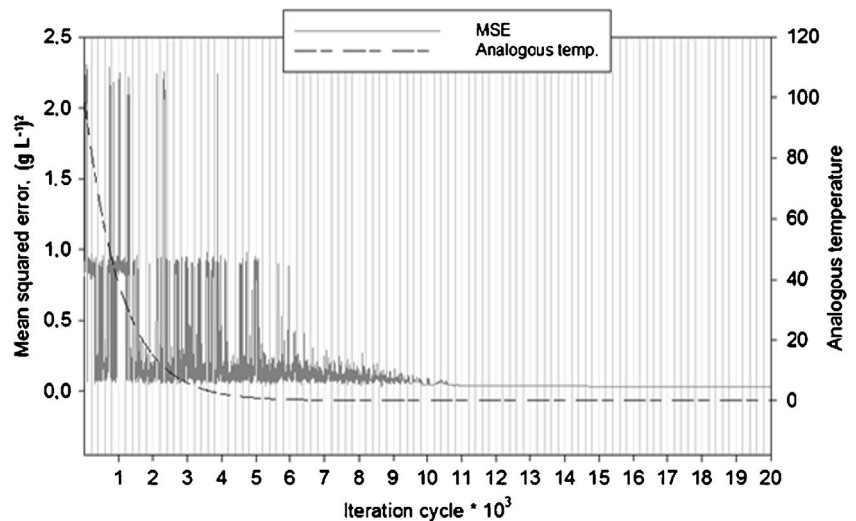
In order to find the optimal perceptron architecture to fit the chord length distribution data, a simulated annealing algorithm was employed. The algorithm creates a random network architecture and evaluates it. Then, a random modification of the architecture is carried out and the network is evaluated again. If the new ensemble produces a better estimation of the biomass concentration, it is accepted. If the new network produces worse results, it can be accepted with a given probability that depends on the Boltzmann probability distribution as shown in Eq. 1, which is itself dependent on an analog temperature.

$$P(\text{state}) = e^{-\frac{\Delta E}{kT}} \tag{1}$$

$P(\text{state})$ represents the probability of accepting a solution, ΔE is the difference between the error of the current network (energy of the state) and the previous one, k is a constant (in real systems the Boltzmann's constant) and T is the temperature. In the present work, the cooling schedule chosen was determined by Eq. 2.

$$T_{i+1} = \alpha T_i \tag{2}$$

T_i is the temperature of the current state. T_{i+1} represents the temperature of the subsequent state and α is a reduction factor, in the case concerned set to 0.99905.

To evaluate each architecture, a Matlab function taking the characteristics of the network was implemented and called within the main programme. This function took as input parameters the characteristics of the network, the input and the output data. The characteristics of the network were codified as strings of 27 bits (Table 1).

The function constructed the network, trained, validated and tested it. The dataset was divided in three subsets to perform each procedure employing 70 % of the data for training (1358 points), 15 % for validation (291 points) and 15 % for testing (291 points). The function returned the network trained

Fig. 8 MSE of the perceptron cycle and analogous temperature in each iteration

and mean squared error of the estimations of biomass concentration corresponding solely to the test subset. Employing only the test set to evaluate the predicting ability of the network favours those networks having a generalisation capacity, thus avoiding overfitting. All networks were trained using the scaled conjugate gradient backpropagation method. The trainings were carried out using the CUDA (Compute Unified Device Architecture) functionalities offered by Matlab to accelerate the running time by employing the parallel computing capabilities of the workstation's GPU.

As mentioned above, the value of the error and the probability value calculated according to Eq. 1 served as a criterion to accept or discard the model. The algorithm was left to carry out 20,000 iterations.

Figure 2 depicts the algorithm implemented to select the best neural network architecture.

Results and discussion

Biomass dry weight and chord length distributions

Dry microalgal biomass concentration ranged between 0.2 and 2.7 g L^{-1}.

To exemplify the results obtained through the FBRM probe, three chord length distributions gathered with the device are presented in Figs. 3, 4 and 5 (0.2, 1.0 and 2.7 g dry biomass L^{-1}). In the figures, the vertical axis represents the number of particles per second detected by the equipment and the horizontal axis the midpoint of the chord length intervals. The profundity axis corresponds to each of the 20 distributions taken in the course of the sampling time.

Table 2 List of feature of the multilayer perceptron selected

| Feature	Value
Number of hidden layers	2
Number of neurons in the first hidden layer	18
Number of neurons in the second hidden layer	14
Bias in the first hidden layer	No
Bias in the second hidden layer	Yes
Bias in the output layer	Yes
Initial value of the weights going from the input to the first hidden layer	0.1
Initial value of the biases in the hidden layers	1
Initial value range of the weights going from the first hidden layer to the second hidden layer or output	1
Transfer function of the first hidden layer	Hyperbolic tangent
Transfer function of the second hidden layer	Logistic
Transfer function of the output layer	Hyperbolic tangent
Ratio of the mean square errors and the mean square weights (regularisation)	0.01
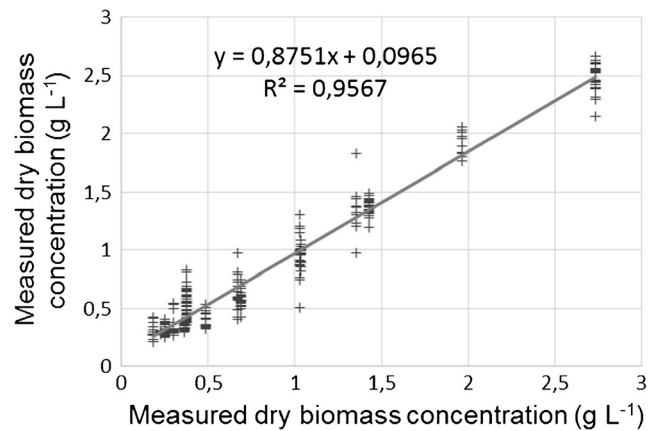
Fig. 9 CLD-estimated biomass concentration vs. gravimetrically determined biomass concentration in the training set

Figure 6 shows three different states (in increasingly lighter hues of grey) produced with the same *C. reinhardtii* culture of 1.0 g L^{-1} dry biomass. The cluster of lines in the nearest position corresponds to the *raw* culture, and the next two ones were obtained with the addition of 2 and 4 ppm of 1 % chitosan, respectively. It can be observed that the lines of the second and third clusters reach higher values of counts per second and that the third cluster is slightly shifted towards higher values of chord length. The increased number of particles detected in the second and third measurements can be ascribed to the fact that the flocculation process aggregates some cells that fall below the lower detection threshold of the equipment into flocs that become visible for the probe. The shifting towards larger values of chord length is due to the formation of larger particles associated to a higher dosage of flocculant.

Figure 7 is a photomicrograph of a flocculated culture sample. Both the floc-formed and isolated microalgal cells are visible. Isolated cells present an approximately spherical shape while flocs are irregular.
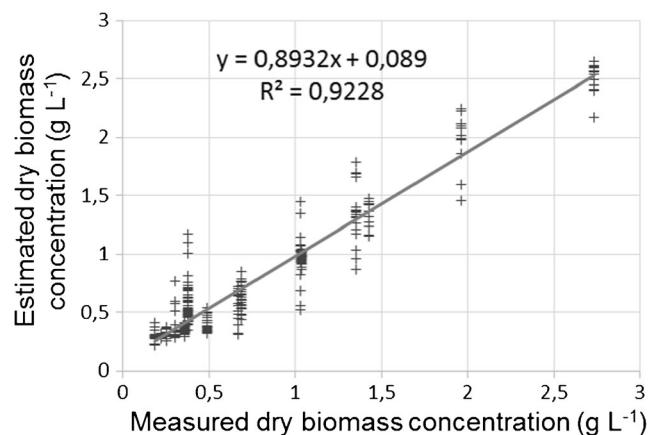

Fig. 10 CLD-estimated biomass concentration vs. gravimetrically determined biomass concentration in the test set

Perceptron selection

Figure 8 shows the evolution of the mean squared error of the multilayer perceptron in the process of selection and the analog temperature over the 20,000 iteration cycles.

The graph shows that at the early stages of iteration, the MSE shows very ample oscillations. At iteration cycle 10,895 (analog temperature value of 0.0032), the error reaches a plateau that corresponds to a MSE of 0.03764 $(g\ L^{-1})^2$. At iteration 18,215, the error reaches its minimum, 0.0309 $(g\ L^{-1})^2$. The network having this minimum error presents the features listed in Table 2.

Figure 9 shows the correlation between the estimated biomass dry concentration and the real one for the train subset.

Figure 10 shows the correlation between the estimated biomass dry concentration and the measured one for the test subset.

Conclusions

A method to estimate the dry biomass concentration of *C. reinhardtii* cultures based on their aggregation state and particle density was implemented. The developement of the method involved selecting the adequate architecture of a multilayer perceptron to translate chord length distribution data gathered though an FBRM probe into microalgal biomass concentration. The optimal architecture of the perceptron was found by applying a systematic selection of the model parameters based on a simulated annealing algorithm that favours the generalisation capacity of the model. The artificial neural network selected was capable to produce very good estimations of the dry biomass concentration of *C. reinhardtii* cultures ($R^2 = 0.9228$ in the test set). In view of the results, it can be ascertained that the method investigated could be a useful tool for the online monitoring of microalgal suspensions if the software is adapted to handle data streams to perform the estimation of biomass concentration instantaneously. With such monitoring system, it would be possible to estimate the concentration of microalgal biomass in a reliable manner without the interferences of factors such as turbidity of the medium, state of aggregation or phase of growth. Such system should be calibrated to account for the particular characteristics of the strain considered (size, aspect ratio) and hydrodynamic conditions of the culture since these affect the flocculation (and auto-flocculation) process and determine the probability of the particles traversing the FBRM probe window.

References

Ali JM, Hussain MA, Tade MO, Zhang J (2015) Artificial intelligence techniques applied as estimator in chemical process systems—a literature survey. Expert Syst Appl 42:5915–5931

Barrett PB, Glennon BA (1999) In-line monitoring of crystal growth in batch crystallization operations. 14th Int. Symp Ind Cryst. Institution of Chemical Engineers, Rugby, UK

Beckmann J, Lehr F, Finazzi G, Hankamer B, Posten C, Wobbe L, Kruse O (2009) Improvement of light to biomass conversion by deregulation of light-harvesting protein translation in *Chlamydomonas reinhardtii*. J Biotech 142:70–77

Blanco A, Negro C, Hooimeijer A, Tijero J (1996) Polymer optimization in paper mills by means of a particle size analyser: an alternative to zeta potential measurements. Appita J 49:113–116

Blanco A, Fuente E, Negro C, Tijero J (2002) Flocculation monitoring: focused beam reflectance measurement as a measurement tool. Can J Chem Eng 80(4):1–7

Chioccioli M, Hankamer B, Ross IL (2014) Flow cytometry pulse width data enables rapid and sensitive estimation of biomass dry weight in the microalgae *Chlamydomonas reinhardtii* and *Chlorella vulgaris*. PLoS ONE 9(5), e97269

Curteanu S, Cartwright H (2011) Neural networks applied in chemistry. I. Determination of the optimal topology of multilayer perceptron neural networks. J Chemometrics 25:527–549

Danquah MK, Uduman N, Bourniquel V, Michael K, Hoadley AFA (2010) Marine microalgae flocculation and focused beam reflectance measurement. Chem Eng 162:935–940

Gorman DS, Levine RP (1965) Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of *Chlamydomonas reinhardtii*. Proc Natl Acad Sci U S A 54:1665–1669

Griffiths MJ, Garcin C, van Hille RP, Harrison STL (2011) Interference by pigment in the estimation of microalgal biomass concentration by optical density. J Microbiol Meth 85:119–123

Hocalar A, Türker M, Karakuzu C, Yüzgeç U (2011) Comparison of different estimation techniques for biomass concentration in large scale yeast fermentation. ISA Trans 50:303–314

Höpfner T, Bluma A, Rudolph G, Lindner P, Scheper T (2010) A review of non-invasive optical-based image analysis systems for continuous bioprocess monitoring. Bioprocess Biosyst Eng 33:247–256

Jarabo R, Monte MC, Fuente E, Santos SF, Negro C (2013) Corn stalk from agricultural residue used as reinforcement fiber in fiber-cement production. Ind Crop Prod 43:832–839

Kieran PM, Jeffers P, Raposo S, Lima-Costa ME, Connolly P, Glennon B (2003) Focused beam reflectance measurement (FBRM) monitoring of particle size and morphology in suspension cultures of *Morinda citrifolia* and *Centaurea calcitrapa*. Biotechnol Lett 25:2023–2028

Li M, Wilkinson D (2005) Determination of non-spherical particle size distribution from chord length measurements. Part 1: theoretical analysis. Chem Eng Sci 60(12):3251–3265

Li H, Kawajiri Y, Grover MA, Rousseau RW (2014) Application of an empirical FBRM model to estimate crystal size distributions in batch crystallization. Cryst Growth Des 14:607–616

McDonald KA, Jackman AP, Hurst S (2001) Characterization of plant suspension cultures using the focused beam reflectance technique. Biotechnol Lett 23:317–324

Pirdashti M, Curteanu S, Kamangar M, Hassim M, Khatami MA (2013) Artificial neural networks: applications in chemical engineering. Rev Chem Eng 29:205–239

Reardon KF, Havlik I, Lindner P, Scheper T (2013) On-line monitoring of large cultivations of microalgae and cyanobacteria. Trends Biotechnol 31:406–414

Strapasson RA, Woiciechowski AL, Letti LAJ, Soccol CR (2014) Some applications of artificial intelligence on biotechnology. J Biotech Biodiver 5:1–11

Uduman N, Bourniquel V, Michael K, Danquah MK, Hoadley AFA (2011) A parametric study of electrocoagulation as a recovery process of marine microalgae for biodiesel production. Chem Eng J 174:249–257

Vaněk M, Hrnčiřík P, Vovsik J, Nahlik J (2004) On-line estimation of biomass concentration using a neural network and information about metabolic state. Bioprocess Biosyst Eng 27:9–15