# Prediction of Confidence Limits for Diacetyl Concentration During Beer Fermentation

Ioan-Cristian Trelea, Sophie Landaud, Eric Latrille, Georges Corrieu

# Prediction of confidence limits for the diacetyl concentration during beer fermentation

Ioan Cristian TRELEA✉, Sophie LANDAUD, Eric LATRILLE, Georges CORRIEU

Unité mixte de recherche de Génie et Microbiologie des Procédés Alimentaires

Institut national agronomique Paris - Grignon, Institut National de la Recherche Agronomique

BP 01, 78850 Thiverval-Grignon, France. E-mail: trelea@grignon.inra.fr

## Summary

Most brewing yeast strains produce diacetyl during the alcoholic fermentation in concentrations well above those tolerated in the finished beer. The conditions and the duration of the "diacetyl rest", of the beer maturation and the process scheduling could be optimized if the diacetyl concentration at the end of the alcoholic fermentation was known in advance. A dynamic model for diacetyl production and reduction has been developed based on experimental data from 14 laboratory scale (15 L) lager beer fermentations carried out in various conditions of temperature (10–16°C), top pressure (50–850 mbar), initial yeast concentration (5–20 million cells / mL) and initial wort gravity (1036–1099 g/L). Uncertainties due to measurement errors, model parameters and batch-to-batch variability were described in a probabilistic framework. The model predicts a probability distribution for the final diacetyl concentration from which a median value and an upper bound, at a specified confidence level, are derived. It is demonstrated that in-line diacetyl measurements at early stages of the fermentation greatly reduce the uncertainty about the final diacetyl level in each specific batch.

## Keywords

alcoholic fermentation, carbon dioxide production, diacetyl, dynamic model, probability distribution

# Introduction

In most lager beers, diacetyl gives an undesired buttery flavor if present in concentrations above 0.05 to 0.1 mg/L. Usually, diacetyl is removed either during the dedicated “diacetyl rest” phase or during the beer maturation phase. The length and the conditions (e.g. temperature) of these phases are greatly influenced by the amount of diacetyl present at the end of the main fermentation. For efficient production scheduling, it is important to predict this concentration as early and as accurately as possible.

Several mathematical models for diacetyl concentration are available in the literature. They are all based on the well-established fact that diacetyl is simultaneously produced and reduced during the fermentation process. The mechanisms they assume or imply are very different, however, and not always satisfactory from a biochemical point of view. This is probably due to the fact that the production and reduction rates could not be measured separately. In the model proposed by Engasser et al. (6), the diacetyl production is proportional to the alcoholic fermentation rate, with a constant yield, while the reduction rate is proportional to the concentration of the reactant in the limiting reaction step. Garcia et al. (7) established a model in which the diacetyl production rate depends on the alcoholic fermentation rate, on the biomass concentration and on the valine consumption rate. The reduction rate involved cell concentration corrected by a cell aging factor. This is not entirely satisfactory because (i) no predictive formula is given for the valine concentration and (ii) a rate limiting step that was shown to be non enzymatic (18) was assumed to depend on the cell concentration. Gee and Ramirez (8) model the total vicinal diketone concentration, i.e. diacetyl plus 2,3-pentanedione. The drawback of this approach is that the organoleptic threshold of 2,3-pentanedione is at least 6 times higher than that of the diacetyl (11) while in our experiments the two concentrations are similar. The total concentration can thus be hardly representative of the beer flavor. In the model, the diketone production rate is proportional to the specific growth rate and to the cell concentration. The reduction rate is proportional to the total diketone and to the cell concentrations which does not correspond to current biochemical knowledge as explained above. Andres-Toro et al. (1) divided the biomass into several compartments, only one of which (the active biomass) is involved in diacetyl synthesis. In their model, the diacetyl synthesis rate is proportional to the fermentable sugar concentration and the reduction rate to the ethanol concentration, which is biochemically rather surprising.

The model proposed in the present work was meant to be useful from a process engineering perspective. It has the following features: (i) It is based on a predictive model for the alcoholic fermentation (17). The progress of

the alcoholic fermentation is routinely monitored in the brewery, via gravity and possibly evolved $CO_2$ and/or refractive index measurements (5,13). In case of discrepancies, the model can be easily adapted for each particular batch (4). (ii) It does not involve quantities usually unmeasured in industry such as valine concentration, active yeast concentration or specific growth rate. (iii). It does not try to describe the intricacy of the diacetyl production mechanism but is compatible with the current understanding of biochemical pathways. (iv) A special emphasis was placed on the reliability of the predictions, on the sources of uncertainty and of the ways to reduce them through in-line measurements. (v) The model was validated in a large range of operating conditions: fermentation temperatures between 10 and 16°C, top pressures between 50 and 850 mbar, initial cell concentrations between 5 and 20 million cells/mL, initial wort densities between 1036 and 1099 kg/m$^3$.

# Experimental

Experiments were carried out in 15 L, 0.5 m high stainless steel tanks (LSL Biolafitte, France) under gentle agitation at 100 rpm. Preliminary experiments showed that mechanical agitation was needed to compensate the absence of the natural agitation that occurs in large scale brewing (10 m high tanks or higher) due to $CO_2$ release. The lager wort and the industrial yeast strain*, Saccharomyces cerevisiae* var. *uvarum*, were provided by the Institut Français de Brasserie et Malterie (IFBM, France). The run R12 was carried out with a different, "Cedarex light hopped wort" provided by Hunton and Fison, UK. Starter cultures were carried out at 20°C in 5 L of wort during 3 days. Temperature was decreased to fermentation temperature 1 day before inoculation and the starter cultures were centrifuged three times (4000 rpm) in physiological saline. The conditions of the experimental runs were selected according to a $2^3$ experimental design, as indicated in Table 1. The factors were the fermentation temperature, the top pressure and the initial yeast concentration. Runs R01-R04 and R06-R09 were performed in extreme operating conditions, while runs R05 and R13-R15 were intended to be repetitions of the central point of the experimental design and R11 a repetition of R03. The run R10 was atypical because of its very high initial wort gravity and run R12 because of the different wort. Due to experimental uncertainty, the initial cell concentration, the initial wort gravity and the lowest top pressure could not be replicated exactly. The actual (measured) values of these factors are indicated in Table 1.

The concentrations of diacetyl and of its precursor, α-acetolactate, were determined by gas chromatography coupled with mass spectrometry (9). The ethanol concentration was determined using a Carlo Erba 5300 gas chromatograph equipped with a stainless steel column (200 mm, ∅0.3 mm) coated with Chromosorb 101 (SGE,

USA). The concentration of fermentable sugar (the sum of the concentrations of fructose, glucose, maltose and maltotriose) was determined using a High Performance Liquid Chromatography system (Waters, USA) with an Aminex HPX-87C column (300 mm, ∅7.8 mm, BioRad, USA) at 85°C. The density of the filtered and degassed wort was determined with a 10 mL pycnometer. The refractive index was measured with an ATAGO refractometer. The evolved $CO_2$ was recorded with a gas meter (Schlumberger, France), delivering a pulse for every liter of gas. Taking into account the amounts of $CO_2$ produced in the considered experiments, the resolution of this measurement was better than 0.5%. The measurements describing the alcoholic fermentation (ethanol, density, $CO_2$ production, refractive index and fermentable sugar) were reconciled using well-established stoichiometric relationships (16). The yeast cell concentration was determined with a particle counter (Coulter Z1, Coultronics, France). Three counts were performed at 3 and 3.5 μm and the logarithmic average of the six counts was taken.

# Dynamic model

## Biochemical background

Relevant pathways involved in diacetyl synthesis and degradation are schematically represented in Figure 1. Quantitatively, the main mass flow is from fermentable sugars (in beer fermentation, these are mainly maltose and glucose) to ethanol and $CO_2$. About 94% of the carbon flows through this pathway, called "alcoholic fermentation" in this text. An intermediate product, important for many cell functions including diacetyl synthesis, is pyruvate. In our experiments, the carbon flow rate from pyruvate to α-acetolactate was about three orders of magnitude lower than from pyruvate to ethanol and $CO_2$. The actual rate is the result of complex interactions and internal cell regulation. The reaction step from α-acetolactate to diacetyl is purely chemical (non-enzymatic). Diacetyl is transformed enzymatically into 2,3-butanediol, whose contribution to the beer flavor is negligible.

From this simplified picture of the biological reality, plausible modeling assumptions can be formulated. (i) Since the rate of the alcoholic fermentation is an important indicator of cell metabolism, the α-acetolactate production rate should be closely related to it. However, the fraction of the carbon flow diverted through this pathway is not necessarily constant. (ii) The reaction steps from diacetyl to 2,3-butanediol being much faster

than from α-acetolactate to diacetyl, the concentration of the diacetyl during the alcoholic fermentation is negligible compared to that of α-acetolactate (9,10). After complete yeast removal by filtration, however, diacetyl can not be reduced further and may accumulate into the finished beer if α-acetolactate is still present. So, what is actually important for the finished beer flavor, is the total concentration of diacetyl plus equivalent α-acetolactate, also called "total" or "potential" diacetyl. It is this total concentration that will be modeled in the paper, and called "diacetyl" for brevity. (iii) In presence of the yeast, the rate of removal of the "total" diacetyl is given by the limiting rate of the chemical reaction step from α-acetolactate to diacetyl and thus independent of the yeast concentration. This is the case for the model of Engasser et al. (6) but not for the other models discussed above (1,7,8).

## Alcoholic fermentation model

The dynamic model for the alcoholic fermentation was developed previously (15,17) and is briefly recalled here for completeness. It was constructed by analogy with classical microbial growth kinetics with substrate limitation and product inhibition (Appendix):

$$\frac{dC_p(t)}{dt} = \nu\left(\theta(t),\, C_d(t)\right)\cdot\frac{S(t)}{K_S + S(t)}\cdot\frac{1}{1+\left(E(t)/K_E\right)^2}\cdot\left(C_p(t) + K_X X_0\right) \qquad \text{Equation 1}$$

$$C_p(0) = 0$$

$$E(t) = Y_{E/C} C_p(t) \qquad \text{Equation 2}$$

$$S(t) = S_0 - Y_{S/C} C_p(t) \qquad \text{Equation 3}$$

In Equation 1, the rate of the alcoholic fermentation was described by the rate of $CO_2$ production $dC_p/dt$. Simultaneously, ethanol ($E$) is produced and fermentable sugars ($S$) are consumed, with constant yields (Equations 2 and 3). The initial fermentation rate, when $C_p = 0$, is taken proportional to the initial yeast concentration $X_0$. The "specific" fermentation rate $\nu$ is given by:

$$\nu\left(\theta,\, C_d\right) = K_\nu \exp\left(K_{\nu\theta}(\theta - \theta_0) - K_{\nu C}(C_d - C_{d0})\right) \qquad \text{Equation 4}$$

For small temperature variations (±3K) compared to the typical absolute fermentation temperature (286K), Equation 4 is a close approximation of the Arrhenius law. A similar dependence was assumed for the dissolved $CO_2$ ($C_d$). For modeling purposes, it was assumed that the produced $CO_2$ ($C_p$) was dissolved in the wort until saturation ($C_{sat}$), and released afterwards ($C_r$):

$$C_d(t) = \min\{C_p(t), \quad C_{sat}(\theta, p)\}$$ Equation 5

$$C_r(t) = \max\{0, \quad C_p(t) - C_{sat}(\theta, p)\}$$ Equation 6

The wort saturation with $CO_2$ was determined using an empirical formula, based on tables of experimental values provided by the Institut Français de Brasserie et Malterie:

$$C_{sat}(\theta, p) = H(\theta)\, p$$ Equation 7

$$H(\theta) = K_0(1 + K_1\theta)\exp(-K_2\theta)$$ Equation 8

In applications, model predictions should be compared with available measurements and model adaptation steps possibly taken (4). Equation 6 assumed that the released $CO_2$ was measured. If the measured quantity was the wort density ($D$) or the refractive index ($R$), the following equations should be used instead:

$$D(t) = D_0 - Y_{D/C} C_p(t)$$ Equation 9

$$R(t) = R_0 - Y_{R/C} C_p(t)$$ Equation 10

Numerical values for the coefficients involved in the alcoholic fermentation model are reported in Table 2. They were either taken from literature or determined from available experimental data. The runs R01-R04 and R06-R09 were used for parameter identification and the other runs for model validation. The model parameters were determined using a standard maximum likelihood method, based on ethanol, fermentable sugar, wort density and wort refractive index measurements simultaneously. This is equivalent to a nonlinear least-squares model fitting provided that each measurement is weighted by the inverse of the standard deviation of its measurement error (16). Numerical computations were performed using the Levenberg-Marquardt minimization method for nonlinear sums of squares implemented in the Matlab software package (3). Reasonable initial values for the model parameters were determined by repeated simulation and graphical comparison of experimental and predicted data.

## Diacetyl production and removal

As discussed previously, in presence of the yeast, the "total" diacetyl concentration is almost equal to the α-acetolactate concentration ($A$). The rate of its conversion to 2,3-butanediol is limited by a non-enzymatic reaction step and hence should depend on the current α-acetolactate concentration, on temperature and on the characteristics of the medium such as composition, pH, redox potential etc. The rate of α-acetolactate production

was assumed to depend on the overall metabolic activity of the yeast, described by the rate of the alcoholic fermentation $dC_p / dt$:

$$\frac{dA(t)}{dt} = Y_{A/C}(t)\frac{dC_p(t)}{dt} - K_A \exp\left(K_{A\theta}(\theta - \theta_0)\right) \cdot A(t) \qquad \text{Equation 11}$$
$$A(0) = 0$$

The characteristics of the microorganism and of the medium are included in the model through the numerical values of the diacetyl yield ($Y_{A/C}$), of the diacetyl reduction constant ($K_A$) and of the temperature sensitivity constant ($K_{A\theta}$). These values are thus specific for each yeast strain – wort type couple. The model of Engasser et al. (6) has the same form as Equation 11 with the sugar consumption rate replacing the $CO_2$ production rate (which is equivalent as the ratio of the two rates is nearly constant) and with a constant yield $Y_{A/C}$. When trying to fit Equation 11 to experimental data from any of the runs R01-R15, however, it turned out that the constant yield hypothesis implied that, as soon as the active fermentation phase was over (sugar depletion), the acetolactate concentration should almost immediately fall to zero. This sharp fall was not observed. Rather, even after complete sugar exhaustion, the acetolactate concentration continued to decline slowly. The prediction of diacetyl level after the end of the alcoholic fermentation being of highest practical importance, the model based on the constant yield hypothesis was judged unsatisfactory. Data presented by Engasser et al. (6) stop before the end of the main fermentation so that the above-mentioned discrepancy in their model could not be assessed. The failure of the constant yield model to describe the data can be explained as follows. The maximum of the α-acetolactate concentration appears nearly at the same time as the maximum of the alcoholic fermentation rate. To describe this, a constant yield model needs high values for both $Y_{A/C}$ and $K_A$, meaning that the α-acetolactate is reduced almost immediately after being produced.

Experimental data could be described by Equation 11 by supposing a variable yield: $Y_{A/C}$ should be maximum at the beginning of the fermentation and fall to zero well before the fermentable sugar depletion. Resulting numeric values for $Y_{A/C}$ and $K_A$ were much less than in the constant yield case, meaning that α-acetolactate was produced during the first half of the fermentation, accumulated into the medium and then declined slowly. The following empirical equation was found to describe the yield variation adequately:

$$\frac{dY_{A/C}(t)}{dt} = -K_Y \frac{dC_p(t)}{dt} \cdot Y_{A/C}(t) \qquad \text{Equation 12}$$
$$Y_{A/C}(0) = Y_0$$

Again, the alcoholic fermentation rate $dC_p/dt$ was used to describe the intensity of the metabolic activity. It should be emphasized that, while Equation 11 is based on biochemical insight, Equation 12 is purely descriptive. The cell regulation mechanisms which make the acetolactate production stop well before fermentable sugar exhaustion have not been investigated in this work. It is worth noting that linking diacetyl production to cell growth instead of alcoholic fermentation progress would result (for the considered database) in an equivalent model since proportionality between yeast growth and alcoholic fermentation progress was observed (Appendix). Proportionality does not necessarily hold in general (e.g. due to yeast settling in naturally agitated tanks) but does often hold in the early stages of the fermentation when diacetyl is produced.

# Sources of uncertainty in model parameter estimation

In order to obtain practically useful predictions of the total diacetyl concentration during the alcoholic fermentation, numerical values for the parameters appearing in Equations 11 and 12 had to be found using some experimental runs and the validity of the predictions checked using the remaining runs. Mathematical predictions of physical reality are always subject to some uncertainty but in the present study the uncertainties were found to be large enough to be worth a detailed study.

**Diacetyl measurement error**

Usual concentrations of diacetyl in beer are less than 1 mg/L and their determination required complex experimental work (9). The results were affected by unavoidable experimental error traditionally described by a Gaussian probability distribution with zero mean and unknown standard deviation $\sigma_A$. Let $I$ denote the set of runs used for model parameter identification, $n_i$ the number of diacetyl measurements and $Y_{0i}$ the initial diacetyl yield in run $i$. Let $w$ be the vector of the unknown model parameters in equations 11 and 12:

$$w = \left[\log K_A \quad K_{A\theta} \quad \log K_Y \quad \log Y_{01} \quad ... \quad \log Y_{0i} \quad ...\right]^T, \qquad i \in I \qquad \text{Equation 13}$$

The logarithms of the unknown scale parameters were determined instead of the parameters themselves. From a practical perspective, this insured positiveness and increased numerical robustness and accuracy. The theory also states that, if the order of magnitude is a priori unknown, parameters should be located uniformly on a logarithmic scale (12).

If $a_{ij}$ is the *measured* total diacetyl concentration in the sample $j$ of the experiment $i$ and $A_{ij}(w)$ is the value *predicted* by the model with unknown parameters set to $w$, then the probability density of observing the given set of measurements, also called the likelihood of the sample, is given by (2):

$$likelihood = \frac{1}{\left(\sigma_A\sqrt{2\pi}\right)^n}\prod_{i\in I}\prod_{j=1}^{n_i}\exp\left(-\frac{\left(A_{ij}(w)-a_{ij}\right)^2}{2\sigma_A^2}\right) \qquad \text{Equation 14}$$

Here $n=\sum_{i\in I} n_i$ is the total number of available measurements. The value $w^*$ which maximizes the likelihood is called the maximum likelihood estimator of the unknown model parameters. It is mathematically equivalent, but numerically more convenient, to minimize the minus logarithm of the likelihood function $L(w)$:

$$L(w) = n\log\sigma_A + \frac{n}{2}\log 2\pi + \frac{1}{2\sigma_A^2}M(w) \qquad \text{Equation 15}$$

$$M(w) = \sum_{i\in I}\sum_{j=1}^{n_i}\left(A_{ij}(w)-a_{ij}\right)^2 \qquad \text{Equation 16}$$

The sum of squares $M(w)$ is independent of $\sigma_A$ and was minimized numerically using a Levenberg-Marquardt algorithm (3), giving:

$$w^* = \arg\min M(w) \qquad \text{Equation 17}$$

Initial values of the parameters required by the numerical minimization algorithm were determined by repeated simulation and graphical comparison of the predicted and measured values. Differentiating the Equation 15 with respect to $\sigma_A$ and setting the derivative to zero yields a slightly biased estimator of the measurement standard deviation. The related, but unbiased version of this estimator is (2):

$$\sigma_A = \sqrt{\frac{M(w^*)}{n-n_w}} \qquad \text{Equation 18}$$

where $n_w$ is the number of the parameters to be estimated and $n-n_w$ is the number of degrees of freedom.

**Model parameter uncertainty**

Dynamic model parameters can only be determined with finite accuracy from Equation 17 due to the presence of measurement noise. When the number of degrees of freedom is sufficiently large (larger than for example 50, which was the case in this study) the maximum likelihood estimator $w^*$ of the unknown parameter vector $w$ is unbiased and has an approximately normal sampling distribution with a covariance matrix $V$ that can be calculated numerically as the inverse of the local information matrix (2):

$$V = \left[ \frac{\partial^2 L}{\partial w \partial w^T}(w^*) \right]^{-1} \qquad \text{Equation 19}$$

**Batch to batch variation of the initial diacetyl yield**

Equation 13 indicates that the model parameters $K_A$, $K_{A\theta}$ and $K_Y$ were supposed common to all runs, i.e. constant for a given combination of wort and yeast strain, while the initial diacetyl yields $Y_{0i}$ were determined specifically for each run. A common diacetyl yield for all runs, or a model based on measured quantities, could not be determined. Repetitions of the same run (within experimental error: e.g. R03-R11, R13-R15) exhibited differences in diacetyl level as large as 1:2 (see the "Results and discussion" section below). Such variations, in otherwise similar fermentation runs, were also observed by other authors (e.g. Figures 2 and 3 in ref. (14)). Exact causes could not be determined. After careful examination across all runs, particularly low or high diacetyl levels could not be correlated satisfactorily to any of the following measurements: temperature, top pressure, initial yeast concentration, yeast growth rate, initial dissolved oxygen concentration, evolution of pH, redox potential and electrical conductivity during the batch, alcoholic fermentation rate, aminoacid uptake. It should be noted that the observed variations in the initial diacetyl yield were far larger than uncertainties due to measurement errors.

In absence of a satisfactory deterministic model, the batch-to-batch variations of the initial diacetyl yields were described by a probability distribution. Among the various distributions tested (normal, log-normal, exponential, gamma), the log-normal distribution was the most plausible in the light of the data (initial yields determined for runs R01-R11 and R13-R15). In particular, the normal distribution predicted too low probabilities for the highest yields; they wouldn't have reasonable chance to appear in a sample of 14 runs. The choice of a log-normal distribution is consistent with Equation 13 and can be interpreted in the light of a central limit theorem: the value of the diacetyl yield in a particular run is influenced by a large number of multiplicative causes. The probability density function of the log-normal distribution is (2):

$$f(Y_0) = \frac{1}{Y_0 \lambda \sqrt{2\pi}} \exp\left( -\frac{1}{2} \left( \frac{\log Y_0 - \mu}{\lambda} \right)^2 \right) \qquad \text{Equation 20}$$

A log-normal distribution for $Y_0$ means that $\log Y_0$ is distributed normally. The parameter $\mu$ represents the expected value of $\log Y_0$ and $\lambda$ represents its standard deviation. However, some care is needed when transposing results from the normal to the log-normal distribution because of its asymmetric shape. For a normal distribution, the mean (expected value), the median (value below and above which fall 50% of the random samples) and the

mode (most probable value) are all the same. For a log-normal distributed variable, the median is still $\exp(\mu)$, but the mean is $\exp(\mu + 0.5 \cdot \lambda^2)$ and the mode is $\exp(\mu - \lambda^2)$. In order to avoid confusion, results will be presented in terms of the median value.

The maximum likelihood estimator for $\mu$ is unbiased, with minimum variance (2):

$$\mu = \frac{1}{n_r} \sum_{i=1}^{n_r} \log Y_{0i} \qquad \text{Equation 21}$$

where $n_r$ is the number of runs used for statistical parameter estimation. The unbiased version of the estimator for $\lambda$ is:

$$\lambda = \sqrt{\frac{1}{n_r - 1} \sum_{i=1}^{n_r} \left( \log Y_{0i} - \mu \right)^2} \qquad \text{Equation 22}$$

**Determination of the initial diacetyl yield for a specific batch**

Equations 13 to 19 are useful in the model identification step when one uses data from several runs to determine fixed model parameters $K_A$, $K_{A\theta}$, $K_Y$ and $\sigma_A$. In doing so, however, one also has to determine initial yields for those runs. Equations 20 to 22 apply when one wishes to determine a priori plausible values for the diacetyl yield for a given wort and yeast strain combination. Another intermediate situation is likely to appear in practice: One already knows the constant model parameters $K_A$, $K_{A\theta}$ and $K_Y$, but wishes to determine the diacetyl yield for a specific batch using available measurements performed with known accuracy $\sigma_A$. In this case, Equation 13 reduces to:

$$w = \left[ \log Y_0 \right] \qquad \text{Equation 23}$$

Equations 14 to 17 and 19 remain valid for this particular case of a single unknown parameter.

# Prediction of the diacetyl concentration in presence of uncertainty

The time evolution of the diacetyl concentration was predicted by the dynamic model consisting of Equations 1 through 12. The alcoholic fermentation model (Equations 1 to 10) turned out to be at least an order of magnitude more accurate than the diacetyl model (Equations 11 and 12). The alcoholic fermentation model was considered deterministic and the corresponding uncertainty was neglected throughout this study. Other sources of uncertainty were taken into account selectively depending on the model usage and on the information at hand.

**Prediction without in-line diacetyl concentration measurements**

In order to predict the diacetyl concentration in a run where no diacetyl measurements would be performed, it was assumed that: (i) Constant model parameters $\log K_A$, $K_{A\theta}$ and $\log K_Y$ (Equation 13) have a joint normal probability distribution with mean $w^*$ (Equation 17) and covariance $V$ (Equation 19). (ii) In absence of any information for that specific batch, the logarithm of the initial diacetyl yield $\log Y_0$ has a normal probability distribution (Equation 20) with mean $\mu$ and standard deviation $\lambda$ (Equations 21 and 22). (iii) The measurement error has a normal probability distribution with zero mean and standard deviation $\sigma_A$ (Equation 18).

Let $A(t,w^*)$ be the diacetyl concentration predicted by the model (Equations 11 and 12) for time $t$, with model parameters set to $w^*$. The actual (or true) diacetyl concentration in a given run is generally different from the predicted one because the model parameters can not be determined exactly. In order to state the accuracy of the model predictions rigorously, confidence limits for the true diacetyl concentration were calculated with the "error propagation" formula:

$$\underline{A}(t) = A(t,w^*) + \sqrt{\frac{\partial A}{\partial w^T}(t,w^*) \cdot V \cdot \frac{\partial A}{\partial w}(t,w^*)} \cdot U_{\alpha/2} \qquad \text{Equation 24}$$

$$\overline{A}(t) = A(t,w^*) + \sqrt{\frac{\partial A}{\partial w^T}(t,w^*) \cdot V \cdot \frac{\partial A}{\partial w}(t,w^*)} \cdot U_{1-\alpha/2} \qquad \text{Equation 25}$$

Here $\underline{A}(t)$ and $\overline{A}(t)$ are lower and upper bounds respectively for the true diacetyl concentration at the confidence level $\alpha$. Stated another way, the true (and generally unknown) diacetyl concentration has the (small) probability $\alpha$. of being either less than $\underline{A}(t)$ or greater than $\overline{A}(t)$. $U$ is the cumulative distribution function of the standard normal distribution (zero mean and unit variance). It should be noted that the first 3 elements of the vector $w$ ($\log K_A$, $K_{A\theta}$ and $\log K_Y$), and the corresponding 3×3 block in matrix $V$ were determined in the preliminary model identification step using runs included in the identification set $I$. The 4th element of $w$ is the expected value for $\log Y_0$, that is $\mu$, and the corresponding diagonal (4,4) element in $V$ is the estimation of its variance, that is $\lambda^2$. The remaining 1×3 and 3×1 blocks in $V$, representing covariances between $\log Y_0$ on one hand and $\log K_A$, $K_{A\theta}$ and $\log K_Y$ on the other hand, are set to zero since they are determined separately.

If diacetyl concentration measurements are performed the measured concentration $a(t)$ is generally different from the actual (or true) one due to unavoidable measurement errors. The confidence limits given by Equations 24 and 25 are not appropriate for the measured values because they do not take into account the measurement error. Lower ($\underline{a}(t)$) and upper ($\overline{a}(t)$) confidence limits of the measured diacetyl concentration were determined as:

$$\underline{a}(t) = A(t, w^*) + \sqrt{\frac{\partial A}{\partial w^T}(t, w^*) \cdot V \cdot \frac{\partial A}{\partial w}(t, w^*) \ + \ \sigma_A^2} \cdot U_{\alpha/2} \qquad \text{Equation 26}$$

$$\overline{a}(t) = A(t, w^*) + \sqrt{\frac{\partial A}{\partial w^T}(t, w^*) \cdot V \cdot \frac{\partial A}{\partial w}(t, w^*) \ + \ \sigma_A^2} \cdot U_{1-\alpha/2} \qquad \text{Equation 27}$$

with the standard the assumption of independent normally distributed measurement noise.

**Prediction with in-line diacetyl concentration measurements**

If diacetyl measurements could be performed during the fermentation, information would be gained about the diacetyl yield for each specific batch. For such cases, the following assumptions were made: (i) The constant model parameters log $K_A$, $K_{A\theta}$ and log $K_Y$ (Equation 13) have the same joint normal probability distribution as above, estimated from previous runs and fixed. (ii) The logarithm of the initial diacetyl yield log $Y_0$ is estimated in real-time from in-line data and has an independent normal probability distribution with mean given by Equation 17 and variance given by Equation 19, both reduced to the single parameter case (Equation 23). (iii) The measurement error has a normal probability distribution with zero mean and standard deviation $\sigma_A$ determined from previous runs and fixed.

Confidence limits for the predicted diacetyl concentration were calculated as before (Equations 24 to 27), except that the 4th element of the parameter vector $w$ was given by Equations 17 and 23 instead of $\mu$ and the (4,4) element of the covariance matrix $V$ was given by Equations 19 and 23 instead of $\lambda^2$.

# Results and discussion

## Diacetyl model identification and validation

The parameters of the dynamic diacetyl model ($K_A$, $K_{A\theta}$ and $K_Y$) as well as the experimental spread of the measurement of the diacetyl concentration ($\sigma_A$) were determined using 8 runs performed under extreme conditions of temperature, top pressure and initial yeast concentration, namely R01-R04 and R06-R09 (Table 1). The median values and the confidence intervals are reported in Table 3.

The prediction of the diacetyl concentration using this model is illustrated in Figure 2. The initial $CO_2$ *evolution* rate is zero, since the $CO_2$ is first dissolved in the wort. This is confirmed by the measured evolution rate.

However, the initial simulated $CO_2$ *production* rate is nonzero, and equals 0.23 $g \cdot L^{-1} \cdot h^{-1}$ in the considered run. During the first 20 hours, the diacetyl production rate (not shown) is high and nearly constant, the acceleration in the alcoholic fermentation rate being compensated by the decrease in the diacetyl yield. The maximum of the diacetyl concentration is reached when the production rate equals the reduction rate. After 35 hours, the diacetyl yield is so low that the diacetyl production rate becomes negligible compared to the reduction rate, and the predicted diacetyl concentration decreases exponentially.

# Verification of the statistical hypothesis

Several statistical assumptions have been made concerning the diacetyl measurement error and the batch-to-batch variation of the diacetyl yield. These assumptions have to be verified in light of the data before drawing conclusions about the uncertainty in the model parameters and its effect on the prediction of the diacetyl concentration.

The confidence interval for the mean of the model residuals was [-0.031 0.015] $g \cdot L^{-1}$, hence the mean was not significantly different from zero. The normality of the residuals was tested using the modified Anderson-Darling statistic (2). The normality hypothesis could not be rejected at a 0.05 significance level. Additional assurance about the normality hypothesis was provided by the statistical plot in Figure 3A. Thus, the assumption about the zero-mean normal distribution of the measurement noise was satisfied meaning that the assumption about the multi-normal distribution of the model parameter vector $w$ was acceptable.

The hypothesis of the log-normal distribution of the initial diacetyl versus $CO_2$ yield ($Y_0$) was also verified using the modified Anderson-Darling statistic (2). The hypothesis could not be rejected at a 0.05 significance level. A graphical verification was provided by the statistical plot in Figure 3B. The parameters of the log-normal distribution of the initial diacetyl yield ($\lambda$ and $\mu$) are reported in Table 3 together with their confidence intervals.

# Comparison of various sources of uncertainty

When predicting the diacetyl concentration for a new fermentation run, three sources of uncertainty were considered: the measurement noise, the imperfect knowledge of the model parameters and the batch-to-batch variation of the initial diacetyl versus $CO_2$ yield ($Y_0$). As far as the initial diacetyl yield was concerned, two practical situations have been distinguished: (i) the diacetyl concentration was not measured and the *a priori* distribution of the yield had to be considered; (ii) in-line diacetyl measurements were performed and an initial

yield could be determined for each specific run thus reducing the uncertainty. In the following, the effect of each source of uncertainty is examined separately in order to state its relative importance.

The effect of the measurement noise is illustrated in Figure 4A. The scatter of the experimental data is particularly obvious between 150 and 250 hours. Confidence intervals bracket experimental data tightly in this region. After 300 hours, experimental data are smoother and in good agreement with model simulations. Large confidence intervals are maintained, however, due to the hypothesis of constant variance of the measurement noise. The first 2 measurements suggest a lag in the diacetyl formation. This lag was not observed systematically and was not included in the present model but might become significant for fermentations performed without mechanical stirring.

The effect of the model parameter uncertainty ($K_A$, $K_{A\theta}$ and $K_Y$) is illustrated in Figure 4B. The first part of the fermentation (25 h) is not affected by these three parameters. The sensitivity to the mentioned parameters, which enter the diacetyl reduction model and the descriptive yield dynamic, is mostly apparent around the maximum of the diacetyl concentration and afterwards. Unlike the preceding case, this is an uncertainty on the true (rather than measured) concentration.

When in-line diacetyl measurements are not performed, the uncertainty on the initial diacetyl yield has to be taken into account through its *a priori* probability distribution. The very large confidence interval shown in Figure 4C should include the true diacetyl concentration for most runs. For many runs, however, these limits are quite conservative. For example, the run R11 (Figure 4C) has one of the lowest yields, and decisions based on the upper limit of the confidence interval, such as the duration of the diacetyl rest phase, would be substantially in error.

A way to determine tighter confidence limits is to perform in-line diacetyl measurements soon after the beginning of the alcoholic fermentation. Even a limited number of measurements significantly improve the estimation of the yield for that specific batch. In the same run R11 (Figure 4D), two measurements were considered, at 10 and 20 % of the total $CO_2$ produced (or, equivalently, of the total consumed fermentable sugar). The upper limit of the confidence interval is reduced by a factor of almost 2 compared to Figure 4C. By chance, the two selected measurements are rather high and the diacetyl concentration still appears slightly overestimated. If the number of measurements taken into account is increased, the confidence interval shrinks further and the overestimation disappears.

## Prediction of plausible ranges of diacetyl concentration

In practice, all sources of uncertainty have to be accounted for simultaneously. It appeared from the previous discussion and from Figure 4 that the uncertainty due to the initial diacetyl yield was dominant but could be substantially reduced by one or two real-time measurements. Hence, it is worth distinguishing between the cases when diacetyl measurements are and are not available. In Figure 5A the combined effect of all uncertainties is shown, supposing no in-line diacetyl measurements were performed. It happened that in this particular run the initial diacetyl yield had a typical value and the median prediction was representative of the true concentration. However, the use of the a priori probability distribution for the diacetyl yield produced quite large confidence intervals. In Figure 5B, the initial diacetyl yield was estimated specifically for the considered run using two measurements. The confidence limits for the true concentration are much tighter.

## Model adaptation to a different wort

The possibility of applying the diacetyl model to the concentrated Cedarex wort was investigated in run R12 (Figure 6). In the alcoholic fermentation model, a correction had to be made by modifying the 'specific' fermentation rate parameter $K_V$ from 0.0474 (Table 2) to 0.0600 $h^{-1}$, reflecting a slight overall acceleration of the alcoholic fermentation in otherwise similar operating conditions. Dependence of the fermentation rate on the wort composition is well known in brewing practice and is generally explained by differences in concentrations of growth factors and/or unsaturated fatty acids. All other parameters listed in Table 2 remained unchanged. No significant difference in the biomass growth was observed in this run compared to the other runs in the database (Figure 7). The diacetyl model had to be modified, however, by allowing diacetyl to be produced later during the alcoholic fermentation, i.e. slowing down the yield decrease: the determined value of the parameter $K_Y$ (using Equation 17) was 0.028 instead of 0.203 $g^{-1}\cdot L$. Parameters $K_A$ and $K_{A\theta}$ were left unchanged (Table 3). Thus, the high maximum diacetyl concentration in this run (2 $mg\cdot L^{-1}$ instead of an average of 0.7 $mg\cdot L^{-1}$ in the other runs) was due to a longer production period rather than to a higher yield. The estimated initial diacetyl yield for run R12 was 0.13 $mg\cdot g^{-1}$, one of the lowest among all runs. The parameters $\lambda$ and $\mu$ describing the probability distribution of the initial yield are expected to change for the Cedarex wort, but this could only be confirmed if more (at least 10) runs were available.

In the light of the data provided by the run R12, it can be speculated that the constant yield model of Engasser et al. (6) might indeed be appropriate for some sorts of wort and/or yeast strains but this would be a very special case and needs further experimental verification.

# Conclusion

A dynamic model predicting the diacetyl concentration during the alcoholic fermentation of beer has been established. In face of high experimental variability, not explained by measured operating conditions and routinely measured wort composition, a probabilistic framework was adopted. The effects of the measurement errors, of the uncertainty in the model parameters and of the batch-to-batch variability were examined separately and in combination. It was demonstrated that the effect of the batch-to-batch variability was dominant. This effect could be substantially reduced for each specific batch, however, by a limited number of in-line diacetyl measurements during the early stages of the alcoholic fermentation.

The model did not attempt to explain the intricacy of the metabolic pathways leading to diacetyl formation but was intended to be useful from a process engineering perspective. It was based on the dynamic of the alcoholic fermentation which is relatively well known and routinely measured in industry. The probabilistic framework provided realistic confidence intervals for the diacetyl concentration. The upper bound of such a confidence interval could be used for taking well-informed decisions concerning the required ‘diacetyl rest’ and the beer production scheduling in general.

The adaptation of the developed model to a new sort of wort required the modification of a limited number of coefficient values. The investigation of the effects of the wort composition, of the yeast strain, of the tank geometry and agitation is the subject of future work.

# Appendix: Relationship between alcoholic fermentation and cell growth

The mathematical form of the Equation 1 was suggested by the classical microbial growth kinetic with substrate limitation and product inhibition:

$$\frac{dX(t)}{dt} = \mu(\theta, C_d, S, E)X(t) \qquad \text{Equation 28}$$

In the considered experiments, near proportionality was observed between cell growth and the progress of the alcoholic fermentation as described by the $CO_2$ release (Figure 7):

$$C_p(t) = Y_{C/X}\left(X(t) - X_0\right) \qquad \text{Equation 29}$$

Strict proportionality implies a constant yield $Y_{C/X}$. Equation 1 can be obtained by substituting $X(t)$ from Equation 29 into Equation 28 and using the notation:

$$\mu(\theta, C_d, S, E) = \nu\left(\theta(t),\, C_d(t)\right) \cdot \frac{S(t)}{K_S + S(t)} \cdot \frac{1}{1 + \left(E(t)/K_E\right)^2} \qquad \text{Equation 30}$$

$$K_X = \frac{1}{Y_{C/X}} \qquad \text{Equation 31}$$

Strict proportionality between yeast growth and alcoholic fermentation progress is not expected to hold in all cases. For example, the substrate limitation and the product (ethanol and dissolved $CO_2$) inhibition constants are non necessarily the same for alcoholic fermentation and growth. Yeast settling in naturally (as opposed to mechanically) agitated tanks may also invalidate Equation 29. This is why Equation 1 is only said to be constructed by analogy with the growth kinetic (Equation 28) and not derived from it. Equation 1 is still expected to remain valid (with suitably chosen limitation and inhibition constants) even if the proportionality expressed by Equation 29 does not hold.

# References


1. Andrés-Toro, B., Giron-Sierra, J.M., Lopez-Orozco, J.A., Fernandez-Condé, C., Peinado, J.M. and Garcia-Ochoa, F. A kinetic model for beer production under industrial operation conditions. Mathematics and Computers in Simulation 48:65-74, 1998.
2. Bury, K. Statistical distributions in engineering. Cambridge University Press:27-48,120-237, 1999.
3. Coleman, T., Branch, M.A. and Grace, A. Optimization toolbox for use with Matlab: User's guide. The MathWorks, Inc.:2.17-2.22, 1999.
4. Corrieu, G., Trelea, I.C. and Perret, B. On-line estimation and prediction of density and ethanol evolution in the brewery. MBAA Technical Quarterly 37(2):173-181, 2000.

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 96:297-302, 1990.
6. Engasser, J.M., Marc, I., Moll, M. and Duteurtre, B. Kinetic modelling of beer fermentation. Proceedings of the European Brewing Convention Congress 579-586, 1981.
7. Garcia, A.I, Garcia, L.A. and Diaz, M. Modelling of diacetyl production during beer fermentation. Journal of the Institute of Brewing, 100:179-183, 1994.
8. Gee, D.A. and Ramirez, W.F. A flavour model for beer fermentation. Journal of the Institute of Brewing, 100:321-329, 1994.
9. Landaud, S., Lieben, P. and Picque, D. Quantitative analysis of diacetyl, pentanedione and their precursors during beer fermentation, by an accurate GC/MS method. Journal of the Institute of Brewing 104:93-99, 1998.
10. Mathis, C., Pons, M.N. and Engasser, J.M. Development of an on-line method for monitoring of vicinal diketones and their precursors in beer fermentation. Analytica Chemica Acta 279:59-66, 1993.
11. Meilgaard, M.C., Reid, D.S. and Wyborski, K.A. Reference standards for the beer flavor terminology system. Journal of the American Society of Brewing Chemists, 36:119-128, 1982.
12. Sivia, D.S. Data analysis: a Bayesian tutorial. Oxford University Press, 112-113, 1998.
13. 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(2):44-50, 1987.
14. Tada, S., Takeuchi, T., Sone, H., Yamano,S., Schofield, M.A., Hammond, J.R.M. and Inue, T. Pilot-scale brewing with industrial scale yeasts which produce the alpha-acetolactate decarboxylase of *Acetobacter aceti ssp. xylinum*. Proceedings of the 25th European Brewing Convention Congress:369-376, 1995.
15. Titica, M., Landaud, S., Trelea, I.C., Latrille, E., Corrieu, G. and Cheruy, A. Modelling of higher alcohol and ester production kinetics based on $CO_2$ emission, with a view to beer flavor control by temperature and top pressure. Journal of the American Society of Brewing Chemists, 54(4):167-174, 2000.
16. Trelea, I.C., Latrille, L., Landaud, S. and Corrieu, G. Reliable estimation of the key variables and of their rates of change in alcoholic fermentation. Bioprocess and biosystems engineering (in press) 2001.
17. Trelea, I.C., Titica, M., Landaud, S., Latrille, E., Corrieu, G. and Cheruy, A. Predictive modelling of brewing fermentation: from knowledge-based to black-box models. Mathematics and Computers in Simulation 56:405-424, 2001.
18. Wainwright, T. Diacetyl – a review. Journal of the Institute of Brewing 79:451-470, 1973.

## Nomenclature

| Symbol | Units | Significance |
|---|---|---|
| $A$ | $mg \cdot L^{-1}$ | "Total" or "potential" diacetyl concentration = diacetyl + equivalent α-acetolactate |
| $A_{ij}$ | $mg \cdot L^{-1}$ | Predicted total diacetyl concentration for sample *j* of the run *i* |
| $\underline{A}$ | $mg \cdot L^{-1}$ | Lower predicted bound for the true total diacetyl concentration, at a specified confidence level |
| $\overline{A}$ | $mg \cdot L^{-1}$ | Upper predicted bound for the true total diacetyl concentration, at a specified confidence level |
| $a_{ij}$ | $mg \cdot L^{-1}$ | Measured total diacetyl concentration in sample *j* of the run *i* |
| $\underline{a}$ | $mg \cdot L^{-1}$ | Lower predicted bound for the measured total diacetyl concentration, at a specified confidence level |
| $\overline{a}$ | $mg \cdot L^{-1}$ | Upper predicted bound for the measured total diacetyl concentration, at a specified confidence level |
| $C_d$ | $g \cdot L^{-1}$ | Carbon dioxide dissolved in the wort |
| $C_{d0}$ | $g \cdot L^{-1}$ | Dissolved carbon dioxide concentration at the operating conditions of the central point of the experimental design |
| $C_p$ | $g \cdot L^{-1}$ | Carbon dioxide produced in the alcoholic fermentation, per liter of wort |
| $C_r$ | $g \cdot L^{-1}$ | Carbon dioxide released during alcoholic fermentation, per liter of wort |
| $C_{sat}$ | $g \cdot L^{-1}$ | Dissolved carbon dioxide concentration at saturation |
| $D$ | $g \cdot L^{-1}$ | Wort density |
| $D_0$ | $g \cdot L^{-1}$ | Initial wort density |
| $E$ | $g \cdot L^{-1}$ | Ethanol concentration |
| $H$ | $g \cdot L^{-1} \cdot mbar^{-1}$ | Solubility of the carbon dioxide in the wort |
| $I$ | none | Set of experimental runs used for model parameter identification |
| $i$ | none | Index of an experimental run in the set *I* |
| $j$ | none | Index of a sample in an experimental run |
| $K_v$ | $h^{-1}$ | Maximum "specific" fermentation rate at the operating conditions of the central point of the experimental design |
| $K_{v\theta}$ | $°C^{-1}$ | Temperature effect on the alcoholic fermentation rate constant |

| | | |
|---|---|---|
| $K_{vC}$ | $g^{-1}\cdot L$ | Dissolved carbon dioxide effect on the alcoholic fermentation rate constant |
| $K_0$ | $g\cdot L^{-1}\cdot mbar^{-1}$ | Solubility of the carbon dioxide in the wort at 0°C |
| $K_1$,$K_2$ | $°C^{-1}$ | Temperature effect on the carbon dioxide solubility constants |
| $K_A$ | $h^{-1}$ | Diacetyl reduction rate constant |
| $K_{A\theta}$ | $°C^{-1}$ | Temperature effect on diacetyl reduction rate constant |
| $K_E$ | $g\cdot L^{-1}$ | Ethanol inhibition constant |
| $K_S$ | $g\cdot L^{-1}$ | Substrate saturation constant |
| $K_X$ | $g\cdot L^{-1}$ | Initial cell concentration constant |
| | $(10^6 \text{ cells})^{-1}$ | |
| $K_Y$ | $g^{-1}\cdot L$ | Diacetyl versus carbon dioxide yield dynamic constant |
| $L$ | none | Minus logarithm of the parameter likelihood function |
| $M$ | $mg^2\cdot L^{-2}$ | Sum of squares of the errors between predicted and measured diacetyl concentrations |
| $n$ | none | Total number of diacetyl measurements used for model parameter identification |
| $n_i$ | none | Number of diacetyl measurements from run $i$ used for model parameter identification |
| $n_r$ | none | Number of runs used for the estimation of the parameters $\lambda$ and $\mu$ |
| $n_w$ | none | Number of elements in vector $w$ |
| $p$ | mbar | Top pressure in the fermentation tank |
| $R$ | none | Wort refractive index |
| $R_0$ | none | Initial wort refractive index |
| $S$ | $g\cdot L^{-1}$ | Fermentable sugar concentration |
| $S_0$ | $g\cdot L^{-1}$ | Initial fermentable sugar concentration |
| $t$ | h | Time since yeast pitching |
| $U_\alpha$ | none | Inverse cumulative probability function for a standard normal distribution, at level $\alpha$. |
| | | Value below which falls a fraction $\alpha$ of the possible values |
| $V$ | NA | Estimated covariance matrix of $w^*$ |
| $w$ | NA | Vector of unknown model parameters |
| $w^*$ | NA | Vector of most probable model parameters |
| $X$ | $10^6 \text{ cells}\cdot mL^{-1}$ | Yeast concentration |
| $X_0$ | $10^6 \text{ cells}\cdot mL^{-1}$ | Initial yeast concentration |

| | | |
|---|---|---|
| $Y_0$ | $mg \cdot g^{-1}$ | Initial diacetyl versus carbon dioxide yield |
| $Y_{0i}$ | $mg \cdot g^{-1}$ | Initial diacetyl versus carbon dioxide yield in experimental run *i* |
| $Y_{A/C}$ | $mg \cdot g^{-1}$ | Diacetyl versus carbon dioxide yield |
| $Y_{D/C}$ | $g \cdot g^{-1}$ | Density versus carbon dioxide yield |
| $Y_{E/C}$ | $g \cdot g^{-1}$ | Ethanol versus carbon dioxide yield |
| $Y_{R/C}$ | $g^{-1} \cdot L$ | Wort refractive index versus carbon dioxide yield |
| $Y_{S/C}$ | $g \cdot g^{-1}$ | Fermentable sugar versus carbon dioxide yield |
| $\theta$ | °C | Wort temperature |
| $\theta_0$ | °C | Fermentation temperature the central point of the experimental design |
| $\lambda$ | none | Standard deviation of the logarithm of $Y_0$ |
| $\mu$ | none | Mean value of the logarithm of $Y_0$ |
| $\mu$ | $h^{-1}$ | Specific cell growth rate |
| $\nu$ | $h^{-1}$ | "Specific" rate of the alcoholic fermentation |
| $\sigma_A$ | $mg \cdot L^{-1}$ | Measurement standard deviation of the diacetyl concentration |

NA = Not applicable

Table 1. Experimental conditions

| Experimental run | Temperature °C | Top pressure mbar | Initial yeast concentration $10^6$ cells $mL^{-1}$ | Initial wort density g $L^{-1}$ |
|---|---|---|---|---|
| R01 | 10 (L) | 800 (H) | 40 (H) | 1070 |
| R02 | 10 (L) | 60 (L) | 19 (H) | 1037 |
| R03 | 10 (L) | 50 (L) | 6 (L) | 1049 |
| R04 | 10 (L) | 800 (H) | 5 (L) | 1050 |
| R05 | 13 (C) | 450 (C) | 10 (C) | 1049 |
| R06 | 16 (H) | 50 (L) | 20 (H) | 1047 |
| R07 | 16 (H) | 800 (H) | 22 (H) | 1051 |
| R08 | 16 (H) | 790 (H) | 5 (L) | 1051 |
| R09 | 16 (H) | 40 (L) | 4 (L) | 1047 |
| R10* | 10 (L) | 800 (H) | 33 (H) | 1099 |
| R11 | 10 (L) | 70 (L) | 6 (L) | 1048 |
| R12** | 13 (C) | 450 (C) | 10 (C) | 1046 |
| R13 | 13 (C) | 450 (C) | 14 (C) | 1051 |
| R14 | 13 (C) | 450 (C) | 10 (C) | 1050 |
| R15 | 13 (C) | 390 (C) | 9 (C) | 1049 |

*Atypical run: higher initial wort density

**Atypical run: different wort

(L) Low value in the experimental design

(C) Central value in the experimental design

(H) High value in the experimental design

Table 2. Numerical values of the coefficients in the alcoholic fermentation model

| Sym-bol | Units | Value with 95% confidence limits | | | Assumed distribution for the estimator | Source |
|---|---|---|---|---|---|---|
| | | Min | Median | Max | | |
| $\theta_0$ | °C | NA | 13 | NA | none | Central condition of the experimental design |
| $C_{d0}$ | $g \cdot L^{-1}$ | NA | 2.76 | NA | none | Central condition of the experimental design |
| $K_v$ | $h^{-1}$ | 0.0415 | 0.0450 | 0.0488 | log-normal | Determined from runs R1-R4 and R6-R9 with a maximum likelihood method |
| $K_{v\theta}$ | $°C^{-1}$ | 0.118 | 0.125 | 0.131 | normal | Idem |
| $K_{vC}$ | $g^{-1} \cdot L$ | 0.020 | 0.055 | 0.090 | normal | Idem |
| $K_E$ | $g \cdot L^{-1}$ | 20.7 | 22.6 | 24.7 | log-normal | Idem |
| $K_S$ | $g \cdot L^{-1}$ | ND | 3 | ND | none | References (1,7,8) |
| $K_X$ | $g \cdot L^{-1} \cdot (10^6 \text{ cells})^{-1}$ | 0.120 | 0.143 | 0.173 | log-normal | Determined from runs R1-R4 and R6-R9 with a maximum likelihood method |
| $K_0$ | $g \cdot L^{-1} \cdot mbar^{-1}$ | ND | $2.83 \cdot 10^{-5}$ | ND | none | Regression with experimental data from tables used by professional brewers |
| $K_1$ | $°C^{-1}$ | ND | $3.66 \cdot 10^{-3}$ | ND | none | Idem |
| $K_2$ | $°C^{-1}$ | ND | $3.35 \cdot 10^{-2}$ | ND | none | Idem |
| $Y_{D/C}$ | $g \cdot g^{-1}$ | NA | 1 | NA | none | Reference (16) |
| $Y_{E/C}$ | $g \cdot g^{-1}$ | 1.013 | 1.028 | 1.043 | normal | Idem |
| $Y_{R/C}$ | $g^{-1} \cdot L$ | $2.440 \cdot 10^{-4}$ | $2.494 \cdot 10^{-4}$ | $2.548 \cdot 10^{-4}$ | normal | Idem |
| $Y_{S/C}$ | $g \cdot g^{-1}$ | 1.834 | 1.884 | 1.934 | normal | Idem |

ND = not determined

NA = not applicable

Table 3. Numerical values of the coefficients in the diacetyl concentration model

| Sym-bol | Units | Value with 95% confidence limits | | | Assumed distribution for the estimator | Degrees of freedom | Source |
|---|---|---|---|---|---|---|---|
| | | Min | Median | Max | | | |
| $K_A$ | $h^{-1}$ | $8.3 \times 10^{-3}$ | $10.2 \times 10^{-3}$ | $12.4 \times 10^{-3}$ | log-normal | 75 | Determined from runs R01-R04 and R06-R09 using Equation 17 |
| $K_{A\theta}$ | $°C^{-1}$ | 0.129 | 0.176 | 0.222 | normal | 75 | Idem |
| $K_Y$ | $g^{-1}\cdot L$ | 0.162 | 0.203 | 0.255 | log-normal | 75 | Idem |
| $\lambda$ | none | 0.24 | 0.34 | 0.54 | chi-2 | 13 | Determined from runs R01-R11 and R13-R15 using Equation 22 |
| $\mu$ | none | -1.78 | -1.59 | -1.39 | student | 13 | Determined from runs R01-R11 and R13-R15 using Equation 21 |
| $\sigma_A$ | $mg\cdot L^{-1}$ | 0.092 | 0.106 | 0.125 | chi-2 | 75 | Determined from runs R01-R04 and R06-R09 using Equation 18 |

# Figure legends

Figure 1. Main biochemical pathways involved in diacetyl synthesis and reduction.

Figure 2. Prediction of the diacetyl concentration in a typical fermentation experiment. Measured diacetyl concentration (o), simulated diacetyl concentration (—), measured $CO_2$ evolution rate (•••), simulated $CO_2$ evolution rate (– –) and simulated diacetyl versus $CO_2$ yield (– · –). The residual (unexplained) variance in the measured diacetyl concentration represents 6% of the total variance.

Figure 3. Graphical verification of the main statistical assumptions. Experimental (o) and theoretical (—) probability distributions. Probability scales are nonlinear, such as to make the theoretical cumulative distribution functions linear. (**A**) Diacetyl model residuals (measurement noise) are likely to come from a normal probability distribution. (**B**) Initial diacetyl yield ($Y_0$) is reasonably well described by a log-normal probability distribution.

Figure 4. Effect of the various sources of uncertainty, taken separately, on the prediction of the diacetyl concentration. Experimental data (o), median concentration (—), 95% confidence limits for the true concentration (– –) and 95% confidence limits for the measured concentration (·····). (**A**) Uncertainty due to measurement errors. (**B**) Uncertainty due to the model parameters $K_A$, $K_{A\theta}$ and $K_Y$. (**C**) Uncertainty due to the batch-to-batch variation of the initial diacetyl versus $CO_2$ yield ($Y_0$). (**D**) Uncertainty due to the initial diacetyl versus $CO_2$ yield ($Y_0$), estimated specifically for the considered batch using the measurements marked by squares (▣).

Figure 5. Prediction of the diacetyl concentration, taking into account the various sources of uncertainty simultaneously. Experimental data (o), median concentration (—), 95% confidence limits for the true concentration (– –) and 95% confidence limits for the measured concentration (·····). (**A**) If no real-time diacetyl measurements were performed, the a priori distribution for the initial diacetyl yield ($Y_0$) had to be taken into account. (**B**) If diacetyl measurements marked by squares (▣) were available, a batch-specific yield and tighter confidence limits for the true concentration could be determined.

Figure 6. Prediction of the diacetyl concentration for the Cedarex wort. Experimental data (o), median concentration (—), 95% confidence limits for the true concentration (– –) and 95% confidence limits for the measured concentration (·····).

Figure 7. Relationship between the biomass concentration and the alcoholic fermentation progress. Data is plotted for released $CO_2$ comprised between 0.01 and 0.95 of the total amount in order to avoid distortions due to dissolved $CO_2$ and inaccurate biomass measurements after yeast flocculation. Within the accuracy of the biomass measurement, the relationship between the biomass concentration and the alcoholic fermentation progress appears linear.

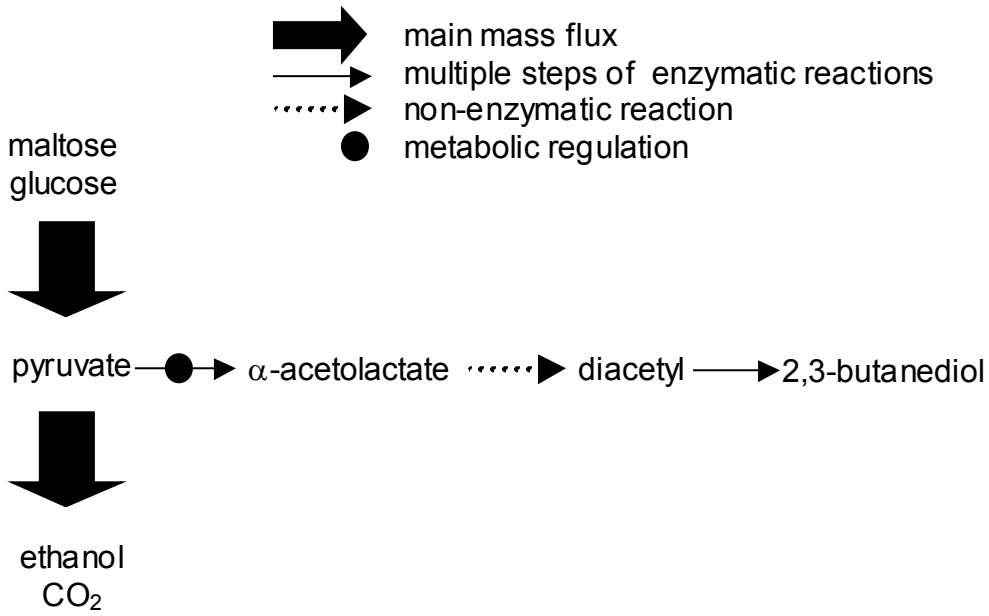



Figure 1. Trelea et al.

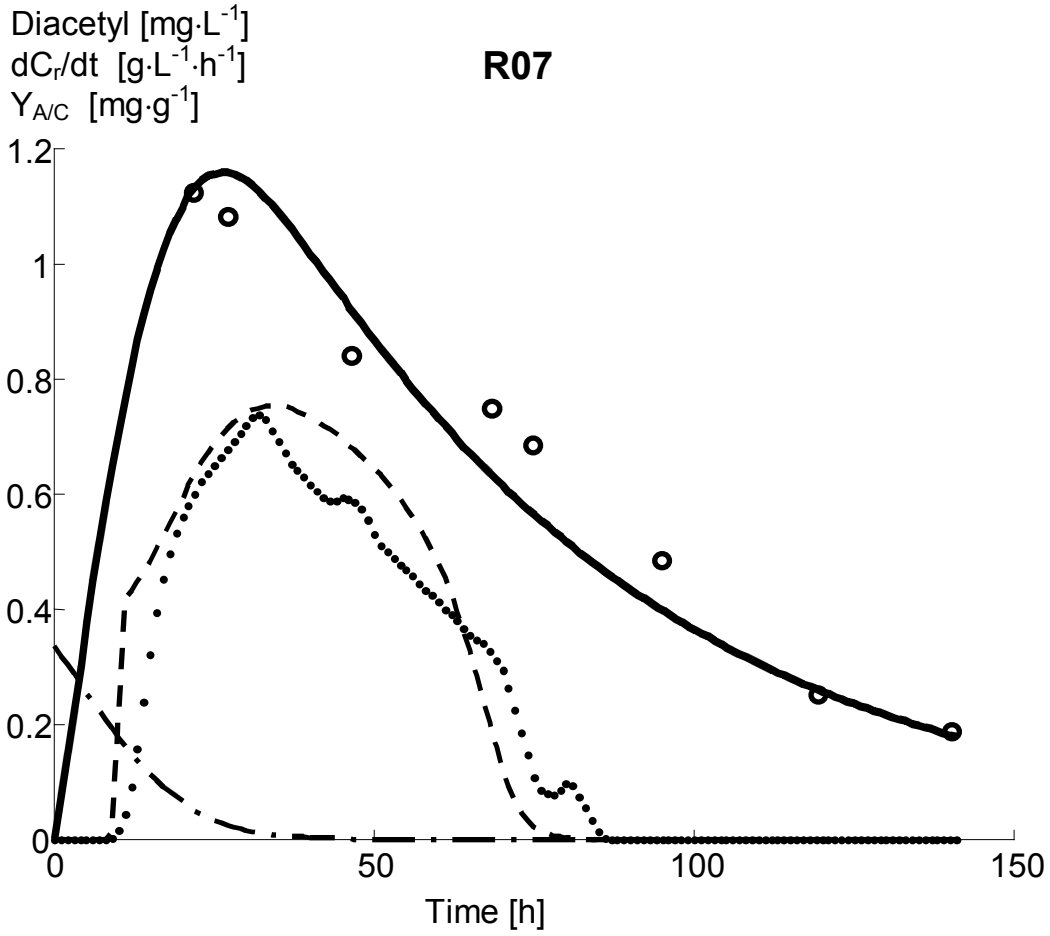
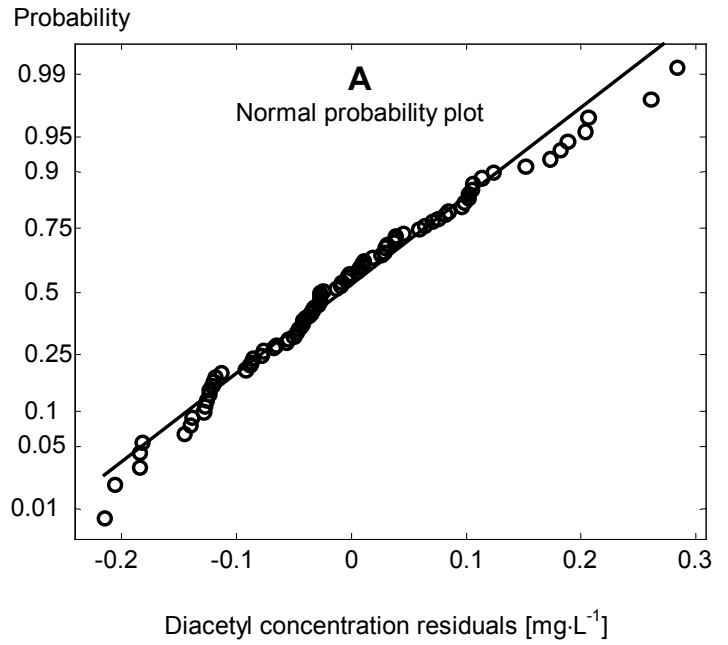



Figure 2. Trelea et al.



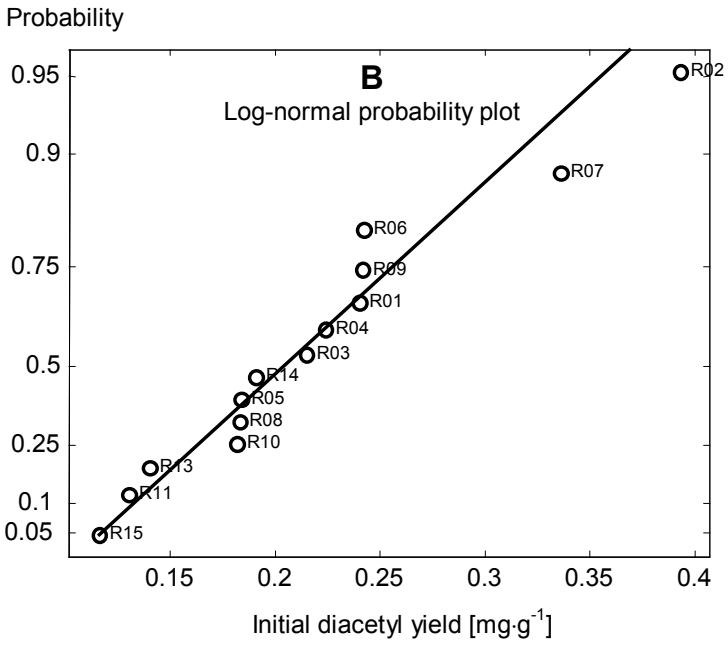




Figure 3. Trelea et al.

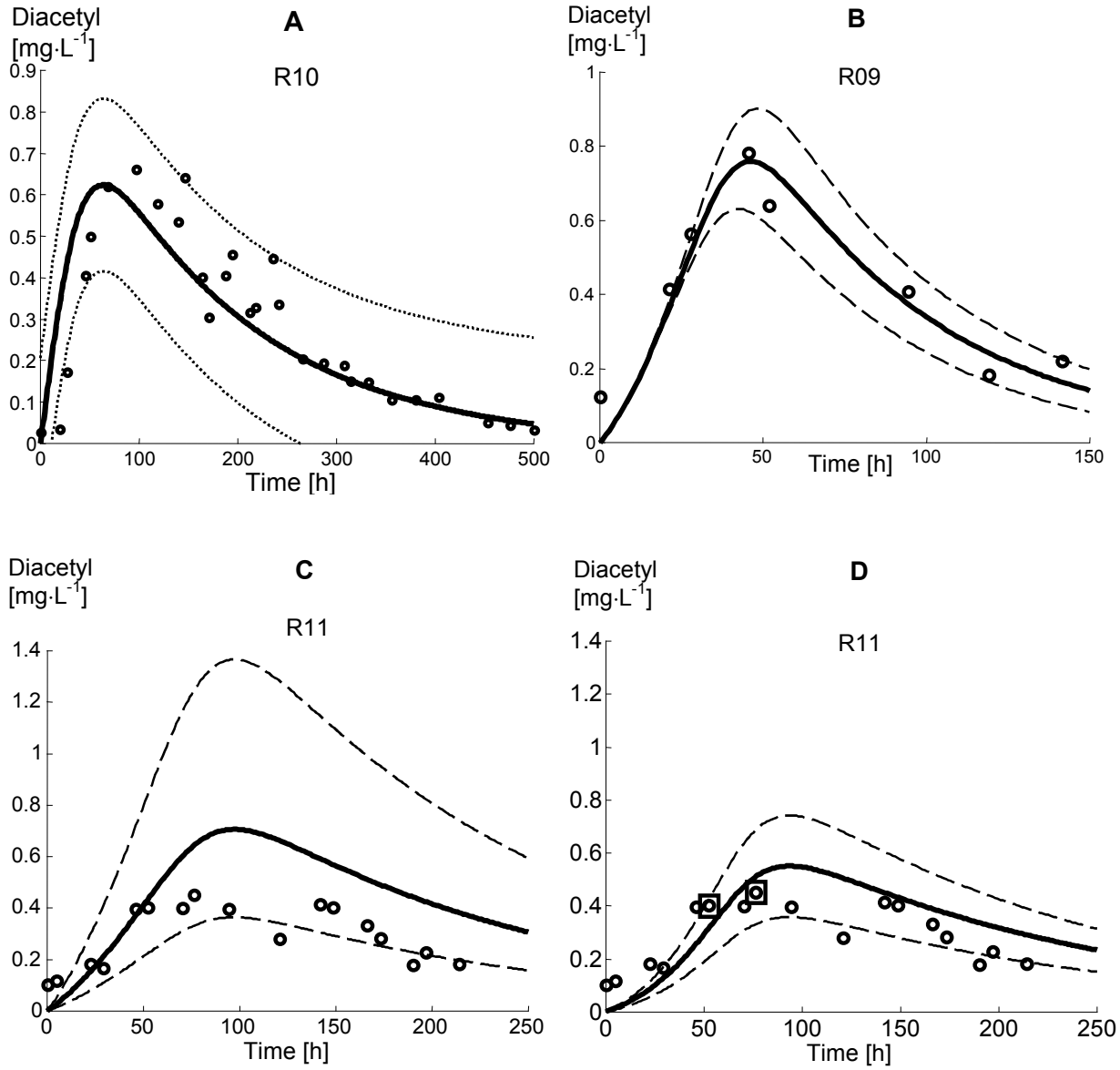



Figure 4. Trelea et al.

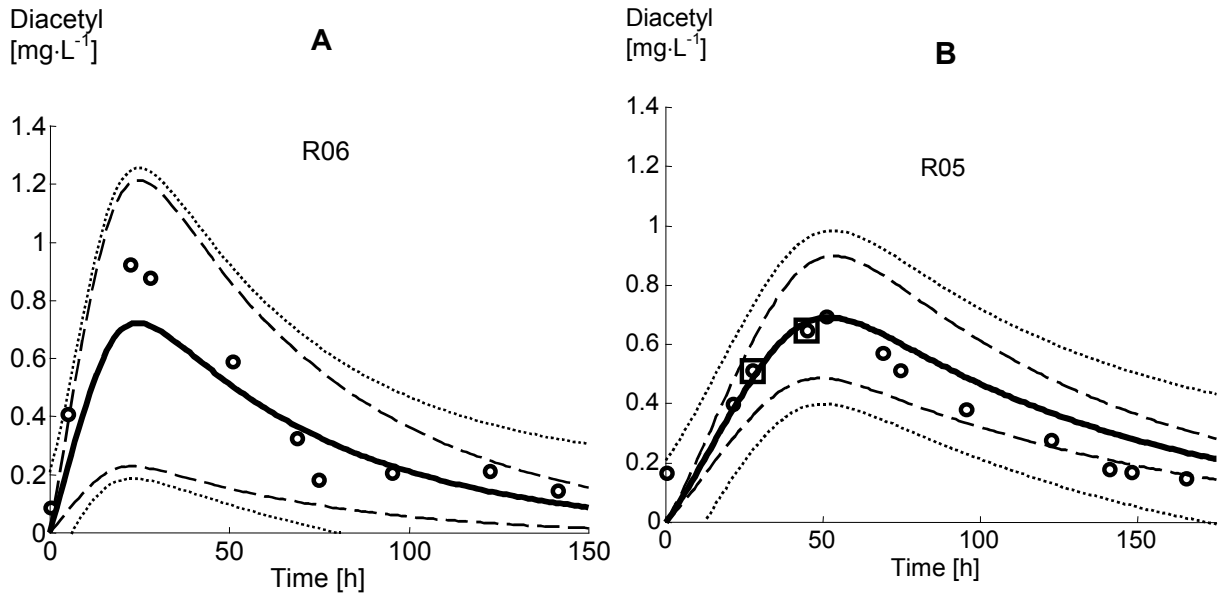



Figure 5. Trelea et al.

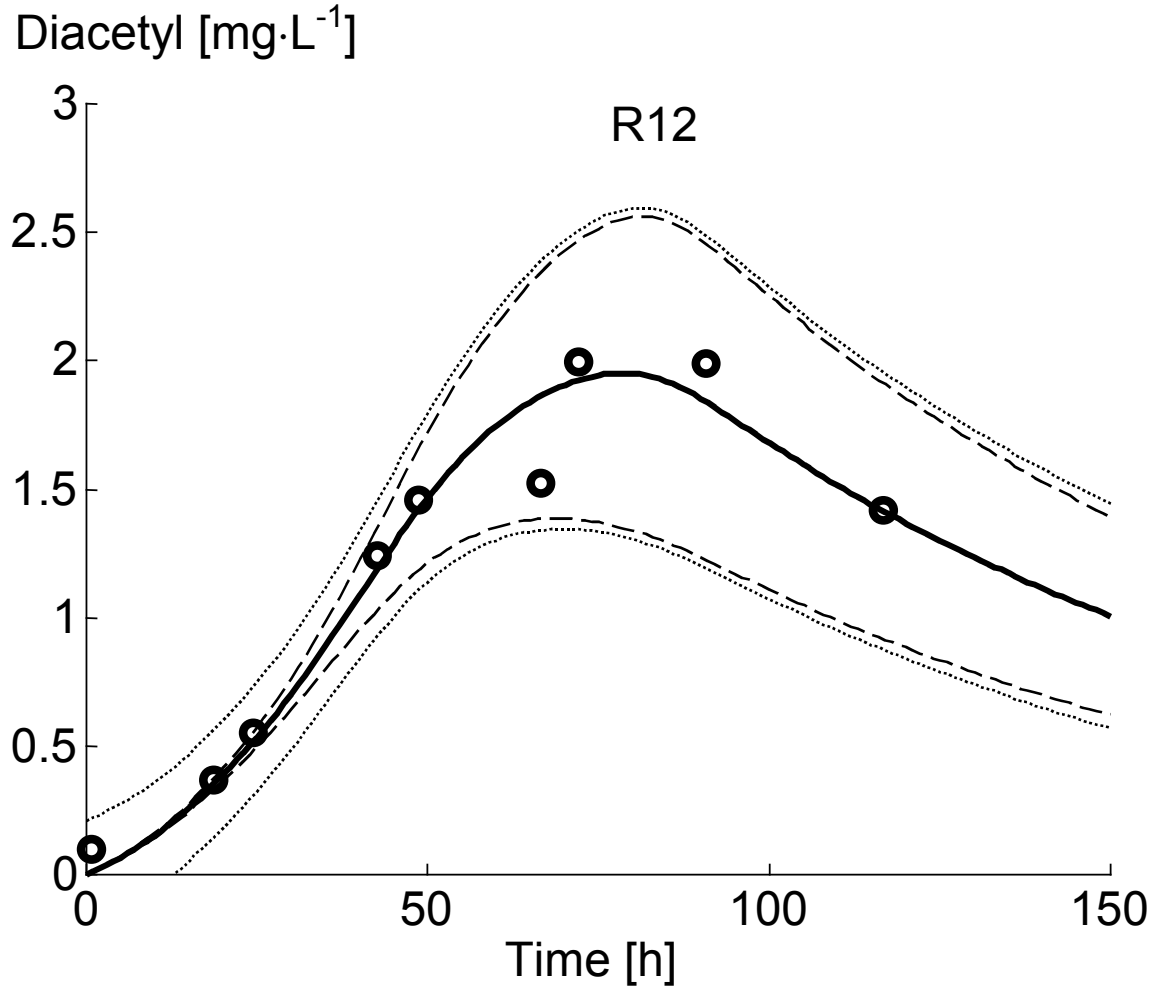



Figure 6. Trelea et al.

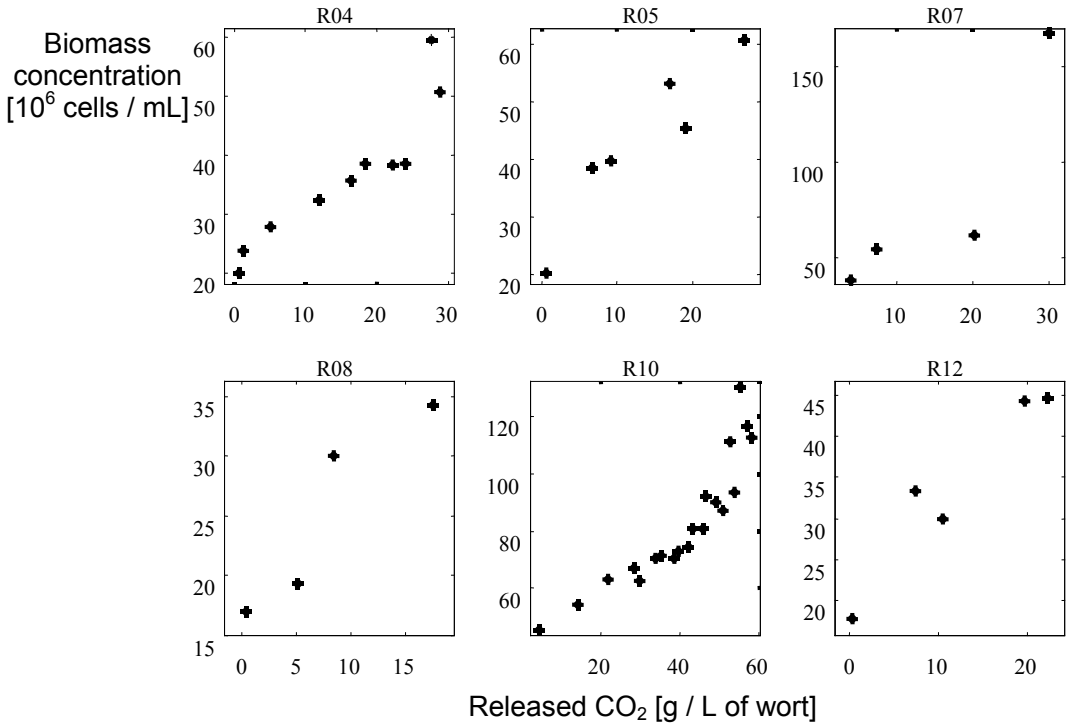



Figure 7. Trelea et al.