Thesis for the Degree of Licentiate of Philosophy

Modeling Yeast Growth Curves

Ilona Pylvänäinen

CHALMERS | GÖTEBORG UNIVERSITY

Department of Mathematical Statistics
Chalmers University of Technology and Göteborg University
SE-412 96 Göteborg, Sweden
August 2003

Modeling Yeast Growth Curves

Ilona Pylvänäinen
Department of Mathematical Statistics
Chalmers University of Technology and Göteborg University

Abstract

This thesis addresses three issues related to growth curve modeling. First, we discuss issues of calibration and blank correction related to using yeast growth data from a microbiological growth analyzer Bioscreen C.

Second, we investigate various parametric approaches to modeling growth curves. We compare the fits of modifications of logistic, Gompertz and Chapman-Richards models for *Saccharomyces cerevisiae* growth curves. The comparisons show that the modified Chapman-Richards model describes our growth data best. Furthermore, we introduce a couple of simultaneous models for two growth curves. They can be used to estimate growth parameters from double samples. These models can easily be generalized to more than two samples.

Third, an EM estimation method for growth parameters in the modified Chapman-Richards function setup is developed. The idea is to take rounding errors in the measured data into account when fitting a growth curve. We compare the estimates from this method with the estimates from the least squares fit of the modified Chapman-Richards model and the estimates from an empirically evaluated ad hoc method used in practice today.

Keywords: Bioscreen, Chapman-Richards model, EM algorithm, Gompertz model, parametric growth curve, logistic model, optical density (OD), *Saccharomyces cerevisiae*

Acknowledgments

I would like to thank my supervisors, Olle Nerman and Marita Olsson, my co-supervisor Anders Blomberg, my pair PhD-project colleague, Elke Ericson, and Jonas Warringer.

Ilona Pylvänäinen, August 2003

Contents

Chapter 1

Introduction

Saccharomyces cerevisiae, better known as baker's yeast, has been domesticated for thousands of years. It is used in baking, brewing and wine making.

S. cerevisiae is also an important model system in modern biology and medicine. It reproduces quickly, and large numbers can be grown in culture in a very small amount of space, in the same way that bacteria can be grown. However, *S. cerevisiae* has the advantage of being a eukaryotic organism, and thus the results of genetic studies with *S. cerevisiae* are more easily applicable to human biology. The collaboration of more than 600 scientists from over 100 laboratories from Europe, USA, Canada and Japan resulted in the publication of the complete genomic sequence of the *S. cerevisiae* in 1996 [8]. It was the first completely sequenced eukaryote.

To complete the characterization of the *S. cerevisiae* genome, the functions of the novel genes need to be determined. The *S. cerevisiae* genome has roughly six thousand genes of which approximately sixty percent have known function. One important approach for characterizing a novel gene is to produce a knock-out mutant[1] lacking the gene, the logic being that the behavior of the mutant, its phenotype, will give important information about the function of the gene. Mutant strains of yeast are produced in many international consortia. During the past few years hundreds of papers on large-scale functional genomics have been published, where these mutant strains play a key role.

[1] A mutant: a strain that differs from the wild type because it carries one or more genetic changes in its DNA. A wild type: reference strain within a specific strain background.

Recently large-scale phenotypic characterizations have received a lot of attention. As a result, a few laboratories have specialized in the large-scale phenotypic analyses of qualitative phenotypes, such as growth or non-growth on agar plates containing a number of different compounds. Although automated to some extent, these methods require a substantial amount of manual work, and may suffer from relying on subjective judgment in the assessment of growth. Besides, these methods do not allow to distinguish the different growth parameters: lag time (time to adapt to the environmental change), growth rate (kinetics of growth), and stationary phase OD increment (related to the efficiency of growth).

Winzeler *et al* [17] showed that large numbers of deletion strains can be pooled and analyzed in parallel by using DNA bar-codes to uniquely mark each strain that misses a gene. In the next step, microarrays are used to follow the abundance of the different bar-codes as cells proliferate. Although a powerful approach, this methodology has some problems. One of the most serious concerns might be the positive and negative interactions between mixed strains that are an inherent consequence of this experimental setup [14].

In an alternative approach, Warringer and Blomberg [14] designed a system for large-scale quantitative phenotypic analysis of *S. cerevisiae* based on a commercially available Bioscreen C Analyzer[2]. In this system it is possible to screen automatically for phenotypic effects for hundreds of different mutants. The analysis of the growth curves is automatic and provides values for growth parameters. The purpose of the system is to build a publicly available phenotypic library containing growth curves and characteristics of all viable *S. cerevisiae* mutants in a wide variety of growth conditions, and to use the library for studying gene functions.

The high throughput combined with a rather good measurement precision in the Bioscreen gives good hope for large-scale analysis of yeast mutants. It also raises demand for more theoretical considerations that could improve the modeling of growth curves and the growth parameter estimation. In this thesis we discuss some of the issues related to using growth data from the Bioscreen: calibration, blank correction, and possible effect of rounding errors. We also compare the fits of several models for *S. cerevisiae* growth curves. Finally, we compare two parametric growth parameter estimation methods

[2]Labsystems Oy, Finland.

and the empirically evaluated ad hoc method developed by Warringer and Blomberg [14], and briefly investigate the effects of a false blank value.

Chapter 2

Background

2.1 How does *S. cerevisiae* grow?

S. cerevisiae divides by budding[1]. The cell cycle begins with a single, unbudded cell. This cell buds, the bud grows to nearly the size of the parent cell, the nucleus divides, and the two cells separate into two unbudded cells. The cycle then begins again for both of the cells. The result is an exponential increase in the number of cells with a doubling time roughly equal to the average cell-division-cycle time. The doubling time varies with the strain, the growth medium, and the temperature. For more details, cf. [12].

When cells are inoculated (seeded), they require a period of preparation before they start dividing. Following this *lag phase*, which may be up to several hours or days long, they enter the *exponential phase* during which their number and mass double at equal time intervals. After a period of growth at a relatively constant rate per cell, some environmental condition, such as lack of nutrient, becomes growth limiting so that the rate of growth diminishes and growth eventually stops. The population number and mass become constant. In the *stationary phase* cells do not divide anymore, but they usually remain viable for several days. An example of a typical growth curve is displayed in Figure 2.1[2].

[1] We work with haploid cells.

[2] This is an ideal growth curve. In growth inhibiting environments growth curves can have different shapes.

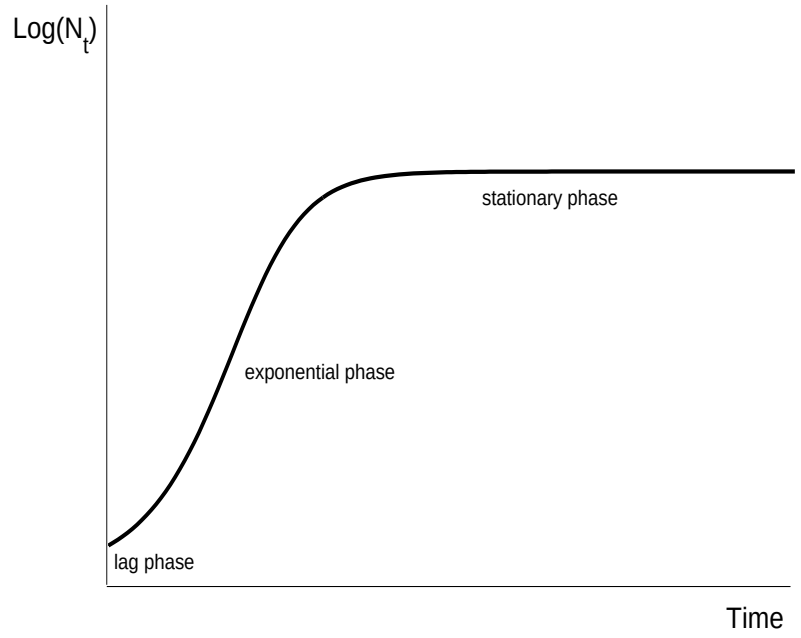

Figure 2.1: *A typical growth curve, where* $\log(N_t)$ *is the logarithm of the number of cells at time* t.

2.2 Optical density

Optical density (absorbance), OD, is a widely used concept to estimate the total number of cells present in a culture. It is a measure of the turbidity of the culture. A cell suspension looks cloudy (turbid) to the eye because cells scatter light passing through suspension. The more cell material present, the more the suspension scatters light and the more turbid it will be. Optical density can be measured with a spectrophotometer, a device that passes light through a cell suspension and detects the amount of unscattered light that passes through. For unicellular organisms, optical density is proportional (within certain limits) to cell number as well as to cell mass. Optical density measurements are quick and easy to perform, and they do not disturb or destroy the sample. They are used widely to follow the rate of growth of cultures, since the same sample can be checked repeatedly [2].

Optical density is defined as

$$\mathrm{OD} = \log_{10}\left(\frac{I_0}{I}\right),$$

where I_0 is the intensity of the incident light and I is the intensity of the transmitted light [11]. The exact optical density of a culture depends on the concentration of the cells present, the species and strain of the microbe present, the growth conditions used, and the wavelength of the light being transmitted. Optical density measurements detect all cells present in a solution, irrespectively of their viability.

2.3 Bioscreen C Analyzer

Bioscreen C Analyzer is a fully automated instrument developed to perform a wide range of microbiology experimentation automatically [1]. The system consists of Bioscreen C reader-analyzer which is simultaneously a dispenser/diluter, incubator and optical density measurement unit, integrated with a computer.

A heating/cooling system provides a wide range of incubation temperatures (from 1^oC to 60^oC), and the temperature can be changed during the experiment. Different shaking intensities and intervals can be chosen (the plates are shaken to provide homogeneous dispersion of cells). Optical density is measured by a wide band (450-580 nm) filter which is rather insensitive to color changes in the sample. For color measurement the Bioscreen has seven different filters with special wavelengths.

There are two 100-well (10×10) disposable Honeycomb multiwell plates in each Bioscreen C instrument. The volume of each well is 400 μl. Each well can be regarded as an individual test vessel. The Bioscreen microbiology reader monitors optical density of the 200 wells simultaneously. The test duration may vary from a single measurement to seven weeks, and the maximum number of measurements per well is 400. This design strongly reduces the time needed for doing experiments compared with traditional manual techniques. In addition, the precision of the Bioscreen measurements is higher than the precision of manual measurements.

2.4 Main data

Altogether 577 strains of *S. cerevisiae* — 576 mutants and one wild type — were run in synthetically defined (SD) medium[3], which is the reference condition, and in 7 different environments where either some chemical was added to the SD medium, or another temperature than the standard 30^oC was used [4]. Optical density was recorded using a Bioscreen C Analyzer. Measurements were taken every 20 minutes during a 48 hour period. Strains were run in quadruplicates (reference condition) or in duplicates (environments), in the same well location and in the same Bioscreen analyzer C instrument during separate days. For more information about the data, cf. [5].

In this thesis, growth data of eight wild types in reference condition in the same run in the same Bioscreen instrument on the same plate, are used as example data in the comparisons of estimation methods and in the investigations of the effects of a false blank value (Tables A.1-A.3). However, several hundreds of growth curves from the 576 mutants experiment were used when trying to find a good model to describe the growth of *S. cerevisiae* and in the comparisons of estimation methods.

In the sequel, for a particular strain in a fixed run, we will call the sequence of the measurement times t_p, $p = 1, ..., 145$, and the measurements z_{t_p}.

2.4.1 Calibration

A technical challenge in automated recording of yeast growth by optical density measurement is the non-linear relation between measured OD value and number of cells at higher cell densities. The yeast cultures should ideally be diluted at higher OD values, but this is not possible in a high throughput setup. Therefore a calibration curve function is needed to transform the non-linear relation to a linear, so that the calibrated OD values will be proportional to the number of cells. Also, a blank representing the background absorption of the plate has to be subtracted from the measured OD values.

[3]The SD medium contains yeast nitrogen base (YNB), ammonium, sulphate, succinic-acid and the necessary amino acids.

[4]The different environments are caffeine, dinitrophenol, methylmethanesulfonate, methylviologen, natrium chloride, and temperatures 39^oC and 41^oC.

Calibration data A 100-well plate and five different Bioscreen instruments were used. First the wells were filled with 350μl sterile water, and the OD was measured once in all Bioscreens. This gave the well and Bioscreen specific blanks. Then the water was poured off the plate and the plate was placed in a 37^{o}C chamber to make all the water evaporate. Stationary phase wild type cells (that had been growing on a shaker in 30^{o}C for over night) were spun down, washed, and suspended in water. From this cell suspension different volumes were taken into tubes. These *undiluted* samples were each diluted ten times in another tube to obtain the *diluted* samples. Then, 45 wells were filled with diluted and another 45 wells with undiluted samples, and the plate was measured once in all Bioscreens.

Since the OD values were measured in five Bioscreens, there are 225 pairwise OD measurements of diluted and undiluted samples. The well and Bioscreen specific blank was subtracted from each of the measured OD values and the blank corrected diluted values were multiplied by the dilution factor (Table A.4). Then, in order to get more robust measurements of the OD, the well specific averages over all Bioscreen instruments were taken so that there were 45 average OD values of the diluted and 45 average OD values of the undiluted samples (Table A.5). After these steps, the well specific averages of the diluted values were regarded as perfect size proportional measurements (for the higher values this is somewhat inconsistent with the resulting calibration curve).

Using the regression method, a curve was fitted with x, the well specific average of the blank corrected undiluted OD, as independent and y, the well specific average of the blank corrected diluted OD multiplied by ten, as dependent variable (Figure 2.2). In doing so, we assume that due to the blank subtraction and multiplication by ten, the amount of variation in y is much larger than the amount of variation in x.

Curve fitting We assume that there is almost a proportional relation between the blank corrected diluted OD values and the blank corrected undiluted OD values approximately up to 0.3. A cubic function

$$y = x + cx^3 \tag{2.1}$$

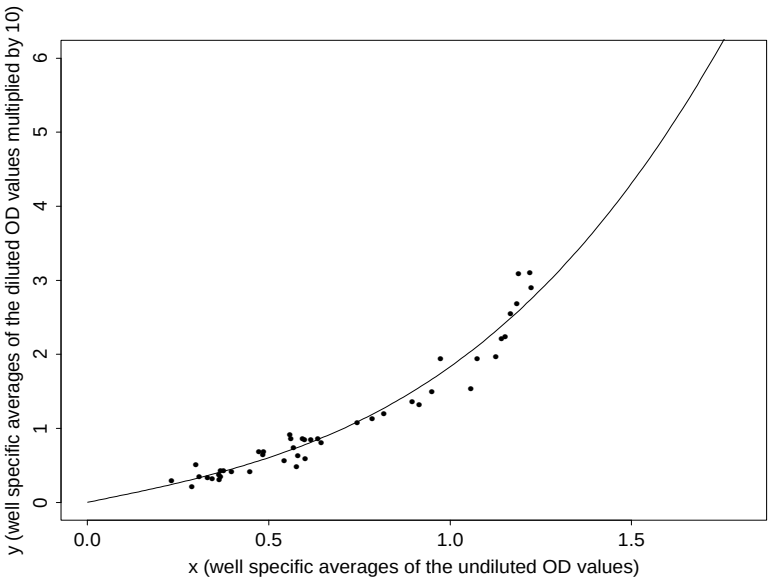

Figure 2.2: *Calibration curve and the data that was used to fit the calibration curve function. The well specific blank values are subtracted and the resulting OD values for the diluted samples are multiplied by ten.*

where c is some constant, was fitted. Using least squares estimates, we obtained the curve[5]

$$y = x + 0.83x^3. \tag{2.2}$$

Since x and y are assumed to be almost equal approximately up to 0.3, the coefficient of x was set to one. Having a second degree term in the polynomial would make the curve too steep in the end, so that when extrapolating for high values of x, the values of y would be too high. To avoid this it was chosen not to have the second degree term in the polynomial.

We measured the same plate in all Bioscreens and plotted the results corresponding to all pairs of Bioscreens against each other. Since the differences between the OD values from the different Bioscreens were rather small,

[5]The value of c was 0.8324057, but here it is rounded to 0.83 for simplicity. In all calculations $c = 0.8324057$ was used.

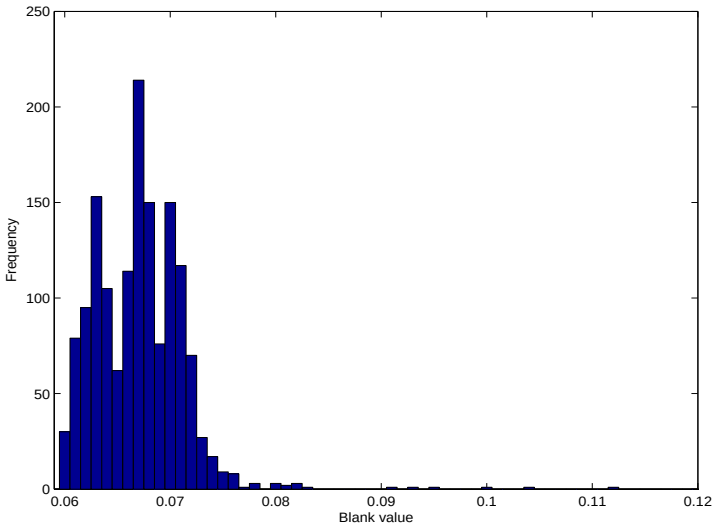

Figure 2.3: *Histogram of the blank values.*

and the lines were close to the 45^o degree line, we decided to use the same calibration curve function for all Bioscreens.

2.4.2 Blank correction

In the 576 mutants experiment a blank equal to 0.067 was used for all wells in all Bioscreens. This blank is the average blank of all wells in all five Bioscreens in two experiments where the OD values of wells containing only sterile water were measured. In these experiments there were altogether 1500 observations which varied between 0.060 and 0.112. The frequency distribution of the blank values in shown in Figure 2.3.

Variances within Bioscreens were rather small (the average of all the within Bioscreen variances was less than 0.00005). There were differences between Bioscreens, the lowest Bioscreen average being 0.063 and the highest being 0.072.

The same blank value was used in all Bioscreens and in all wells because in practice it is not possible to measure Bioscreen and well specific blanks for each run. Neither can the Bioscreen averages from the blank experiments be used as Bioscreen specific blanks, because the blank depends also on the

disposable plates. In the calibration data it is however important to use the well and Bioscreen specific blanks because the errors are multiplied by ten.

2.4.3 Discussion

It would have been possible to fit a calibration curve function assuming that there is measurement error in both x and y, but then the error structures should have been modeled more carefully. The calibration curve fitting could have also been done in two steps. First, to fit the function as we did. Second, to replace the small y values (for example values corresponding to $x < 0.35$) by the values from the calibration curve function and fit the curve again. This approach could be motivated by the observation that the measurement precision of x is much higher than the measurement precision of y, and the small x values are rather accurate.

We do not really know how well the calibration curve function works for high OD values. In the data that was used to fit it, the highest undiluted OD value is 1.22, but in our data there are OD values up to 1.7.

We are aware that the calibration curve function and the use of the same blank value in all Bioscreens and in all wells are debatable. The effects of a false blank value will be discussed in Chapter 7.

Chapter 3

The Warringer method

Warringer *et al*'s [15] method for calculating lag time, growth rate, and stationary phase OD increment, is referred to as *the Warringer method*.

3.1 Calculation of growth parameters

After blank subtraction and calibration of the raw data the growth curves are smoothened to reduce contributions from noise by averaging over three consecutive measurements (no averaging of initial value). In addition, artifacts arising from non-biological catastrophic events, such as impurities in the well, are avoided by setting measurements that are smaller than previous measurements to missing values, *i.e.* if $z_{t_p} > z_{t_{p+1}}$ then $z_{t_{p+1}}$ is considered missing [1]. The resulting (discrete) growth curve is named *smoothened growth curve*, and its value at time t_p is denoted by x_{t_p}.

The lag time: Smoothened growth curves are log transformed

$$y_{t_p} = \log(x_{t_p}).$$

[1] Missing values are ignored in the calculations. For example, if the average of the first six values is to be calculated and there are missing values, the average of the non-missing values of the six first values is used.

Let n denote the number of time points. Define $k_p(t)$ as the line going through points (t_p, y_{t_p}) and $(t_{p+7}, y_{t_{p+7}})$, which are $\frac{7}{3}$ hours apart

$$k_p(t) = \frac{y_{t_{p+7}} - y_{t_p}}{\frac{7}{3}} \cdot (t - t_p) + y_{t_p}, \tag{3.1}$$

$p = 1, ..., n-7$, and define h as a horizontal line corresponding to the average of the first six values of the logarithm transformed smoothened growth curve,

$$h = \frac{\sum_{t_p=1}^{6} y_{t_p}}{6}.$$

The intercept λ_p of each $k_p(t)$ with h is obtained by solving

$$\begin{aligned} h &= \frac{y_{t_{p+7}} - y_{t_p}}{\frac{7}{3}} \cdot (\lambda_p - t_p) + y_{t_p} \\ \Rightarrow \lambda_p &= \frac{\left(h - y_{t_p} + \frac{y_{t_{p+7}} - y_{t_p}}{\frac{7}{3}} \cdot t_p\right)}{\left(\frac{y_{t_{p+7}} - y_{t_p}}{\frac{7}{3}}\right)}. \end{aligned}$$

The average of the two highest values of λ_p, $p = 1, ..., n-7$, is taken as the lag time

$$\lambda = \frac{\lambda_1^* + \lambda_2^*}{2},$$

where λ_1^* and λ_2^* are the highest and the second highest values of λ_p. The lag time calculation is illustrated in Figure 3.1.

The growth rate: Smoothened growth curves are again log transformed

$$y_{t_p} = \log(x_{t_p}).$$

Slopes are calculated between every pair of values spaced 2/3 hours apart along the curve (no slopes are calculated from the first eight time points),

$$\mu_p = \frac{y_{t_{p+2}} - y_{t_p}}{2/3},$$

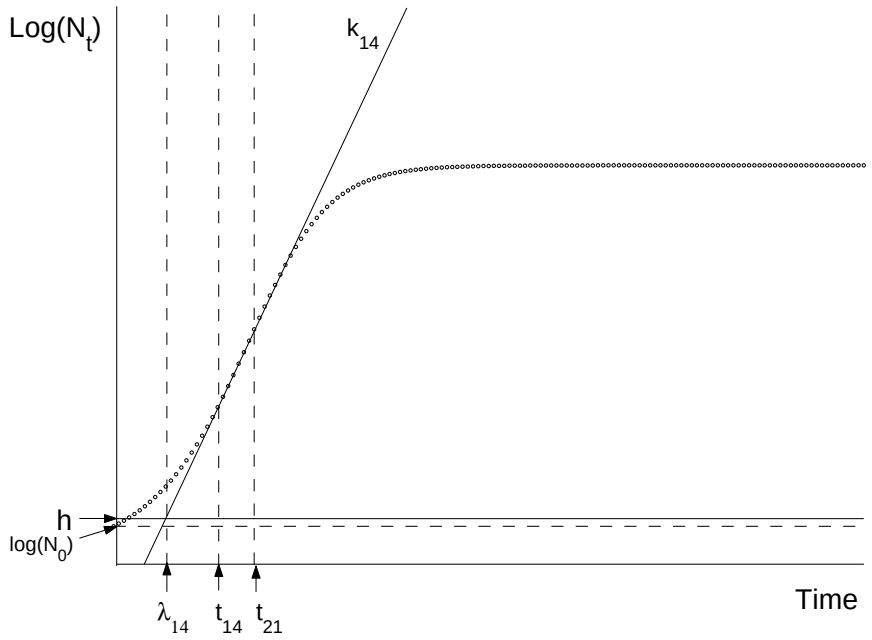

Figure 3.1: *A growth curve. The calculation of the lag time is illustrated with an example of* λ_{14}*, the time axis value at the intercept of the line going through time points 14 (*$6\frac{2}{3}$ *hours) and 21 (*9 *hours) and* h*, the line corresponding to the average of the first six values of the logarithm transformed smoothened growth curve. The size of the population at time* t*,* $\log(N_t)$*, is measured by the blank subtracted, calibrated, smoothened, logarithm transformed OD value.*

where $p = 9, ..., n-2$, and n is the number of time points. Of the seven highest slopes, the top two are discarded and an average is calculated from the following five

$$\mu = \frac{\sum_{i=3}^{7} \mu_i^*}{5},$$

where μ_i^* is the ith highest value of μ_p.

The stationary phase OD increment: The difference between the average of the last six values and the average of the first six values of the smoothened (not logarithm transformed) growth curve is taken as the sta-

tionary phase OD increment,

$$Y = \frac{\sum_{p=n-5}^{n} x_{t_p}}{6} - \frac{\sum_{p=1}^{6} x_{t_p}}{6}.$$

Smoothened growth curves for which no stationary phase can be observed (defined as smoothened growth curves with a coefficient of variation for the last six values higher than 2%), are discarded.

3.2 Discussion

The Warringer method presented here is the latest version of the methods that Warringer *et al* [15] have used for calculating growth parameters during the last years. When developing the original method [14], Warringer and Blomberg visually inspected 600 growth curves. However, after becoming more experienced in how the method works in practice for a larger dataset, Warringer *et al* [15] modified it.

The method can be questioned. For example, why disregard the first eight time points in the growth rate calculation? The values in the beginning often exhibit small deviations from a smooth curve, and also some growth curves have a steep ascent in the beginning. Each of these factors could result in a too high growth rate estimate, especially in case of a slow growth. In addition, there should be no reason to use the first eight time points in the growth rate estimation since that is when the cells are in the lag phase or in the transition from the lag phase to the exponential phase.

Why calculate the growth rate as the average of the $3^{rd} - 7^{th}$ highest slopes? The two highest slopes are disregarded to provide a safety margin, because when slopes are calculated locally there are often too high slopes due to the values which, even after smoothening, exhibit small deviations from a smooth curve. Taking the average of the following five slopes as the growth rate makes the estimate of the growth rate more robust and hopefully less influenced by the oscillating observations.

However, disregarding the first eight time points and taking the average of the $3^{rd} - 7^{th}$ highest slopes as the growth rate does not guarantee that when there are oscillating observations, they do not make the growth rate higher than it should be. Although simple to use and often successful in practice, the Warringer method seems to overestimate the growth rate when the growth

is slow and when the values oscillate along the curve. On the other hand, the method may underestimate the growth rate when the growth curves are smooth. Most of these problems could be successfully addressed by a more statistically minded approach. In the sequel, various parametric models and their estimation procedures are discussed.

Chapter 4

Growth models

An adequate growth model is needed to describe growth curves and to concentrate the information in measured data into a number of meaningful parameters. A parametric model is also needed to fit a curve using an EM algorithm approach and for simulating growth curves.

In this chapter we compare as models for yeast growth the following commonly used functions: the logistic [18], Gompertz [18], Richards [10], and Chapman-Richards [9] functions. All of them model the relative population size $\log(N_t/N_0)$, where N_0 is the initial size of the population and N_t is the size of the population at time t. This can be a problem because the curves cannot pass through 0 at $t = 0$. Therefore we modify them in order to model $\log(N_t)$ instead.

4.1 Comparisons of growth models

Most of the functions describing a sigmoidal growth curve contain parameters that do not have a clear biological interpretation and it can be difficult to estimate initial values for the parameters. To address this problem Zwietering *et al* [18] re-parameterized the logistic, Gompertz, Richards, Schnute and Stannard growth curve functions. They showed that the modified functions of Richards, Schnute and Stannard are the same. In the following we denote by A the asymptote, the maximum value of the growth reached (on the logarithmic scale); by μ the maximum growth rate, tangent of the growth curve at the inflection point; and by λ the lag time, the time axis intercept

of the tangent at the inflection point on the growth curve.

For easy reference we give the growth curve functions with their re-parameterized forms here:

Logistic The logistic function is

$$v_t = \log\left(\frac{N_t}{N_0}\right) = \frac{a}{1+\exp(b-ct)}, \tag{4.1}$$

$$= \frac{A}{1+\exp\left\{\frac{4\mu}{A}(\lambda-t)+2\right\}}, \tag{4.2}$$

where $a, c, A, \mu, \lambda > 0$.

Gompertz The Gompertz function is

$$v_t = \log\left(\frac{N_t}{N_0}\right) = \frac{a}{\exp\{\exp(b-ct)\}}, \tag{4.3}$$

$$= \frac{A}{\exp\left\{\exp\left[\frac{\mu\cdot\exp(1)}{A}(\lambda-t)+1\right]\right\}}, \tag{4.4}$$

where where $a, c, A, \mu, \lambda > 0$.

Richards The Richards function is

$$v_t = \log\left(\frac{N_t}{N_0}\right)$$

$$= \frac{a}{(1+\nu\cdot\exp\{k(\tau-t)\})^{\frac{1}{\nu}}}, \tag{4.5}$$

$$= \frac{A}{\left\{1+\nu\cdot\exp\left[\frac{\mu}{A}\cdot(1+\nu)^{(1+\frac{1}{\nu})}\cdot(\lambda-t)+(1+\nu)\right]\right\}^{\frac{1}{\nu}}}, \tag{4.6}$$

where $a, k, A, \mu, \lambda > 0$ and $\nu \neq 0$.

Chapman-Richards The Chapman-Richards function [9] is

$$v_t = \log\left(\frac{N_t}{N_0}\right) = \beta_0 \left[1 - \beta_1 \cdot \exp(-\beta_2 \cdot t)\right]^{1/(1-\beta_3)}, \tag{4.7}$$

where

$$\beta_0, \beta_1, \beta_2 > 0 \text{ and } \beta_3 < 1$$

or

$$\beta_0, \beta_2 > 0,\ \beta_1 < 0 \text{ and } \beta_3 > 1.$$

Re-parameterizing the Chapman-Richards function so that it contains biological parameters as in Zwietering *et al* [18] (the re-parameterizing is done the same way as will be presented for a modified Chapman-Richards function in detail in Section 4.1.3), gives

$$v_t = \log\left(\frac{N_t}{N_0}\right) = A \cdot \left[1 - (1-\beta_3) \cdot \exp\left\{\frac{\beta_3^{\frac{\beta_3}{\beta_3 - 1}} \cdot \mu}{A} \cdot (\lambda - t) + \beta_3\right\}\right]^{\frac{1}{1-\beta_3}} \tag{4.8}$$

where

$$\begin{aligned} A &= \beta_0 \\ \mu &= \beta_0 \beta_2 \beta_3^{\frac{\beta_3}{1-\beta_3}} \\ \lambda &= \frac{\log\left(\frac{\beta_1}{1-\beta_3}\right) - \beta_3}{\beta_2}. \end{aligned}$$

Substitution of ν by $\beta_3 - 1$ in the re-parameterized Richards function (4.6) would result in the re-parameterized Chapman-Richards function (4.8). In fact, the Chapman-Richards model is also known as the Richards model.

When $\beta_3 = 2/3$, the function (4.7) results in the von Bertalanffy model [13]. Richards [10] showed that the function is also equivalent to the monomolecular, the logistic and the Gompertz models for specific values of β_3. When $\beta_3 = 0$, it reduces to the monomolecular growth model, and when

$\beta_3 = 2$, it becomes the logistic model. The limiting form of the function when $\beta_3 \to 1$ is the Gompertz.

The Chapman-Richards model is very flexible. It can model both exponential and sigmoidal growth. This high flexibility is, however, combined with disadvantages as well. The parameters $(\beta_1, \beta_2, \beta_3)$ affect the growth curve in a highly collinear manner which can produce problems during nonlinear regression. In practice, problems of convergence often occur.

4.1.1 Modified growth models

All models mentioned above have a problem at $t = 0$ because $v_t > 0$ for all t (although v_0 is close to 0). Therefore we modified them in the spirit of Garthright [6], *i.e.* instead of modeling $\log(N_t/N_0)$, we model $\log(N_t)$. That is, we introduce a new parameter D, and set

$$g_t = \log(N_t) = y_t + D. \tag{4.9}$$

For logistic

$$y_t = \frac{a}{1 + \exp(b - ct)}, \tag{4.10}$$

for Gompertz

$$y_t = \frac{a}{\exp\{\exp(b - ct)\}}, \tag{4.11}$$

and for Chapman-Richards

$$y_t = \beta_0 \left[1 - \beta_1 \cdot \exp(-\beta_2 \cdot t)\right]^{1/(1-\beta_3)}, \tag{4.12}$$

and D is approximately $\log(N_0) - y_0$. By adding the parameter D, fitting problems that would occur whenever y_0 is noticeably above zero, are avoided. In the sequel, when we write logistic, Gomperz or Chapman-Richards, we refer to their modified versions as presented in this section. Note that the interpretations of the parameters of the re-parameterized functions are no longer the same as in Section 4.1.

4.1.2 Fitting the data

We fitted a nonlinear regression model via least squares using the *nls*-function in Splus[1]. The fits of several hundred growth curves from different runs and different Bioscreen instruments were compared visually.

The Chapman-Richards model is sensitive to how the initial parameters are chosen. However, by using the least squares estimates of the logistic or the Gompertz model as initial parameters for the Chapman-Richards, no convergence or other computational problems were encountered.

Figure 4.1 shows typical curve fits of the three models compared. The Chapman-Richards nearly always gives the best fit as it is to be expected since it encompasses both the logistic and the Gompertz models. The Gompertz model overestimated the slope, and moreover, it did not give a sufficient fit at any other part of the curve but the stationary phase. The logistic model gave a better fit than the Gompertz, but the Chapman-Richards function was clearly the best to describe our growth data. However, even when using the Chapman-Richards model there might be minor problems with the fit in the beginning of the curve and in the transition from the exponential phase to the stationary phase. Also, the stationary phase OD value might be slightly underestimated.

We are interested in fitting normal growth curves rather than problematic growth curves, such as rapidly collapsing curves[2]. The discussion above concentrated on normal growth curves, and here we continue by saying a few words about fitting abnormal growth curves.

Not only did the Chapman-Richards model give the best fit, but it also had less convergence problems than the other two models. If the shape of a growth curve differed much from sigmoidal, the logistic model and in particular the Gompertz model fits did not converge.

Figure 4.2 depicts some curve fits where the Chapman-Richards model fit converged but the logistic and the Gompertz model fits did not. Examples of abnormal curves where none of the model fits converged are shown in Figure 4.3. The Chapman-Richards model cannot be considered sufficient to

[1]MathSoft, Inc. The Gauss-Newton algorithm and the default convergence criteria were used. In case of convergence problems, changing the convergence criteria did not help.

[2]The curves collapse most often due to the existence of air bubbles that disturb the OD measurement.

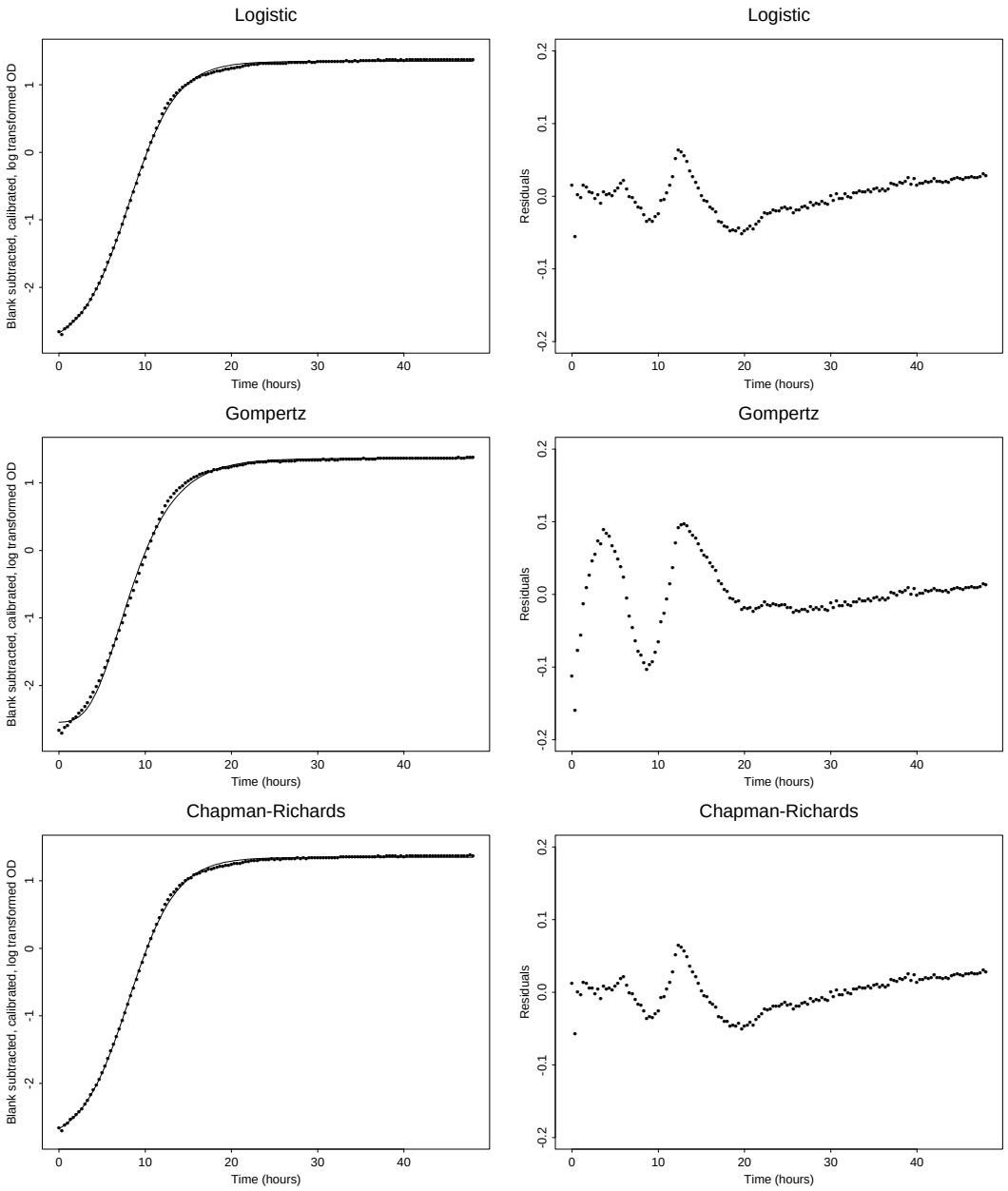

Figure 4.1: *Fits of the growth curve of the well 2 in the example data with the logistic, Gompertz and Chapman-Richards models. The corresponding residual plots are on the right.*

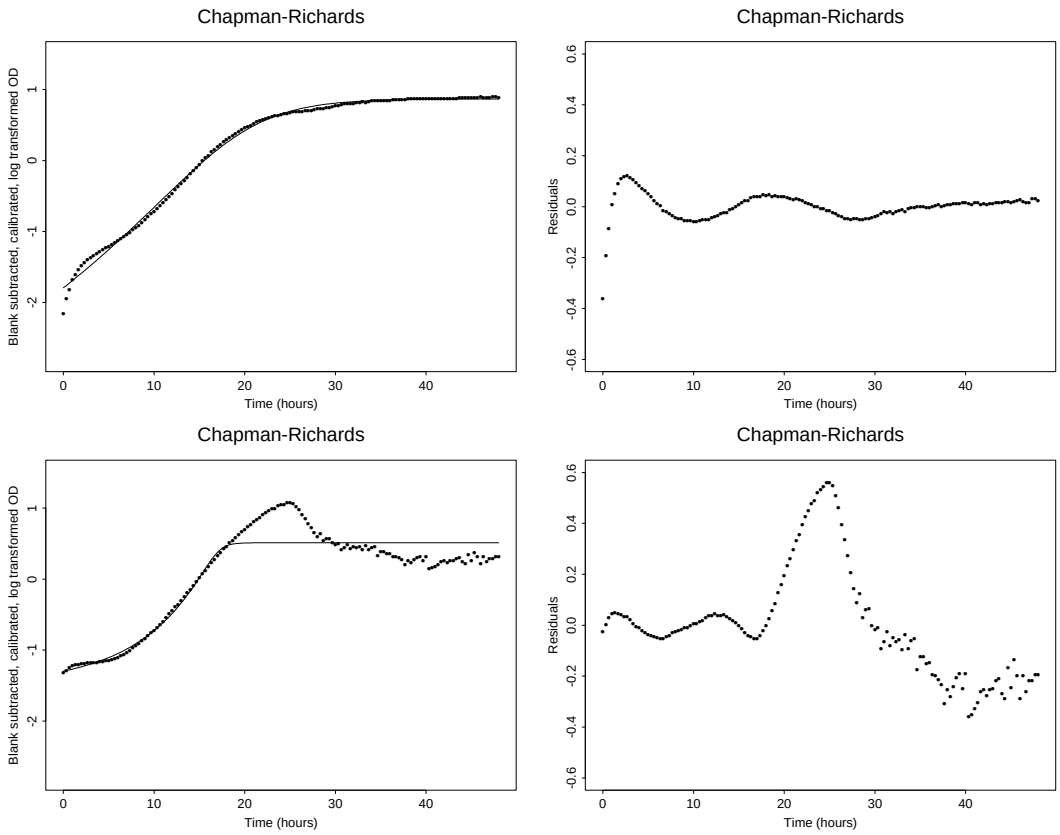

Figure 4.2: *Fits of abnormal growth curves with the Chapman-Richards model. The corresponding residual plots are on the right.*

describe abnormal growth curves. However, given the diversity of forms abnormal curves assume, it is very difficult to find a model that fits sufficiently well with all types of growth curves.

4.1.3 Growth parameters

Next, the growth parameters of the Chapman-Richards model

$$g_t = \log(N_t) = \beta_0 \left[1 - \beta_1 \cdot \exp(-\beta_2 \cdot t)\right]^{1/(1-\beta_3)} + D \tag{4.13}$$

are derived. Because of modeling $\log(N_t)$ instead of $\log(N_t/N_0)$ and adding the parameter D, the growth parameters A, μ, and λ that Zwietering *et al* use

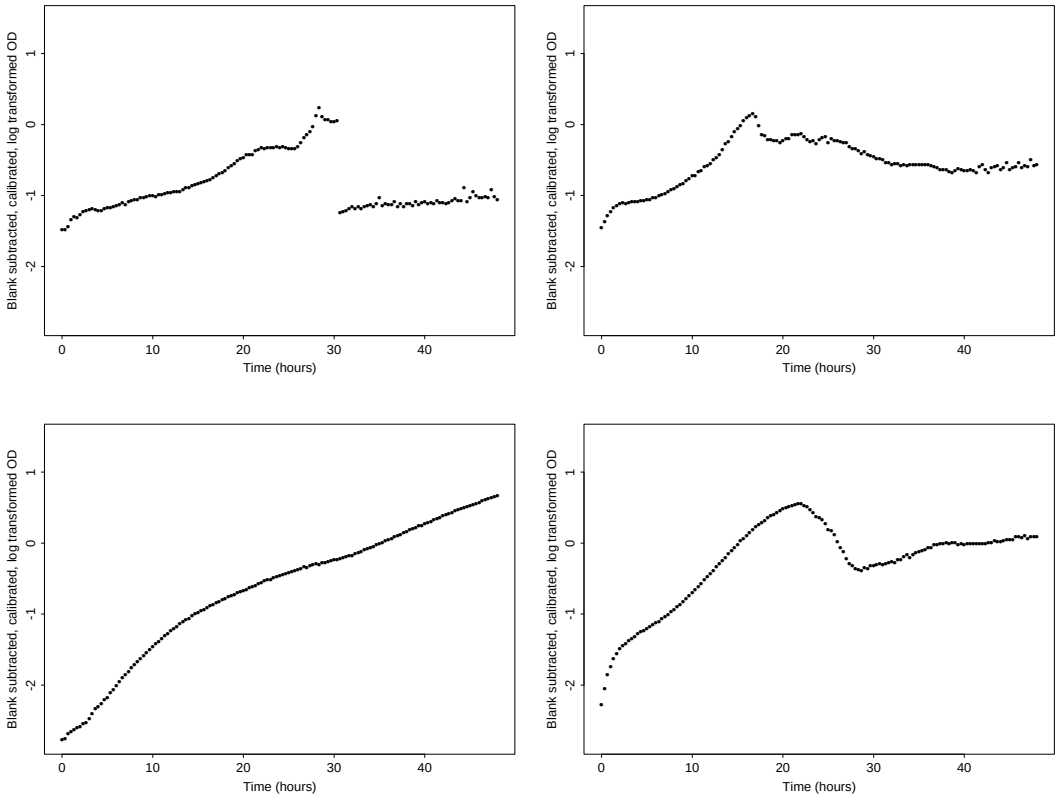

Figure 4.3: *Abnormal growth curves.*

are no longer the same as the parameters we want to estimate. In addition, the stationary phase OD increment we estimate differs from the parameter A of Zwietering *et al* in that it is the increment on the non-logarithmic scale. The growth parameter derivation is illustrated in Figure 4.4.

The stationary phase OD increment: The stationary phase OD increment, the final OD minus the initial OD, is

$$\begin{aligned} Y &= \exp\{\beta_0 + D\} - \exp\{g_0\} \\ &= \exp\{\beta_0 + D\} - \exp\left\{\beta_0(1-\beta_1)^{\frac{1}{1-\beta_3}} + D\right\}. \end{aligned} \tag{4.14}$$

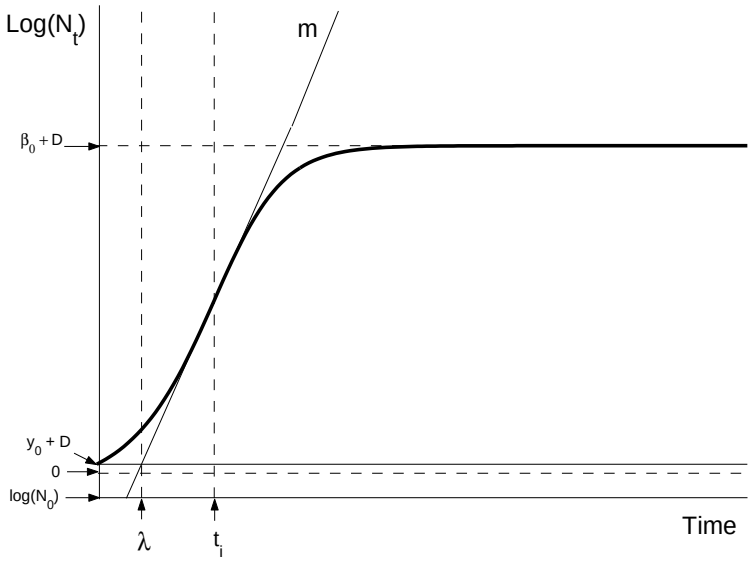

Figure 4.4: *A growth curve. An illustration of the growth parameter calculation in the Chapman-Richards model.*

(We have idealized slightly so that we think of the final OD not when the experiment ends but in infinite time.)

The growth rate: The (maximum) growth rate, μ, is defined as the tangent of the growth curve at the inflection point. The inflection point t_i is obtained by calculating the second derivative of the function (4.13) with respect to t, setting it to zero and solving it for t:

$$\frac{dg_t}{dt} = \frac{\beta_0\beta_1\beta_2 \exp\{-\beta_2 t\}\left(1-\beta_1 \exp\{-\beta_2 t\}\right)^{\frac{1}{1-\beta_3}-1}}{1-\beta_3} \tag{4.15}$$

$$\begin{aligned}\frac{d^2g_t}{dt^2} &= \frac{\beta_0\beta_1^2\beta_2^2\left(\frac{1}{1-\beta_3}-1\right)\exp\{-2\beta_2 t\}(1-\beta_1 \exp\{-\beta_2 t\})^{\frac{1}{1-\beta_3}-2}}{1-\beta_3} \\ &- \frac{\beta_0\beta_1\beta_2^2 \exp\{-\beta_2 t\}\left(1-\beta_1 \exp\{-\beta_2 t\}\right)^{\frac{1}{1-\beta_3}-1}}{1-\beta_3}.\end{aligned} \tag{4.16}$$

Equating this to zero gives the solution

$$t_i = \frac{\log(\frac{\beta_1}{1-\beta_3})}{\beta_2}. \tag{4.17}$$

The growth rate is finally derived by calculating the first derivative at the inflection point,

$$\mu = \left(\frac{dg_t}{dt}\right)_{t_i} = \beta_0\beta_2\beta_3^{\frac{\beta_3}{1-\beta_3}}. \tag{4.18}$$

Since we work on the logarithmic scale this corresponds to the relative growth rate on the absolute scale.

The lag time: The tangent line through the inflection point is

$$m = \mu t + \beta_0\beta_3^{\frac{1}{1-\beta_3}} - \mu t_i + D. \tag{4.19}$$

The lag time, λ, is the time axis value at the intercept of the tangent with $y_0 + D$:

$$y_0 + D = \mu\lambda + \beta_0\beta_3^{\frac{1}{1-\beta_3}} - \mu t_i + D \tag{4.20}$$

$$\begin{aligned} \Rightarrow \lambda &= \frac{y_0 - \beta_0\beta_3^{\frac{1}{1-\beta_3}} + \mu t_i}{\mu} \\ &= \frac{\beta_0(1-\beta_1)^{\frac{1}{1-\beta_3}} - \beta_0\beta_3^{\frac{1}{1-\beta_3}} + \mu\frac{\log(\frac{\beta_1}{1-\beta_3})}{\beta_2}}{\mu}. \end{aligned} \tag{4.21}$$

We were not able to rewrite the function (4.13) so that it would only contain the growth parameters and D and β_3. However, it is not really necessary to re-parameterize it because the initial parameters can be estimated using the estimates from the least squares fit of the model for $\log(N_t/N_0)$, function (4.8).

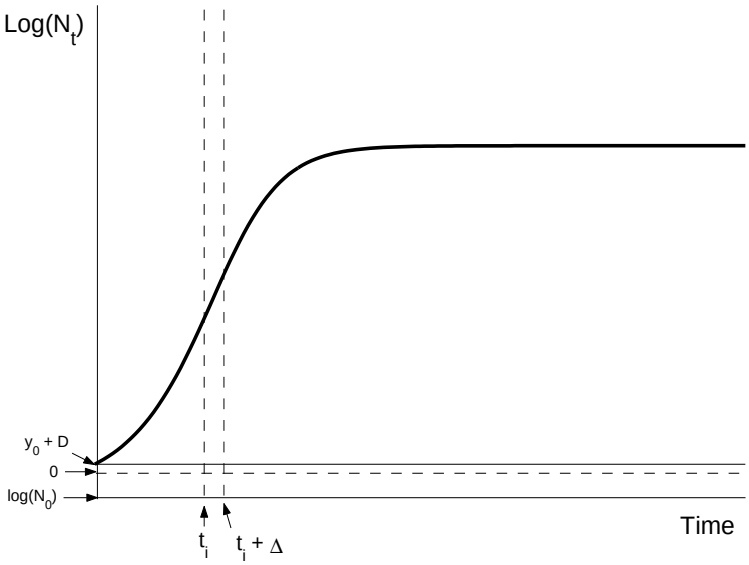

Figure 4.5: *A growth curve. An illustration of the three part model.*

4.2 A three part model

The curve fit in the beginning of the curve and in the transition from the exponential phase to the stationary phase is often not good. Even with the Chapman-Richards model the fit is sometimes rather poor in these parts of the curve. In addition, since the models are sigmoidal, the linear part of the curve may be poorly estimated. This is the case with the Gompertz model.

To try to overcome these problems we tried to fit a model consisting of three parts: the beginning of the curve until the inflection point, the linear part following the inflection point, and the rest after the linear part. One of the functions, the logistic, the Gompertz, or the Chapman-Richards, is used but with the exception that the linear part in the middle is modeled as a straight line. That is, we have

$$g_t^* = \begin{cases} g_t, & t \le t_i, \\ g_{t_i} + \mu(t - t_i), & t_i \le t \le t_i + \Delta, \\ g_{t-\Delta} + \mu \cdot \Delta, & t \ge t_i + \Delta, \end{cases} \tag{4.22}$$

where Δ is the time span of the linear part ($\Delta \geq 0$) and g_t is the logistic, the Gompertz, or the Chapman-Richards function as given in (4.9). The three part model is illustrated in Figure 4.5.

4.2.1 Fitting the data

We fitted a nonlinear regression model via least squares like in Section 4.1.2, and compared the fits of several hundred growth curves visually. Abnormal growth curves were excluded from the comparisons.

When using the logistic or the Chapman-Richards function as g_t, the time span of the linear part always approached zero, but the least squares fit never converged. With the Gompertz function as g_t there were no convergence problems. The estimated time span of the linear part was with nearly ideally growing cells approximately 3-4 hours. The curve fit was somewhat better than with the Gompertz model where no linear part is added in the middle, at least in the exponential phase, but still not perfect in particular in the beginning of the curves (Figure 4.6).

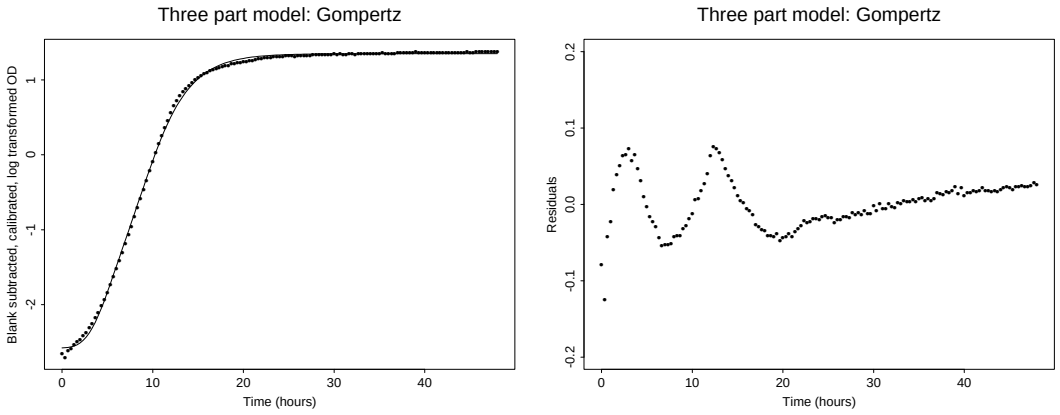

Figure 4.6: *The growth curve of the well 2 in the example data fitted with the three part model (using the Gompertz function). The corresponding residual plot is on the right.*

4.3 Simultaneous models for two growth curves

Often in case of double samples, the start OD values of the two samples vary, but the end OD values are almost the same, and the growth curves have the same shape except for the length of the exponential phase. This is a natural phenomenon, because in the sample with less cells in the beginning, there are more nutrients per cell, and thus the population can grow for a longer time before it runs out of nutrients. However, even the absolute amount of nutrients can vary between samples, and then the samples can have different start and final OD values, but the shapes of the growth curves (apart from the length of the exponential phase) tend to be nearly the same. In such cases modeling the growth curves simultaneously would probably give a better estimate of the growth than for example taking averages of growth parameters of two separately modeled curves.

4.3.1 Model I

We tried to model two growth curves simultaneously using the three part model so that all parameters except the time span of the linear part (Δ) and D, are the same for both of the curves.

The model of the curve with a smaller increment on the logarithmic scale (*i.e.* the difference between the logarithm of the start OD and the logarithm of the end OD) is

$$g_t^{*(1)} = \begin{cases} g_t^{(1)}, & t \leq t_i, \\ g_{t_i}^{(1)} + \mu(t - t_i), & t_i \leq t \leq t_i + \Delta_1, \\ g_{t-\Delta_1}^{(1)} + \mu \cdot \Delta_1, & t \geq t_i + \Delta_1, \end{cases} \tag{4.23}$$

where $g_t^{(1)}$ is the Chapman-Richards function

$$g_t^{(1)} = \beta_0 \left[1 - \beta_1 \cdot \exp(-\beta_2 \cdot t)\right]^{1/(1-\beta_3)} + D_1, \tag{4.24}$$

and the model of the curve with a larger increment on the logarithmic scale is

$$g_t^{*(2)} = \begin{cases} g_t^{(2)}, & t \le t_i, \\ g_{t_i}^{(2)} + \mu(t - t_i), & t_i \le t \le t_i + \Delta_2, \\ g_{t-\Delta_2}^{(2)} + \mu \cdot \Delta_2, & t \ge t_i + \Delta_2, \end{cases} \tag{4.25}$$

where

$$g_t^{(2)} = \beta_0 \left[1 - \beta_1 \cdot \exp(-\beta_2 \cdot t)\right]^{1/(1-\beta_3)} + D_2. \tag{4.26}$$

4.3.2 Model II

We also tried to fit a model where the asymptotes of the curves (4.23) and (4.25) were forced to be the same. Now

$$\beta_0 + \Delta_1\beta_0\beta_2\beta_3^{\frac{\beta_3}{1-\beta_3}} + D_1 = \beta_0 + \Delta_2\beta_0\beta_2\beta_3^{\frac{\beta_3}{1-\beta_3}} + D_2 \tag{4.27}$$

$$\Rightarrow D_2 = D_1 + (\Delta_1 - \Delta_2)\beta_0\beta_2\beta_3^{\frac{\beta_3}{1-\beta_3}} \tag{4.28}$$

$$= D_1 + (\Delta_1 - \Delta_2)\mu. \tag{4.29}$$

Thus, the Chapman-Richards function (4.26) in the model of the curve with a lower initial OD value can be written as

$$g_t^{(2)} = \beta_0 \left[1 - \beta_1 \cdot \exp(-\beta_2 \cdot t)\right]^{1/(1-\beta_3)} + D_1 + (\Delta_1 - \Delta_2)\mu. \tag{4.30}$$

In our data, the differences in the stationary phase OD increment of double samples of normally growing cells are small, in general less than 1%. This gives us reason to believe that a simultaneous model, where the asymptotes are forced to be the same, could be a good compromise model for two growth curves. It would be more natural to force the stationary phase OD increments to be the same, but forcing the asymptotes to be the same is almost equal to it and easier to implement.

4.3.3 Fitting the data

The models were fitted to the data using the least squares method. With both of the models the estimates of Δ_1 always approached zero but the least squares fits never converged. When the Δ_1 was fixed to zero, there were no convergence problems and the curve fits were good, see for example Figure 4.7.

The fits of the two models were often very similar, since the asymptotes are nearly the same in most cases. In the few cases were the asymptotes really differed, the fit was naturally better with model I, see for example Figure 4.8 (for this pair of growth curves the least squares fit of model II did not converge).

We also tried these models with the logistic and the Gompertz functions instead of the Chapman-Richards function. With the logistic function the least squares fit only converged if Δ_1 was fixed to zero, but with the Gompertz function there were no convergence problems, however, the fit was not adequate.

4.4 Discussion

With rather normal growth curves, the Chapman-Richards model always gave a reasonably good fit. However, the residual plots imply that there is a systematic error in the model, and that the Chapman-Richards model is not ideal for our data. On the other hand, since the small deviations of the data from the theoretical model are mostly in the transition from the exponential phase to the stationary phase, the growth parameter estimation should not suffer from the model not being exact.

One reason for why the three part model fit never converged when the logistic or the Chapman-Richards functions were used, could be that the logistic and the Chapman-Richards models already are flexible enough to model the linear part well, before adding the extra linear part. Since the Chapman-Richards model gives a good fit, adding a linear part in the middle may not be relevant when fitting one curve at a time, but in simultaneous modeling of two (or more) curves it could be essential. The problem of the three part model fit not converging can be overcome in simultaneous modeling by fixing Δ_1 to zero, and thus forcing Δ_2 to be modeled.

If we wish to pick up what is representative for a double sample, a simultaneous model may give better and more stable estimates than modeling the curves separately. We introduced two simultaneous models, the first one is more natural, the second one has less parameters. The results from fitting the simultaneous models to our data are promising, the fit is nearly as good as when modeling one curve at a time. These models can easily be generalized, and they could be a natural complement to the phenotypic library Warringer *et al* [16] try to build.

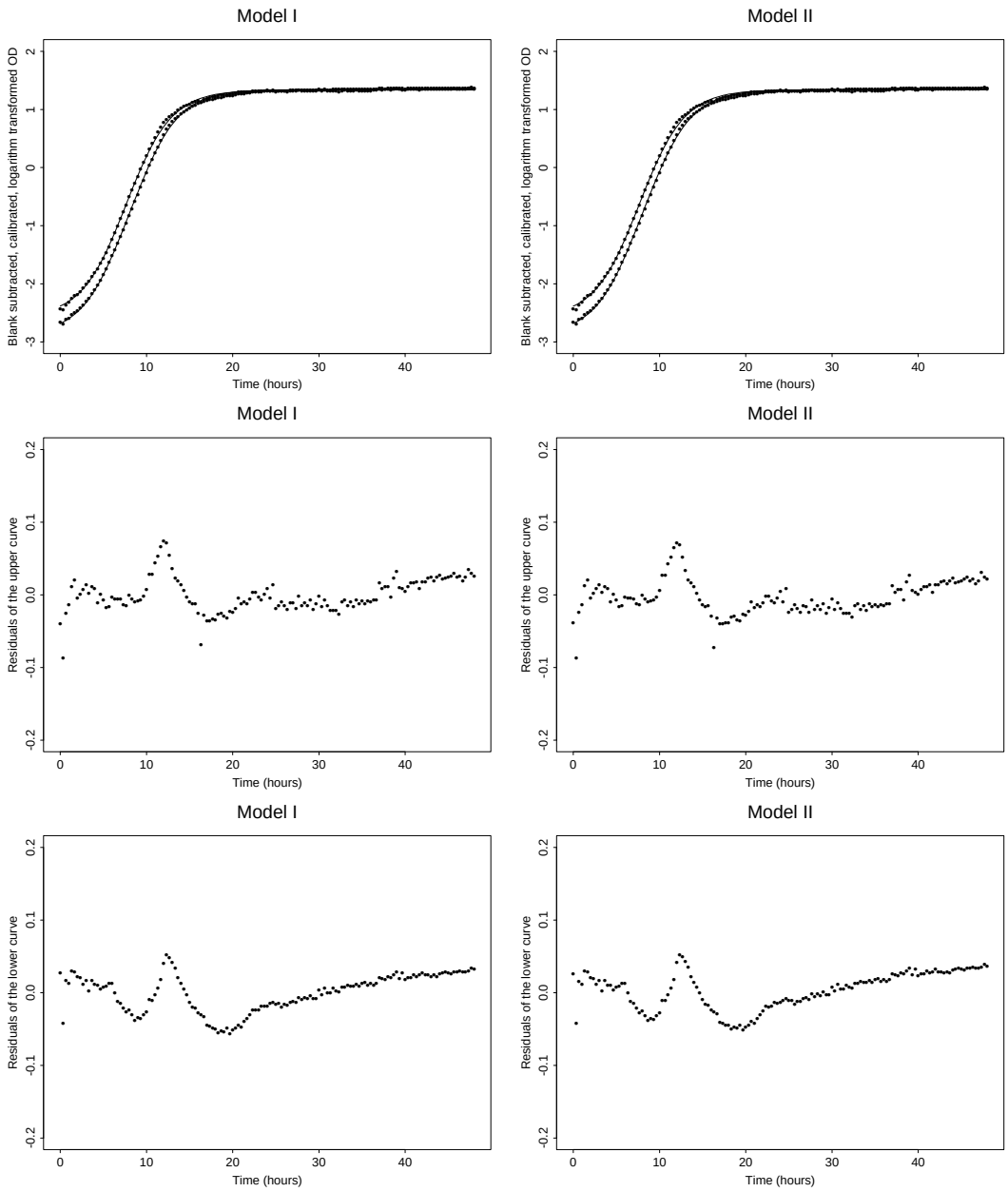

Figure 4.7: *Two curves (example data wells 4 and 6) are fitted using the simultaneous models (Δ_1 is fixed to zero). The residual plots of the fit of the upper curve are in the middle and the residual plots of the fit of lower curve are at the bottom.*

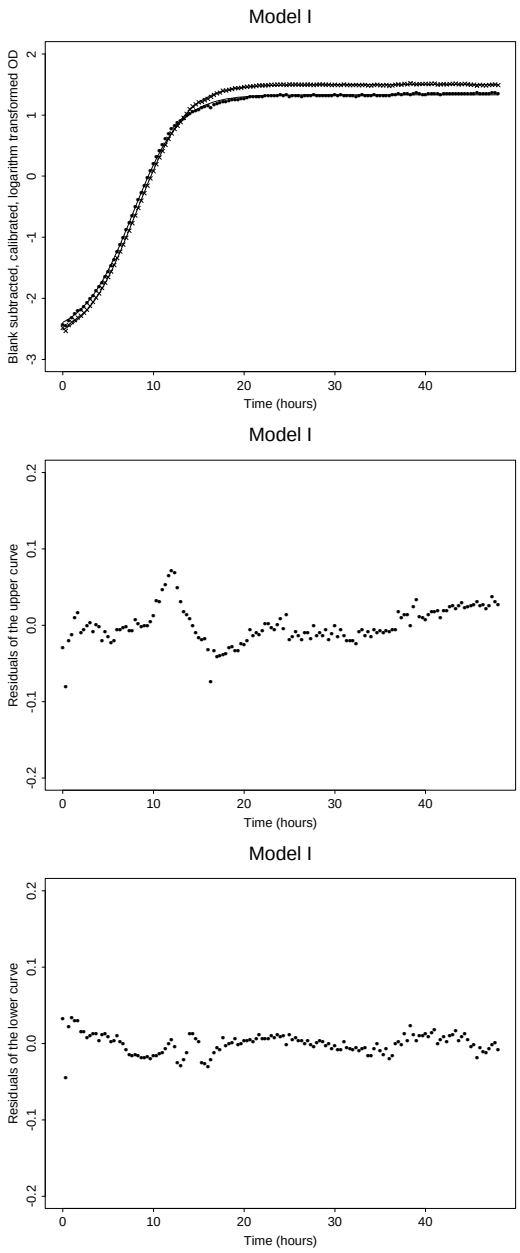

Figure 4.8: *Two curves are fitted using the simultaneous model I (Δ_1 is fixed to zero). The residual plot of the fit of the curve with a lower end OD is in the middle and the residual plot of the fit of the curve with a larger end OD is at the bottom.*

Chapter 5

Rounding error

The OD values from the Bioscreen are rounded to the three closest decimals. This implies that the observed value gives information about an interval within which the non-rounded value has to lie. To take the rounding into account when modeling the growth curve, an EM-algorithm is developed to estimate the parameters of the theoretical growth curve.

5.1 Model

The deviations of the measurements from the theoretical growth curve are assumed to be independent and normally distributed at each time point. Our model can be written as

$$f\left(x_{t_p};\theta,\sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left\{-\frac{\left[x_{t_p} - g_{t_p}(\theta)\right]^2}{2\sigma^2}\right\}, \tag{5.1}$$

where

- x_{t_p} is the theoretical, true value of the growth curve at time t_p

  $x_{t_p} \sim N\left(g_{t_p}(\theta),\sigma\right)$, where $g_{t_p}(\theta)$ is given by the modified version of the Chapman-Richards growth model (4.13)

- $\theta = (\beta_0, \beta_1, \beta_2, \beta_3, D)$, the parameters in the growth model (4.13).

The log-likelihood function is

$$\log L\left[\mathbf{x};\theta,\sigma^2\right] = \frac{n}{2}\log\left(\frac{1}{2\pi\sigma^2}\right) - \frac{1}{2\sigma^2}\sum_{p=1}^{n}[x_{t_p} - g_{t_p}(\theta)]^2. \tag{5.2}$$

5.2 EM algorithm

The term EM algorithm was introduced by Dempster, Laird and Rubin in 1977 [4]. It is a broadly applicable approach to the iterative computation of maximum likelihood estimates (MLE) in incomplete data problems, or problems where the observed data can be thought of as incomplete in some sense. Each iteration of the EM-algorithm involves two steps, *expectation step* (E-step) and *maximization step* (M-step). The E-step finds the conditional expectation of the missing data given the observed data and current estimated parameters (or initial estimated parameters). The E-step can be thought of as filling in the missing data. Once the filled complete data are constructed, then the parameters are estimated in the M-step by performing maximum likelihood estimating procedure, and so on, proceeding iteratively until convergence is reached [4] [7].

5.2.1 Method

The intervals within which the OD values from the Bioscreen lie, are in the EM-setting the incomplete, observed data $\mathbf{y}$, and the true (but unobservable) OD values are the complete data $\mathbf{x}$. Recall that for a given OD value z_{t_p} from the Bioscreen, the calibrated value is $z_{t_p} + 0.83z_{t_p}^3$. Taking the blank and the rounding error into account, we get that the observed data intervals $\mathbf{y} = (\mathbf{a}, \mathbf{b})$ are

$$\begin{aligned}\mathbf{a} &= \log[(\mathbf{z} - 0.0005 - 0.067) + 0.83 \cdot (\mathbf{z} - 0.0005 - 0.067)^3] \\ \mathbf{b} &= \log[(\mathbf{z} + 0.0005 - 0.067) + 0.83 \cdot (\mathbf{z} + 0.0005 - 0.067)^3],\end{aligned}$$

where $\mathbf{z}$ is a vector of OD values from the Bioscreen over time.

E-step In the E-step we compute the conditional expected value

$$Q(\theta,\sigma^2|\theta^{(k)},\sigma^{2(k)}) = E\{\log L(\theta,\sigma;\mathbf{x})|\mathbf{y},\theta^{(k)},\sigma^{2(k)}\}, \tag{5.3}$$

where $\theta^{(k)}$ and $\sigma^{2(k)}$ are the estimates of θ and σ^2 in the k:th iteration step, and $\mathbf{y} = (\mathbf{a}, \mathbf{b})$. The expectation in (5.3) can be written as

$$\begin{aligned} Q(\theta, \sigma^2 | \theta^{(k)}, \sigma^{2(k)}) &= \frac{n}{2} \log(\frac{1}{2\pi\sigma^2}) \\ &- \frac{1}{2\sigma^2} \sum_{p=1}^{n} \{E[x_{t_p}^2 | y_{t_p}, \theta^{(k)}, \sigma^{2(k)}] - (E[x_{t_p} | y_{t_p}, \theta^{(k)}, \sigma^{2(k)}])^2 \\ &+ (E[x_{t_p} | y_{t_p}, \theta^{(k)}, \sigma^{2(k)}] - g_{t_p}(\theta))^2\}, \end{aligned}$$

where

$$E[x_{t_p} | y_{t_p}, \theta^{(k)}, \sigma^{2(k)}] = \frac{\frac{\sigma^{(k)}}{\sqrt{2\pi}} \left\{ e^{-\frac{1}{2}(a_{t_p}^*)^2} - e^{-\frac{1}{2}(b_{t_p}^*)^2} \right\}}{\Phi(b_{t_p}^*) - \Phi(a_{t_p}^*)} + g_{t_p}(\theta^{(k)})$$

and

$$\begin{aligned} E[x_{t_p}^2 | y_{t_p}, \theta^{(k)}, \sigma^{2(k)}] &= \sigma^{2(k)} \frac{\left\{ \frac{1}{\sqrt{2\pi}} \left((a_{t_p}^*) e^{-\frac{1}{2}(a_{t_p}^*)^2} - (b_{t_p}^*) e^{-\frac{1}{2}(b_{t_p}^*)^2} \right) + \Phi(b_{t_p}^*) - \Phi(a_{t_p}^*) \right\}}{\Phi(b_{t_p}^*) - \Phi(a_{t_p}^*)} \\ &+ 2 g_{t_p}(\theta^{(k)}) \sigma^{(k)} \frac{\frac{1}{\sqrt{2\pi}} \left\{ e^{-\frac{1}{2}(a_{t_p}^*)^2} - e^{-\frac{1}{2}(b_{t_p}^*)^2} \right\}}{\Phi(b_{t_p}^*) - \Phi(a_{t_p}^*)} + g_{t_p}^2(\theta^{(k)}), \end{aligned}$$

where $a_{t_p}^* = \frac{a_{t_p} - g_{t_p}(\theta^{(k)})}{\sigma^{(k)}}$ and $b_{t_p}^* = \frac{b_{t_p} - g_{t_p}(\theta^{(k)})}{\sigma^{(k)}}$.

M-step In the M-step the conditional expected value $Q(\theta, \sigma^2 | \theta^{(k)}, \sigma^{2(k)})$ is maximized to get updates $\theta^{(k+1)}$ and $\sigma^{2(k+1)}$. Thus in the M-step we calculate

$$\begin{aligned} \sigma^{2(k+1)} &= \frac{1}{n} \sum_{p=1}^{n} \{E[x_{t_p}^2 | y_{t_p}, \theta^{(k)}, \sigma^{2(k)}] - (E[x_{t_p} | y_{t_p}, \theta^{(k)}, \sigma^{2(k)}])^2 \\ &+ (E[x_{t_p} | y_{t_p}, \theta^{(k)}, \sigma^{2(k)}] - g_{t_p}(\theta^{(k)}))^2\} \end{aligned}$$

and $\theta^{(k+1)}$ are the least squares estimates, those values that minimize

$$\sum_{p=1}^{n} \left(E[x_{t_p} | y_{t_p}, \theta^{(k)}, \sigma^{2(k)}] - g_{t_p}(\theta) \right)^2.$$

5.3 Discussion

The residual plots of the curve fits (in Chapter 4) imply that the deviations from the theoretical growth curve are systematic rather than random. Thus, our assumption about the deviations of the measurements from the theoretical growth curve being independent and normally distributed at each time point is probably violated.

In the next chapter we will use the EM algorithm to see if the rounding has an effect on the growth parameter estimates from the least squares fit of the Chapman-Richards model.

Chapter 6

Evaluation and comparison of estimation methods

To investigate the rounding effect we compare the estimates of the growth parameters[1] from two parametric methods: the least squares fit of the Chapman-Richards model, and the least squares fit of the Chapman-Richards model with rounding correction (*i.e.* the EM algorithm approach presented in Chapter 5). Furthermore, we generate data to investigate the rounding effect when using the Warringer method. We also compare the estimates from the Warringer method with the estimates from the least squares fit of the Chapman-Richards model.

6.1 Simulations

First, measurements without rounding errors, x_{t_p}, were simulated from the model presented in Section 5.1 using different sets of values for parameters $\beta_0, \beta_1, \beta_2, \beta_3$, and D. They were then transformed to correspond to the OD values from the Bioscreen, z_{t_p}, *i.e.* the OD values with rounding error but before subtracting the blank, the calibration, and the logarithm transformation. Finally, a blank equal to 0.067 was subtracted from z_{t_p}, the calibration was done, and the growth parameters were estimated.

[1]The growth parameters Y, μ and λ are calculated using the estimates of the model parameters $\theta = (\beta_0, \beta_1, \beta_2, \beta_3, D)$.

The transformation from x_{t_p} to z_{t_p} is done in the following way:

$$\begin{aligned} z^*_{t_p} &= b - \frac{\left(\frac{2}{3}\right)^{1/3}}{\sqrt{c}\left(9e^{x_{t_p}}\sqrt{c} + \sqrt{3}\sqrt{4 + 27e^{2x_{t_p}}c}\right)^{1/3}} \\ &+ \frac{\left(9e^{x_{t_p}}\sqrt{c} + \sqrt{3}\sqrt{4 + 27e^{2x_{t_p}}c}\right)^{1/3}}{2^{1/3} \cdot 3^{2/3}\sqrt{c}} \end{aligned} \tag{6.1}$$

where b is the blank, 0.067, and c is the calibration curve coefficient, 0.83. Note that $z^*_{t_p}$ are the non-rounded values. The z_{t_p} are the rounded (to the three closest decimals) values of $z^*_{t_p}$. The equation of $z^*_{t_p}$ (6.1) is obtained by solving the equation of the operations done on the OD values from the Bioscreen (subtracting the blank, the calibration and the logarithm transformation, but here non-rounded values are used),

$$x_{t_p} = \log[(z^*_{t_p} - b) + c(z^*_{t_p} - b)^3], \tag{6.2}$$

for $z^*_{t_p}$.

Standard deviations $\sigma = 0.01$ and $\sigma = 0.02$ were used in the simulations (because the estimates of σ from the EM algorithm based on data of ideally growing cells were between 0.01 and 0.02). The sets of parameter values were chosen from Chapman-Richards model fits to our data as representatives of different types of standard growth curves. The corresponding growth curves (without measurement error) are displayed in Figure 6.1.

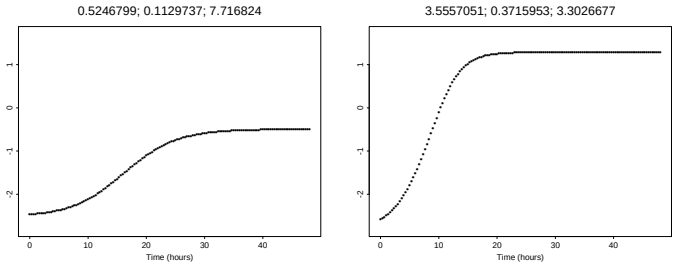

Figure 6.1: *Measurements generated from the Chapman-Richards model with Y, μ and λ as given on top of the graphs, respectively.*

The rounding correction had only a minimal effect on the growth parameter estimates (Tables 6.1 - 6.2). The estimates from the Warringer method differed from those from the least squares fit of the Chapman-Richards model.

6.2 Generated measurements

The effect of the rounding error when estimating the growth parameters with the Warringer method was investigated. Measurements without rounding errors were generated from the Chapman-Richards model (4.13) using different sets of values for parameters $\beta_0, \beta_1, \beta_2, \beta_3$, and D, and rounded to the three closest decimals. The sets of parameter values were chosen from Chapman-Richards model fits to our data as representatives of different types of standard growth curves. The corresponding growth curves are displayed in Figure 6.2. Growth parameters were calculated from the non-rounded and from the rounded measurements (Table 6.3). Because the true parameters are known and there is no measurement error involved, these estimates can also be used to investigate how the estimated growth parameters from the Warringer method systematically differ from the true parameters.

The effect of the rounding is small. The largest difference between the estimates from the non-rounded and rounded values is approximately 1.5% for growth rate, 1% for yield, and less than 0.5% for lag time, and that is when the growth is very slow. The lag time is almost always overestimated. The stationary phase OD increment seems to be rather well estimated, it is only slightly underestimated. The estimated growth rates are close to the true ones, only a little lower.

6.3 Example data

We also made a comparison of the three estimation methods based on the example data. The differences between the estimates from the least squares fit of the Chapman-Richards model with rounding correction (*i.e.* the EM based estimates) and without rounding correction were very small, while the estimates from the Warringer method differed from the estimates from the least squares fit of the Chapman-Richards model (Table 6.4). The estimates from the Warringer method were larger for stationary phase OD increment

and lag time, and smaller for growth rate, compared with the estimates from the least squares fit of the Chapman-Richards model.

6.4 Conclusions

The rounding error was investigated using different approaches for the Warringer method and for the least squares fit of the Chapman-Richards model [2]. The result was that the effect of the rounding error is small no matter which one of the estimation methods is used, even when the growth is slow and the stationary phase OD increment is small.

The estimates from the Warringer method differ from the estimates from the least squares fit of the Chapman-Richards model. In general, when cells grow fast, the Warringer method underestimates the growth rate, and when cells grow slowly, it overestimates the growth rate. This can be explained, at least partly, by the calculation of local slopes and using the average of the $3^{rd} - 7^{th}$ highest. Also, it seems that the Warringer method overestimates the lag time. The Chapman-Richards model slightly underestimates the stationary phase OD increment.

Sometimes the Warringer method did not return a value for the stationary phase OD increment because the coefficient of variation for the last six values was higher than 2%. We do not have any such criterion when we estimate stationary phase OD increment using the Chapman-Richards model, although the least squares fit is likely not to converge if no stationary phase is reached. If this method is to be used in real experiments, some criterion for the stationary phase OD increment, such as the one in the Warringer method, may be needed.

[2]The Warringer method cannot be used in the EM algorithm since it is not based on a parametric model.

Table 6.1: *The means, standard deviations and minimums and maximums of the growth parameters in 100 simulations. At times the Warringer method returned missing values for stationary phase OD increment due to the coefficient of variation of the last six OD values of the smoothened growth curve being greater than 2%. (EM = Chapman-Richards with rounding correction, C-R = Chapman-Richards, W = Warringer).*

σ		Data simulated from C-R with param.	Estimation method	Mean	Std	Min	Max
	Y	0.5246799	EM	0.5246785	0.001177880	0.5216767	0.5276762
0.01	μ	0.1129737	EM	0.1129744	0.000492475	0.1120249	0.1139237
	λ	7.7168240	EM	7.7208976	0.058447948	7.6030571	7.8317932
	Y	0.5246799	C-R	0.5246803	0.001178216	0.5216808	0.5276732
0.01	μ	0.1129737	C-R	0.1129764	0.000492351	0.1120267	0.1139238
	λ	7.7168240	C-R	7.7211330	0.058497122	7.6032077	7.8318705
	Y	0.5246799	W	0.5222199	0.002541194	0.5162084	0.5272015
0.01	μ	0.1129737	W	0.1171654	0.002205720	0.1123454	0.1229769
	λ	7.7168240	W	7.9839066	0.186447483	7.6795980	8.4064849
	Y	0.5246799	EM	0.5248723	0.002212928	0.5194164	0.5303139
0.02	μ	0.1129737	EM	0.1129791	0.000943188	0.1111062	0.1148820
	λ	7.716824	EM	7.7192973	0.118623742	7.4659903	8.0249051
	Y	0.5246799	C-R	0.5248732	0.002212608	0.5194179	0.5303114
0.02	μ	0.1129737	C-R	0.1129797	0.000943445	0.1111148	0.1148827
	λ	7.7168240	C-R	7.7193223	0.118704204	7.4648347	8.0251376
	Y	0.5246799	W	0.5236434	0.004446626	0.5121418	0.5346346
0.02	μ	0.1129737	W	0.1284940	0.005411177	0.1179393	0.1403039
	λ	7.7168240	W	8.3077034	0.295204830	7.6181917	9.2846129

Table 6.2: *The means, standard deviations and minimums and maximums of the growth parameters in 100 simulations. At times the Warringer method returned missing values for stationary phase OD increment due to the coefficient of variation of the last six values of the smoothened growth curve being greater than 2%. (EM = Chapman-Richards with rounding correction, C-R = Chapman-Richards, W = Warringer).*

σ		Data simulated from C-R with param.	Estimation method	Mean	Std	Min	Max
	Y	3.5557051	EM	3.5553129	0.003789803	3.5463612	3.5633842
0.01	μ	0.3715953	EM	0.3717154	0.000966882	0.3696886	0.3741025
	λ	3.3026677	EM	3.3001536	0.024666876	3.2423069	3.3678401
	Y	3.5557051	C-R	3.5553133	0.003790054	3.5463610	3.5633861
0.01	μ	0.3715953	C-R	0.3717157	0.000968463	0.3697145	0.3741589
	λ	3.3026677	C-R	3.3001555	0.024758640	3.2427645	3.3663842
	Y	3.5557051	W	3.5712918	0.014001769	3.5300364	3.6020469
0.01	μ	0.3715953	W	0.3621711	0.002973581	0.3549762	0.3694788
	λ	3.3026677	W	3.4226579	0.047174098	3.3250369	3.5317702
	Y	3.5557051	EM	3.5560257	0.009397656	3.5388849	3.5763836
0.02	μ	0.3715953	EM	0.3712789	0.002242866	0.3657885	0.3761153
	λ	3.3026677	EM	3.2919396	0.057181790	3.1644820	3.4128011
	Y	3.5557051	C-R	3.5560254	0.009396689	3.5388892	3.5763834
0.02	μ	0.3715953	C-R	0.3712769	0.002243350	0.3658019	0.3761285
	λ	3.3026677	C-R	3.2918770	0.057175684	3.1648278	3.4131014
	Y	3.5557051	W	3.5912798	0.027330418	3.5133286	3.6351853
0.02	μ	0.3715953	W	0.3660302	0.007186311	0.3532417	0.3810974
	λ	3.3026677	W	3.4647121	0.086332581	3.3495159	3.6955502

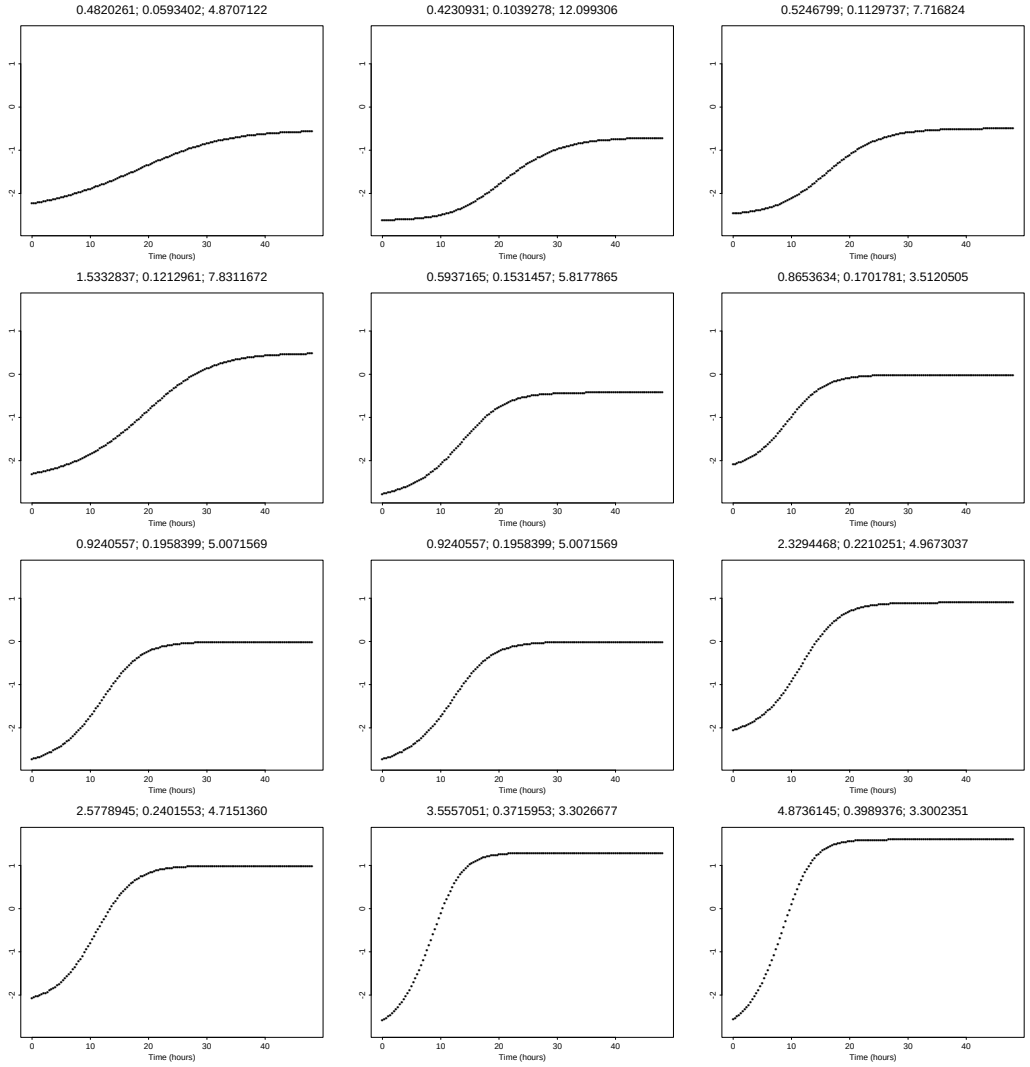

Figure 6.2: *Measurements generated from the Chapman-Richards model with Y, μ and λ as given on top of the graphs, respectively.*

Table 6.3: *Growth parameters calculated using the Warringer method both on true values and on rounded values. The data is simulated from Chapman-Richards model.*

	Data from C-R with param.	The Warringer method using true values	The Warringer method using rounded values
Y	0.4820261	0.4601377	0.4603697
μ	0.0593402	0.0591697	0.0600618
λ	4.8707122	5.1885236	5.2380668
Y	0.4230931	0.4171988	0.4174565
μ	0.1039278	0.1030358	0.1029922
λ	12.099306	12.0595279	12.0531383
Y	0.5246799	0.5224131	0.5227353
μ	0.1129737	0.1119046	0.1126244
λ	7.7168240	7.7415196	7.7667872
Y	1.5332837	1.5056579	1.5052135
μ	0.1212961	0.1209835	0.1211536
λ	7.8311672	7.9622812	7.9773934
Y	2.0845279	2.0735540	2.0740747
μ	0.1487075	0.1482091	0.1486831
λ	5.6931074	5.8590869	5.8794212
Y	0.5937165	0.5919584	0.5927003
μ	0.1531457	0.1512645	0.1513532
λ	5.8177865	5.9019797	5.9254058
Y	0.8653634	0.8607245	0.8610007
μ	0.1701781	0.1672007	0.1663950
λ	3.5120505	3.6362405	3.6231189
Y	0.9240557	0.9216105	0.9222128
μ	0.1958399	0.1929351	0.1927450
λ	5.0071569	5.0993609	5.1001583
Y	2.3294468	2.3244855	2.3253794
μ	0.2210251	0.2193446	0.2202977
λ	4.9673037	5.0383771	5.0587525
Y	2.5778945	2.5728839	2.5714388
μ	0.2401553	0.2382022	0.2383848
λ	4.7151360	4.7767173	4.7865178
Y	3.5557051	3.5490465	3.5495761
μ	0.3715953	0.3672698	0.3675677
λ	3.3026677	3.4096894	3.4107767
Y	4.8736145	4.8662401	4.8661553
μ	0.3989376	0.3943077	0.3942303
λ	3.3002351	3.4066480	3.4163534

Table 6.4: *Growth parameters calculated on the example data. (EM = Chapman-Richards with rounding correction, C-R = Chapman-Richards, W = Warringer).*

Well		EM	C-R	W
	Y	3.763375	3.763352	3.849816
1	μ	0.380356	0.380297	0.363401
	λ	3.085568	3.083988	3.165851
	Y	3.775514	3.775502	3.872240
2	μ	0.383319	0.383285	0.367711
	λ	3.142825	3.141887	3.167670
	Y	3.776889	3.776868	3.854175
3	μ	0.368232	0.368173	0.361570
	λ	3.138799	3.137151	3.252444
	Y	3.786103	3.786082	3.860301
4	μ	0.378505	0.378446	0.370203
	λ	2.999469	2.997863	3.092076
	Y	3.769556	3.769531	3.845761
5	μ	0.373667	0.373588	0.363375
	λ	2.885867	2.883659	2.982461
	Y	3.611292	3.611263	3.676320
6	μ	0.383948	0.383903	0.372793
	λ	3.373912	3.372873	3.390100
	Y	3.671790	3.671773	3.800783
7	μ	0.377573	0.377523	0.365013
	λ	3.066510	3.065183	3.146775
	Y	3.725725	3.725693	3.805750
8	μ	0.380361	0.380297	0.367649
	λ	3.333461	3.331654	3.415210

Chapter 7

Effects of a false blank value

A blank equal to 0.067 is always subtracted from the raw data. However, it may not be the true blank. For example in the blank experiments, where the average was 0.067, the values varied between 0.06 and 0.112. The following comparisons were done in order to see what effects a false blank can have and, when the blank is unknown, how the estimates of the growth parameters vary depending on which blank value is used.

7.1 Comparisons of different blanks

Blank comparisons were done using simulated measurements and the example data.

Simulations Measurements corresponding to the OD values from the Bioscreen, z_{t_p}, were simulated in the same way as described in Section 6.1. The difference here is that b is the true blank. True blank values 0.060, 0.067 and 0.112 were used. After this, a blank equal to 0.067 was subtracted from z_{t_p}, calibration was done and growth parameters were estimated using the EM algorithm, the least squares fit of the Chapman-Richards model and the Warringer method. For each true blank value, 100 simulations with a standard deviation 0.01 ($\sigma = 0.01$) were done.

Example data Growth parameters were estimated from the example data with different blanks. That is, we looked at how the estimates of the growth

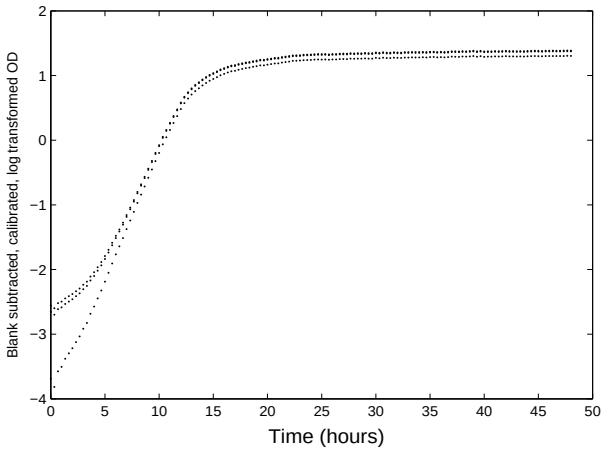

Figure 7.1: *Example data, well 2. Growth curve with three different blanks: 0.060 (the highest curve), 0.067 (the middle curve) and 0.112 (the lowest curve).*

parameters would change if instead of using the blank equal to 0.067, a blank equal to 0.060 or 0.112 was used. The true blank value for the example data is not known.

7.2 Results

In the simulations, the estimates of the growth parameters varied noticeably depending on how much the true blank differed from 0.067. For example when the measurements were simulated from the model with $Y = 0.5246799$, $\mu = 0.1129737$, $\lambda = 7.7168240$ and $\sigma = 0.01$, the differences between the averages of the 100 simulations with the true blank 0.060 and the true blank 0.112, were for the stationary phase OD increment more than 6%, for the growth rate more than 20%, and for the lag time more than 12%, with all the estimation methods (Table 7.2). The corresponding percentages for measurements simulated form the model with $Y = 3.5557051$, $\mu = 0.3715953$, $\lambda = 3.3026677$ and $\sigma = 0.01$, were for the growth rate and the stationary phase OD increment almost 7%, and for the lag time approximately 20% (Table 7.1).

Also, in case of the example data, the different blanks give rather different

results (for example, see Table 7.3). Already by looking at growth curves with different blanks it is clear that the blank effect is not negligible (for example, see Figure 7.1). It does not seem that any of the estimation methods is more robust than any other against the blank effect.

It should be emphasized here that as large blank values as 0.112 are rare. In our blank experiments of 1500 observations only three were greater than 0.10, and 13 were greater than 0.08. Some of the larger measurements can be suspected to be large due to measurement errors, and the blank variation is likely to be smaller in reality than in our blank experiments. However, the blank variation is probably still alarmingly large and it would be important to understand the sources of it. We choose to postpone these investigations to the future due to time pressure.

Table 7.1: *The means, standard deviations, minimums and maximums of estimated growth parameters in 100 simulations. The measurements are simulated from the model presented in 5.1 where* $Y = 3.5557051$, $\mu = 0.3715953$, *and* $\lambda = 3.3026677$, *and* $\sigma = 0.01$. *At times the Warringer method returned missing values for stationary phase OD increment due to the coefficient of variation of the last six values of the smoothened growth curve being greater than 2%. (EM = Chapman-Richards with rounding correction, C-R = Chapman-Richards, W = Warringer).*

True blank value		Estimation method	Mean	Std	Min	Max
	Y		3.5224872	0.003813197	3.5125864	3.5311656
0.060	μ	EM	0.3753898	0.000883317	0.3733319	0.3773137
	λ		3.1420529	0.023902633	3.0994190	3.1911947
	Y		3.5224942	0.003813773	3.5125756	3.5311856
0.060	μ	C-R	0.3754049	0.000875916	0.3733830	0.3773094
	λ		3.1424383	0.023683350	3.1008812	3.1915630
	Y		3.5352652	0.012569833	3.4922953	3.5588934
0.060	μ	W	0.3661708	0.003261082	0.3571747	0.3726249
	λ		3.2837268	0.044814831	3.1876985	3.4013437
	Y		3.5556879	0.004433155	3.5469946	3.5652130
0.067	μ	EM	0.3716792	0.000956086	0.3692808	0.3738117
	λ		3.3028659	0.022124241	3.2600500	3.3478470
	Y		3.5556864	0.004433365	3.5470057	3.5652065
0.067	μ	C-R	0.3716739	0.000945796	0.3692995	0.3738260
	λ		3.3027211	0.021851381	3.2597162	3.3479981
	Y		3.5704150	0.012427128	3.5398420	3.5896838
0.067	μ	W	0.3629409	0.003841431	0.3548266	0.3715655
	λ		3.4257975	0.042096797	3.3360639	3.5082715
	Y		3.7789007	0.004283199	3.7675689	3.7903827
0.112	μ	EM	0.3519059	0.001006282	0.3496131	0.3540373
	λ		4.0266374	0.019923210	3.9963059	4.0833707
	Y		3.7788953	0.004283203	3.7675608	3.7903762
0.112	μ	C-R	0.3518903	0.001002466	0.3496039	0.3540105
	λ		4.0262528	0.019833098	3.9963406	4.0830106
	Y		3.7957745	0.014278441	3.7532447	3.8244297
0.112	μ	W	0.3428292	0.004727404	0.3328409	0.3542226
	λ		4.0753099	0.052306070	3.9659379	4.2298268

Table 7.2: *The means, standard deviations, minimums and maximums of estimated growth parameters in 100 simulations. The measurements are simulated from the model presented in 5.1 where* $Y = 0.5246799$, $\mu = 0.1129737$, *and* $\lambda = 7.7168240$, *and* $\sigma = 0.01$. *At times the Warringer method returned missing values for stationary phase OD increment due to the coefficient of variation of the last six values of the smoothened growth curve being greater than 2%. (EM = Chapman-Richards with rounding correction, C-R = Chapman-Richards, W = Warringer).*

True blank value		Estimation method	Mean	Std	Min	Max
0.060	Y	EM	0.5206361	0.0011242444	0.5184967	0.5235836
	μ		0.1164835	0.0004848789	0.1155191	0.1176426
	λ		7.5250377	0.0696293694	7.3731741	7.6759135
0.060	Y	C-R	0.5206381	0.0011271283	0.5184892	0.5235811
	μ		0.1164842	0.0004857834	0.1155190	0.1176349
	λ		7.5250352	0.0697184151	7.3719633	7.6759906
0.060	Y	W	0.5182365	0.0023175253	0.5127455	0.5235591
	μ		0.1213942	0.0025626826	0.1163936	0.1285476
	λ		7.7968587	0.1732424030	7.4869154	8.2818585
0.067	Y	EM	0.5244643	0.0011183728	0.5215310	0.5270312
	μ		0.1129788	0.0004952809	0.1119664	0.1139611
	λ		7.7222852	0.0625590980	7.5530610	7.8910218
0.067	Y	C-R	0.5244658	0.0011174913	0.5215335	0.5270270
	μ		0.1129801	0.0004956970	0.1119708	0.1139644
	λ		7.7224043	0.0626375534	7.5537542	7.8913969
0.067	Y	W	0.5219238	0.0029848540	0.5139984	0.5274568
	μ		0.1182217	0.0025952907	0.1139984	0.1260067
	λ		8.0219723	0.1672777994	7.7086495	8.5242180
0.112	Y	EM	0.55403927	0.0012439754	0.5515443	0.5564671
	μ		0.09633300	0.0003331582	0.0953123	0.0969755
	λ		8.62091951	0.0438856438	8.5420464	8.7054781
0.112	Y	C-R	0.55403709	0.0012442041	0.5515389	0.5564600
	μ		0.09633186	0.0003330993	0.0953096	0.0969747
	λ		8.62079022	0.0438726813	8.5416130	8.7051940
0.112	Y	W	0.55225289	0.0024419866	0.5463495	0.5562697
	μ		0.10081863	0.0024019458	0.0940262	0.1071308
	λ		8.92196752	0.1525300400	8.6715689	9.3847679

Table 7.3: *Example data, well 2. Estimated growth parameters with different blank values. (EM = Chapman-Richards with rounding correction, C-R = Chapman-Richards, W = Warringer).*

Blank		EM	C-R	W
0.060	Y	3.8107493	3.8107408	3.9087347
	μ	0.3787928	0.3787623	0.3646417
	λ	3.2894732	3.2886534	3.3173462
0.067	Y	3.7755136	3.7755034	3.8722402
	μ	0.3833194	0.3832851	0.3677111
	λ	3.1428247	3.1418830	3.1676701
0.112	Y	3.5545208	3.5544533	3.6456281
	μ	0.4240683	0.4239482	0.4070372
	λ	1.3997353	1.3967100	1.6775939

Chapter 8

Discussion

Although the least squares fit of the Chapman-Richards model as well as the the simultaneous models for two curves give promising results, more empirical comparisons (such as the ones that Warringer and Blomberg [14] did) are needed in order to assess the practical relevance of the mathematical models.

The Chapman-Richards model works well for a wide range of standard growth curves, but for growth curves of abnormal shapes the fit can be poor. However, given the diversity of forms abnormal curves assume, it is very difficult if not impossible to find a parametric model that fits sufficiently well all types of growth curves. In some applications a model that gives a good fit with rather normal growth curves might be good enough. If a fit is poor, or the estimation does not converge, a warning could be returned, and in this way abnormal growth curves could be separated (automatically) from the rest, and then be investigated with more care, or discarded.

Probably a more flexible model would be needed. An ideal model would be to have all the three growth parameters, the lag time, the growth rate and the stationary phase OD increment, well modeled. It could be a model consisting of three parts, each part idealizing the model for the specific parameter, and yet the different parts being smoothly connected. Buchanan *et al* [3] did introduce a three phase linear model that divides the growth curves into three phases: the lag and stationary phases where the growth rate is zero, and the exponential phase where the logarithm of the cell population increases linearly with time. It is a simple model that does not make

the transitions from the lag phase to the exponential growth and from the exponential growth to the stationary phase in a smooth curve. Although probably useful in many applications, we considered it too rough for our purposes.

There are some concerns related to the growth parameter estimation that do not directly depend on the model used. Warringer and Blomberg [14] stressed that the stationary phase OD increment should be viewed with some caution as an indicator of efficiency of growth. First, the relation between the biomass and the OD measured can differ quite substantially between different strains. Second, it is not known if the end of the growth phase is always the result of complete utilization of the carbon source glucose or due to other limitations.

Perhaps the most serious concern related to the growth parameter estimation is whether the definition of the lag time used is appropriate. There is currently no generally accepted definition for the boundary between the lag and the exponential phases. If the lag time is defined using the tangent line through the inflection point, it will be proportionally shorter for slowly growing cells than for fast growing cells. Another problem might be that if the OD measurements are not started soon after the sample has been prepared, the lag time is in reality longer than what can be seen from the growth curve.

The effect of a false blank value on the estimated growth parameters is alarmingly large, although some of it will probably disappear in the later analysis of the growth parameters due to our experimental setup. In the further processing of the estimated growth parameters we normalize the mutant values by the wild type values. These wild types are grown in the same Bioscreen and in the same run as the mutants. Thus, the mutants and the wild types that are used in the normalizing are likely to have almost the same blank. However, since a false blank can have different effects depending on whether the growth is slow or fast, the blank effect on the mutants may not be the same as on the wild types, and thus it cannot be assumed that the normalizing takes the blank effect away completely.

Future research could focus on analysis of the variations of the parametric estimates of lag time, growth rate and stationary phase OD increment. The only way to obtain realistic estimates of variation that depends on the measurement process and on the experiment are systematic empirical studies. These studies should be done in close collaboration with the

biologists. Also non-parametric estimation methods could be considered. Growth curves could be smoothened using *e.g.* kernel, spline, and k nearest neighbor smoothers, and the growth parameters could be estimated non-parametrically from the smoothened growth curves. Confidence intervals for the non-parametric growth parameter estimates would be obtained by bootstrapping technique.

Bibliography

[1] *User's Manual Bioscreen C.* Labsystems Oy, 1997.

[2] T.D. Brock, M.T. Madigan, J. Martinko, and J. Parker. *Biology of Microorganisms.* Prentice Hall International, Inc., 1994.

[3] R.L. Buchanan, R.C. Whiting, and W.C. Damert. When is Simple Good Enough: a Comparison of the Gompertz, Baranyi, and Three-Phase Linear Models for Fitting Bacterial Growth Curves. *Food Microbiology*, 14:313–326, 1997.

[4] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, B 39:1–38, 1977.

[5] E. Ericson, J. Warringer, I. Pylvänäinen, L. Fernandez, O. Nerman, and A. Blomberg. Analysis of 576 Saccharomyces cerevisiae Deletion Strains Revealed Significant Phenotypes for as Many Unknown as Known Genes. *In preparation.*

[6] W.E. Garthright. Refinements in the Prediction of Microbial Growth Curves. *Food Microbiology*, 8:239–248, 1991.

[7] G.J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions.* Wiley series in probability and statistics. John Wiley & Sons, 1997.

[8] H.W. Mewes, K. Albermann, M. Bähr, D. Frishman, A Gleissner, J. Hani, K. Heumann, K. Kleinen, A Maierl, S.G Olivier, F. Pfeiffer, and A. Zollner. Overview of the Yeast Genome. *Nature*, 387 (supplement):3–74, 1997.

[9] L.V. Pienaar and K.J. Turnbull. The Chapman-Richards generalization of von Bertalanffy's Growth Model for Basal Area Growth and Yield in Even-Aged Stands. *Forest Science*, 19(1):2–22, 1973.

[10] F.J. Richards. A Flexible Growth Function for Empirical Use. *Journal of Experimental Botany*, 10(29):290–300, 1959.

[11] P. Singleton and D. Sainsbury, editors. *Dictionary of Microbiology and Molecular Biology*. John Wiley & Sons, 3rd edition, 2001.

[12] J.N. Strathern, E.W. Jones, and J.R. Broach, editors. *The Molecular Biology of the Yeast Saccharomyces, Life Cycle and Inheritance*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1981.

[13] L. Von Bertalanffy. *Teoretische Biologie*. A Franke AG Verlag, Bern, 1951.

[14] J. Warringer and A. Blomberg. Automated Screening in Environmental Arrays Allows Analysis of Quantitative Phenotypic Profiles in Saccharomyces Cerevisiae. *Yeast*, 20(1):69–78, 2003.

[15] J. Warringer, E. Ericson, L. Fernandez, O. Nerman, and A. Blomberg. Yeast Phenomics on a Genome-Wide Scale. *Submitted*.

[16] J. Warringer, E. Ericson, I. Pylvänäinen, O. Nerman, and A. Blomberg. Revealing Quantitative Phenotypic Profiles. *Submitted*.

[17] E.A. Winzeler, Daniel D. Shoemaker, A. Astromoff, H. Liang, K. Anderson, B. Andre, R. Bangham, R. Benito, J.D. Boeke, H. Bussey, A.M Chu, C. Connelly, K Davis, F. Dietrich, S.V. Dow, M. El Bakkoury, F. Foury, S.H. Friend, E. Gentalen, G. Giaever, J.H. Hegemann, T. Jones, M. Laub, H. Liao, N. Liebundguth, D.J. Lockhart, A. Lucau-Danila, M. Lussier, N. M'Rabet, P. Menard, M. Mittmann, C. Pai, C. Rebischung, J. L. Revuelta, L. Riles, C.J. Roberts, P. Ross-MacDonald, B. Scherens, M. Snyder, S. Sookhai-Mahadeo, R. K. Storms, S. Véronneau, M. Voet, G. Volckaert, T.R. Ward, R. Wysocki, G.S. Yen, K. Yu, K. Zimmermann, P. Philippsen, M. Johnston, and R.W. Davis. Functional Characterization of the S. cerevisiae Genome by Gene Deletion and Parallel Analysis. *Science*, 285:901–906, 1999.

[18] M.H. Zwietering, I. Jongenburger, F.M. Rombouts, and K. van't Riet. Modeling of the Bacterial Growth Curve. *Applied and Environmental Microbiology*, 56(6):1875–1881, 1990.

Appendix A

Tables

Table A.1: *Example data, OD values directly from the Bioscreen (continues).*

Time point	well 1	well 2	well 3	well 4	well 5	well 6	well 7	well 8
0	0.143	0.137	0.155	0.147	0.144	0.138	0.139	0.130
1	0.139	0.134	0.152	0.145	0.145	0.135	0.136	0.127
2	0.146	0.140	0.157	0.150	0.151	0.141	0.142	0.132
3	0.149	0.142	0.162	0.154	0.154	0.144	0.145	0.135
4	0.153	0.146	0.167	0.158	0.159	0.147	0.149	0.138
5	0.158	0.149	0.172	0.163	0.163	0.151	0.152	0.141
6	0.161	0.152	0.176	0.167	0.167	0.153	0.155	0.144
7	0.165	0.156	0.182	0.172	0.173	0.156	0.160	0.147
8	0.170	0.160	0.187	0.177	0.179	0.159	0.164	0.152
9	0.176	0.166	0.193	0.184	0.184	0.164	0.170	0.156
10	0.183	0.171	0.201	0.191	0.193	0.169	0.177	0.161
11	0.191	0.180	0.209	0.200	0.203	0.175	0.185	0.167
12	0.200	0.188	0.219	0.211	0.214	0.183	0.193	0.174
13	0.210	0.198	0.229	0.221	0.225	0.191	0.203	0.181
14	0.224	0.209	0.242	0.234	0.237	0.202	0.216	0.191
15	0.237	0.223	0.256	0.250	0.254	0.214	0.229	0.203
16	0.252	0.239	0.271	0.267	0.269	0.229	0.246	0.216
17	0.273	0.258	0.292	0.289	0.291	0.247	0.264	0.230
18	0.294	0.279	0.316	0.312	0.316	0.264	0.283	0.247
19	0.318	0.300	0.341	0.337	0.341	0.286	0.305	0.265
20	0.347	0.324	0.371	0.368	0.370	0.310	0.330	0.286
21	0.378	0.353	0.402	0.399	0.403	0.342	0.362	0.310
22	0.411	0.384	0.436	0.433	0.438	0.374	0.394	0.338
23	0.448	0.418	0.473	0.472	0.475	0.408	0.431	0.369
24	0.491	0.457	0.519	0.517	0.519	0.445	0.467	0.401
25	0.532	0.496	0.561	0.560	0.561	0.485	0.509	0.437
26	0.576	0.537	0.605	0.607	0.606	0.528	0.553	0.475
27	0.623	0.584	0.652	0.655	0.656	0.574	0.599	0.518
28	0.671	0.631	0.699	0.705	0.705	0.621	0.648	0.563
29	0.722	0.682	0.750	0.756	0.757	0.667	0.697	0.610
30	0.774	0.733	0.802	0.809	0.810	0.711	0.747	0.659
31	0.831	0.790	0.860	0.867	0.865	0.756	0.797	0.709
32	0.878	0.839	0.907	0.913	0.916	0.805	0.849	0.760
33	0.930	0.891	0.957	0.965	0.966	0.852	0.899	0.810
34	0.980	0.942	1.005	1.015	1.015	0.901	0.949	0.862
35	1.032	0.993	1.053	1.064	1.063	0.953	0.999	0.912
36	1.084	1.048	1.105	1.115	1.113	1.002	1.046	0.962
37	1.129	1.095	1.146	1.152	1.151	1.050	1.090	1.013
38	1.159	1.132	1.174	1.179	1.178	1.091	1.124	1.058
39	1.186	1.165	1.203	1.207	1.208	1.127	1.151	1.104
40	1.212	1.194	1.230	1.234	1.233	1.156	1.180	1.138
41	1.235	1.218	1.251	1.253	1.254	1.186	1.205	1.167
42	1.258	1.242	1.273	1.275	1.276	1.211	1.231	1.195
43	1.279	1.264	1.294	1.298	1.297	1.234	1.254	1.221
44	1.294	1.283	1.308	1.312	1.314	1.256	1.272	1.243
45	1.311	1.299	1.322	1.325	1.327	1.276	1.289	1.266
46	1.325	1.315	1.337	1.337	1.344	1.293	1.305	1.284
47	1.341	1.332	1.355	1.355	1.361	1.308	1.318	1.301
48	1.350	1.344	1.362	1.364	1.370	1.323	1.331	1.316
49	1.362	1.357	1.373	1.373	1.381	1.336	1.342	1.332

Table A.2: *(continued) Example data, OD values directly from the Bioscreen.*

Time point	well 1	well 2	well 3	well 4	well 5	well 6	well 7	well 8
50	1.371	1.368	1.384	1.385	1.391	1.351	1.353	1.346
51	1.376	1.372	1.392	1.393	1.401	1.356	1.358	1.356
52	1.388	1.382	1.399	1.402	1.409	1.370	1.369	1.369
53	1.395	1.389	1.406	1.410	1.415	1.375	1.372	1.375
54	1.402	1.397	1.411	1.415	1.421	1.384	1.379	1.385
55	1.414	1.401	1.417	1.422	1.425	1.394	1.387	1.395
56	1.422	1.409	1.426	1.429	1.435	1.394	1.388	1.398
57	1.425	1.414	1.432	1.434	1.440	1.399	1.388	1.407
58	1.427	1.422	1.436	1.440	1.445	1.409	1.402	1.416
59	1.435	1.422	1.440	1.444	1.449	1.406	1.399	1.420
60	1.439	1.429	1.444	1.450	1.453	1.414	1.409	1.426
61	1.443	1.434	1.449	1.454	1.459	1.417	1.409	1.431
62	1.452	1.440	1.452	1.459	1.462	1.425	1.425	1.439
63	1.454	1.441	1.459	1.464	1.463	1.426	1.425	1.440
64	1.458	1.448	1.463	1.465	1.467	1.430	1.429	1.445
65	1.462	1.453	1.469	1.468	1.472	1.434	1.433	1.448
66	1.465	1.458	1.474	1.471	1.476	1.431	1.431	1.448
67	1.472	1.464	1.480	1.474	1.478	1.440	1.442	1.454
68	1.473	1.465	1.478	1.474	1.477	1.438	1.443	1.455
69	1.472	1.467	1.480	1.477	1.477	1.444	1.449	1.457
70	1.476	1.471	1.484	1.480	1.480	1.445	1.453	1.463
71	1.474	1.472	1.486	1.477	1.482	1.448	1.455	1.463
72	1.482	1.473	1.486	1.476	1.480	1.453	1.461	1.471
73	1.479	1.476	1.485	1.479	1.483	1.453	1.461	1.470
74	1.485	1.478	1.489	1.481	1.486	1.458	1.463	1.469
75	1.476	1.477	1.486	1.481	1.483	1.451	1.455	1.469
76	1.475	1.478	1.486	1.483	1.486	1.456	1.463	1.472
77	1.478	1.475	1.488	1.486	1.486	1.452	1.457	1.468
78	1.480	1.478	1.489	1.486	1.485	1.455	1.459	1.468
79	1.478	1.478	1.488	1.486	1.485	1.449	1.457	1.468
80	1.482	1.481	1.490	1.486	1.486	1.457	1.465	1.474
81	1.482	1.482	1.490	1.491	1.486	1.446	1.467	1.464
82	1.480	1.481	1.487	1.489	1.486	1.453	1.462	1.472
83	1.488	1.486	1.492	1.492	1.489	1.455	1.468	1.473
84	1.484	1.484	1.492	1.493	1.486	1.450	1.464	1.470
85	1.486	1.486	1.493	1.494	1.488	1.454	1.473	1.471
86	1.482	1.485	1.488	1.491	1.488	1.456	1.466	1.475
87	1.490	1.488	1.494	1.496	1.491	1.460	1.468	1.477
88	1.482	1.486	1.490	1.493	1.488	1.454	1.467	1.474
89	1.488	1.486	1.494	1.495	1.488	1.448	1.464	1.473
90	1.495	1.493	1.497	1.504	1.494	1.462	1.476	1.480
91	1.487	1.489	1.496	1.496	1.489	1.450	1.468	1.476
92	1.491	1.495	1.498	1.499	1.494	1.459	1.478	1.480
93	1.487	1.491	1.494	1.500	1.489	1.463	1.481	1.480
94	1.485	1.491	1.495	1.499	1.490	1.452	1.472	1.478
95	1.488	1.495	1.498	1.500	1.495	1.448	1.467	1.478
96	1.491	1.493	1.497	1.502	1.497	1.453	1.470	1.480
97	1.489	1.492	1.498	1.502	1.495	1.454	1.467	1.478
98	1.495	1.496	1.502	1.503	1.497	1.458	1.472	1.480
99	1.497	1.496	1.501	1.504	1.499	1.471	1.477	1.488

Table A.3: *(continued) Example data, OD values directly from the Bioscreen.*

Time point	well 1	well 2	well 3	well 4	well 5	well 6	well 7	well 8
100	1.495	1.498	1.500	1.503	1.498	1.473	1.480	1.487
101	1.496	1.497	1.500	1.502	1.499	1.471	1.481	1.486
102	1.496	1.497	1.500	1.503	1.499	1.472	1.481	1.486
103	1.500	1.499	1.499	1.504	1.501	1.472	1.483	1.490
104	1.498	1.497	1.501	1.504	1.502	1.469	1.478	1.490
105	1.500	1.500	1.504	1.505	1.503	1.471	1.477	1.490
106	1.498	1.501	1.503	1.507	1.501	1.471	1.477	1.489
107	1.498	1.498	1.500	1.503	1.497	1.475	1.481	1.488
108	1.496	1.500	1.502	1.502	1.498	1.472	1.477	1.491
109	1.497	1.498	1.499	1.506	1.501	1.476	1.485	1.493
110	1.497	1.500	1.500	1.503	1.502	1.478	1.486	1.493
111	1.503	1.505	1.504	1.507	1.500	1.479	1.490	1.494
112	1.504	1.504	1.506	1.506	1.501	1.476	1.483	1.493
113	1.501	1.503	1.503	1.506	1.501	1.481	1.498	1.492
114	1.503	1.506	1.505	1.504	1.501	1.480	1.491	1.493
115	1.499	1.505	1.502	1.504	1.501	1.475	1.485	1.492
116	1.505	1.507	1.508	1.506	1.503	1.480	1.491	1.494
117	1.508	1.510	1.509	1.506	1.501	1.479	1.494	1.495
118	1.502	1.504	1.504	1.504	1.502	1.477	1.491	1.494
119	1.503	1.509	1.507	1.505	1.504	1.479	1.489	1.495
120	1.501	1.503	1.504	1.506	1.499	1.475	1.485	1.495
121	1.502	1.505	1.508	1.504	1.500	1.476	1.487	1.494
122	1.501	1.505	1.507	1.506	1.503	1.477	1.489	1.493
123	1.506	1.507	1.507	1.507	1.502	1.480	1.493	1.496
124	1.504	1.506	1.510	1.506	1.504	1.474	1.484	1.494
125	1.500	1.507	1.506	1.509	1.502	1.477	1.488	1.495
126	1.504	1.509	1.512	1.507	1.503	1.475	1.484	1.496
127	1.502	1.507	1.509	1.507	1.504	1.478	1.493	1.498
128	1.508	1.507	1.509	1.506	1.502	1.480	1.496	1.497
129	1.507	1.506	1.511	1.505	1.502	1.476	1.492	1.497
130	1.505	1.507	1.508	1.510	1.503	1.476	1.495	1.496
131	1.508	1.506	1.509	1.511	1.504	1.478	1.500	1.494
132	1.506	1.508	1.512	1.507	1.502	1.478	1.500	1.499
133	1.506	1.509	1.509	1.508	1.505	1.478	1.498	1.498
134	1.506	1.510	1.509	1.510	1.504	1.478	1.497	1.497
135	1.508	1.509	1.509	1.510	1.506	1.477	1.496	1.496
136	1.507	1.508	1.511	1.507	1.504	1.476	1.496	1.498
137	1.507	1.510	1.510	1.509	1.506	1.478	1.499	1.499
138	1.509	1.510	1.510	1.505	1.508	1.479	1.497	1.498
139	1.510	1.511	1.508	1.510	1.508	1.481	1.499	1.500
140	1.504	1.510	1.508	1.510	1.508	1.477	1.499	1.500
141	1.508	1.510	1.509	1.510	1.506	1.477	1.500	1.498
142	1.513	1.511	1.512	1.509	1.510	1.480	1.504	1.499
143	1.512	1.513	1.513	1.513	1.509	1.481	1.501	1.498
144	1.510	1.512	1.514	1.512	1.509	1.480	1.500	1.501

Table A.4: *Calibration curve function data (run in June 3, 2002). d=diluted, ud=undiluted, the abbreviations for the specific Bioscreen instruments are given in the parenthesis. Well specific blank values are withdrawn and the undiluted values are multiplied by the dilution factor 10.*

ud (B)	d (B)	ud (C)	d (C)	ud (D)	d (D)	ud (E)	d (E)	ud (F)	d (F)
1.211	2.5	1.181	2.78	1.255	3.09	1.196	2.9	1.276	3.27
1.2	2.65	1.174	2.94	1.243	3.34	1.215	3.13	1.265	3.48
1.158	2.63	1.134	2.92	1.217	3.29	1.204	3.13	1.231	3.5
1.151	2.26	1.127	2.5	1.212	2.93	1.206	2.76	1.223	2.98
1.134	2.18	1.112	2.41	1.191	2.73	1.192	2.6	1.205	2.86
1.108	1.84	1.094	2.14	1.166	2.33	1.171	2.24	1.17	2.52
1.09	1.63	1.071	1.87	1.152	2.11	1.163	2	1.156	2.23
1.011	1.58	1.036	1.82	1.096	2.05	1.125	1.97	1.107	2.32
0.98	1.24	1.018	1.4	1.094	1.66	1.109	1.62	1.084	1.77
0.844	1.19	0.875	1.42	0.997	1.6	1.025	1.59	1.008	1.7
0.858	1.32	0.894	1.32	0.927	1.34	0.862	1.44	0.936	1.39
0.866	1.06	0.902	1.14	0.952	1.46	0.905	1.45	0.95	1.47
0.766	0.9	0.8	1.03	0.858	1.26	0.814	1.35	0.849	1.43
0.919	1.55	0.951	1.72	1.011	2.07	0.973	2.15	1.013	2.21
0.684	0.82	0.727	0.95	0.78	1.17	0.751	1.21	0.779	1.23
0.717	0.88	0.788	1	0.817	1.2	0.79	1.28	0.816	1.32
0.547	0.66	0.628	0.78	0.643	0.9	0.613	0.93	0.647	0.99
0.577	0.66	0.527	0.74	0.608	0.87	0.615	0.92	0.663	1.03
0.521	0.64	0.552	0.76	0.615	0.91	0.619	0.94	0.657	1.03
0.435	0.49	0.459	0.56	0.5	0.72	0.508	0.72	0.515	0.76
0.577	0.66	0.614	0.77	0.648	0.9	0.637	0.93	0.699	1.03
0.515	0.49	0.553	0.58	0.602	0.68	0.6	0.66	0.632	0.75
0.532	0.45	0.563	0.52	0.624	0.64	0.634	0.65	0.648	0.72
0.506	0.36	0.542	0.43	0.599	0.52	0.608	0.52	0.628	0.57
0.477	0.45	0.512	0.52	0.563	0.59	0.57	0.62	0.592	0.68
0.23	0.18	0.262	0.22	0.291	0.29	0.306	0.19	0.353	0.22
0.186	0.23	0.21	0.26	0.238	0.32	0.25	0.33	0.275	0.35
0.269	0.25	0.293	0.27	0.324	0.37	0.303	0.41	0.351	0.43
0.289	0.25	0.315	0.28	0.351	0.36	0.334	0.4	0.37	0.42
0.307	0.24	0.326	0.27	0.365	0.35	0.348	0.39	0.375	0.36
0.32	0.27	0.344	0.31	0.385	0.4	0.363	0.43	0.399	0.44
0.334	0.31	0.361	0.36	0.397	0.46	0.376	0.51	0.411	0.51
0.356	0.31	0.397	0.35	0.415	0.43	0.387	0.48	0.43	0.55
0.329	0.25	0.362	0.32	0.385	0.34	0.359	0.39	0.399	0.42
0.323	0.24	0.354	0.28	0.372	0.33	0.355	0.33	0.412	0.4
0.33	0.33	0.344	0.35	0.375	0.49	0.365	0.47	0.423	0.54
0.266	0.37	0.284	0.43	0.306	0.54	0.306	0.57	0.335	0.62
0.404	0.31	0.428	0.35	0.469	0.46	0.431	0.46	0.506	0.51
0.425	0.48	0.455	0.56	0.502	0.71	0.465	0.78	0.517	0.89
0.439	0.49	0.465	0.61	0.514	0.73	0.488	0.76	0.523	0.86
0.518	0.7	0.531	0.83	0.588	1	0.559	0.96	0.595	1.11
0.523	0.68	0.538	0.8	0.595	0.92	0.555	0.9	0.594	1.03
0.528	0.61	0.545	0.73	0.598	0.76	0.569	0.72	0.602	0.91
0.595	0.65	0.62	0.76	0.673	0.83	0.652	0.83	0.68	0.98
1.112	1.86	1.109	2.11	1.167	2.38	1.174	2.29	1.197	2.55

Table A.5: *Well specific means of the calibration curve function data. (Well specific blanks are withdrawn and the diluted OD-values are multiplied by the dilution factor 10).*

Well specific means of the undiluted samples	Well specific means of the diluted samples
1.22	2.91
1.22	3.11
1.19	3.09
1.18	2.69
1.17	2.56
1.14	2.21
1.13	1.97
1.07	1.95
1.06	1.54
0.95	1.5
0.895	1.36
0.915	1.32
0.817	1.19
0.973	1.94
0.744	1.08
0.786	1.14
0.616	0.852
0.598	0.844
0.593	0.856
0.483	0.65
0.635	0.858
0.58	0.632
0.6	0.596
0.577	0.48
0.543	0.572
0.288	0.22
0.232	0.298
0.308	0.346
0.332	0.342
0.344	0.322
0.362	0.37
0.376	0.43
0.397	0.424
0.367	0.344
0.363	0.316
0.367	0.436
0.299	0.506
0.448	0.418
0.473	0.684
0.486	0.69
0.558	0.92
0.561	0.866
0.568	0.746
0.644	0.81
1.15	2.24