ORIGINAL PAPER

Cole–Cole, linear and multivariate modeling of capacitance data for on-line monitoring of biomass

Michal Dabros · Danielle Dennewald ·
David J. Currie · Mark H. Lee · Robert W. Todd ·
Ian W. Marison · Urs von Stockar

Abstract This work evaluates three techniques of calibrating capacitance (dielectric) spectrometers used for on-line monitoring of biomass: modeling of cell properties using the theoretical Cole–Cole equation, linear regression of dual-frequency capacitance measurements on biomass concentration, and multivariate (PLS) modeling of scanning dielectric spectra. The performance and robustness of each technique is assessed during a sequence of validation batches in two experimental settings of differing signal noise. In more noisy conditions, the Cole–Cole model had significantly higher biomass concentration prediction errors than the linear and multivariate models. The PLS model was the most robust in handling signal noise. In less noisy conditions, the three models performed similarly. Estimates of the mean cell size were done additionally using the Cole–Cole and PLS models, the latter technique giving more satisfactory results.

Introduction

Over the last few decades, the field of biotechnology has gained significant importance, both in industry and on a purely academic level. The number of related processes, products and applications has increased at an exponential rate, paving the way for cutting-edge research in the discipline and heightening the technology's economic position. Growing efforts are made to optimize the processes, increase their efficiency and productivity, improve the quality of the desired product, and thus, increase product safety and manufacturing profitability. One way to achieve these improvements is through accurate bioprocess monitoring and control [1–3]. As biomass is one of the key parameters in biotechnological processes, monitoring this variable in real-time is highly desirable [4]. On-line monitoring of biomass concentration allows control of culture conditions in order to obtain the desired (or constant) cell density or to choose the optimal moment to induce the production of a recombinant protein. Real-time measurements of the average cell size or volume provide information about the morphology of the microorganisms and can serve, for example, as an on-line indicator of osmotic stress on the cells [5].

On-line monitoring of biomass is a fairly recent field that is still undergoing considerable development. Various methods of monitoring biomass have been explored, and they are usually classified into two groups: indirect and direct methods. Indirect methods do not measure the

M. Dabros · D. Dennewald · U. von Stockar (✉)
Laboratory of Chemical and Biological Engineering (LGCB),
Ecole Polytechnique Fédérale de Lausanne (EPFL), Station 6,
1015 Lausanne, Switzerland
e-mail: urs.vonstockar@epfl.ch

M. Dabros
e-mail: Michal.dabros@epfl.ch

D. J. Currie · M. H. Lee
Department of Computer Science, University of Wales,
Aberystwyth, UK

I. W. Marison
School of Biotechnology, Dublin City University, Glasnevin,
Dublin 9, Ireland

R. W. Todd
Aber Instruments Ltd, Science Park, Aberystwyth, UK

biomass concentration itself. Instead, they monitor other parameters that can be related to it, for example the concentration of compounds that are produced or consumed during biomass growth. The most commonly used parameters are oxygen uptake rate (OUR) and carbon dioxide evolution rate (CER). Knowing the specific rates for the cell type used in the culture, the biomass concentration can be estimated. The major problem with this method is that the specific uptake or evolution rates are assumed to be constant during the culture, which is not always the case since these specific rates may fluctuate with the physiological state of the cells [6]. These methods are, thus, based on only partially valid assumptions and may lead to significant errors in the predictions. In the field of spectroscopy, indirect biomass monitoring has been reported using fluorescence measurements [4, 7]. Direct determination of biomass concentration is done either by biological quantification methods (viable cell counting or petri plating) or by exploring the physical properties of the cells. For applications involving on-line monitoring of biomass, only physical methods give the required real-time measurement. Physical methods are based mostly on the quantification of optical, acoustic, magnetic or electrical properties. The most frequently used optical method is optical density (OD) measurement. Unfortunately, its use in in-situ applications is limited because the measurement is very sensitive to bubbles, cell aggregation and non-cellular scattering particles present in the suspension [8]. Success in on-line, in-situ monitoring of biomass with near-infrared probes has been shown by several authors [9–11]. Optical monitoring of cell concentration and average cell volume can also be achieved with in-situ microscopy (ISM) [5, 12]. Amongst acoustic methods, the technique known as acoustic resonance densitometry (ARD) is based on the relationship between the resonant frequency of the suspension and the fluid density. After subtracting the density of the supernatant fluid, the cell concentration is determined by correlating it linearly to the fluid density [6]. The drawbacks of this approach are its poor sensitivity and the significant dependence on temperature, medium characteristics and, again, the presence of bubbles. Nuclear magnetic resonance (NMR) techniques are used mostly for fundamental research and not for bioprocess monitoring [1]. The main disadvantages of these methods are the long measurement time needed to obtain acceptable sensitivity and the relatively expensive equipment. Finally, over the last few years, satisfactory results and enhanced in-situ applicability were achieved using dielectric spectroscopy [4, 8, 13–15].

Dielectric spectroscopy, also known as capacitance spectroscopy, is based on the principle that under the influence of an electrical field, cells suspended in a conducting medium act as capacitors and are capable of storing electrical charge. The overall capacitance of the suspension, observed over the so-called β-dispersion range of frequencies (typically between 3 and 10 MHz) is directly proportional to the total volume of viable cells affected by the field [15–17]. In addition, by collecting capacitance readings over the characteristic range of frequencies of the electrical current, observations can be made as to the size of the cells in the suspension [18, 19]. One of the main advantages of this technology is that only viable biomass is measured while cell debris, necromass and other non-cellular particles are not [15]. Capacitance probes are typically non-invasive and in-situ sterilizable, making them attractive for bioprocesses. Measuring frequency in currently available instruments is in the order of several scans per minute, thus quick enough to follow adequately the kinetics of most cell cultures and appropriate for on-line process control. Many successful monitoring applications of dielectric spectroscopy have been reported, mainly involving animal cell cultures [18–22] and microbial fermentations [14, 17, 19, 23–25].

One of the predicaments with dielectric spectroscopy is that signal characteristics and measurement reproducibility are strongly influenced by factors like electrode polarization or variable medium conductivity, as well as by the physical setup of the bioreactor, exact positioning of the probe and the proximity of stationary (baffles) and moving (agitator) metal components. For this reason, dielectric instruments are usually calibrated in-situ, maintaining constant experimental settings. Choosing appropriate data pre-treatment routines and developing accurate and robust calibration models is essential to ensure good prediction performance of the instruments in on-line conditions. The desired compromise is to attain a suitable equilibrium between prediction precision and the model's robustness in future applications.

The goal of this work is to evaluate the performance of three techniques of calibrating capacitance measurements for on-line biomass monitoring. The first method involves the application of the Cole–Cole equation, a reference theoretical representation of the dielectric behavior of cell suspensions. The remaining two methods are based on purely empirical modeling of dielectric signals: direct linear correlation of capacitance measurements to biomass concentration and multivariate modeling of capacitance spectra for the estimation of biomass concentration and mean cell size.

The article begins with a brief theoretical introduction outlining the major characteristics of capacitance spectroscopy. The following section presents the three calibration approaches proposed in this study. The experimental segment of the work involves a quantitative assessment of the models' performance in strictly predictive conditions. Aerobic batch cultures of *Kluyveromyces marxianus* and *Saccharomyces cerevisiae* in two

laboratory-scale bioreactors are used as case studies for validation and performance evaluation. The discussion aims to provide a methodology of choosing the appropriate calibration approach in order to increase the overall utility of capacitance spectroscopy in on-line biomass monitoring and control and eliminate the need for frequent instrument recalibration and post-run measurement adjustments.

Principles of dielectric spectroscopy

The theoretical principles of dielectric spectroscopy in the context of biotechnology have been described in detail by various authors [8, 13, 15–17], so only the major points are summarized here. In essence, dielectric studies are based on quantifying the response of a material to an electric field applied to it. The response is typically described by the material's conductivity and permittivity. Conductivity (σ), measured in S/m, quantifies of the ability of the material to conduct the electrical charge. Permittivity (ε), measured in F/m, is the amount of charge that is stored by the material due to the polarization of its components. Permittivity of the material is often expressed as relative to the permittivity of vacuum ($\varepsilon_0 = 8.854 \times 10^{-12}$ F/m), giving the dimensionless relative permittivity (also called dielectric constant), $\varepsilon_T = \varepsilon/\varepsilon_0$. By dividing conductivity and permittivity by the probe constant (d/A in m^{-1}, the ratio of the distance between the electrodes and the electrodes' area), one obtains the corresponding conductance (G in S) and capacitance (C in F) of the material, respectively.

The permittivity of a material tends to fall (and its conductivity to rise) in a series of step-like shifts as the frequency of the electrical field rises. These step changes, called dispersions, are due to losses of certain characteristic polarization abilities of the substance [8, 13]. In the case of cell suspensions, three major dispersions are identified: the α-, β- and γ-dispersions. The α-dispersion is caused predominantly by the activity of ions by the charged surfaces of cells and particles. The γ-dispersion is due mainly to the bipolar rotation of water molecules. Of particular interest

in biomass quantification is the β-dispersion, resulting from the build-up of electrical charge at the cell membranes.

Under the influence of an electric field applied to a cell suspension, the ions present in the electrolytic medium migrate towards the electrodes. The cytoplasm of the cells is also conducting but due to the presence of the non-conducting plasma membrane, the charged ions inside the cells are constrained to the cell volume. Trapped inside the membrane, the ions accumulate at the sides of the cell, and the cell becomes polarized as shown in Fig. 1. Clearly, only cells with undamaged membranes capable of electrical insulation contribute to the increase in capacitance. Most dead cells autolyse shortly after death and their membranes rupture, while non-cellular material cannot store electrical charge. Thus, only viable biomass is measured. Each living cell in the suspension assumes the behavior of a tiny electrical capacitor and the overall capacitance of the suspension rises as a function of the total biovolume (i.e. the volume fraction of the suspension which is enclosed by an intact membrane).

Measuring the capacitance over a predetermined range of electrical field frequencies is the basic idea of scanning dielectric spectroscopy. The typical frequency range used in bioprocesses monitoring is in the order of 0.1–10 MHz, where the β-dispersion occurs. At the lower frequencies of this range, there is enough time for electrical charge to build up at the cell membranes. However, at the high-frequency end of the spectrum, the electrical field changes direction too rapidly for the cell membrane to polarize, and the biomass no longer contributes to the measured capacitance. The net rise from the background capacitance (C_∞) at high frequencies to the increased capacitance at low frequencies is expressed as ΔC and can be attributed to the charge-storing properties of the biomass. The frequency corresponding to half of the measured ΔC is called the characteristic frequency (f_c). A typical capacitance spectrum obtained with scanning dielectric spectroscopy is illustrated in Fig. 2.

One of the most common theoretical ways of describing the dielectric properties of cell suspensions and the β-dispersion is to use the Cole–Cole equation [26]. The
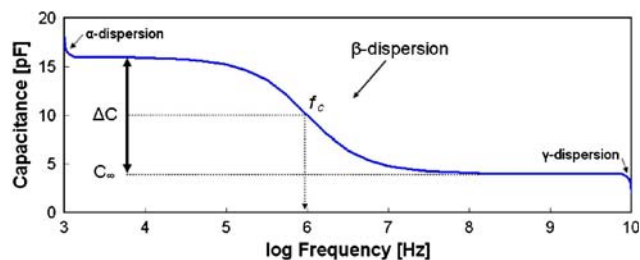
Fig. 1 Cell polarization at the plasma membrane
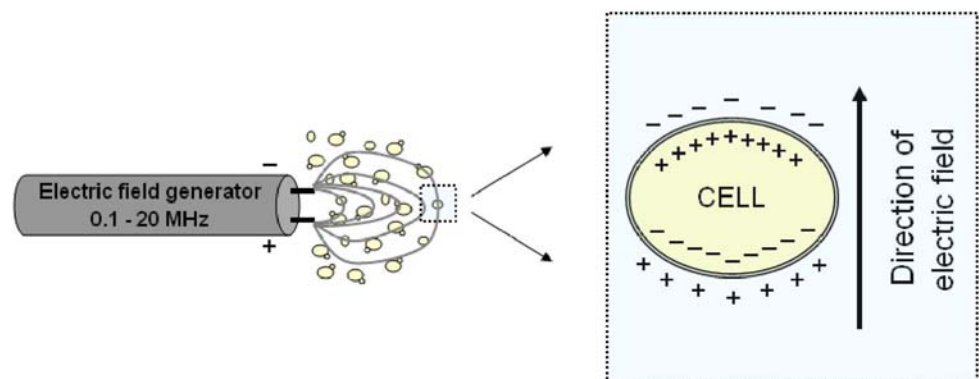
Fig. 2 Typical capacitance spectrum obtained with scanning dielectric spectroscopy

Fig. 3 The shape of β-dispersion with changing values of the Cole–Cole α

Cole–Cole equation is itself based on the Debye equation [27] which is derived from theoretical modeling of a polarized object. The Debye equation assumes that the polarization of materials decays exponentially when the applied electric field is removed. The Cole–Cole equation recognizes that in reality not many systems obey the Debye model and introduces an empirical fitting parameter, the so-called Cole–Cole α, which has the effect of broadening the dispersion. The equation models the shape of the β-dispersion graph in terms of its magnitude, ΔC, its characteristic frequency, f_c, its high frequency component, C_∞, and the Cole–Cole α:

$$C(f) = \frac{\Delta C \left(1 + \left(\frac{f}{f_c}\right)^{(1-\alpha)} \sin\left(\frac{\pi}{2}\alpha\right)\right)}{\left(1 + \left(\frac{f}{f_c}\right)^{(2-2\alpha)} + 2\left(\frac{f}{f_c}\right)^{(1-\alpha)} \sin\left(\frac{\pi}{2}\alpha\right)\right)} + C_\infty \quad (1)$$

The Debye equation can be regained from the above Cole–Cole expression by setting α equal to zero. Fig. 3 shows the effects of changing the value of the Cole–Cole parameter. The four curves are calculated using constant parameter values of $\Delta C = 20$ pF, $f_c = 5$ MHz and $C_\infty = 10$ pF and then altering the Cole–Cole α from 0 to 0.6. It can be clearly seen that increasing this value has the effect of broadening the dispersion significantly.

The physical origin of the empirical value α is disputed and no single convincing explanation for its value has been discovered. It seems likely that the value of α is often due to a variety of effects, some of which are more or less important in different systems. The following are some of the suggested origins of this parameter's value: distribution of cell shapes and sizes [28], morphology of extra-cellular spaces [29], mobility of membranous proteins [30, 31] and the fractal nature of dielectric relaxation [32]. Typical values of α for biological cells are in the order of 0.1 to 0.2 [13].

Modeling of capacitance data

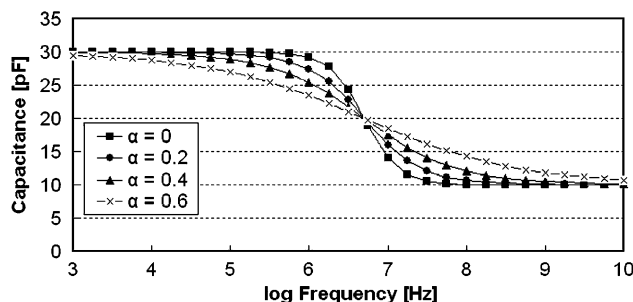Because of the sensitivity of dielectric signals to equipment setup and process conditions, capacitance instruments are usually calibrated in-situ. A calibration experiment representative of future applications is carried out, and samples are collected at regular intervals to provide standards for the calibration model. The three techniques used for modeling the capacitance data and calibrating the spectrometer are described below.

Physical modeling of dielectric properties of cell suspensions

The physical modeling algorithm predicts cell size and concentration using a three-stage approach [33]. In the first step, the Cole–Cole equation is fitted to experimental permittivity data [34]. The permittivity formulation is used instead of the capacitance one since permittivity is a property of the material whilst capacitance depends on both material and geometry.

Four variables are fixed at this stage: $\Delta\varepsilon$, the difference between low and high frequency permittivity; ω_c, the characteristic angular frequency in radians per second; α, the Cole–Cole parameter; and ε_∞, the high frequency (background) permittivity.

$$\varepsilon(\omega) = \frac{\Delta\varepsilon \left(1 + \left(\frac{\omega}{\omega_c}\right)^{(1-\alpha)} \sin\left(\frac{\pi}{2}\alpha\right)\right)}{\left(1 + \left(\frac{\omega}{\omega_c}\right)^{(2-2\alpha)} + 2\left(\frac{\omega}{\omega_c}\right)^{(1-\alpha)} \sin\left(\frac{\pi}{2}\alpha\right)\right)} + \varepsilon_\infty \quad (2)$$

In the second step, the Cole–Cole equation for conductivity is fitted to experimental conductivity data [35]:

$$\sigma(\omega) = \frac{-\Delta\sigma \left(1 + \left(\frac{\omega}{\omega_{c,2}}\right)^{(1-\alpha_2)} \sin\left(\frac{\pi}{2}\alpha_2\right)\right)}{\left(1 + \left(\frac{\omega}{\omega_{c,2}}\right)^{(2-2\alpha_2)} + 2\left(\frac{\omega}{\omega_{c,2}}\right)^{(1-\alpha_2)} \sin\left(\frac{\pi}{2}\alpha_2\right)\right)} + (\sigma_L + \Delta\sigma). \quad (3)$$

Fixed at this stage are the following four variables: $\Delta\sigma$, the difference between high and low conductivity; $\omega_{c,2}$ is the characteristic angular frequency for conductivity; α_2, the Cole–Cole parameter for conductivity; and σ_L, the low

frequency conductivity. It should be noted that the values of the angular frequency and the Cole–Cole α are different in Eqs. 2 and 3. Fig. 4 summarizes, schematically, the eight parameters determined at the first two stages of the algorithm.

In the third step of the algorithm, the values of the cell radius (r) and the cell number density (N_v) are obtained using the iterative process described below. To relate the magnitude of $\Delta\varepsilon$ and the characteristic frequency to the properties of the cells, the following pair of equations based on the Pauly-Schwan spherical cell model can be used [15, 19, 36]:

$$\Delta\varepsilon = \frac{3 N_v \pi r^4 C_m}{\varepsilon_0} \qquad (4)$$

$$\omega_c = \frac{1}{r C_m \left(\frac{1}{\sigma_i} + \frac{1}{2\sigma_e} \right)} \qquad (5)$$

where r is the cell radius, N_v is the number density (cells per unit volume), C_m is membrane capacitance per unit area, σ_i is the internal conductivity of the cell and σ_e is the conductivity of the suspending medium. The values of $\Delta\varepsilon$ and ω_c are known from the first stage. The parameters C_m and σ_i are either known from a reference source or can be determined through calibration. The last remaining unknown, σ_e, can be calculated using the following model [36]:

$$\sigma_e = \frac{\sigma_L}{(1 - P)^{1.5}} \qquad (6)$$

where P is the volume fraction of cells ($P = \frac{4}{3}\pi r^3 N_v$) and σ_L is known from the second stage. Thus, the problem contains three non-linear equations (Eq. 4–6) and three unknowns: r, N_v and σ_e. The iteration starts by choosing an initial value of σ_e. A good initial estimate is the value of the high-frequency conductivity:

$$\sigma_e \approx \sigma_L + \Delta\sigma. \qquad (7)$$

Note that at low volume fractions (little biomass), $\Delta\sigma$ will be close to zero and the external conductivity will be similar to the low-frequency conductivity $\sigma_e \approx \sigma_L$. Having the initial value for of σ_e, the cell radius is calculated using Eq. 5 and the number density is obtained from Eq. 4. In the next step of the iteration, σ_e is calculated using Eq. 6 and the procedure is repeated until convergence.

To improve the algorithm's predictions further, constraints can be added to limit the solutions to some predefined plausible ranges. For example, the estimates of the cell radius can be confined to fall within the range expected for the particular cell type.

Linear modeling of dielectric signals

Linear modeling is the simplest and most common empirical calibration method used in dielectric spectroscopy. The approach is based on determining a linear correlation between capacitance measured in the β-dispersion region and biomass concentration or cell number. The frequency at which the capacitance is measured depends on the organism. Typically, excitation frequencies of around 1,000 kHz are used for bacteria cells and between 500 and 600 kHz for yeasts and mammalian cells [20, 37]. A somewhat more sophisticated technique is the dual-frequency method, where the background capacitance of the medium (C_∞) measured at an elevated frequency (>10 MHz) is systematically subtracted from the measured capacitance values. In this case, the linear correlation is derived between values of delta-capacitance (see ΔC in Fig. 2) and biomass concentration. This approach corrects for potential baseline shifts during the process.

Multivariate modeling of scanning dielectric signals

Scanning dielectric spectroscopy creates the opportunity for more advanced modeling techniques based on multivariate analysis and chemometrics. Chemometric regression methods, described in detail elsewhere [38–41] work by decomposing multivariate data sets into a reduced form containing more informative principal components

Fig. 4 Parameters established by fitting the Cole–Cole permittivity and conductivity equations to experimental data
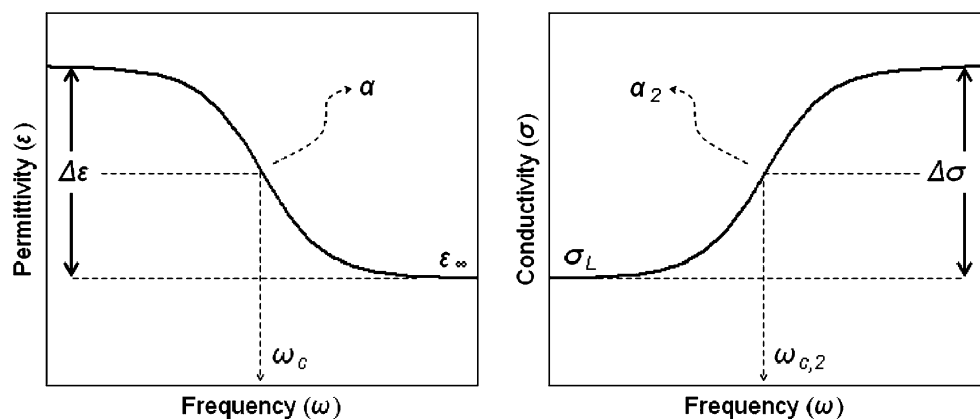
that describe the major trends present in the data. Applying this technique in spectroscopy allows the extraction of latent patterns from spectra and using this information to model specific variables that are difficult to quantify directly, often due to their intrinsic interactions. For example, since the dielectric properties of a cell suspension are dependent actually on the amount of biomass volume present in the system, linear correlations between capacitance and biomass weight concentration could fail if the cells change size during the process. Multivariate analysis can be used to model biomass concentration and cell size as two separate variables by exploiting the distinctive shapes of the capacitance spectra. A parameter of particular importance is the position the characteristic frequency, which should ideally be indicative of the morphology and average size of cells in the system [18, 19, 42]. Smaller cells are more readily polarized so the characteristic frequency of suspensions containing these cells will be greater than that of a suspension of larger cells (Fig. 5). Using a partial least squares (PLS) model, Cannizzaro et al. [18, 23] succeeded in estimating the median size of yeast and mammalian cells, as well as proposed a way of detecting important changes in the process by analyzing the scores and loadings of the model and graphing capacitance phase plots.

Just like in linear modeling, the residual capacitance at a high frequency (C_∞) can be subtracted from capacitance values at lower frequencies to eliminate the effects of potential baseline shifts.

Experimental

In total, eight aerobic batch experiments were performed in this study using two bioreactor settings and two types of wild type yeast obtained from the Centraalbureau voor Schimmelcultures (Utrecht, NL): the Crabtree-negative strain CBS 5670 *Kluyveromyces marxianus* and the Crabtree-positive strain CBS 8066 *Saccharomyces cerevisiae*.
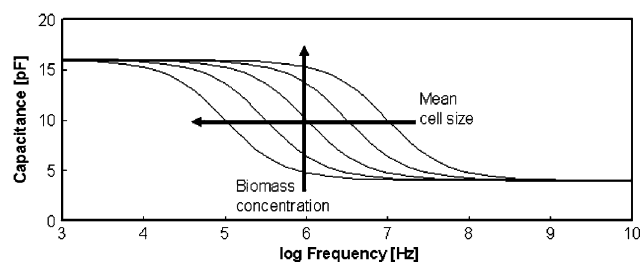

Fig. 5 Idealistic representation of how biomass concentration and mean cell size is estimated in scanning dielectric spectroscopy using multivariate modeling

Pre-cultures and growth medium

Source cells were stored at -80 °C in 1.8 ml aliquots. For each batch, the reaction inoculum was obtained by adding one aliquot into a 1-l Erlenmeyer flask containing 100 ml of a sterile complex pre-culture medium (10 g/l yeast extract OXOID, 10 g/l peptone BACTO and 20 g/l glucose) and incubating it for 24 h at 30 °C and 200 rpm. The defined culture medium was sterilized by filtration and contained, per liter: 20 g glucose, 5 g $(NH_4)_2SO_4$, 3 g KH_2PO_4, 0.5 g $MgSO_4 \times 7H_2O$, as well as trace elements and vitamins (adapted from Verduyn et al. [43] and Cannizzaro et al. [44]). The medium was supplemented with 0.5 ml/l of a standard antifoam agent to prevent foaming.

Culture conditions

The four batch cultures of *K. marxianus* (KMB01–KMB04) were cultivated in a fully automated 2-l laboratory bioRC1 calorimeter from Mettler Toledo (Greifensee, Switzerland). The batch cultures of *S. cerevisiae* (SCB01–SCB04) were grown in a 3.6-l laboratory bioreactor from Bioengineering (Wald, Switzerland). Both vessels were equipped with a Rushton-type agitator, baffles, temperature and pH probes and control mechanisms, gas inlet and outlet ports, a base inlet port and a sampling port. The reactors were sterilized in-situ at 121 °C for 20 min. All cultures were grown at 30 °C with an agitation speed of 800 rpm and an inlet air flow rate of 1.3 *vvm*. A solution of 2 M NaOH or KOH was used to control the pH at 5; no acid control was necessary.

Reference measurements

Samples of about 10 ml were collected at intervals between 1 and 2 h using an in-house developed automated sampling robot, BioSampler 2002 [23]. Dry cell weight (DCW) was determined by putting 8 ml of the culture medium through a pre-weighed 0.22 μm pore filter, drying the filter and subsequently reweighing it. Optical density measurements were performed as a backup method at 600 nm using the Spectronic Helios-Epsilon spectrophotometer from Thermo (Waltham, MA, USA).

Cell size distribution data were obtained using a Coulter Counter Model ZM equipped with a Channelyzer 256 (Coulter Electronics Limited, UK). An orifice tube of 70 μm was used, and the instrument settings were the following: current: 200 μA, gain: 2, attenuation: 8, $K_c = 6.811$. The instrument was calibrated with latex beads of 5.06 μm diameter before using it the first time and the orifice tube was rinsed before each use. For each measurement, the mean cell volume and diameter were calculated using the

following equations obtained from the instrument's reference manual:

$$V_{\mathrm{ave}} = \frac{1}{N_t} \sum_{i=4}^{255} V_i N_i \tag{8}$$

$$d_{\mathrm{ave}} = \sqrt[3]{\frac{6 \, V_{\mathrm{ave}}}{\pi}} \tag{9}$$

where N_t is the total number of cells in all channels, N_i is the number of cells in channel i and V_i is the volume corresponding to channel i. To eliminate noise caused by small particles, channels 1 through 3 were omitted. Each sample was analyzed twice, and an average value was taken of the two measurements.

Capacitance spectrometer

The dielectric instrument used in this study is the Biomass Monitor 210 from Aber Instruments (Aberystwyth, UK). The spectrometer was equipped with a 12 mm sterilizable probe containing four annular electrodes. This configuration is particularly favorable, as four-terminal probes (as opposed to two-terminal probes) may reduce electrode polarization [17, 36]. The probe was introduced directly into the reactor and sterilized in-situ. The biomass monitor was switched on 3 h before starting the experiments to allow stabilization of the signal. During the cultures, 25 frequencies from 0.1 to 20 MHz were scanned every 15 s and the capacitance as well as the conductivity of the cell suspension was registered at each frequency. A program developed in-house using LabView (National Instruments, Austin, TX) was used to collect and store the measured data. Due to the significant level of noise, all data were smoothed with respect to time using the Savitzky-Golay algorithm over a moving window of 81 points (20 min).

Spectrometer calibration results

The first batch of each cell strain (KMB01 and SCB01) served to collect calibration data sets in-situ. Fourteen calibration samples were used for KMB01 and seventeen for SCB01. The reference measurements of biomass concentration and mean cell diameter were obtained using the methods described above. The performance of all calibration models was evaluated by calculating the standard error of calibration (SEC):

$$\mathrm{SEC} = \sqrt{\frac{\sum_{i=1}^{n_C} (\hat{y}_i - y_i)^2}{n_C}} \tag{10}$$

where $\hat{y}$ and y are the reference and model-predicted values, respectively, and n_C is the number of calibration samples.

Cole–Cole model

The Cole–Cole modeling algorithm was designed and implemented in the Java programming language using the Levenberg-Marquardt non-linear least squares technique [45]. The values of $\Delta\varepsilon$ and $\Delta\sigma$ were constrained to be positive, the ω_c values were confined to the range 0–1×10^{10} rad/s and the Cole–Cole α values were limited to within 0–0.5. The cell radius was constrained to 2–3 microns for *K. marxianus* and 2–3.5 microns for *S. cerevisiae*, while the number density was constrained to be greater than 1×10^{10} cells/m^3. The values of C_m and σ_i were determined using the calibration data sets and then applied to the remaining validation data sets.

The standard error of calibration obtained with the Cole–Cole model during KMB01 was 1.44 g/l for biomass and 0.32 μm for cell diameter. For SCB01, these values were 0.45 g/l and 0.40 μm, respectively. Figure 6 shows the fit obtained for these calibration batches.

Linear model

Linear calibration models were built using the Excel spreadsheet program. The dual-frequency mode was used and delta-capacitance values (ΔC) were obtained by subtracting the capacitance reading at the background frequency of 15.56 MHz (C_∞) from the capacitance reading at the excitation frequency of 370 kHz. Calibration was performed by obtaining a linear correlation coefficient between biomass dry cell weight and the corresponding values of ΔC. Negative concentration values were zeroed. The standard error of calibration for the linear models was 0.64 g/l for KMB01 and 0.42 g/l for SCB01. Figure 7 shows the fit obtained for the two calibration cultures.

Multivariate model

Multivariate modeling was carried out in Matlab (The MathWorks, Inc., Natick, MA, USA). Capacitance spectra were collected over 18 frequency points between 370 kHz and 15.56 MHz and the background capacitance reading at 15.56 MHz (C_∞) was subtracted from all capacitance readings at lower frequencies. All data were mean-centered. A PLS model was built for biomass dry cell weight and mean cell diameter using the PLS_Toolbox 4.1 (Eigenvector Research, Inc., Wenatchee, WA, USA). Two latent variables were used and explained 99.1% of the variance in the calibration spectra from KMB01 and 99.6% in those from SCB01. Negative concentration values were zeroed. The standard errors of calibration for the multivariate models were 0.37 g/l (biomass concentration) and 0.09 μm (cell diameter) for KMB01. For SCB01, these values were, respectively, 0.31 g/l and 0.20 μm. The fit

Fig. 6 Cole–Cole calibration models for biomass concentration (*triangles*) and mean cell diameter (*circles*) developed for *K. marxianus* (**a**) and *S. cerevisiae* (**b**)
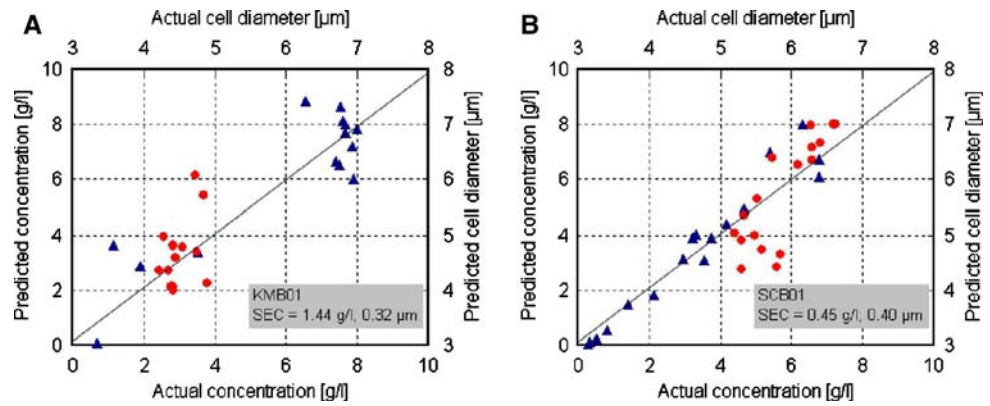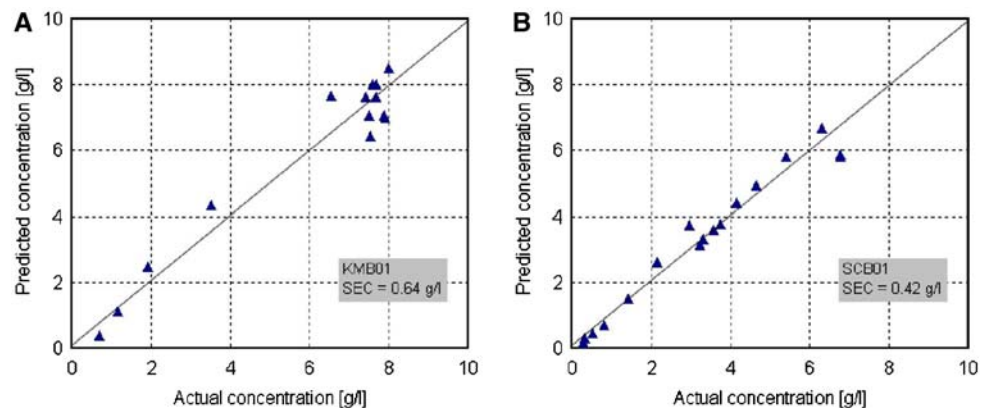

Fig. 7 Linear calibration models for biomass concentration developed for *K. marxianus* (A) and *S. cerevisiae* (B)


obtained with the PLS model for the two strains can be seen in Fig. 8.

Signal noise

From the calibration results of all three models, it appeared that signal noise in the smaller bioRC1 reactor (*K. marxianus* batches) was stronger than in the larger Bioengineering reactor (*S. cerevisiae* cultures). To verify this observation, the standard deviation of the filter residuals obtained for both cultures with the Savitzky-Golay smoothing algorithm was plotted in Fig. 9 as a function of frequency.

The noise in batch KMB01 was considerably higher, most likely due to the limited space in the bioRC1 vessel resulting in increased interference of reactor components with the probe's field of activity. Indeed, the radius of the probe's field of detection is typically around 3 cm according to the manufacturer, while in the bioRC1 reactor, the probe was located only about 1 cm from the reactor walls, 1 cm from the baffles and 2 cm from agitator. In the larger Bioengineering vessel, the capacitance probe was somewhat more isolated: 1 cm from the reactor walls and 2.5 cm from the metal baffles and agitator. It should also be noted that in both reactors the signal noise decreases with

Fig. 8 Multivariate (PLS) calibration models for biomass concentration (*triangles*) and mean cell diameter (*circles*) developed for *K. marxianus* (**a**) and *S. cerevisiae* (**b**)
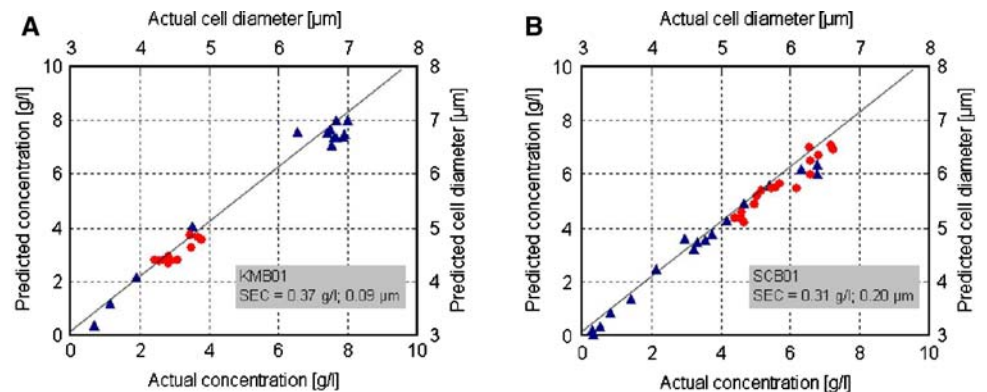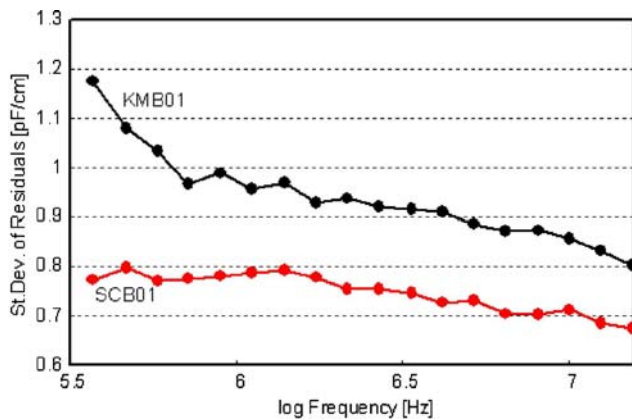
Fig. 9 Standard deviation of the filter residuals obtained for the two calibration cultures

increasing frequency. This phenomenon is due to the more acute influence of electrode polarization on capacitance measurements and increased sensitivity to phase noise at lower excitation frequencies.

Validation results

Following the calibration cultures, the three succeeding batches for each strain (KMB02-04 and SCB02-04) were used for evaluating the models. Baseline synchronization with respect to the calibration culture was performed at the moment of inoculation. The predictive performance of all models was evaluated by calculating the standard error of prediction (SEP) using validation samples obtained for each cell strain during the validation cultures:

$$SEP = \sqrt{\frac{\sum_{i=1}^{n_P} (\hat{y}_i - y_i)^2}{n_P}} \tag{11}$$

where $\hat{y}$ and y are the reference and predicted values, respectively, and n_P is the number of validation samples.
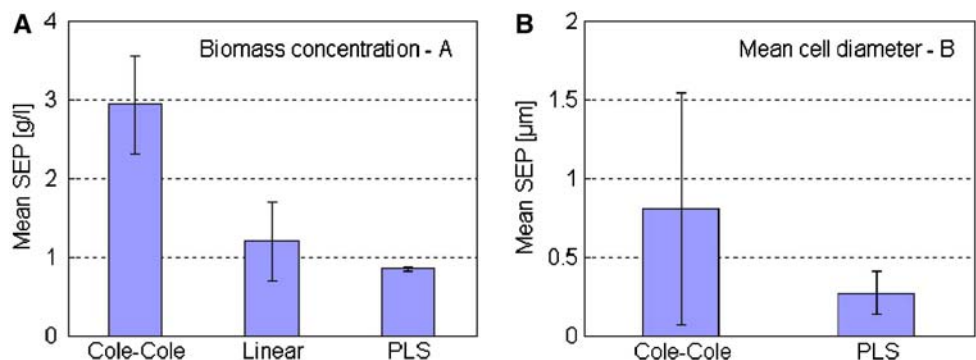
Kluyveromyces marxianus batches

The average standard errors of prediction obtained for the three validation batches of *K. marxianus* are shown in Fig. 10.

In the estimation of biomass concentration, the Cole–Cole model produced prediction errors that were significantly higher than those obtained with the linear and multivariate models. The linear model was somewhat less accurate and robust than the PLS model. A possible reason for this is that signal noise was stronger at lower values of the frequency range (see Fig. 9), where the linear model had been calibrated. The multivariate model, being calibrated over the entire spectrum was influenced by signal noise to a lesser extent. Finally, the PLS model was also more accurate than the Cole–Cole model in the prediction of mean cell diameter.

The predicted profiles of biomass concentration and mean cell diameter obtained for the *K. marxianus* cultures are shown in Fig. 11.

Saccharomyces cerevisiae batches

The average standard errors of prediction obtained for the three validation batches of *S. cerevisiae* are shown in Fig. 12.

In the prediction of biomass concentration, the prediction errors of all three models were considerably lower for the *S. cerevisiae* batches, compared to the errors obtained for the *K. marxianus* batches. This is most likely due to the lower level of signal noise observed in the larger Bioengineering reactor. The Cole–Cole algorithm performed nearly as well as the linear and multivariate methods, although the increased standard deviation of the prediction error could point to a lower level of robustness. Also owing to the relatively constant noise across the frequency range (see Fig. 9), the linear model performed on par with the PLS model. The PLS model was again more accurate and more stable than the Cole–Cole model in the prediction of mean cell diameter.

Fig. 10 Mean standard error of prediction obtained for the *K. marxianus* validation batches using the three models for biomass concentration (**a**) and mean cell diameter (**b**)

The predicted profiles of biomass concentration and mean cell diameter obtained for the *S. cerevisiae* cultures are shown in Fig. 13.

Discussion and outlook

On-line monitoring of biomass using capacitance spectroscopy offers great potential for process development, optimization and control studies. Yet, as is the case in other spectroscopic methods, the development of accurate and robust calibration routines often remains the Achilles' heel of the technique. In this study, two direct calibration techniques based on linear and multivariate (PLS) modeling of capacitance data were compared to the theoretical model of cell suspensions based on the Cole–Cole equation. Validation results involving six yeast fermentations revealed that the linear and PLS models were generally more robust and outperformed the Cole–Cole model, especially in the more noisy conditions of a smaller bioreactor. For the estimation of biomass concentration, the linear and multivariate models provided similar results in experimental settings involving lower levels of signal noise. However, in more noisy conditions, the PLS model
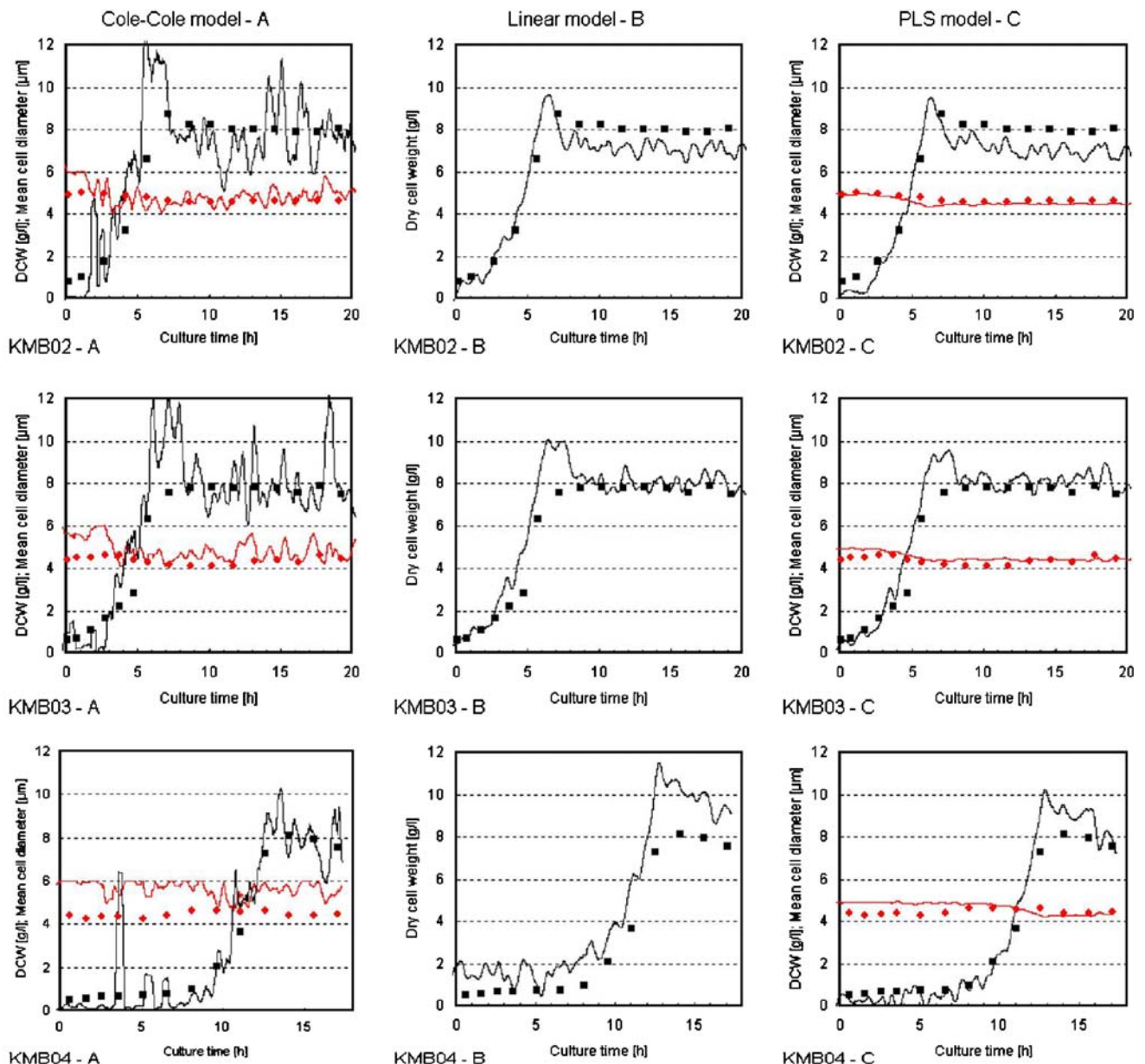

Fig. 11 Predicted profiles of biomass concentration (square reference points) and mean cell diameter (round reference points) obtained for the three validation batches of *K. marxianus* with the Cole–Cole model (**a**), the linear model (**b**) and the PLS model (**c**)

Fig. 12 Mean standard error of prediction obtained for the *S. cerevisiae* validation batches using the three models for biomass concentration (**a**) and mean cell diameter (**b**)
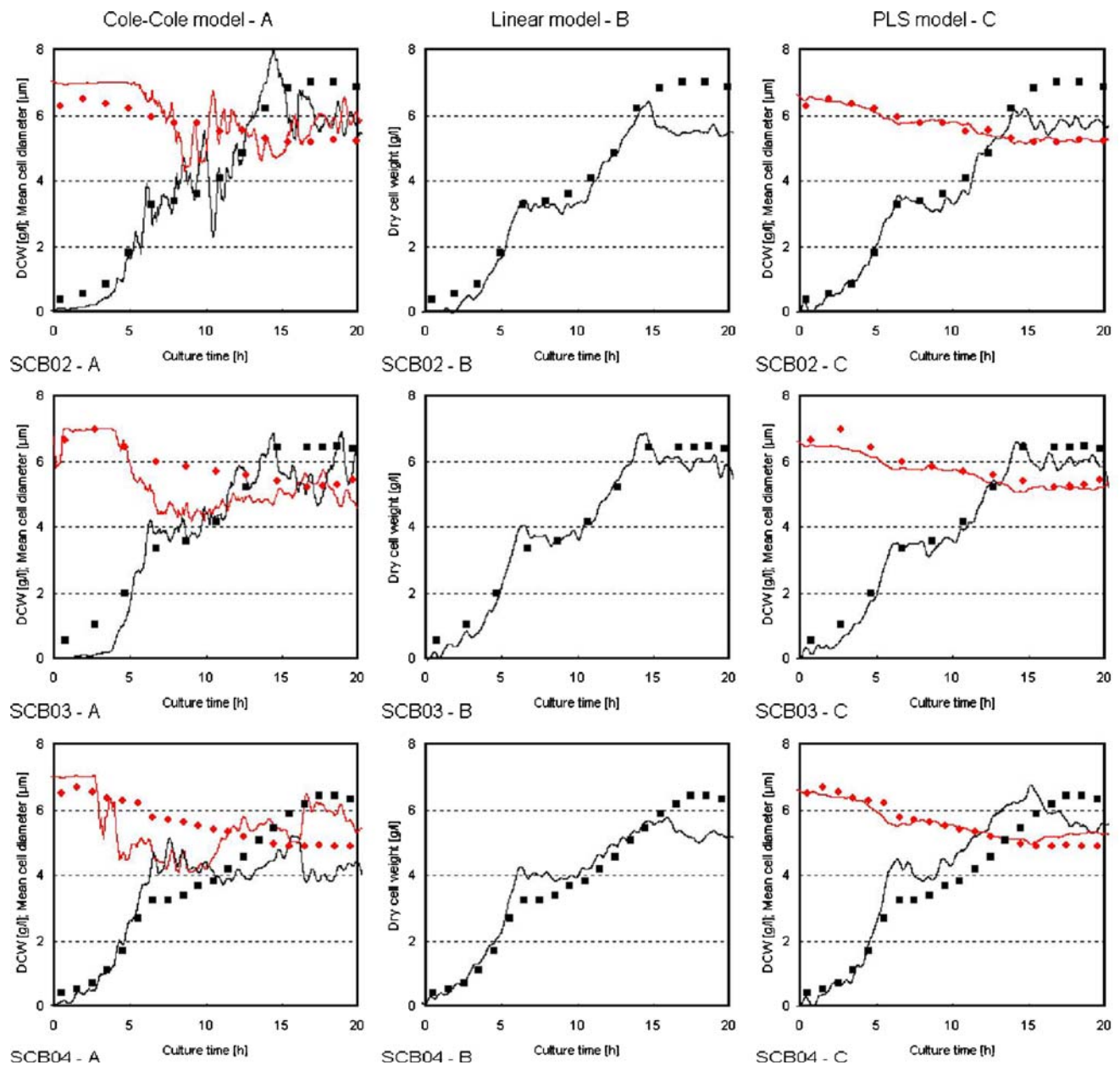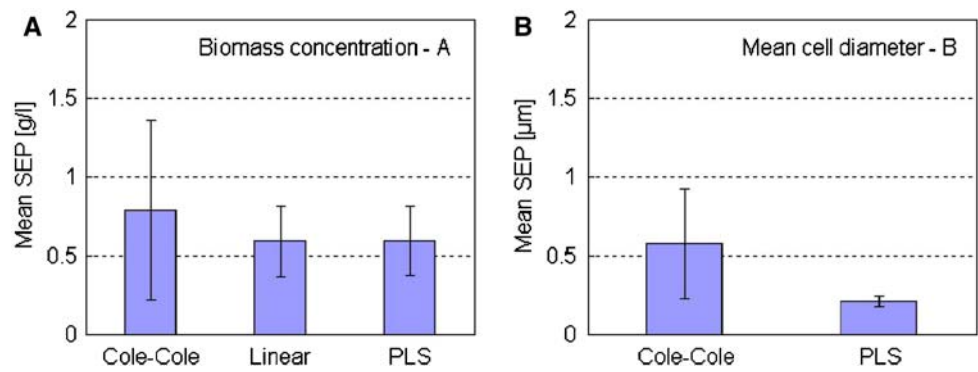

Fig. 13 Predicted profiles of biomass concentration (square reference points) and mean cell diameter (round reference points) obtained for the three validation batches of *S. cerevisiae* with the Cole–Cole model (**a**), the linear model (**b**) and the PLS model (**c**)

showed greater robustness and had lower prediction errors. The linear calibration was more sensitive to noise since the correlation was established at a low excitation frequency where the effects of electrode polarization, phase noise and other interferences are more pronounced. The PLS model was built over the entire frequency domain, which resulted in a greater noise averaging capacity. Predictions of the average cell diameter were also achieved using the Cole–Cole and PLS models, with the latter technique giving more accurate results.

In addition to the comparison of calibration methods, the study exposed the impact of the reactor size on noise in the dielectric signal. In the smaller vessel studied, where the probe was closer to various reactor components, the obtained signal was considerably noisier and more difficult to model than in the larger reactor, where the probe had more space. Hence, in small-scale in-situ applications of capacitance spectroscopy, care should be taken when positioning the probe so that it is placed as distant as possible from the reactor's walls, baffles, agitator and other components, especially metal objects, as their presence close to the probe's field of detection may cause interference in the signal. Because of this problem, it is expected that the performance of dielectric spectroscopy in microreactor applications may be limited.

The modeling of the average cell size needs to be studied more carefully and over a wider range of cell sizes. Furthermore, future studies should seek to exploit the latent information behind the characteristic shape of the capacitance spectra. The width of the inflection point (described by the Cole–Cole α) has been shown to provide some insight into the distribution of cell sizes in the suspension [28]. With sufficiently low signal noise and a higher frequency resolution, cell size distribution could be modeled using an appropriate distribution function and multivariate analysis.

References

1. Vojinovic V, Cabral JMS, Fonseca LP (2006) Real-time bioprocess monitoring. Part I: In situ sensors. Sensors and Actuators B Chemical 114(2):1083–1091
2. Schügerl K (2001) Progress in monitoring, modeling and control of bioprocesses during the last 20 years. J Biotechnol 85(2):149–173
3. von Stockar U, Valentinotti S, Marison I, Cannizzaro C, Herwig C (2003) Know-how and know-why in biochemical engineering. Biotechnol Adv 21(5):417–430
4. Olsson L, Nielsen J (1997) On-line and in situ monitoring of biomass in submerged cultivations. Trends Biotechnol 15(12):517–522
5. Camisard V, Brienne JP, Baussart H, Hammann J, Suhr H (2002) Inline characterization of cell concentration and cell volume in agitated bioreactors using in situ microscopy: Application to volume variation induced by osmotic stress. Biotechnol Bioeng 78(1):73–80
6. Konstantinov K, Chuppa S, Sajan E, Tsai Y, Yoon S, Golini F (1994) Real-time biomass-concentration monitoring in animal-cell cultures. Trends Biotechnol 12(8):324–333
7. Surribas A, Montesinos JL, Valero FF (2006) Biomass estimation using fluorescence measurements in Pichia pastoris bioprocess. J Chem Technol Biotechnol 81(1):23–28
8. Kell DB, Markx GH, Davey CL, Todd RW (1990) Real-time monitoring of cellular biomass-methods and applications. Trac-Trends Analyt Chem 9(6):190–194
9. Tamburini E, Vaccari G, Tosi S, Trilli A (2003) Near-infrared spectroscopy: A tool for monitoring submerged fermentation processes using an immersion optical-fiber probe. Appl Spectrosc 57(2):132–138
10. Arnold SA, Gaensakoo R, Harvey LM, McNeil B (2002) Use of at-line and in-situ near-infrared spectroscopy to monitor biomass in an industrial fed-batch *Escherichia coli* process. Biotechnol Bioeng 80(4):405–413
11. Hall JW, McNeil B, Rollins MJ, Draper I, Thompson BG, Macaloney G (1996) Near-infrared spectroscopic determination of acetate, ammonium, biomass, and glycerol in an industrial Escherichia coli fermentation. Appl Spectrosc 50(1):102–108
12. Joeris K, Frerichs JG, Konstantinov K, Scheper T (2002) In-situ microscopy: online process monitoring of mammalian cell cultures. Cytotechnology 38(1–2):129–134
13. Markx GH, Davey CL (1999) The dielectric properties of biological cells at radiofrequencies: applications in biotechnology. Enzyme Microb Technol 25(3–5):161–171
14. Mishima K, Mimura A, Takahara Y, Asami K, Hanai T (1991) On-line monitoring of cell concentrations by dielectric measurements. J. Ferment Bioeng 72(4):291–295
15. Yardley YE, Kell DB, Barrett J, Davey CL (2000) On-line, real-time measurements of cellular biomass using dielectric spectroscopy. Biotechnology & Genetic Engineering Reviews, vol 17. Intercept Ltd. Scientific Technical & Medical Publishers, Andover, pp 3–35
16. Davey CL, Davey HM, Kell DB, Todd RW (1993) Introduction to the dielectric estimation of cellular biomass in real-time, with special emphasis on measurements at high-volume fractions. Anal Chim Acta 279(1):155–161
17. Harris CM, Todd RW, Bungard SJ, Lovitt RW, Morris JG, Kell DB (1987) Dielectric permittivity of microbial suspensions at radio frequencies: a novel method for the real-time estimation of microbial biomass. Enzyme Microb Technol 9(3):181–186
18. Cannizzaro C, Gugerli R, Marison I, von Stockar U (2003) On-line biomass monitoring of CHO perfusion culture with scanning dielectric spectroscopy. Biotechnol Bioeng 84(5):597–610
19. Siano SA (1997) Biomass measurement by inductive permittivity. Biotechnol Bioeng 55(2):289–304
20. Ducommun P, Kadouri A, von Stockar U, Marison IW (2002) On-line determination of animal cell concentration in two industrial high-density culture processes by dielectric spectroscopy. Biotechnol Bioeng 77(3):316–323
21. Cerckel I, Garcia A, Degouys V, Dubois D, Fabry L, Miller AOA (1993) Dielectric-spectroscopy of mammalian-cells .1. evaluation of the biomass of hela-cell and cho-cell in suspension by low-frequency dielectric-spectroscopy. Cytotechnology 13(3):185–193
22. Davey CL, Kell DB, Kemp RB, Meredith RWJ (1988) On the audio- and radio-frequency dielectric behaviour of anchorage-independent, mouse L929-derived LS fibroblasts. Bioelectrochem Bioenerg 20(1–3):83–98

23. Cannizzaro C (2002) Spectroscopic monitoring of bioprocesses: A study of carotenoid production by *Phaffia Rhodozyma* Yeast [PhD thesis]. Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

24. November EJ, Van Impe JF (2000) Evaluation of on-line viable biomass measurements during fermentations of Candida utilis. Bioprocess Eng 23(5):473–477

25. Asami K, Yonezawa T (1996) Dielectric behavior of wild-type yeast and vacuole-deficient mutant over a frequency range of 10 kHz to 10 GHz. Biophys J 71(4):2192–2200

26. Cole KS, Cole RH (1941) Dispersion and absorption in dielectrics I. Alternating current characteristics. J Chem Phys 9(4):341–351

27. Debye P (1929) Polar Molecules. The Chemical Catalog Company Inc, New York

28. Markx GH, Davey CL, Kell DB (1991) To what extent is the magnitude of the cole–cole-alpha of the beta-dielectric dispersion of cell-suspensions explicable in terms of the cell-size distribution. Bioelectrochem Bioenerg 25(2):195–211

29. Ivorra A, Genesca M, Sola A, Palacios L, Villa R, Hotter G, Aguilo J (2005) Bioimpedance dispersion width as a parameter to monitor living tissues. Physiol Meas 26(2):S165–S173

30. Kell DB, Harris CM (1985) On the dielectrically observable consequences of the diffusional motions of lipids and proteins in membranes. 1. theory and overview. Eur Biophys J 12(4):181–197

31. Harris CM, Kell DB (1985) On the dielectrically observable consequences of the diffusional motions of lipids and proteins in membranes. 2. experiments with microbial-cells, protoplasts and membrane-vesicles. Eur Biophys J 13(1):11–24

32. Ryabov YE, Feldman Y (2002) Novel approach to the analysis of the non-Debye dielectric spectrum broadening. Physica a-Stat Mech Appl 314(1–4):370–378

33. Currie DJ, Lee MH, Todd RW (2006) Prediction of physical properties of yeast cell suspensions using dielectric spectroscopy. conference on electrical insulation and dielectric phenomena; pp 672–675.

34. Davey CL, Markx GH, Kell DB (1993) On the dielectric method of monitoring cellular viability. Pure Appl Chem 65(9):1921–1926

35. Davey CL (1993) The theory of the b-dielectric dispersion and its use in the estimation of cellular biomass. Aber Instruments Ltd, Aberystwyth, UK

36. Davey CL, Davey HM, Kell DB (1992) On the dielectric properties of cell suspensions at high-volume fractions. Bioelectrochem Bioenerg 28(1–2):319–340

37. Aber-Instruments(2005). Biomass Monitor 210 User Manual. Aberystwyth, UK

38. Brereton RG (2000) Introduction to multivariate calibration in analytical chemistry. Analyst 125(11):2125–2154

39. Beebe KR, Kowalski BR (1987) An introduction to multivariate calibration and analysis. Anal Chem 59(17):A1007–A1017

40. Brereton RG (2007) Applied chemometrics for scientists. Wiley, Chichester, UK

41. Haaland DM, Thomas EV (1988) Partial least-squares methods for spectral analyses. 1. relation to other quantitative calibration methods and the extraction of qualitative information. Anal Chem 60(11):1193–1202

42. Yardley JE, Todd R, Nicholson DJ, Barrett J, Kell DB, Davey CL (2000) Correction of the influence of baseline artefacts and electrode polarisation on dielectric spectra. Bioelectrochemistry 51(1):53–65

43. Verduyn C, Postma E, Scheffers WA, Vandijken JP (1992) Effect of benzoic-acid on metabolic fluxes in yeasts-a continuous-culture study on the regulation of respiration and alcoholic fermentation. Yeast 8(7):501–517

44. Cannizzaro C, Valentinotti S, von Stockar U (2004) Control of yeast fed-batch process through regulation of extracellular ethanol concentration. Bioprocess Biosyst Eng 26(6):377–383

45. Marquardt DW (1963) An algorithm for least-squares estimation of nonlinear parameters. J Soc Ind Appl Math 11(2):431–441