Predictive modelling of brewing fermentation: from knowledge-based to black-box models

Ioan-Cristian Trelea, Mariana Titica, Sophie Landaud, Eric Latrille, Georges Corrieu, Arlette Cheruy

Predictive modelling of brewing fermentation: from knowledge-based to black-box models

Ioan Cristian TRELEA[1], Mariana TITICA[2], Sophie LANDAUD[1], Eric LATRILLE[1], Georges CORRIEU[1], Arlette CHERUY[2]

[1]Laboratoire de Génie et Microbiologie des Procédés Alimentaires, INRA - INAPG, 78850 Thiverval - Grignon, France. E-mail: trelea@platon.grignon.inra.fr

[2]Laboratoire d'Automatique de Grenoble, ENSIEG, Rue de la Houille Blanche, Domaine Universitaire, BP 46, 38402 Saint Martin d'Hères, France

Abstract

Advanced monitoring, fault detection, automatic control and optimisation of the beer fermentation process require on-line prediction and off-line simulation of key variables. Three dynamic models for the beer fermentation process are proposed and validated in laboratory scale: a model based on biological knowledge of the fermentation process, an empirical model based on the shape of the experimental curves and a black-box model based on an artificial neural network. The models take into account the fermentation temperature, the top pressure and the initial yeast concentration, and predict the wort density, the residual sugar concentration, the ethanol concentration, and the released CO_2. The models were compared in terms of prediction accuracy, robustness and generalisation ability (interpolation and extrapolation), reliability of parameter identification and interpretation of the parameter values. Not surprisingly, in the case of a relatively limited experimental data (10 experiments in various operating conditions), models that include more process knowledge appear equally

accurate but more reliable than the neural network. The achieved prediction accuracy was 5% for the released CO_2 volume, 10% for the density and the ethanol concentration and 20% for the residual sugar concentration.

Keywords: beer, density, dynamic model, neural network

1. Introduction

Though beer fermentation is well known for a long time, fermentation practices in most production plants are essentially unchanged from traditional batch processes. Industrial batch fermentation is usually carried out in open loop conditions, using a predefined and empirically established temperature profile (which is often confidential). The off-line measurement of the wort density, well correlated with the fermentable sugar concentration, and expressed as an equivalent sucrose concentration (Plato), is widely used on industrial scale for the monitoring of the fermentation process.

Currently, one of the most important problems in brewing industry is to achieve desirable beer quality and to eliminate batch-to-batch differences in beer composition.

Optimisation and control of the fermentation process is a way to comply with these demands. Modern techniques of process control and optimisation require an adequate mathematical bioprocess description, which is the basic ingredient for development of model based control strategies.

The aim of this work is to propose such models of alcoholic beer fermentation, for control and optimisation purposes. These models have to describe the effect of the operating conditions on

the evolution of the fermentation state: CO_2 production, substrate concentration, ethanol concentration, and wort density.

The fermentation rate was estimated through the CO_2 production rate for two reasons:

- investigations on laboratory and pilot scales have confirmed that the rate of CO_2 evolution is a good indicator of yeast growth, fermentable sugars consumption and ethanol production rate [1,2]
- the amount of CO_2 evolved can be accurately measured by on-line sensors, available on industrial scale [3,4]

The existing mathematical models of beer fermentation do not meet these specifications. The models based on CO_2 are static models, consisting in correlation CO_2 - wort density [5]. The predictive models describe the fermentation rate as a sugar uptake rate [6] or yeast growth rate [7], for which sensors for on-line measurement do not exist in industry. Furthermore, these models include only the temperature effect on the fermentation rate.

Dynamic models considered in this paper are of the form:

$$\frac{dC(t)}{dt} = f(C(t), T(t), P(t), X_0) \tag{1}$$

in the continuous time case, and

$$C(t+\tau) = g(C(t), T(t), P(t), X_0) \tag{2}$$

in the discrete time formulation. Here f and g are non-linear functions, t is the time, τ the sampling interval, C the evolved amount of CO_2, T the temperature, P the top pressure and X_0 the initial yeast concentration.

These models were coupled with static linear relationships between CO_2 and sugar consumption, ethanol production and density change.

In the present study, three modelling methods, chosen according to the process characteristics (process knowledge, existing on-line measurements, available experimental data) where studied comparatively:

—Biochemical knowledge based model. The model structure was chosen among known laws (e.g. Monod growth law). This kind of structures are interesting because the model parameters have some biological significance, but they are strongly non-linear and can pose problems in terms of parameter identification.

—Empirical model. The structure was chosen based on the shape of the experimental curves. The parameters have a clear meaning in terms of curve shape, but a rather loose biological significance. This kind of model has good mathematical properties, from an identification point of view.

—Neural network model. Neural networks [8] are general function approximators, widely used in recent years. They are purely black box models: their structure is not linked to any particular curve shape and the coefficients have no special interpretation.

The paper is organised as follows. The first section describes the experimental pilot plant on which our experimental data were obtained. Then we present the identification from experimental data of the static part of the model: linear correlations between the produced CO_2, the ethanol concentration, the sugar concentration and the wort density. Finally, three dynamic modelling methods are presented and the resulting models analysed comparatively, in terms of their relative performance.

2. Materials and methods

2.1. Fermentation tank

Fermentations were carried out in a stainless steel bioreactor (15 L capacity - LSL-Biolafitte, France). It included : (1) a temperature sensor (standardized platinum probe - 100 Ω at 0°C); (2) a volumetric flow meter (Schlumberger, France) measuring the volume of CO_2 evolved, delivering one electrical pulse for each litre of gas; (3) a top pressure sensor (Wika-type 98, Germany).

An electronic interface board (OPTO 22, France) was used to digitize and store data supplied by the sensors. Home made software (PPOUF) was installed on a PC-compatible micro-computer for the real time data acquisition, supervision and control. The temperature was controlled with ± 0.07°C accuracy and the top pressure with ± 10 mbar accuracy.

2.2. Wort and yeast strain

The lager wort (approximately 12 Plato) was provided by the Institut Français de Brasserie et Malterie (IFBM, France). Before inoculation, the wort (11 L for each experiment) was aerated until saturation. Gentle agitation was used (100 rpm) to prevent early yeast flocculation.

An industrial lager-type yeast strain, *Saccharomyces cerevisiae (uvarum)*, provided by IFBM (France), was used. Starter cultures were carried out at 20°C in 5 L of lager wort during 3 days. Temperature was decreased at fermentation temperature one day before inoculation. Starter cultures were centrifuged three times (4000 rpm) in cold physiological saline before inoculation.

2.3. Experimental design

The experimental design (Figure 1) considered the effect of three operating factors: the fermentation temperature T, the top pressure P and the initial yeast concentration X_0. The ranges were chosen representative of beer fermentation conditions in industry. Experiment R5 was performed in intermediate operating conditions T_{R5} = 13°C, P_{R5} = 450 mbar, $X_{0\ R5}$ = 10 million cells/mL. The other eight experiments explored all possible combinations of "extreme" conditions: temperature levels of T_{min} = 10 C and T_{max} = 16°C, pressure levels of P_{min} = 50 mbar and P_{max} = 800 mbar and initial yeast concentration levels of X_{0min} = 5 millions cells/mL and X_{0max} = 20 millions cells/mL of wort.

2.4. Measurement of key variables

During the fermentation experiments, samples were taken every 12 hours until the fermentation was complete in order to determine wort density, residual sugar concentration and ethanol concentration.

Density. The density of filtered and degazified wort was determined by a 10 mL pycnometer. The real density value d, in kg m^{-3}, was converted to Plato (D), which is traditionally used in brewery. The following relationship was used [9]:

$$D = (d - 999.448) / 4.08745 \quad (3)$$

Fermentable sugars. The concentration of fermentable sugars, maltotriose, maltose, glucose and fructose were monitored using High Performance Liquid Chromatography (HPLC) system (Waters, USA) with an Aminex HPX-87C column (300 mm × 7.8 mm i.d.) (BioRad, USA) at 85°C.

Ethanol. The ethanol concentration was determined using a Carlo Erba 5300 gas chromatograph equipped with a stainless steel column (200 mm × 0.3 mm i.d.) coated with Chromosorb 101 (SGE, USA).

Evolved CO_2. The evolved CO_2 was measured by a volumetric flow meter, which delivered a pulse for every litre of gas. The instantaneous release rate was computed off-line as a moving average of the evolved gas volume over a symmetric 10 hour period (5 hours backwards and 5 hours forwards in time). The CO_2 release rate was corrected in order to take into account the wort volume variation caused by sampling, as well as the gas accumulation in the head space.

Dissolved CO_2. The fraction of the produced CO_2 which is dissolved in the wort (and not released), can not be neglected, especially in low temperature and high pressure conditions. The dissolved CO_2, in litres per litres of wort, was calculated using a formula established at IFBM:

$$C_d(T,P) = 2.96 \cdot 10^{-6}(P+1000)(T+273.16)e^{-0.0335T} \quad (4)$$

with temperature T in Celsius and pressure P in mbar. It was assumed that CO_2 was released after the wort is saturated, which caused an apparent lag.

3. Relationships between fermentable sugar consumption, ethanol production, density decrease and evolved CO_2 during alcoholic beer fermentation

The on-line measurement of evolved CO_2 can be used for monitoring of the alcoholic beer fermentation, because of the linear relationships established between the amount of produced CO_2, the sugar assimilated by the yeast and the ethanol produced [10]:

$$S(t) = S_0 - Y_{S/C} \cdot C_p(t) \quad (5)$$

$$E(t) = Y_{E/C} \cdot C_p(t) \quad (6)$$

C_p represents the produced (as opposed to evolved or released) CO_2, S represents the fermentable sugars concentration, S_0 the initial sugar concentration and E the ethanol concentration. $Y_{S/C}$ and $Y_{E/C}$ are constant yield coefficients.

In brewery, the concentration of fermentable sugars (S) is expressed in terms of wort density (D), traditionally used as descriptor of the fermentation process. It has been shown that a linear correlation exists between the CO_2 produced and the density decrease:

$$D(t) = D_0 - Y_{D/C} \cdot C_p(t) \quad (7)$$

D_0 is the initial wort density and $Y_{D/C}$ a constant yield coefficient.

Values of yield coefficients were accurately identified from the set of experimental data, as shown in Table 1. The yield coefficients did not depend on the operating conditions. The final amount of evolved CO_2 per litre of wort and the final ethanol concentration were proportional to the initial concentration of fermentable sugars.

4. Biochemical knowledge based model

Knowledge based models consist in a set of mathematical equations describing the phenomena occurring during wort fermentation (main metabolic pathways, growth limitation and/or inhibition etc.). The model structure is chosen among existing relationships describing

the biological kinetics. The main advantage of this type of model is to account for the biological phenomena, but its structure may be complex and difficult to verify and validate.

In our case, the model must describe through the amount of produced CO_2, the yeast growth from sugars and the inhibition effect of ethanol. Besides, it must include the operating conditions effects on these phenomena.

The selected model structure, describing both substrate limitation and inhibition by ethanol, is:

$$\begin{aligned} &\frac{dC_p(t)}{dt} = v_{\max} \cdot \frac{S(t)}{K_S + S(t)} \cdot \frac{1}{1 + K_I E^2(t)} \cdot \left[C_p(t) + C_0 \cdot X_0 \right] \\ &C_p(0) = 0 \end{aligned} \tag{8}$$

Here v_{max} is the maximum CO_2 production rate, K_S (Monod constant) accounts for the substrate saturation effect and K_I for the ethanol inhibition. $v_{max}C_0X_0$ represents the initial CO_2 production rate and X_0 is the initial yeast concentration. Sugar (S) and ethanol (E) concentrations are related to produced CO_2 (C_p) by constant yield coefficients (equations (5) and (6), and Table 1).

Owing to the poor identification properties of K_I, K_S and C_0 parameters (as illustrated below), and the low model sensitivity to variations of parameter K_S [11], we assumed an operating conditions dependence only on the maximum "specific" rate parameter (v_{max}) and ignored any similar effects on other kinetic parameters. This dependence was expressed by a linear function of operating conditions and their interactions, as follows:

$$v_{\max} = a_T T_n + a_P P_n + a_X X_{0n} + a_{TP} T_n P_n + a_{TX} T_n X_{0n} + a_{PX} P_n X_{0n} + a_0 \tag{9}$$

Here T_n, P_n and X_{0n} are normalized values of T, P, X_0 respectively, ranging from -1 to 1 in our experimental design:

$$T_n = \frac{2T - \left(T_{\max} + T_{\min}\right)}{T_{\max} - T_{\min}}, \quad P_n = \frac{2P - \left(P_{\max} + P_{\min}\right)}{P_{\max} - P_{\min}}, \; X_n = \frac{2X_0 - \left(X_{0\max} + X_{0\min}\right)}{X_{0\max} - X_{0\min}} \qquad (10)$$

Factor normalisation provides an independent evaluation of principal and interacting factor effects: absolute value and sign of coefficients *a* allows classification of the factors, according to their relative importance (the largest coefficient corresponds to the most influent factor or factor interaction, positive sign expresses a positive effect and vice versa).

Thus, the parameter identification was performed in two steps:

- estimation of the kinetic parameters of CO_2 production rate from experiments carried out in constant operating conditions;
- determination of the function coefficients in equation (9) for the kinetic parameter v_{max} from the results of the previous step.

In the first step, the kinetic parameters were identified for every run of the experimental design. Next, a global model parameters identification was made, with constant K_S, K_I and C_0 values and variable v_{max}, using previous parameters values as initial conditions. The algorithm used to fit the model parameters consisted in a numerical minimization of a non-linear function: the sum of squares of the errors between measured and predicted released CO_2. The error bars for the parameters were estimated from the curvature of the cost function (residual standard deviation) near the optimum, with the standard assumption that a limited Taylor series expansion to second order approximates the cost function well enough. This is generally true only for "small" variations around the optimum value of the parameters. Thus, large error bars are not reliable *per se*, but simply indicate that the given parameter can not be determined properly from the available experimental data.

The model (8) gives the dynamic of total amount of produced CO_2, which represents the sum of dissolved and released CO_2 (which was actually measured). Thus the released fraction of CO_2 (C) is calculated as follows :

$$C(t) = \max\left\{0,\ C_p(t) - C_d(T,P)\right\} \tag{11}$$

where the dissolved CO_2 fraction (C_d) is given by equation (4).

In the second step, estimated values of kinetic parameter v_{max} were used for the identification of the coefficients of the function (9). Model parameters were adjusted using all experimental design runs, except the run R5, on which validation of the resulting model was carried out.

Therefore, the retained global model is the following :

$$\frac{dC_p(t)}{dt} = v_{\max}(T,P,X_0)\cdot\frac{S\left(C_p(t)\right)}{K_s + S\left(C_p(t)\right)}\cdot\frac{1}{1+K_I E\left(C_p(t)\right)^2}\cdot[C_p(t) + C_0\cdot X_0] \tag{12}$$
$$C_p(0) = 0$$

5. Empirical model

This model is based on the observation that CO_2 production rate curves had similar shapes, with three characteristic phases. Depending on the operation conditions, the production rate curves were more or less expanded along the time axis and had different peak values. The area below each curve represents the total amount of CO_2 produced. It is proportional to the initial amount of fermentable sugars in the wort (wort density) and does not depend on the operating conditions (temperature, pressure, yeast concentration).

Our empirical model has the following form:

$$\frac{dC_p(t)}{dt} = k_1\left[1-(1-k_2X_0)e^{-k_3C_p(t)}\right]e^{-k_4C_p(t)}\left[1-e^{-k_5(C_p(t)-C_f)}\right]e^{k_6T_n(t)}e^{-k_7P_n(t)}e^{k_8X_{0n}} \qquad (13)$$
$$C_p(0) = 0$$

Each factor has a well-defined interpretation in terms of the fermentation curve shape (Figure 2). For constant temperature and pressure, the initial fermentation rate is proportional to the initial concentration of yeast cells in the wort. In equation (13), this is expressed by the factor k_2X_0: for C_p =0, the first square bracket reduces to k_2X_0. The acceleration phase is basically due to the yeast growth. The shape of the curve in the acceleration phase is given by the factor $e^{-k_3C_p}$: the larger the coefficient k_3, the shorter the acceleration phase. The gradual inhibition phase is due to ethanol accumulation in the wort, which inhibits yeast metabolism. In equation (13), the ethanol inhibition is introduced by the factor $e^{-k_4C_p}$: the larger the coefficient k_4, the sharper the inhibition phase. The final, abrupt, inhibition phase is due to fermentable sugars limitation. Because of the proportionality between sugars consumption and CO_2 production, the fermentation rate becomes zero when a known amount of CO_2 (C_f) has been produced. C_f can be determined in advance, from the initial wort density, using equations (5) and (7). In equation (13), the final inhibition phase is introduced via the factor $e^{-k_5(C_p-C_f)}$.

It was assumed that the operating conditions accelerate or slow down the fermentation in the same way in all phases. The effect of the operating conditions, in normalized units, is expressed by the coefficients k_6, k_7 and k_8 in equation (13).

The model coefficients k_1 to k_8 were estimated using a nonlinear least squares optimization procedure from experimental runs R1–R4 and R6–R9. The run R5 was used for validation. The error bars were estimated using the same technique as for the biochemical model.

Remark 1. The mathematical form of the factors describing various fermentation phases is arbitrary to a large extent. Many other formulae which look like Figure 2 could be invented.

The same is true for the factors which take into account the operating conditions. In our experimental design, a natural choice for these factors would be linear dependencies, of the form:

$$\left(1+k_6 T_n\right)\left(1-k_7 P_n\right)\left(1+k_8 X_{0n}\right) \tag{14}$$

This is very similar to the selected exponential form, since for $|x| << 1$, $e^x \approx 1+x$. We preferred the exponential form for its numerical robustness. During numerical optimization of the model coefficients, trial steps were taken in which the CO_2 production rate occasionally became negative (when using equation (14)) and the whole algorithm broke down.

Remark 2. Each coefficient has a well-defined effect, so manual adjustment by trial and error is perfectly feasible, avoiding the need of a sophisticated numerical routine. Despite their apparent similarity, k_2 and k_8 have distinct effects: k_2 mainly controls the duration of the apparent lag phase (via the initial fermentation rate), while k_8 has a global speedup or slowdown effect.

6. Neural network model

Artificial neural networks are general function approximators, widely used in high-dimensional interpolation and classification problems. Many textbooks on neural networks exist. Bishop [12] proposed a concise introduction to the field. The ability of the neural networks to approximate any "reasonable" multivariate function can be used to construct nonlinear discrete-time dynamic models [13,14], as in equation (2).

The dynamic neural network used to predict the released CO_2 is shown in Figure 3. The network has four *inputs*: the fraction of the total CO_2 released at the current time $C(t)/C_f$, the

current wort temperature $T(t)$, the initial yeast concentration X_0 and the current pressure $P(t)$. A dummy fifth unitary input represents the bias. The *hidden layer* consists of three processing units (*neurons*) which compute weighted sums of the inputs before applying sigmoid (*squashing*) nonlinear functions. In matrix-vector notation, the output of the hidden layer can be written as:

$$h(t) = sig\left(W_1\left[C(t)/C_f \quad T_n(t) \quad P(t)_n \quad X_{0n} \quad 1\right]^t\right) \tag{15}$$

where $sig(x) = 1/(1+e^{-x})$ is the sigmoid function (applied to each vector element) and W_1 is the 3×5 matrix of the hidden layer *weights* (coefficients). The *output layer* has a single processing unit which computes a weighted sum of hidden layer outputs and applies to it the same sigmoid function. The network output is the predicted released CO_2 fraction one sampling time step ahead. The time step was taken as $\tau = 5$ hours. Mathematically, the network output is given by:

$$C(t+\tau)/C_f = sig\left(W_2\left[h^t \quad 1\right]^t\right) \tag{16}$$

where W_2 is the 1×4 matrix of the output layer weights.

In order to achieve prediction for arbitrary time horizons, the predicted CO_2 was re-introduced at the network input.

A network structure as the one in Figure 3 is able to approximate continuous mappings with arbitrary accuracy, provided that there are enough hidden units. Reliable theoretical guidelines for selecting the number of hidden units are still lacking, so most authors select this number on a trial and error basis. The network coefficients (W_1 and W_2) are computed with a Levenberg-Marquardt nonlinear optimization algorithm, aiming at a minimum prediction error, on a set of experimental data. In neural network terminology, weight optimization is

called *training*, and the experimental data used in this process form the *training set*. As for the other models, the training set was formed of the experimental runs R1–R4 and R6–R9. In order to avoid overfitting (learning of a particular data set instead of the underlying mapping), a second data base (*validation set*) is used to verify the network performance. The validation set included the run R5.

7. Results and discussion

7.1. Wort density and evolved CO_2 prediction

As previously mentioned, the most used descriptor of the fermentation process in brewing industry is the wort density, which indirectly reflects the remaining fermentable sugar concentration. From the process control point of view, the evolved CO_2 volume is much more convenient for on-line measurement. Since the CO_2 production is proportional to sugar uptake, a linear relationship exists between the initial wort density ant the final volume of the produced CO_2. However, the exact coefficients of this relationship can not be inferred from the yield coefficients reported in Table 1, because the wort contains many other constituents which modify its density. The relationship between the initial wort density and the final CO_2 volume has to be determined for each particular wort type. In out case we found:

$$C_f = 1.2698 \cdot D_0 + 2.5781 \qquad (17)$$

Thus, equations (5), (6), (7) and (17), together with the dynamic models of CO_2 evolution (12), (13), (15) and (16) allow us to make predictions of the evolution of all key variables of the beer fermentation process (density, CO_2, ethanol, sugar) from the initial wort density, initial yeast concentration, temperature and top pressure. The predicted wort density and the

evolved CO_2 volume, for the entire set of explored operating conditions, are shown in Figure 4, together with the experimental data. The prediction errors, expressed as residual standard deviations, are reported respectively in Tables 2 and 3.

The temperature effect is the most obvious. Decreasing the wort temperature from 16 to 10°C, increased the fermentation time by a factor of 2. Increasing the top pressure, slightly slowed down the fermentation, probably due to the increased amount of dissolved CO_2 [15]. Increasing the initial yeast concentration decreased the apparent time lag. The yeast generation time in our experiments was about 12 h at 16°C and 24 h at 10°C. This is consistent with a double apparent time lag in Figure 4, when comparing R1-R4, R2-R3, R7-R8 and R6-R9.

It can be seen that the overall performance of the three models is quite similar. The visual fit is reasonably good and the prediction error is of the order of the measurement uncertainty. This means that, from the point of view of the prediction accuracy, any of the three models could be equally useful in applications. When analysing the prediction error, it is important to distinguish between data used for model identification, namely runs R1–R4 and R6–R9, and independent validation data, namely run R5. Experiment R5 was performed in intermediate operating conditions, so the conclusions are only valid in interpolation. The prediction accuracy for the run R5 is globally similar (sometimes slightly worse, sometimes slightly better), than for the rest of the runs. The three models are thus expected to behave reasonably well for operating conditions situated inside the cube shown in Figure 1. Given their mathematical form, this was expected for the biochemical and the empirical models, but was not necessarily true for the neural network model. Before using the neural network model in applications, additional tests, for other combinations of operating conditions, are strongly recommended.

7.2. CO_2 evolution rate prediction

During the on-line monitoring the beer fermentation process, the CO_2 evolution rate is a better indicator of the fermentation phase than either the total evolved CO_2 volume or the current wort density (Figure 2). Thus, accurate prediction of the fermentation rate may valuable in some applications. However, good visual fit between experimental and predicted CO_2 volume does not necessarily imply equally good agreement for the CO_2 evolution rate, as illustrated in Figure 5 for some typical experiments. As a general remark, the CO_2 evolution rate is predicted less accurately than the cumulated CO_2: the error is of order of 20%, compared to 5% for the CO_2 (Tables 2 and 3). This is partly due to the fact that the CO_2 release rate is measured indirectly and less accurately, as explained above, but mainly to the model identification method: the parameters of all three models are determined aiming at a minimum prediction error on cumulated CO_2 volume. If the prediction of the fermentation rate is an issue, parameter identification should be based on the CO_2 release rate.

It can also be noted in Figure 5 that the neural network model predicts unusual peaks in the fermentation rate. This comes from the network structure (Figure 3), based on sigmoid activation functions, whose derivatives have a similar peaked shape. Better results could be obtained if the CO_2 evolution rate was introduced in the network learning criterion. The biochemical and the empirical models do not suffer from the same drawback because they make explicit use of the CO_2 evolution rate in their formulation (equations 12 and 13).

7.3. Prediction of ethanol and fermentable sugars concentrations

In some applications, such as production of low-alcoholic beer, explicit knowledge of the ethanol and fermentable sugars concentrations may be very useful. Equations (5) and (6) make such predictions straightforward, once the produced CO_2 volume is known. The initial sugar

concentration in the wort is determined from the initial wort density, using equation (17) and the yield coefficient given in Table 1.

Some examples of prediction of ethanol and residual sugars concentrations are given in Figure 6. The overall error in sugars concentration prediction (Tables 2 and 3) appears to be significantly higher (20%) than for ethanol (10%), density (10%) and CO_2 (5%). There is no fundamental reason for this, since stoechiometric relationships between sugar consumption and ethanol and CO_2 production are well known. The most probable explanation is the difficulty to distinguish between fermentable and non fermentable sugars by analytical methods. In High Performance Liquid Chromatography (HPLC), the characteristic peak of the fermentable sugar maltotriose is very close to peaks produced by sugars containing more than three glucose units, which are not catabolised by the yeast.

7.4. Robustness to changes in initial wort density

The experiments shown in Figure 1, except R1 and R2, were carried out with the same initial wort density of 12 Plato, which corresponds to standard lager beer. Experiment R1 used wort with 17 Plato and experiment R2 with 10 Plato. In this density range, the fermentation process is essentially unchanged, and all three models behave equally well.

An additional experiment R10 was performed in the same operating conditions (temperature, top pressure, initial yeast concentration) as R1, but with a much higher wort density (24 Plato). On one hand, the biochemical and the empirical models gave remarkably similar predictions, despite their widely different mathematical form. This was probably due to the fact that both models take into account the main fermentation phases: yeast growth, ethanol inhibition and sugar limitation. However, after 200 hours new phenomena appeared to come

into play, slowing down the fermentation more than the extrapolated ethanol inhibition could suggest.

On the other hand, the neural network model had no build-in knowledge of the fermentation process. The use of a normalized value of the CO_2 ($C(t)/C_f$) hide the reality of a different initial wort density. The operating conditions (temperature, pressure, initial yeast concentration) being similar with experiment R1, a similar fermentation kinetic was predicted, in *normalized* CO_2: the total amount of CO_2 should be released in 160 hours. When converting these normalized values to physical units, the fermentation rate appeared significantly overestimated, as in Figure 7. In principle, this drawback of the neural network model could be fixed by introducing an additional input, representing the initial wort density. In practice, to get reliable predictions, experimental data should be available for a large number of combinations of input variables, uniformly distributed in the input space (Figure 1). With an additional input, the number of necessary experiments would become prohibitive.

7.5. Model parameter identification

Besides the predictive ability of a model, the significance and the reliability of its parameters are also important. The significance of the model parameters reflects the modelling assumptions and depends on the model structure. The reliability of an identified numerical value for a parameter depends both on the model structure and on the available experimental data. In this work, the parameters of all three models were determined with a Levenberg-Marquardt nonlinear optimization algorithm, aiming at the best possible fit between measured and predicted values of the evolved CO_2, for experiments R1–R4 and R6–R9. The numerical values for the 10 coefficients appearing in the biochemical dynamic model are reported in Table 4, for the 8 coefficients of the empirical model in Table 5, and the 19 weights of the

neural network in Table 6. As a general rule, parameters with analogous roles in the biochemical and empirical models appeared to have similar identification properties.

As already mentioned, the parameters of the biochemical model were identified in two steps. In the first step, it appeared that the maximum CO_2 production rate (ν_{max} in equation (8)) could be inferred reliably for every experimental run, while the other parameters (K_S, K_I, and C_0) could not. This is illustrated in Figure 8 for the $\nu_{max} - K_S$ pair and the experiment R5: the long flat "valley" parallel with the K_S axis means that many values of K_S result in a similar fit. This is why ν_{max} was identified separately for each experiment, while K_S, K_I and C_0 were identified globally on the whole database, in order to get reasonable error bars (Table 4). In the second step, the dependence of ν_{max} on the operating conditions and their interactions was established. It appears from Table 4 that the average fermentation rate a_0 and its dependence on the wort temperature a_T can be determined most reliably, which is consistent both with prior biological knowledge and with visual inspection of Figure 4. The top pressure (a_P), the initial yeast concentration (a_X), and the temperature – yeast concentration interaction (a_{TX}) also have some effect, while the temperature – pressure (a_{TP}) and pressure – yeast concentration (a_{PX}) interactions can be safely neglected.

In the case of the empirical model, all parameters were determined in a single step of numeric optimization. Initial values for all parameters were estimated by graphical interpretation of the CO_2 production rate curve shape (Figure 2). With some thought and a few dozens of trials and errors, good values for all parameters can be determined even without numeric optimisation. The temperature effect on the fermentation rate (k_6) can be determined most reliably, followed by the yeast concentration (k_8) and the top pressure effect (k_7), as well as the average fermentation rate (k_1). The effect of the yeast concentration on the initial fermentation rate (k_2), the ethanol inhibition (k_4) and the substrate limitation (k_5) are determined less accurately.

In the case of the neural network model, the weight identification starts with random initial values. In general, there is no guarantee that the numerically computed optimum is a global one, neither that the resulting network will behave correctly for combinations of input values it has never seen before (even in interpolation). The best number of hidden neurons (and hence the network weights, Figure 3) also has to be determined by trial and error. In this work, networks having between 1 and 5 hidden neurons were tested. 15 random weight initializations were made for each number of hidden neurons, and weight optimization performed using experiments R1–R4 and R6–R9. The network with the best generalization performance (as tested on run R5) was selected. Typically, for the considered number of input variables, neural networks perform well when at least a few tens of *examples* (combinations of operating conditions) are available, uniformly distributed in the input space. In our case, the database was very limited, so it is not surprising that "complex" networks, with 4 hidden neurons or more were unable to interpolate correctly inside the cube shown in Figure 1, even if regularization techniques [16] were used. The flexibility of the networks with 1 or 2 neurons appeared too limited, so the structure shown in Figure 3 was selected. The analysis of individual weights of a neural network is not very relevant.

On one hand, the identification of a neural network model requires relatively heavy computations: optimization in a high dimensional space, many random weight sets, many candidate structures. On the other hand, the simulation of an existing network is fast and straightforward, (iteration of the equations (15) and (16)), while the simulation of the other two models requires a numerical solution of an ordinary differential equation.

8. Conclusion

For the monitoring, control and optimization of the beer fermentation process, the key variables are the wort density (reflecting the sugars concentration in the wort), the ethanol concentration and the evolved CO_2 volume, which is the only variable easy to measure on line. Because of the stoechiometric proportions, any of these variables can be expressed as a linear function of any other. The operating factors available for fermentation process control are the wort temperature, the top pressure and the initial yeast concentration. Three dynamic models of CO_2 production are presented, which take into account these variables: a model based on biochemical knowledge, an empirical model based on the shape of the experimental kinetic curves and an artificial neural network. The models are compared for their prediction accuracy, generalization ability, parameter identification and interpretation.

As far as the prediction accuracy is concerned, the three models were similar. The evolved CO_2 had the lower prediction error (5%), followed by wort density (10%), ethanol concentration (10%), CO_2 evolution rate (20%) and residual sugars concentration (20%). These differences were due to measurement accuracy and to the optimization criterion for the model identification (based on cumulated CO_2 volume).

The generalization ability concerned interpolation in the space of the operating conditions and extrapolation for higher wort density. From this point of view, the biochemical and the empirical model were clearly preferable. The neural network had to be explicitly selected for good interpolation properties, and uniformly good interpolation was never guaranteed. In contrast, the mathematical form of the other two models made their behaviour much more predictable. Extrapolation proved to be a difficult task for either model, but the neural network was still significantly worse than the other two models.

The parameters of the biochemical model have some biological significance, and those of the empirical model have a simple graphical interpretation. The model structure selection requires prior knowledge of the problem being solved. In contrast, the neural network is a pure black box representation. Its structure is selected by heavy, but standard numerical tests.

As a general conclusion, it might be said that whenever physical, chemical or biological insight is available, it is advisable to build this knowledge into the model structure, to ensure predictable, even if not very accurate, behavior in previously unseen situations. This is especially important if the available experimental data is limited and the dimension of the input space is high. On the other hand, if a large database is available, and examples are well distributed in the input space, a "blind" statistical approach (such as neural computation) might be a quick way to get a relatively reliable model.

Nomenclature

$a_T \ldots a_0$		Biochemical model parameters
C	L L^{-1}	Evolved CO_2 volume, per litre of wort
C_0	10^{-6} mL	Initial CO_2 production rate coefficient
C_d	L L^{-1}	Dissolved CO_2 volume, per litre of wort
C_f	L L^{-1}	Final produced CO_2 volume, per litre of wort
C_p	L L^{-1}	Produced CO_2 volume, per litre of wort
d	kg m^{-3}	Wort density, expressed in standard physical units
D	Plato	Wort density, expressed as equivalent sucrose concentration
D_0	Plato	Initial wort density, expressed as equivalent sucrose concentration
E	mL $(100mL)^{-1}$	Ethanol concentration
h		Output of the neural network hidden layer
$k_1 \ldots k_8$		Empirical model coefficients
K_I	g^{-2} L^2	Product inhibition coefficient
K_S	g L^{-1}	Substrate limitation coefficient
P	mbar	Top pressure in the fermentation tank
P_{min}, P_{max}	mbar	Minimum and maximum top pressure in the experimental design
P_n		Normalised dimensionless pressure
S	g L^{-1}	Fermentable sugars concentration
S_0	g L^{-1}	Initial fermentable sugars concentration
t	h	Time
T	C	Wort temperature

T_{min}, T_{max}	C	Minimum and maximum temperature in the experimental design
T_n		Normalised dimensionless temperature
W_1		Matrix of neural network coefficients, for the hidden layer
W_2		Matrix of neural network coefficients, for the output layer
X_0	10^6 mL^{-1}	Initial yeast concentration, in millions of cells per millilitre
X_{0n}		Normalised dimensionless initial yeast concentration
X_{0min}, X_{0max}	10^6 mL^{-1}	Minimum and maximum initial yeast concentration in the experimental design
$Y_{D/C}$	Plato	Yield coefficient of density decrease versus CO_2 production
$Y_{E/C}$	mL (100mL)$^{-1}$	Yield coefficient of ethanol versus CO_2 production
$Y_{S/C}$	g L^{-1}	Yield coefficient of sugar consumption versus CO_2 production
τ	h	Sampling time
ν_{max}	h^{-1}	Maximum specific CO_2 production rate

References

[1] Stassi, P., Rice, J.F., Munroe, J.H. and Chicoye, E., Use of CO_2 evolution rate for the study and control of fermentation. MBAA Technical Quarterly, 24(1987) 44-50.

[2] Corrieu, G., Perret, B., Merican, F. Dispositif de perte de charge et de mesure de débit et application à un procédé de contrôle de l'évolution de la fermentation brassicole. French patent No°96 03823 (1996), publication No 27 46914.

[3] Stassi, P., Fehring, J., Ball, C., Goetzke, G. and Ryder, D., Optimization of fermentor operations using a Fermentor Instrumentation System. MBAA Technical Quarterly, 32(1995) 57-65.

[4] Perret, B., Trelea, I.C. and Corrieu, G., On-line estimation and prediction of density and ethanol evolution in the brewery. MBAA Convention, sept. 12-14, Keystone Colorado, (1999).

[5] Daoud, I.S. and Searle, B.A., On-line monitoring of brewery fermentation by measurement of CO_2 evolution rate. Journal of Institute of Brewing, 96(1990) 297-302.

[6] Douglas A.G., Ramirez W.F. A flavor model for beer fermentation. J. Inst. Brew. 100(1994) 321-329.

[7] Andres-Toro, B., Giron-Sierra, J.M., Lopez-Orozco, J.A., Fernandez-Conde, C., Peinado, J.M. and Garcia-Ochoa, F., A kinetic model for beer production under industrial fermentation conditions. Mathematics and Computers in Simulation, 48(1998) 65-74.

[8] Hornik, K., Stinchcombe, M. and White, H.. Multilayer feedforward networks are universal approximators. Neural Networks 2(1989) 359-366.

[9] Goldiner, F., and Kampf, W. Rohrzucher-, Alkohol-, Stammwurze- und Korrectionstafel. Instut fur Garungsgewerbe, Berlin (1966).

[10] Pandiella, S., Garcia, L., Diaz, M. and Daoud, I. Monitoring the carbon dioxide during beer fermentation. MBAA Technical Quarterly, 32(1995) 126-131.

[11] Vialas C., Cheruy, A., Gentil, S. An experimental approach to improve the Monod model identification. IFAC Modeling and Control of Biotech. Processes, Holland (1985).

12] Bishop, C. M. Neural networks and their applications. Rev. Sci. Instrum. 65(1994) 1803-1832.

[13] Narendra, K. S. and Parthasarathy, K. Identification and control of dynamical systems using neural networks. IEEE Transactions on Neural Networks 1(1990) 4-26.

[14]Acuna, G., Latrille, E., Beal, C. and Corrieu, G., Static and dynamic neural network models for estimating biomass concentration during thermophilic lactic acid bacteria batch cultures. Journal of Fermentation and Bioengineering 85 (1998) 615-622.

[15] Rodney, P. J., and Greenfield, P. F. Effect of carbon dioxide on yeast growth and fermentation. Enzyme Microb. Technol 4(1982) 210-223.

[16] MacKay, D.J.C. Bayesian interpolation. Neural Computation 4(1992) 415-447.

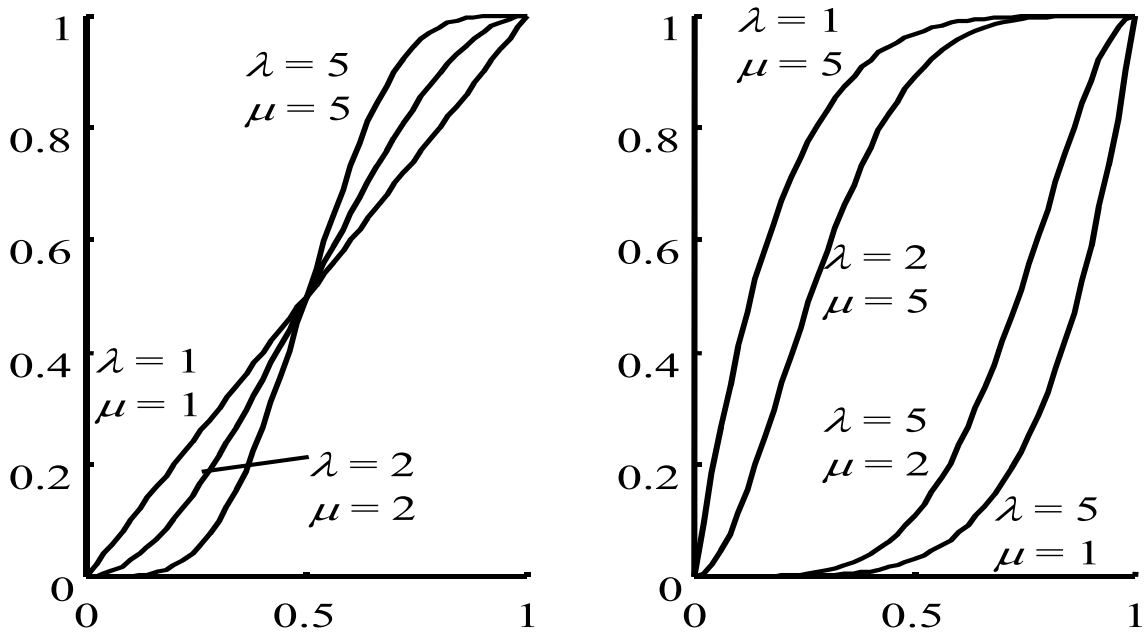

Figure 1

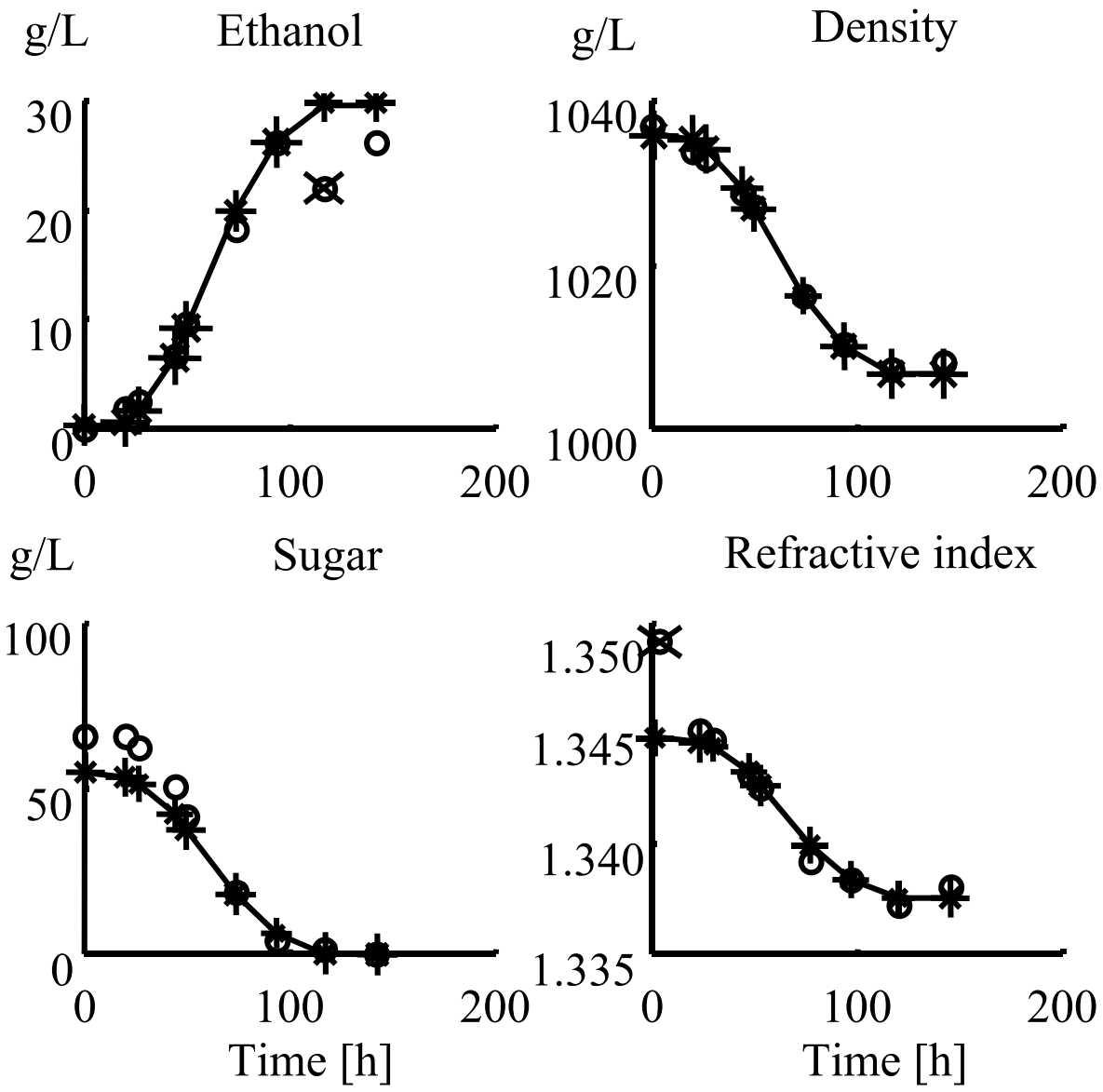
g/L
Ethanol
30
20
10
0
0
100
200
g/L
Density
1040
1020
1000
0
100
200
g/L
Sugar
100
50
0
0
100
200
Time [h]
Refractive index
1.350
1.345
1.340
1.335
0
100
200
Time [h]

Figure 2

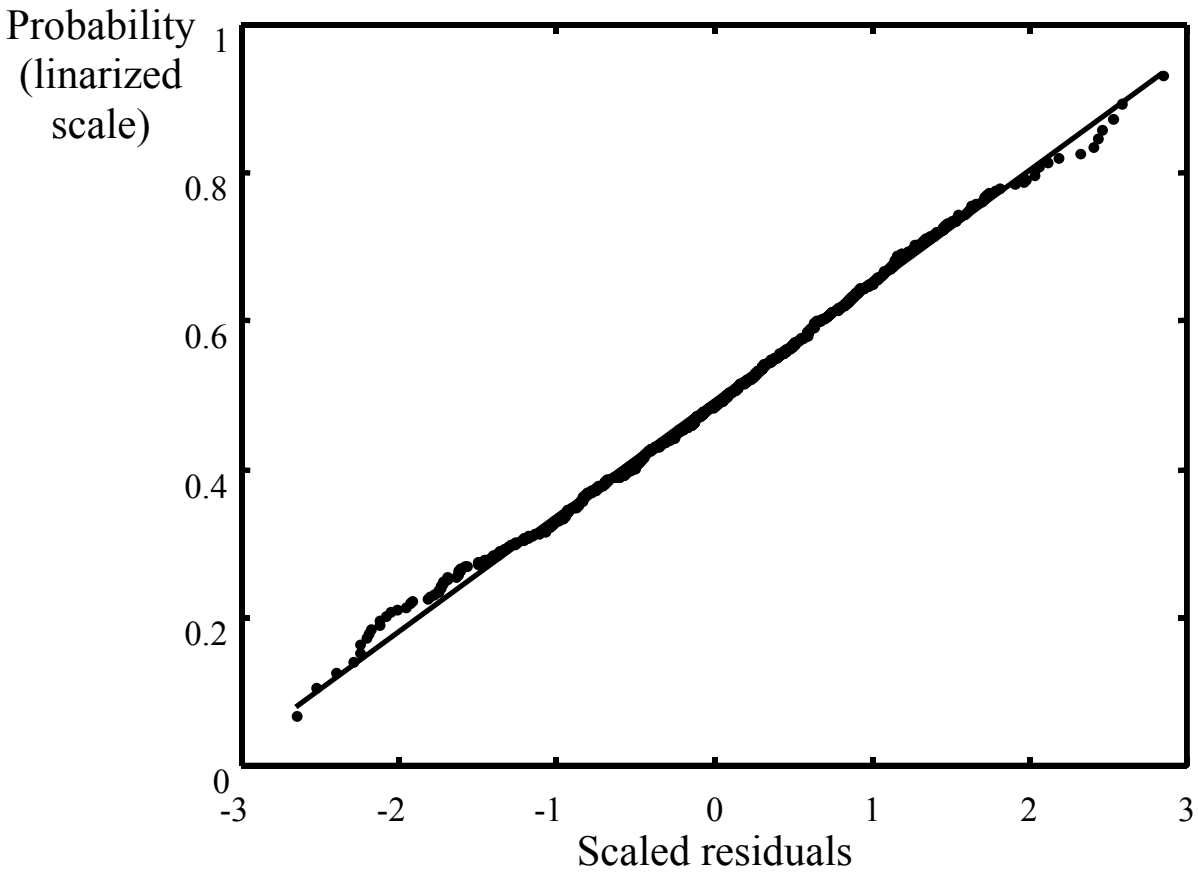
Probability
(linarized
scale)
1
0.8
0.6
0.4
0.2
0
-3
-2
-1
0
1
2
3
Scaled residuals

Figure 3

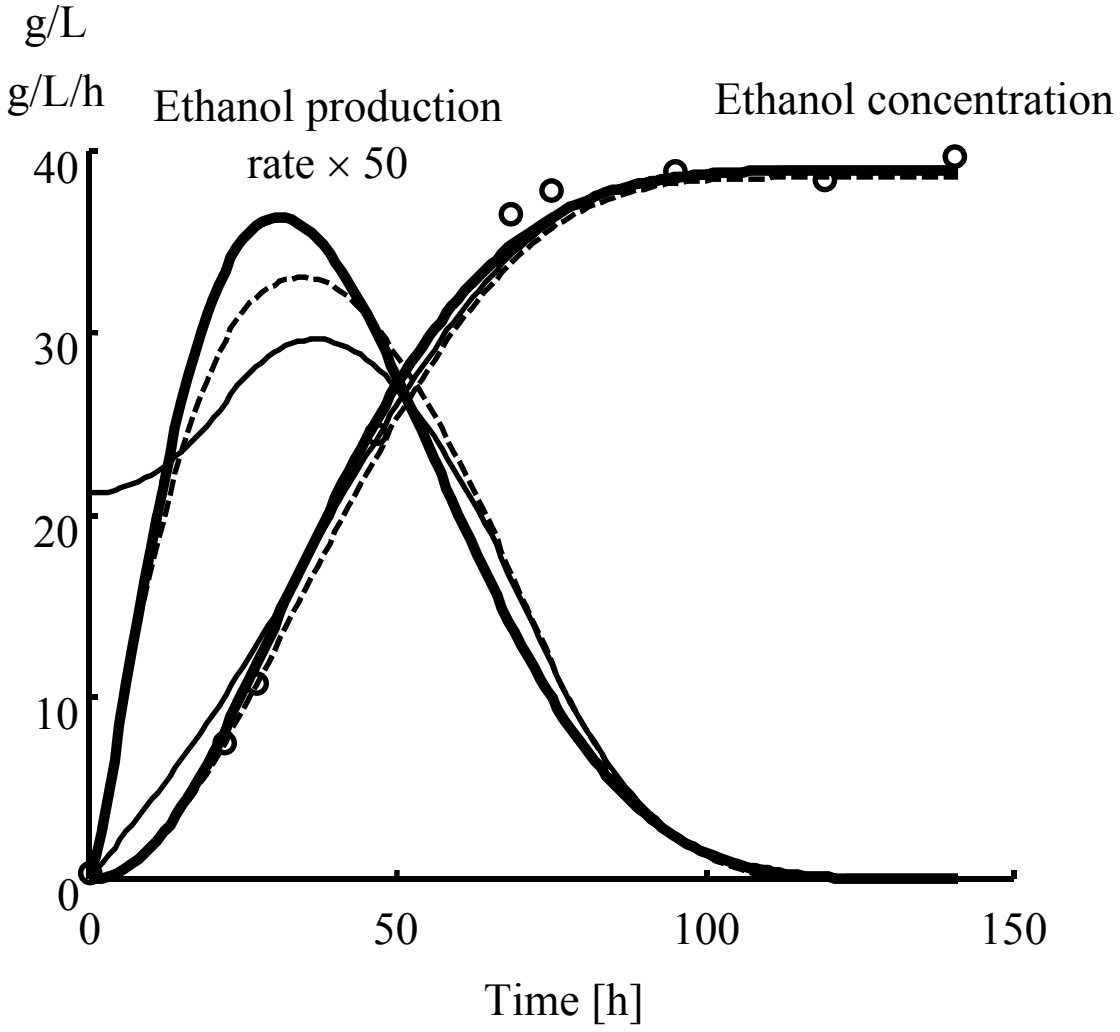
g/L
g/L/h
Ethanol production
rate × 50
Ethanol concentration
40
30
20
10
0
0
50
100
150
Time [h]

Figure 4

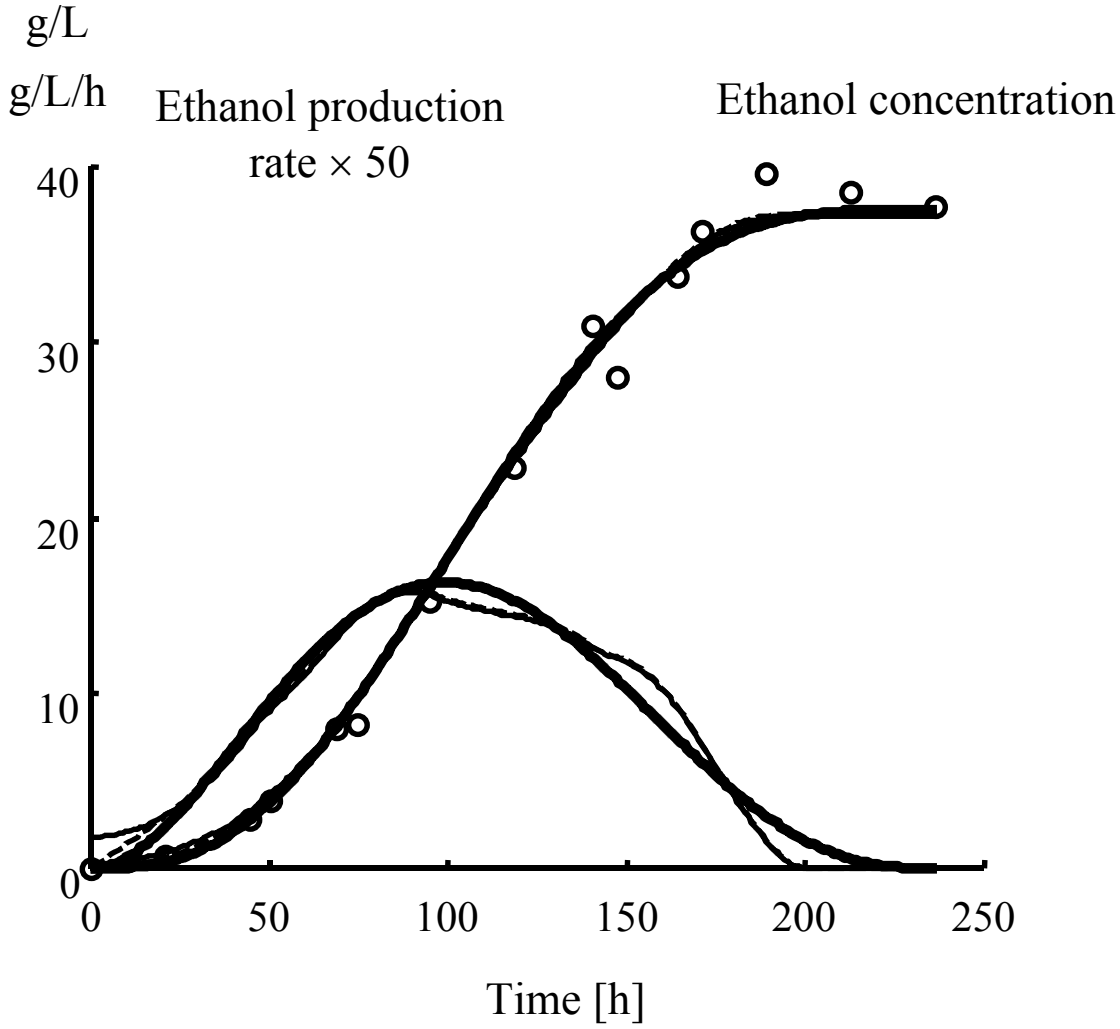
g/L
g/L/h
Ethanol production
rate × 50
Ethanol concentration
40
30
20
10
0
0
50
100
150
200
250
Time [h]

Figure 5

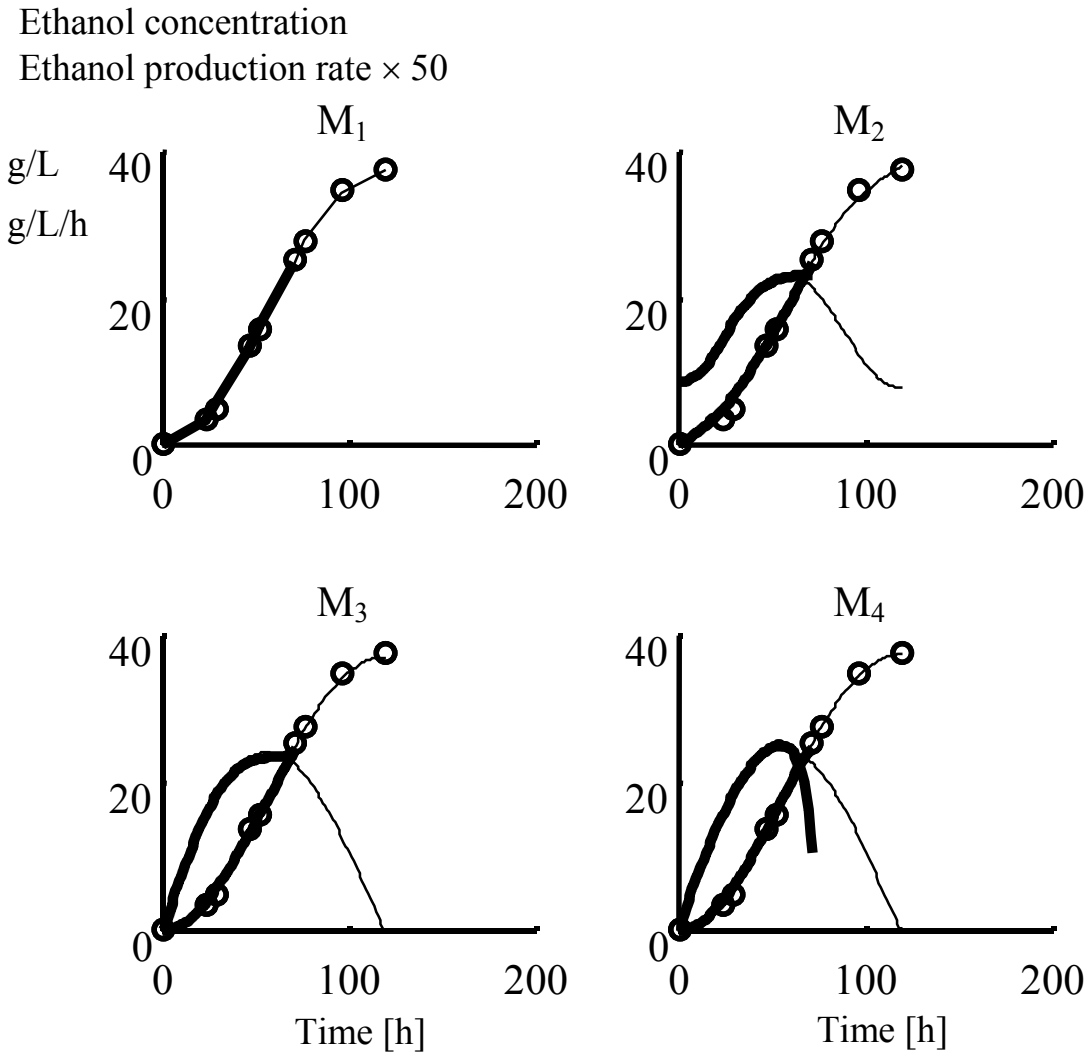
Ethanol concentration
Ethanol production rate × 50
M1
M2
M3
M4
g/L
g/L/h
40
20
0
0
100
200
Time [h]

Figure 6

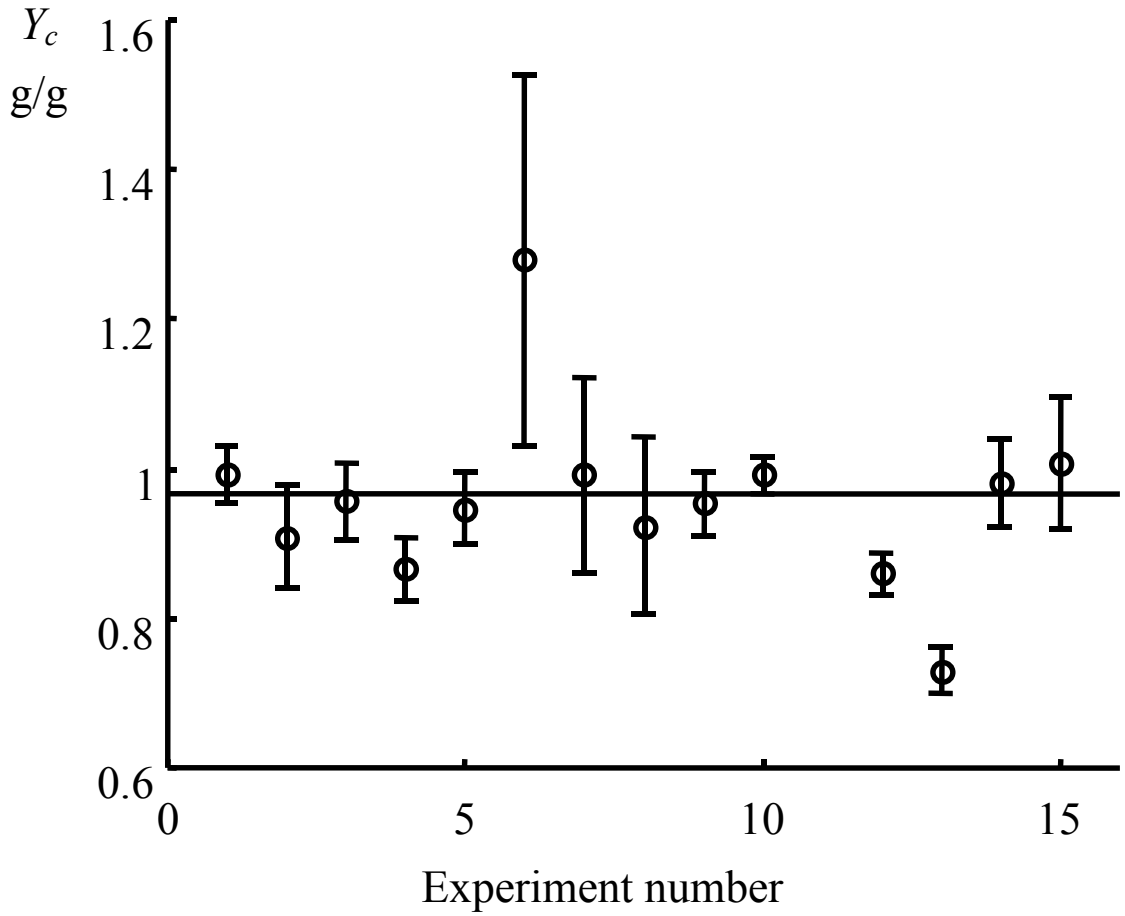

Figure 7