

THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

A Parametric Approach to Yeast Growth Curve Estimation and Standardization

Ilona Pylvänäinen

CHALMERS | GÖTEBORG UNIVERSITY



Department of Mathematical Sciences
Division of Mathematical Statistics
Chalmers University of Technology and Göteborg University
SE-412 96 Göteborg, Sweden

A Parametric Approach to Yeast Growth Curve Estimation and Standardization
ILONA PYLVÄNÄINEN
ISBN 91-7291-628-1

© ILONA PYLVÄNÄINEN, 2005.

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 2310
ISSN 0346-718X

Department of Mathematical Sciences
Division of Mathematical Statistics
Chalmers University of Technology and Göteborg University
SE-412 96 Göteborg
Sweden
Telephone +46 (0)31-772 1000

A Parametric Approach to Yeast Growth Curve Estimation and Standardization

ILONA PYLVÄNÄINEN

Department of Mathematical Statistics

Chalmers University of Technology and Göteborg University

Abstract

The purpose of this thesis is to contribute to the understanding of yeast growth. It builds upon a dataset consisting of growth curves of 576 *Saccharomyces cerevisiae* mutants in eight different environments. The data will be a part of a publicly available phenotypic library, PROPHECY, containing growth curves and characteristics of viable *S. cerevisiae* mutants in a wide variety of growth conditions.

We compare the fits of modifications of logistic, Gompertz, and Chapman-Richards models for the growth curves. The comparisons indicate that the modified Chapman-Richards model describes our growth data best. Relevant information about the behavior of the mutants is obtained by estimating the physiologically important growth parameters: the lag time (time to adapt to the environmental change), the maximum relative growth rate, and the efficiency of growth. We introduce an alternative parameterization of the modified Chapman-Richards model that uses these growth parameters and investigate its uniqueness and parameter restrictions. We also show convexity of its logarithmic parameter space.

One of our findings is that the lag time and the growth rate depend strongly on the initial population size. However, in large-scale experiments with hundreds of strains, it is difficult to have the same constant initial population size. To address this problem and to enable easy visualization of the data, we develop a method to standardize growth curves with respect to the initial population size. The idea is to use a modified Chapman-Richards curve to predict what the behavior of a growth curve would have been, had the population had a fixed standard initial size. As a result, the initial population size correlation with lag time and growth rate reduces remarkably. We also introduce two ways to construct a summary curve from several standardized growth curves.

We suggest a set of filtering methods, based on the standardized and summary curves, in order to detect experiments and individual curves that are atypical or spurious. Finally, we compare the variability of wild type normalized mutant growth parameters from the modified Chapman-Richards, standardized, and summary curves. The variances are typically slightly smaller with the standardizing and summarizing methods than with the direct Chapman-Richards approach.

Keywords: Bioscreen, Chapman-Richards model, growth curve, growth rate, lag time, optical density (OD), *Saccharomyces cerevisiae*, standardized curve, stationary phase OD increment, summary curve

MSC2000 classification: 62P10

Acknowledgments

I would like to warmly thank my advisor Olle Nerman for all his time, help, encouragement, patience, and optimism during the past five years. Many thanks also to my co-advisor Marita Olsson for her continuous support and careful reading of this thesis.

I would also like to thank my biology colleagues at the Department of Cell and Molecular Biology, Göteborg University. Warm thanks to my co-advisor Anders Blomberg, who originally suggested the standardization of growth curves and has given valuable comments along the way. Many thanks to Elke Ericson for always being helpful and for kindly and quickly answering all of my questions, and to Jonas Warringer and Luciano Fernandez-Ricaud.

I would also like to thank Marianne Månsson for carefully checking the Convexity properties section of this thesis.

Ilona Pylvänäinen, April 2005

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 1 |
| 2 | Background | 5 |
| 2.1 | How does <i>S. cerevisiae</i> grow? | 5 |
| 2.2 | Optical density | 5 |
| 2.3 | Bioscreen C Analyzer | 7 |
| 2.4 | Motivating dataset | 7 |
| 2.4.1 | Calibration | 8 |
| 2.4.2 | Blank correction | 10 |
| 2.4.3 | Discussion | 11 |
| 3 | Growth models | 13 |
| 3.1 | Traditional growth models and their suggested biological parameterizations | 13 |
| 3.2 | Modified growth models | 16 |
| 3.2.1 | Growth parameters | 16 |
| 3.2.2 | Derivation of the growth parameters of the Chapman-Richards model | 17 |
| 3.2.3 | Comparing the fits of the modified growth models | 20 |
| 3.3 | A three part model | 21 |
| 3.3.1 | Fitting the three part model to the data | 23 |
| 3.4 | Discussion | 23 |
| 4 | An alternative parameterization of the Chapman-Richards model | 41 |
| 4.1 | Uniqueness | 42 |
| 4.2 | Convexity properties | 48 |
| 5 | Standardizing curves | 51 |
| 5.1 | Standardizing upwards | 52 |
| 5.1.1 | Standardizing two or more curves simultaneously | 55 |

| | | |
|----------|---|------------|
| 5.2 | Standardizing downwards | 57 |
| 5.3 | Fitting the standardization models to the data | 60 |
| 5.4 | Discussion | 61 |
| 6 | Summarizing curves | 71 |
| 6.1 | Method I | 71 |
| 6.2 | Method II | 73 |
| 6.3 | Fitting the data | 74 |
| 6.4 | Discussion | 75 |
| 7 | Quality filters | 81 |
| 7.1 | Quality filters for runs | 83 |
| 7.1.1 | Testing on data | 83 |
| 7.2 | Quality filters for wild type curves | 85 |
| 7.2.1 | Testing on data | 87 |
| 7.3 | Quality filters for mutant curves | 87 |
| 8 | The effect of standardization and summarizing on logarithmic strain coefficients (LSC) | 97 |
| 8.1 | LSC | 98 |
| 8.1.1 | The variance of runwise LSC | 98 |
| 8.2 | Discussion | 99 |
| 9 | Conclusions | 101 |
| A | Figures | 109 |
| B | Tables | 111 |
| C | Gompertz augmented Chapman-Richards model | 115 |
| D | Discussion of Conjecture 1 | 119 |
| E | Proofs | 123 |
| F | Simultaneous models for two growth curves | 135 |
| F.1 | Model I | 135 |
| F.2 | Model II | 136 |
| F.3 | Fitting the simultaneous models to the data | 137 |

Chapter 1

Introduction

Saccharomyces cerevisiae, better known as baker's yeast, has been domesticated thousands of years ago. 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 of cells can be grown in culture in a very small space, in the same way as bacteria can be grown. However, *S. cerevisiae* has the advantage of being a eukaryotic organism, and thus the results from genetic studies with *S. cerevisiae* are more easily applicable to human biology. The collaboration of more than 600 scientists from over 100 laboratories in Europe, USA, Canada, and Japan resulted in a publication of the complete genomic sequence of the *S. cerevisiae* in 1996 [10]. 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 seventy percent have a known function [14]. One important approach for characterizing a novel gene is to produce a knock-out mutant¹ 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 several 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.

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

¹A 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.

to distinguish the three physiologically relevant growth parameters: lag time (time to adapt to the environmental change), maximum relative growth rate (kinetics of growth), and stationary phase OD increment (related to the efficiency of growth).

Winzeler *et al* [28] showed that large numbers of deletion strains can be pooled, grown together 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 being a powerful approach, this methodology has some drawbacks. 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 [25].

In an alternative approach, Warringer and Blomberg [25] designed a system for large-scale quantitative phenotypic analysis of *S. cerevisiae* based on a commercially available Bioscreen C Analyzer². 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 estimates for growth parameters. The purpose of the system is to build a publicly available phenotypic library, PROPHECY³, containing growth curves and characteristics of viable *S. cerevisiae* mutants in a wide variety of growth conditions, and to use the library for studying gene functions. PROPHECY is publicly accessible at <http://prophecy.lundberg.gu.se> and it is continuously updated with growth data [7].

Warringer *et al* [27] used the system for phenotypic analysis of a set of 14 deletion strains in *S. cerevisiae*. Applying 96 conditions and analyzing 3000 growth curves, statistically significant phenotypes for nearly all strains in the screen were detected. These quantitative phenotypes portray aberrant growth behavior considering all three growth parameters, thus capturing defects in multiple, independent aspects of growth. Ericson *et al* [6] applied the system on quantitative phenotypic analysis of 576 *S. cerevisiae* mutants in eight different environments. Statistically significant phenotypes were revealed for over sixty percent of the analyzed genes. A functional role for the majority of the genes had not been reported earlier [14].

These developments are important initial steps towards large-scale analysis of mutants based on rigorous statistical grounds. However, more analytical tools need to be put in place before the methodology becomes fully operational. It is the aim of this thesis to address several issues related to growth curve modeling and growth parameter estimation. We hope that the results we obtain will contribute to establishing of a rigorous modeling basis that will facilitate the phenotypic analysis of large numbers of mutants.

In Chapter 2 we introduce the data that motivated the thesis and briefly discuss the issues of calibration and blank correction related to the yeast growth data from the

²Labsystems Oy, Finland

³PROfiling of PHEnotypic Characteristics in Yeast

Bioscreen. In Chapter 3 we compare the fits of modifications of logistic, Gompertz, and Chapman-Richards models for *S. cerevisiae* growth curves. The comparisons show that of these the modified Chapman-Richards model describes our growth data best. In Chapter 4 we give an alternative biological parameterization to the modified Chapman-Richards model, and investigate the basic theoretical properties of this parameterization.

The lag time and the growth rate depend strongly on the initial population size. However, in large-scale experiments with hundreds of mutants, it is difficult to keep the initial population size constant. To address this problem and to enable easy visualization of the data, we introduce a method to standardize growth curves with respect to the initial population size in Chapter 5. The idea is to predict what the behavior of a growth curve would have been, had the population had a standard initial population size. In Chapter 6 we present two ways to construct a summary curve from curves from parallel experiments.

In Chapter 7 we suggest a set of methods based on the standardized and summary curves to filter out curves or whole experiments that are atypical or spurious. Finally, in Chapter 8 we compare the variability of wild type normalized mutant growth parameters from the modified Chapman-Richards, standardized, and summary curves.

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 starts over for both of the cells. The result is an exponential increase in the number of cells. The doubling time varies with the strain, the growth medium, and the temperature. For more details, cf. [20].

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 number of cells and the cell 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 logarithmic growth curve is displayed in Figure 2.1.²

2.2 Optical density

Optical density (absorbance), OD, is a widely used concept in the estimation of 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 the light passing

¹We work with haploid cells.

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

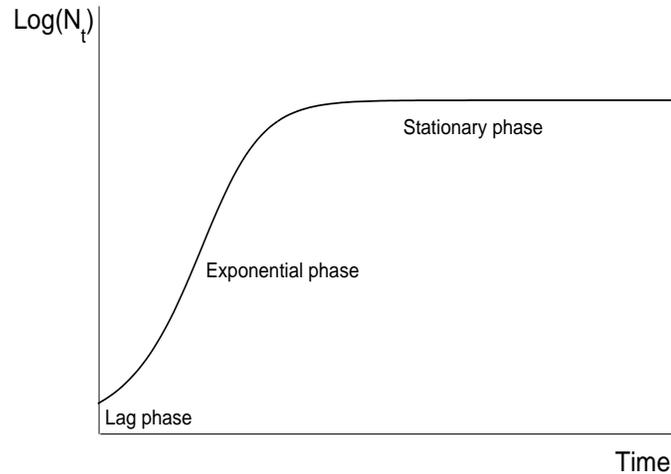


Figure 2.1: A typical logarithmic growth curve, where N_t is the number of cells at time t .

through the suspension. The more cell material is present, the more the suspension scatters the 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 goes through. For unicellular organisms, optical density is proportional (within certain limits) to the number of cells as well as to the 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 monitor the rate of growth of cultures, since the same sample can be checked repeatedly [2].

Optical density is defined as

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

where I_0 is the intensity of the incident light and I is the intensity of the transmitted light [17]. 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 sense all cells present in a solution, irrespectively of their viability.

Since the cell sizes affect the absorption capacity, the OD measurements are never perfectly proportional to the number of cells or to the cell mass. This error affects even the measurements done in the exponential phase since the cell size distribution in a culture depends on the age distribution which in turn depends on the rate of growth. For the sake of simplicity in the sequel, we choose to ignore this problem, both in the calibration (Section 2.4.1) and in the interpretation of the data.

2.3 Bioscreen C Analyzer

Bioscreen C Analyzer is an instrument developed to perform a wide range of microbiology experimentation automatically [1]. It 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 $1^{\circ}C$ to $60^{\circ}C$). 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.

There are two 100-well (10×10) disposable Honeycomb multiwell plates in each Bioscreen C instrument. The volume of each well is $400 \mu l$. 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 of measurements, and the maximum number of measurements per well is 400. This design strongly reduces the time and work 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 Motivating dataset

Altogether 577 strains of *S. cerevisiae* — 576 mutants and one wild type — were run in synthetically defined (SD) medium³, which is the reference condition, and in seven different environments where either some chemical was added to the SD medium, or another temperature than the standard $30^{\circ}C$ was used. The different environments and their abbreviations are given in Table 2.1. Optical density was recorded using a Bioscreen C Analyzer. Measurements were taken every 20 minutes during a 48 hour period, *i.e.* at 145 time points. Strains were run in quadruplicates (reference condition) or in duplicates (environments), in the same well location and in the same Bioscreen C Analyzer instrument during different days. The wild type positions on the plates were randomized once, with one per quadrant. The positioning of the wild types and mutants on the plates and in the Bioscreen instruments is shown in Figure A.1 in Appendix A. In the sequel, by run we refer to each 48 hour period of OD measurements of 192 mutants and 8 wild types in a specific Bioscreen instrument.

All data are smoothened so that each OD value lower than the previous value (*i.e.* the OD value at the previous time point) is set the previous value. This is biologically reasonable since the measured OD values tend to be too small rather than too large, mostly due to air bubbles. For more information about the data, see [6]. When we

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

Table 2.1: *The environments of the motivating dataset and their abbreviations.*

| Environment | Abbreviation |
|------------------------|--------------|
| Temperature 39°C | 39°C |
| Temperature 41°C | 41°C |
| Dinitrophenol | DN |
| Caffeine | CA |
| Sodium chloride | NA |
| Methylviologen | MV |
| Methylmethanesulfonate | MM |
| Reference condition | NO |

refer to a specific run, we write the environment abbreviation (for 39°C and 41°C only the numbers are written), then the Bioscreen instrument (C, D or E), and then the date, *e.g.* 39D0307 stands for the run in environment 39°C, in Bioscreen instrument D, on March 7.

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 the current 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 would 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.

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 each Bioscreen. This gave us the well and Bioscreen specific blanks. Then, the water was poured off the plate and the plate was placed in a 37°C chamber to make all the water evaporate. Stationary phase wild type cells (that had been growing on a shaker in 30°C over

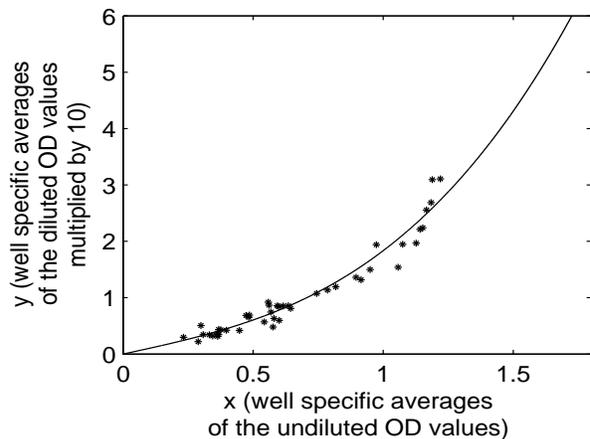


Figure 2.2: *Calibration curve and the data that were used to fit the calibration curve function. The well specific blank values are subtracted and the resulting OD values for the diluted samples are multiplied by ten.*

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 each Bioscreen.

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 B.1 in Appendix B). 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 B.2 in Appendix B). 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).

Curve fitting

Using regression, a curve was fitted with the well specific average of the blank corrected undiluted OD (x) as independent and the well specific average of the blank corrected diluted OD multiplied by ten (y) as dependent variable (Figure 2.2). Therefore, 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 .

We assume that the blank corrected diluted OD values and the blank corrected undiluted OD values are almost equal approximately up to 0.3. A cubic function

$$y = x + cx^3$$

was fitted.⁴ Using least squares estimation, we obtained the curve⁵

$$y = x + 0.83x^3. \tag{2.1}$$

Having a second degree term in the polynomial would make the curve too steep in the right end, so that when extrapolating for high values of x , the values of y would be too high.

We measured the same plate in each Bioscreen 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, and the lines were close to the 45° degree line, we decided to use the same calibration curve function for all Bioscreens. All data in this thesis are calibrated using the function (2.1) where now x is the blank corrected OD value from the Bioscreen and y is the resulting calibrated blank corrected OD value (more about the blank correction in the next section).

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 histogram of the blank values is 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.

⁴Since x and y are assumed to be almost equal approximately up to 0.3, the coefficient of x was set to one.

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

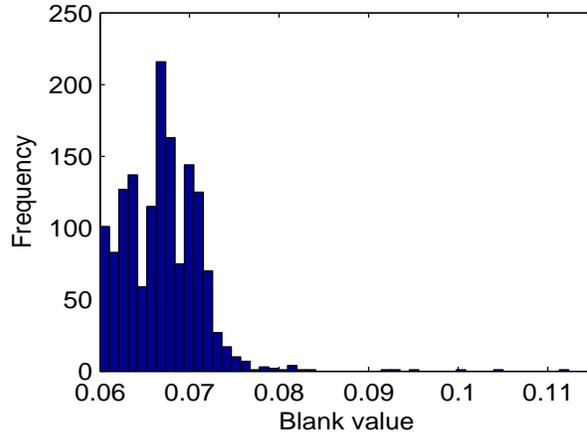


Figure 2.3: *Histogram of the blank values from two experiments where the OD values of wells containing only sterile water were measured. There are in total 1500 observations.*

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 alternatively have been done in two steps. First, to fit the function as we did. Second, to replace the small y values (*e.g.* values corresponding to $x < 0.35$) by the values from the first step 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 that 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 dataset that it is based upon, the highest undiluted OD value is 1.22, but in the motivating dataset (and in most of the data collected in PROPHECY) there are OD values up to 1.7. Also, we are aware that the use of the same blank value in all Bioscreens and in all wells is questionable. The effect of a false blank value was found to be alarmingly large, although some of it may disappear in the later analysis of the growth parameters due to our experimental setup [12]. The few really extreme blank measurements are hopefully measurement errors, rather than true blanks.

Chapter 3

Growth models

An adequate growth model is useful for describing growth curves and for concentrating the information in measured data into a number of meaningful parameters. Also, a parametric model will be needed when standardizing growth curves with respect to the initial OD, as we will see in Chapter 5.

In this chapter we compare the following commonly used functions as models for yeast growth: logistic [30], Gompertz [9], Richards [13], and Chapman-Richards [11]. 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 . Modeling $\log(N_t/N_0)$ can be a problem because the curves cannot pass through 0 at $t = 0$. Therefore we adopt the ideas of Garthright [8] and modify the functions in order to model $\log(N_t)$ instead.

3.1 Traditional growth models and their suggested biological parameterizations

Most of the commonly used functions for describing a sigmoidal¹ growth curve utilize parameters that do not have a clear biological interpretation and it can be difficult to give initial values for the parameters in the model fitting algorithms. To address this problem Zwietering *et al* [30] re-parameterized the logistic, Gompertz, Richards, Schnute [16], and Stannard [19] growth curve functions. They showed that the modified functions of Richards, Schnute, and Stannard are basically the same. The new parameters in the re-parameterized functions are: A_z the asymptote, the maximum value of the growth reached (on the logarithmic scale); μ the maximum relative population growth rate, the slope of the tangent of the logarithmic growth curve at the

¹A sigmoidal growth curve is an increasing curve which first has a convex shape and then a concave shape.

inflection point; and λ_z the lag time, the time axis intercept of the tangent at the inflection point on the logarithmic growth curve. We use the notations A_z and λ_z for the growth parameters in the Zwietering's re-parameterized functions to distinguish them from the modified growth parameters that we will actually use and estimate (Section 3.2.2).

For easy reference we give the growth curve functions together with their re-parameterized forms here. Note that we always assume that measurements start at time zero, so that $t \geq 0$.

Logistic: The logistic growth function is

$$\begin{aligned} v_t = \log\left(\frac{N_t}{N_0}\right) &= \frac{\beta_0}{1 - \beta_1 e^{-\beta_2 t}} \\ &= \frac{A_z}{1 + e^{\frac{4\mu}{A_z}(\lambda_z - t) + 2}} \quad , \end{aligned}$$

where $\beta_0, \beta_2, A_z, \mu, \lambda_z > 0$, and $\beta_1 < -1$.

Gompertz: The Gompertz function is

$$\begin{aligned} v_t = \log\left(\frac{N_t}{N_0}\right) &= \beta_0 e^{-e^{b - \beta_2 t}} \\ &= A_z e^{-e^{\frac{\mu e}{A_z}(\lambda_z - t) + 1}} \quad , \end{aligned}$$

where $\beta_0, b, \beta_2, A_z, \mu, \lambda_z > 0$.

Richards: The Richards function is

$$\begin{aligned} v_t = \log\left(\frac{N_t}{N_0}\right) &= \frac{\beta_0}{[1 + \nu e^{k(\tau - t)}]^{\frac{1}{\nu}}} \\ &= \frac{A_z}{\left[1 + \nu e^{\frac{\mu}{A_z}(1 + \nu)^{(1 + \frac{1}{\nu})}(\lambda_z - t) + (1 + \nu)}\right]^{\frac{1}{\nu}}} \quad , \end{aligned} \quad (3.1)$$

where $\beta_0, k, A_z, \mu, \lambda_z > 0$, and $\nu \neq 0$.

Chapman-Richards: The Chapman-Richards function [11] is

$$v_t = \log\left(\frac{N_t}{N_0}\right) = \beta_0 \left[1 - \beta_1 e^{-\beta_2 t}\right]^{1/(1 - \beta_3)} \quad , \quad (3.2)$$

where

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

or

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

The restrictions $1 - \beta_3 < \beta_1 < 1$ and $\beta_1 < 1 - \beta_3$ are made in order to have the inflection time point of the curve later than at time zero. Re-parameterizing the Chapman-Richards function so that it contains biological parameters as in Zwietering *et al* [30] (the re-parameterizing is done in the same way as the re-parameterization of the modified Chapman-Richards function, which will be presented in detail in Section 3.2.1), gives

$$v_t = \log \left(\frac{N_t}{N_0} \right) = A_z \left[1 - (1 - \beta_3) e^{\frac{\beta_3}{A_z} \frac{\beta_3 - 1}{\beta_3} \mu (\lambda_z - t) + \beta_3} \right]^{\frac{1}{1 - \beta_3}}, \quad (3.3)$$

where

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

Substituting ν by $\beta_3 - 1$ in the re-parameterized Richards function (3.1) would result in the re-parameterized Chapman-Richards function (3.3). In fact, the Chapman-Richards model is also known as the Richards model.

When $\beta_3 = 2/3$, the function (3.2) results in the von Bertalanffy function [22]. Richards [13] showed that the function is also equivalent to the logistic model when $\beta_3 = 2$. The restriction that we have adopted, that the inflection time point should be positive, restricts the values of β_1 and β_3 so that the otherwise possible $\beta_3 = 0$ is not allowed. However, with $\beta_3 = 0$ and $0 < \beta_1 < 1$, the function corresponds to the monomolecular growth model [21]. The limiting form of the function when β_3 tends to 1 and β_1 tends to 0 in a subordinated rate, is the Gompertz (for more details, cf. Appendix C). We will not discuss the details of the von Bertalanffy and monomolecular models.

The Chapman-Richards model is very flexible. It can be fitted to both exponential and sigmoidal growth patterns. 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 cause convergence problems in the curve fitting algorithms.

3.2 Modified growth models

All models described above have a problem at $t = 0$ because $v_t > 0$ for all t (although v_0 is close to 0). Therefore we modify them in the spirit of Garthright [8], *i.e.* instead of modeling $\log(N_t/N_0)$, we model $\log(N_t)$. That is, we introduce a new parameter $D < 0$, and set

$$g_t = \log(N_t) = y_t + D, \quad (3.4)$$

where D is $\log(N_0) - y_0$. We then have for the logistic curve,

$$y_t = \frac{\beta_0}{1 - \beta_1 e^{-\beta_2 t}}, \quad (3.5)$$

for the Gompertz curve,

$$y_t = \beta_0 e^{-e^{b-\beta_2 t}}, \quad (3.6)$$

and for the Chapman-Richards curve,

$$y_t = \beta_0 \left[1 - \beta_1 e^{-\beta_2 t} \right]^{1/(1-\beta_3)}. \quad (3.7)$$

By adding the parameter D , fitting problems that would occur whenever y_0 is noticeably above zero, are avoided.

Convention 1 *In the sequel, when we write logistic, Gompertz or Chapman-Richards, we refer to their modified versions as presented in this section.*

3.2.1 Growth parameters

To obtain information about the growth behavior of the cells, we estimate the following physiologically important growth parameters: the lag (or adaptation) time λ , the (maximum relative) growth rate μ , and the stationary phase OD increment Y .

The lag time is traditionally defined as the time required to adjust cell metabolism to conditions permissive for reproduction [23]. For instance, a longer lag time in certain chemical environment may indicate that it takes a longer time for the cells to produce a defense against the chemical, and thus a longer time to be able to start growing. The (maximum relative) growth rate is the maximum derivative of

the logarithmic growth curve g_t . From the growth rate the doubling time, the time required for the population to double, can easily be calculated as $\log(2)/\mu$.² A smaller growth rate in some environment may for example indicate that the DNA replication takes a longer time in that environment, or that the rate of cell death is larger than in the reference condition. The amount of time required for a population to reach a specific size is, for a range of relatively large sizes, approximately determined by the initial population size, the lag time, and the doubling time. Therefore both lag time and growth are important in safety related food microbiology, for example.

The cell density in the stationary phase reflects the achieved biomass increase, given a limited amount of energy, *i.e.* the efficiency of growth. We estimate the efficiency of growth by the stationary phase OD increment, the difference between the final OD and the initial OD. For example, a smaller stationary phase OD increment in some environment may indicate that in that particular environment the cells cannot use the existing energy as effectively as in the reference condition.

3.2.2 Derivation of the growth parameters of the Chapman-Richards model

Next, the growth parameters of the Chapman-Richards model

$$g_t = \log(N_t) = \beta_0 \left[1 - \beta_1 e^{-\beta_2 t} \right]^{1/(1-\beta_3)} + D \quad (3.8)$$

are derived. Because of modeling $\log(N_t)$ instead of $\log(N_t/N_0)$ and adding the parameter D , the growth parameters A_z and λ_z that Zwietering *et al* use are not the parameters we want to estimate. In addition, the stationary phase OD increment we estimate differs from the parameter A_z of Zwietering *et al* in that it is the increment on the non-logarithmic scale. The growth parameter derivation is illustrated in Figure 3.1.

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

$$\begin{aligned} Y &= e^{\beta_0+D} - e^{g_0} \\ &= e^{\beta_0+D} - e^{\beta_0(1-\beta_1)^{\frac{1}{1-\beta_3}} + D}. \end{aligned}$$

(We have idealized slightly in that we think of the final OD to be not that at the end of experiment but the value after infinite time.) The stationary phase OD increment should only be estimated for curves that have reached, or almost reached, the stationary phase at the last time point.

²Note that in the fitted Chapman-Richards curve there is no exact exponential phase, but if there was one with the relative growth rate μ , the doubling time would be $\log(2)/\mu$.

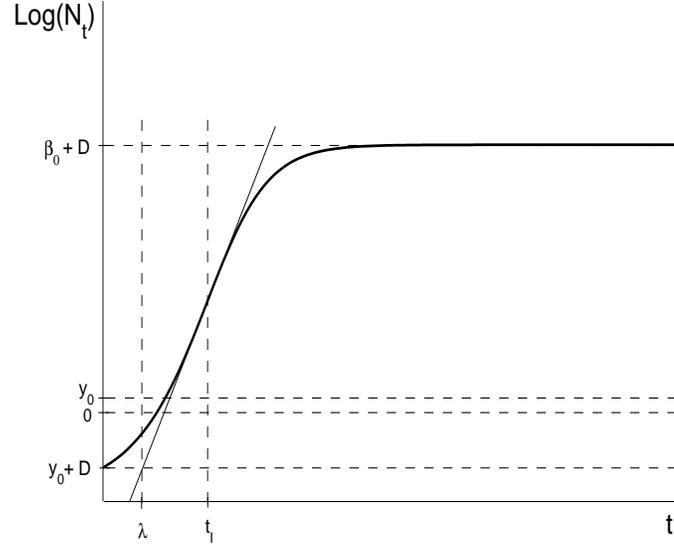


Figure 3.1: An illustration of the growth parameter calculation in the Chapman-Richards model. Here N_t is the population size at time t , t_I is the inflection time point, y_0 is given by (3.7) (at $t = 0$), $D = \log(N_0) - y_0$, and λ is the lag time.

The growth rate: The (maximum relative) growth rate, μ , is defined as the slope of the tangent of the logarithmic growth curve g_t at its inflection point. The inflection time point t_I is obtained by calculating the second derivative of the function (3.8) with respect to t , setting this to zero and solving with respect to t . The first derivative is

$$\frac{dg_t}{dt} = \frac{\beta_0 \beta_1 \beta_2 e^{-\beta_2 t} (1 - \beta_1 e^{-\beta_2 t})^{\frac{1}{1-\beta_3}-1}}{1 - \beta_3}$$

while the second derivative is given by

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

Equating this to zero gives the solution

$$t_I = \frac{\log\left(\frac{\beta_1}{1-\beta_3}\right)}{\beta_2}.$$

The growth rate parameter μ is finally derived by calculating the first derivative at this inflection time point t_I :

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

Since we work on the logarithmic size scale, μ corresponds to the maximum relative growth rate on the absolute scale.

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

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

The lag time λ , is the time axis value at the intercept of this tangent line with the base line $y_0 + D$, so that

$$y_0 + D = \mu \lambda + \beta_0 \beta_3^{\frac{1}{1-\beta_3}} - \mu t_I + D. \quad (3.9)$$

Solving (3.9) with respect to λ yields:

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

We were not able to rewrite the function (3.8) so that it would only contain the growth parameters and D and β_3 . However, if needed, the initial values for the parameters (in the model fitting algorithms) can be estimated using the estimates from the least squares fit of the model for $\log(N_t/N_0)$, function (3.3). Furthermore, in Chapter 4 we will see that the Chapman-Richards model can be expressed as a function of the initial OD denoted by s , the derivative d_0 at time zero, λ , μ , and β_3 , even if we cannot write down the function explicitly.

The growth parameters of the logistic and Gompertz models are derived analogously. The growth parameters are

$$\lambda = \frac{\frac{4}{1-\beta_1} - \log\left(-\frac{1}{\beta_1}\right) - 2}{\beta_2},$$

$$\mu = \frac{\beta_0 \beta_2}{4},$$

$$Y = e^{\beta_0 + D} - e^{\frac{\beta_0}{1-\beta_1} + D},$$

for the logistic model, and

$$\begin{aligned}\lambda &= \frac{\frac{b}{e} + e^{-e^b} - \frac{1}{e}}{\frac{\beta_2}{e}}, \\ \mu &= \frac{\beta_0 \beta_2}{e}, \\ Y &= e^{\beta_0 + D} - e^{\beta_0 e^{-e^b} + D},\end{aligned}$$

for the Gompertz model.

3.2.3 Comparing the fits of the modified growth models

We compare the fits of the modified growth models on the smoothed, blank corrected, and calibrated data described in Section 2.4, *i.e.* hundreds of growth curves from different environments. Nonlinear regression models were fitted via least squares in the 145 measurement points, using the large-scale algorithm in the *lsqnonlin*-function in Matlab.³ It is a subspace trust region method based on the interior-reflective Newton method described in [3], [4]. Our experience shows that the solutions are not sensitive to the choice of the start values. For the sake of reproducibility, we give the exact start values that we used for the parameters in the model fitting algorithms: $\beta_0 = 4.5$, $\beta_1 = -50$, $\beta_2 = 0.3$, $D = -3$ for the logistic; $\beta_0 = 4.5$, $b = 3.2$, $\beta_2 = 0.3$, $D = -3$ for the Gompertz; and $\beta_0 = 4.5$, $\beta_1 = -50$, $\beta_2 = 0.3$, $\beta_3 = 3$, $D = -3$ for the Chapman-Richards.

The fits are compared visually and by looking at the coefficient of determination,

$$r^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{t_p=1}^{145} (g_{t_p}^* - x_{t_p})^2}{\sum_{t_p=1}^{145} (x_{t_p} - \bar{x})^2}, \quad (3.10)$$

where $g_{t_p}^*$ is the fitted curve value at time point t_p , x_{t_p} is the observed⁴ value at time point t_p , and $\bar{x} = \frac{\sum_{t_p=1}^{145} x_{t_p}}{145}$.

Figures 3.2-3.5 show typical curves fitted by the three models compared. As expected, the Chapman-Richards method nearly always gives the best fit, since it encompasses both the logistic and the Gompertz models. The Gompertz model overestimates the slope, and moreover, it does not give a sufficiently good fit at any part of the curve. The logistic model gives a better fit than the Gompertz. However, the residual plots imply that there is a small systematic error in the Chapman-Richards

³The Matlab functions are available upon request.

⁴Smoothened, blank corrected, and calibrated OD value.

model as well. The minor systematic deviations of the data from the theoretical model are in the beginning of the curve and in the transition from the exponential phase to the stationary phase.

We are primarily interested in modeling typical growth curves rather than problematic growth curves. Hence, the discussion above considers typical growth curves. However, we would like to say a few words about fitting atypical growth curves, three examples are given in Figure 3.6. The Chapman-Richards model gives clearly the best fit also for atypical curves although it cannot be considered sufficient to describe them. The top curve in Figure 3.6 is an example of an outcome of technical artifacts. The middle curve is a typical example of a curve in the Methylmethanesulfonate environment. The Chapman-Richards model should not be used for the curves in this environment. The bottom curve shows occasionally observed atypical behavior in the very beginning of an experiment. Given the diversity of forms atypical curves assume, it is very difficult to find a model that fits sufficiently well to all types of growth curves. However, even if the model cannot be considered sufficient to describe atypical curves, it could be possible to use the information of the fit, *e.g.* the coefficient of determination, to filter out bad curves. We will do this in Chapter 7.

It is natural that the Chapman-Richards model gives the best fit of the data since it encompasses the other two models and it has more parameters than the other two models. This does not necessarily mean that the model fits well to the data, the model could be overfitting. As the number of parameters in a model increases, the model curve can bend in more complicated ways. If the number of parameters in our model is larger than necessary to catch the main characteristics of the "true" growth curve, the risk of overfitting increases. Similarly, if we use models with less parameters than necessary, the risk of underfitting increases; the models may not be flexible enough to match the actual growth curve well enough. However, since there are so many measurements for each curve, we do not have reason to believe that we have any overfitting problem here.

3.3 A three part model

From the residual plots of the fit of hundreds of growth curves, we see that the fit in the beginning of the curve and in the transition from the exponential phase to the stationary phase, is often not good. Even the fit of the Chapman-Richards model 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 especially with the Gompertz model.

The desire to overcome the problems mentioned above was one of the reasons why we wanted to fit a model which divides the growth curve into three parts. The other reason was to try to neutralize correlation between the initial OD and the lag time,

and between the initial OD and the growth rate.

It has been reported that the initial OD may influence the rate of growth [5]. This is a natural phenomenon, because in a sample with more cells in the beginning, there are less nutrients per cell, and thus the population can grow for a shorter time (than a population with less cells in the beginning) before it runs out of nutrients. It may not even reach the maximum growth rate. The growth in the beginning, when there are still enough nutrients for all the cells, does not tend to be affected by the initial OD.

We investigated the correlation between initial OD (the calibrated and blank corrected OD value at the time zero) and growth parameters on a dataset containing 99 wild types in the reference condition. The initial OD values vary between 0.015 and 0.106 (Figure 3.7). The dataset comes from an experiment where the effect of the initial OD was studied, and thus the range of the initial OD values is wide on purpose. The growth parameters are calculated as given in Section 3.2.2 (using the Chapman-Richards model). There is a strong negative correlation between the lag time and initial OD, and between the growth rate and initial OD (Figure 3.8). However, there is hardly any correlation between the initial OD and the stationary phase OD increment. Figures 3.9-3.11 show the histograms of the initial OD values in each environment and over all environments in the motivating dataset. The averages and coefficient of variations of the initial OD values in each run are given in Table 3.1.

We construct 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.⁵ 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

$$g_t^* = \begin{cases} g_t, & t \leq t_I, \\ g_{t_I} + \mu(t - t_I), & t_I \leq t \leq t_I + \Delta, \\ g_{t_I - \Delta} + \mu\Delta, & t \geq t_I + \Delta, \end{cases} \quad (3.11)$$

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 (3.4). The three part model is illustrated in Figure 3.12.

⁵We still call the cut point inflection point.

3.3.1 Fitting the three part model to the data

We fitted the three part model as a nonlinear regression model via least squares as in Section 3.2.3, to the same data.⁶ The start values for the parameters in the model fitting algorithms were the same as in Section 3.2.3 and the start value for Δ was 0. Atypical growth curves are excluded from the comparisons. Examples of curve fits with the three part model are given in Figures 3.13-3.16. Figures 3.2-3.5 show the same data fitted by the ordinary models.

The Gompertz model gains the most from adding the linear part in the middle. For almost all curves the estimate of Δ is larger than one hour, and the fit of the model improves remarkably compared to the ordinary Gompertz model. With the logistic growth function as g_t , the estimate of Δ is zero for more than 50% of the curves. For the rest of the curves the fit is in general improved by adding a linear part in the middle. However, the ordinary Chapman-Richards model (3.8) gives a better fit than the three part model with logistic or Gompertz function.

The estimate of Δ is smallest when using the Chapman-Richards function in the three part model. For over 90% of the curves it is zero, and for over 95% less than one hour. Even for the curves with the estimate of Δ larger than one hour, the fit of the the three part model is often similar to the fit of the ordinary Chapman-Richards model. Although in some cases the fit of the three part model is clearly better, it does not neutralize the correlation between the initial OD and lag time and the correlation between the initial OD and growth rate (Figure 3.17). In Chapter 5 we will introduce another method to neutralize the effect of the initial OD.

3.4 Discussion

With rather typical "normal" growth curves, the Chapman-Richards model always gives 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.

The three part model with logistic and Gompertz functions was clearly better than the logistic and Gompertz models themselves, but not better than the ordinary Chapman-Richards model. When compared to the Chapman-Richards model, the three part model with the Chapman-Richards function gave a better fit in few cases, and in the rest of the cases the fit was equal to that of the Chapman-Richards model. Since the tree part model is more complicated than the Chapman-Richards model,

⁶The Matlab functions are available upon request.

adding a linear part in the middle may not be relevant here. However, in Chapter 5 we will see that it is essential in the standardization of curves.

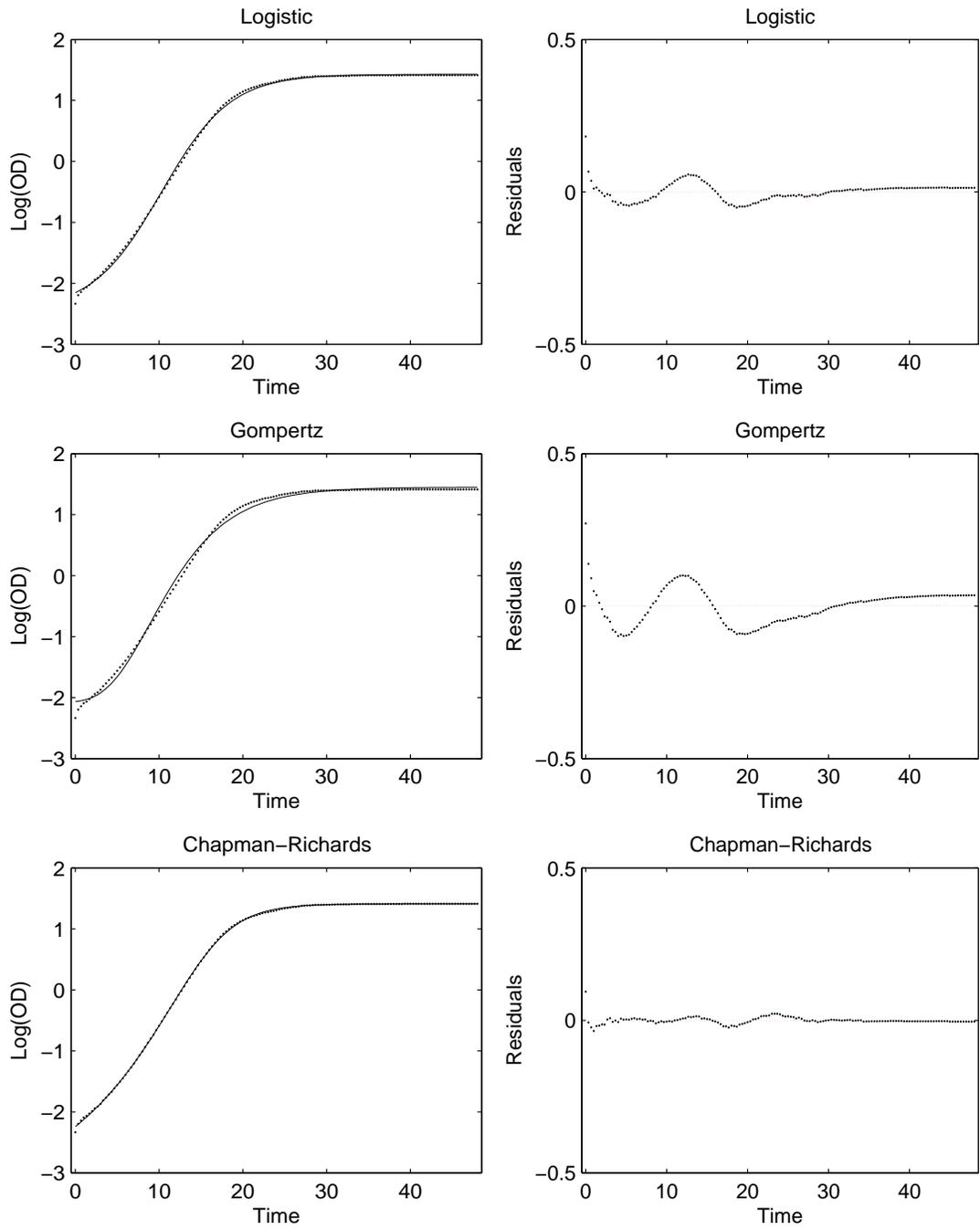


Figure 3.2: *The logistic, Gompertz and Chapman-Richards models are fitted to the data NOD0305, well 3. The corresponding residual plots are on the right.*

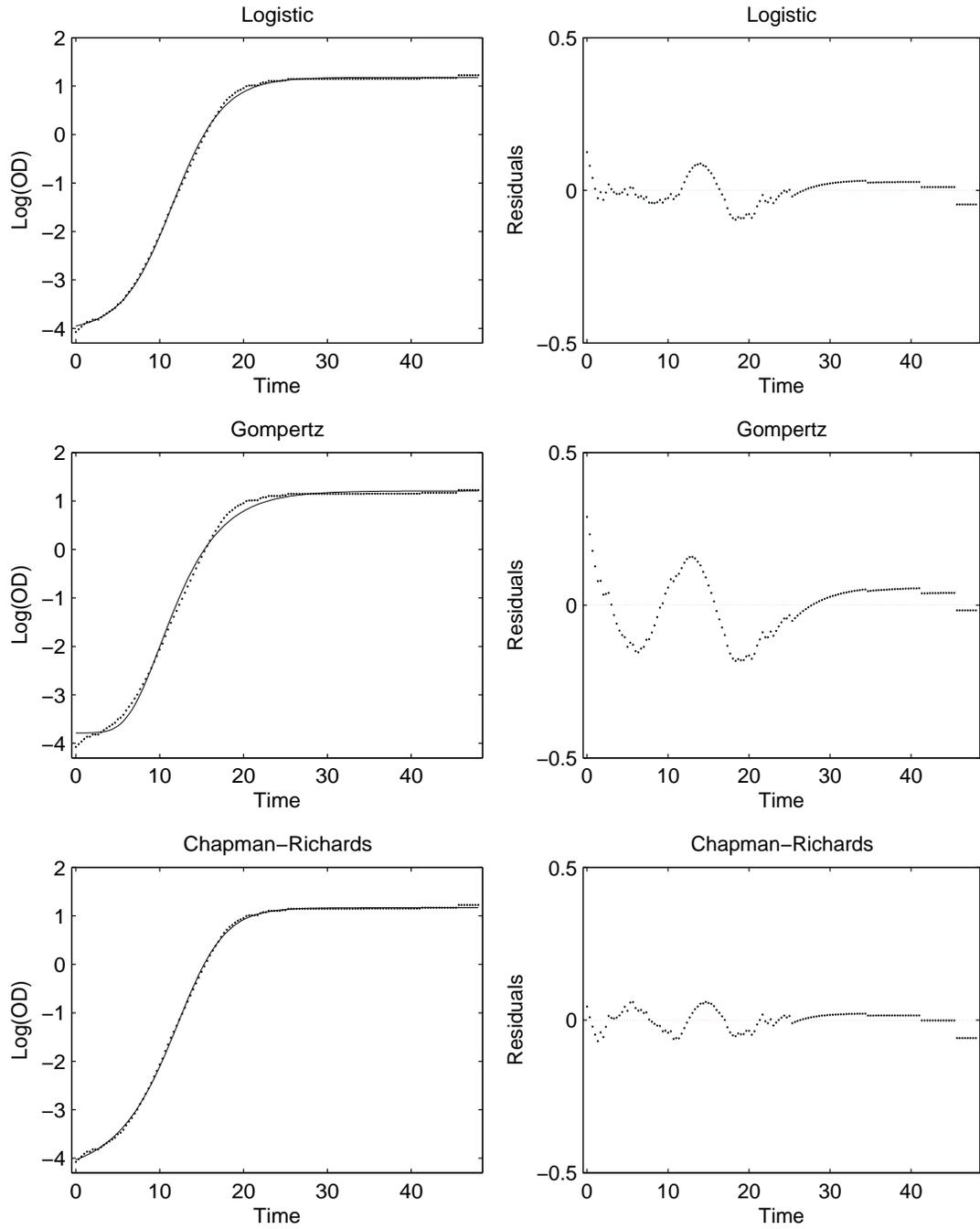


Figure 3.3: *The logistic, Gompertz and Chapman-Richards models are fitted to the data NOC0426, well 7. The corresponding residual plots are on the right.*

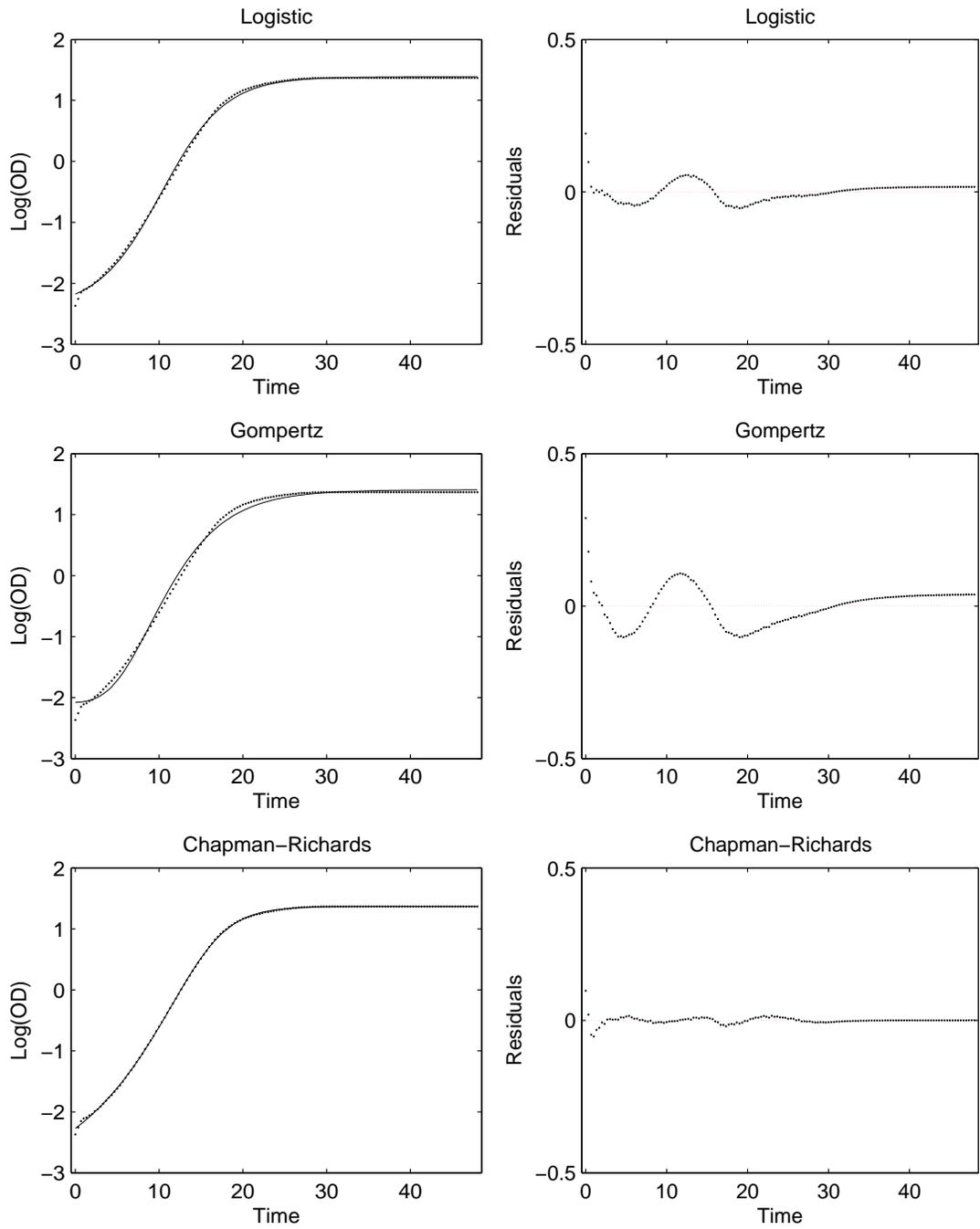


Figure 3.4: The logistic, Gompertz and Chapman-Richards models are fitted to the data NOD0326, well 3. The corresponding residual plots are on the right.

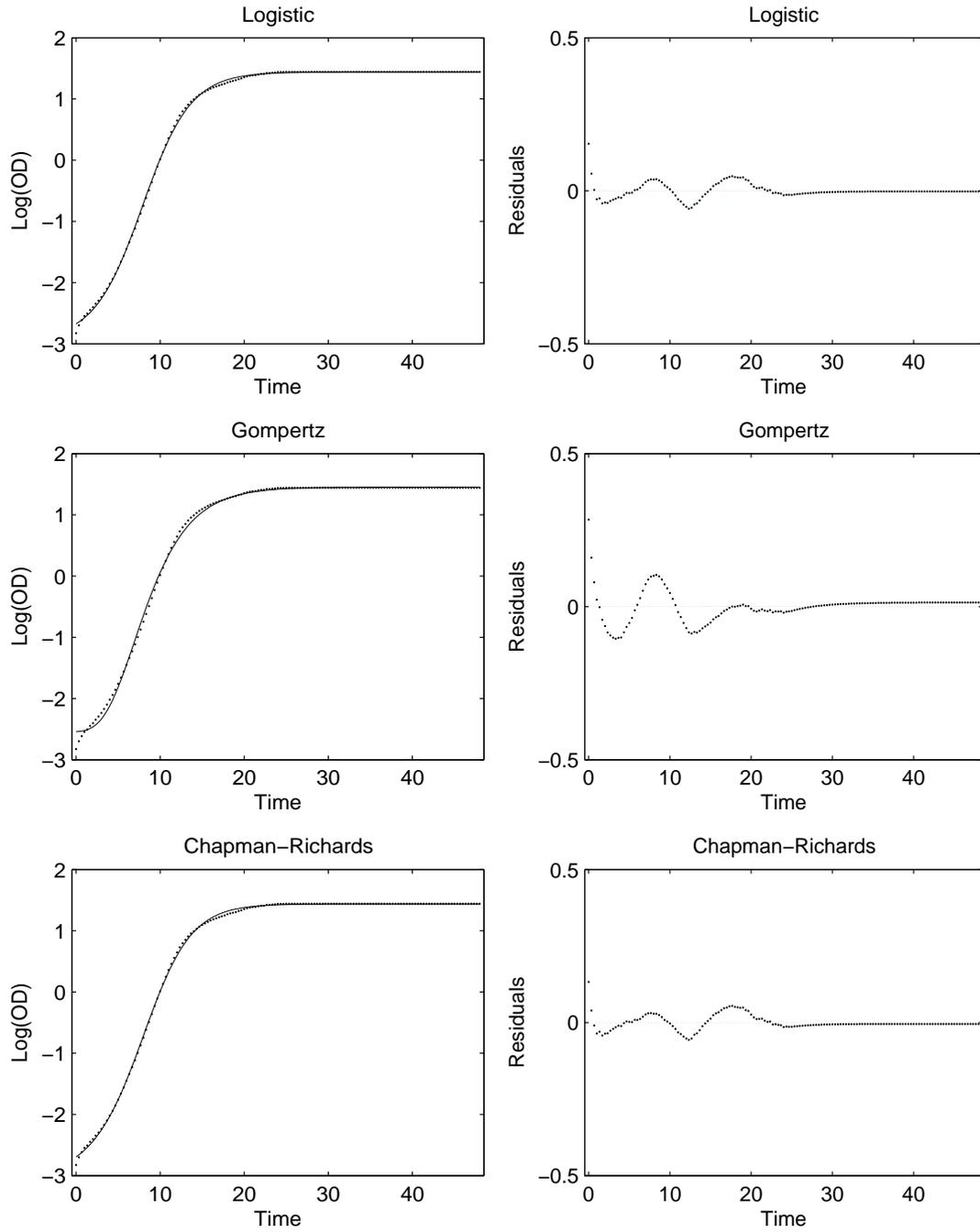


Figure 3.5: The logistic, Gompertz and Chapman-Richards models are fitted to the data NOD0406, well 1. The corresponding residual plots are on the right.

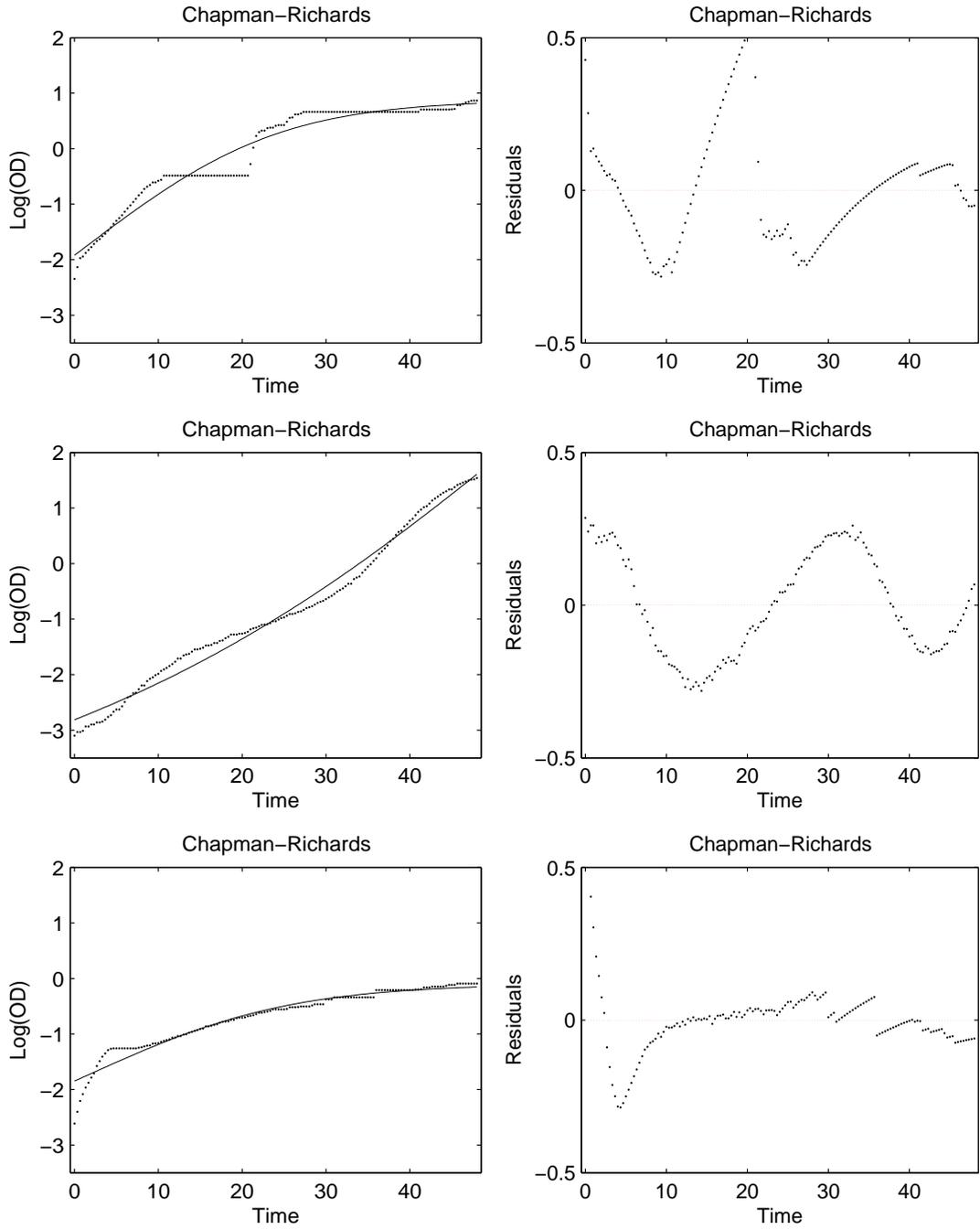


Figure 3.6: *Some atypical growth curves (starting from the top: 39C0309, well 62; MMC0408, well 6; 41E0314, well 20) and fitted Chapman-Richards models. The corresponding residual plots are on the right.*

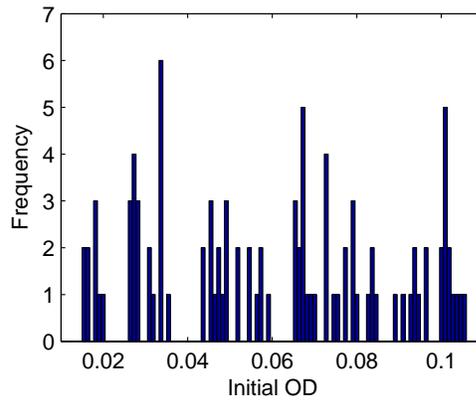


Figure 3.7: *The initial OD values of the 99 wild types in reference condition. The values are blank corrected and calibrated. The dataset comes from an experiment where the effect of the initial OD value was studied and thus the range of the initial OD values is wide in purpose.*

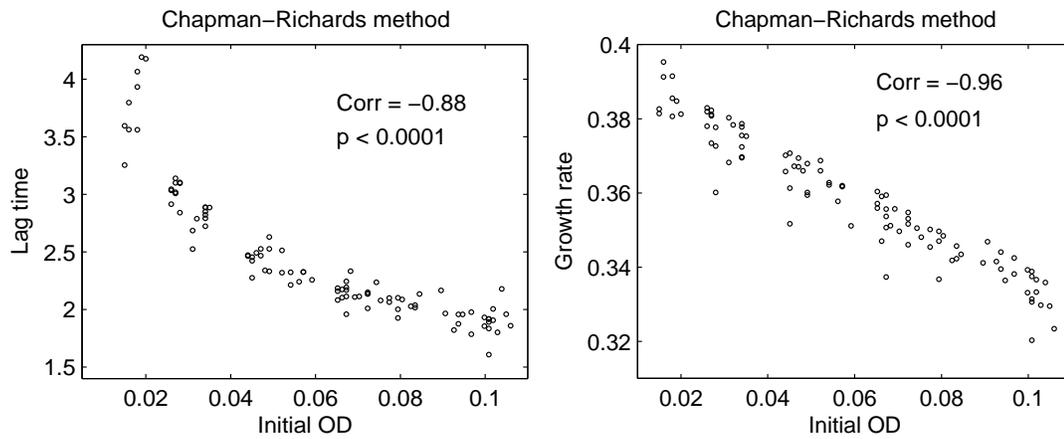


Figure 3.8: *The initial OD values of the 99 wild types in reference condition plotted against lag time and growth rate estimates from the Chapman-Richards method.*

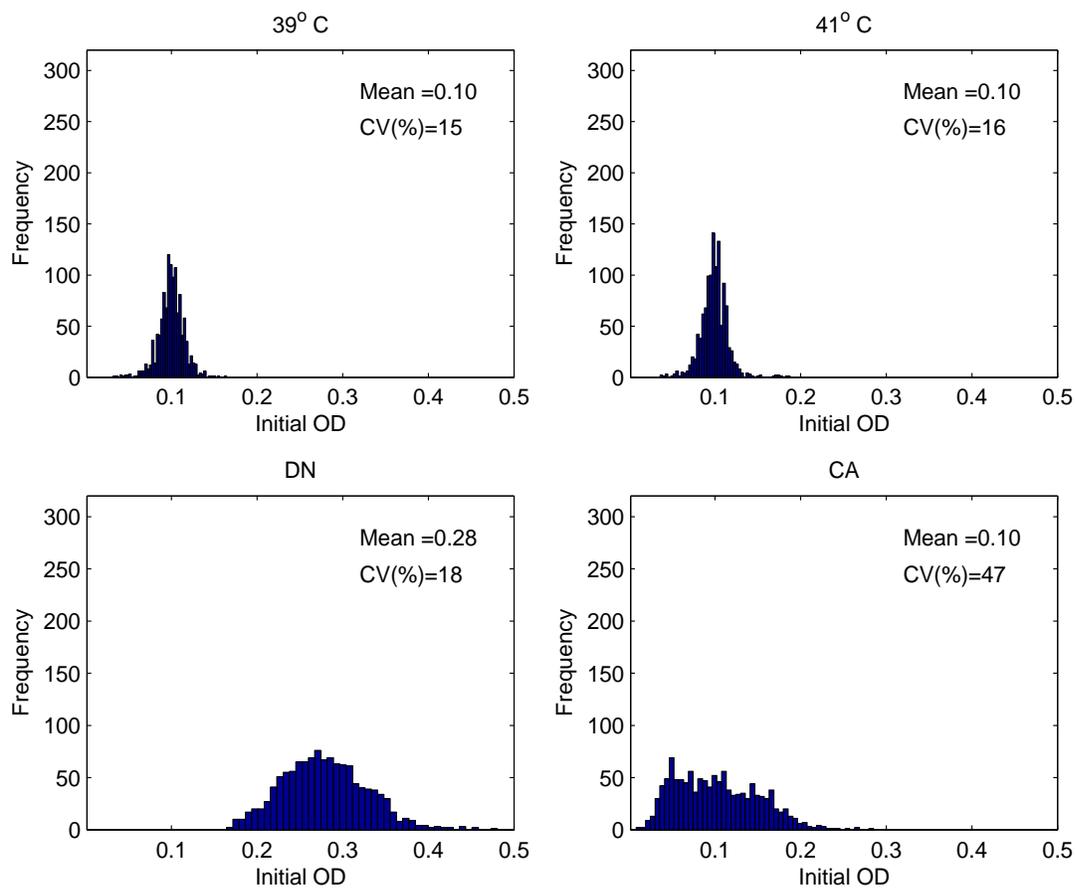


Figure 3.9: *The initial OD values of all mutants and wild types in each environment. The values are blank corrected and calibrated.*

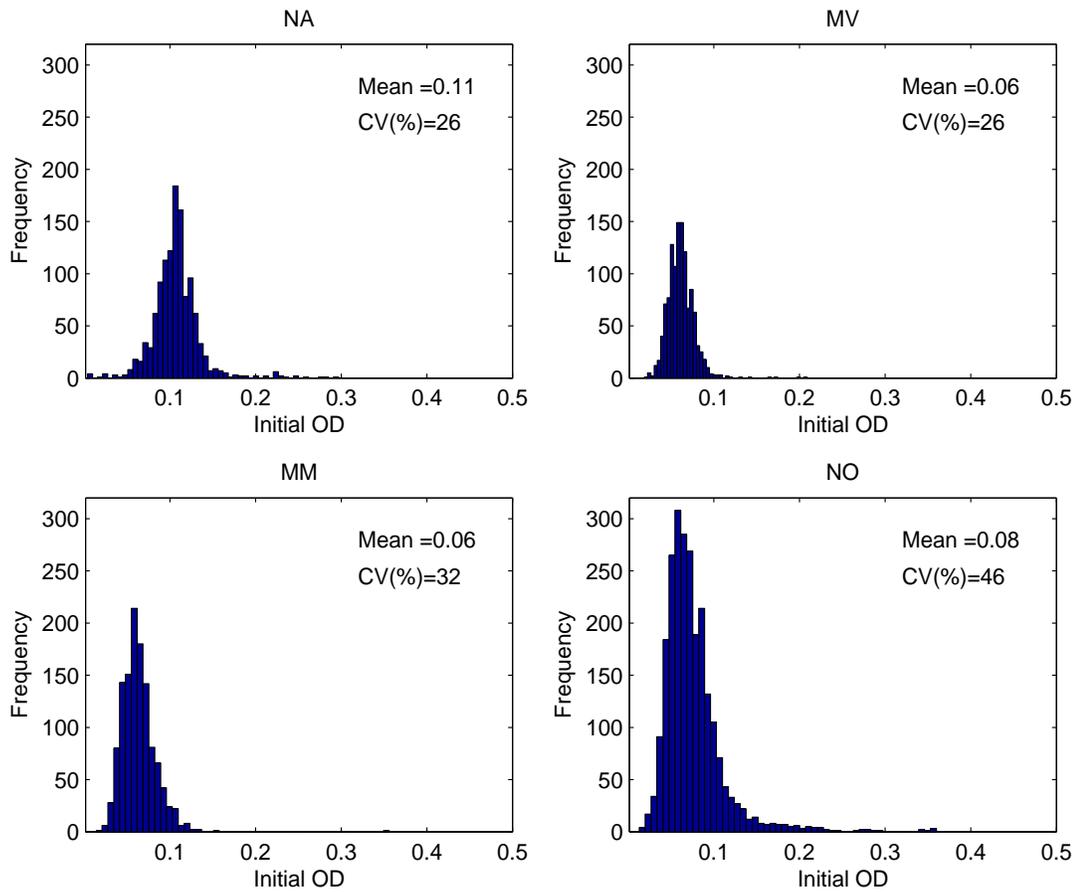


Figure 3.10: *The initial OD values of all mutants and wild types in each environment. The values are blank corrected and calibrated.*

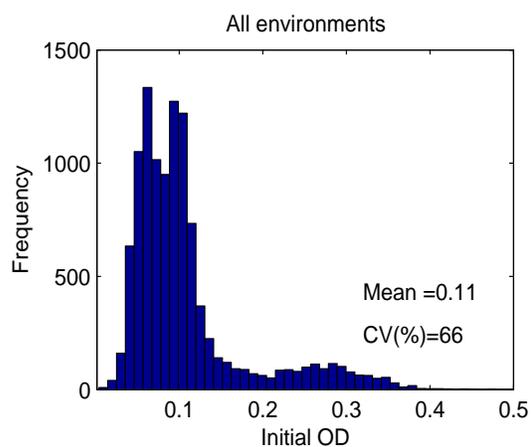


Figure 3.11: *The initial OD values of all mutants and wild types in all environments. The values are blank corrected and calibrated.*

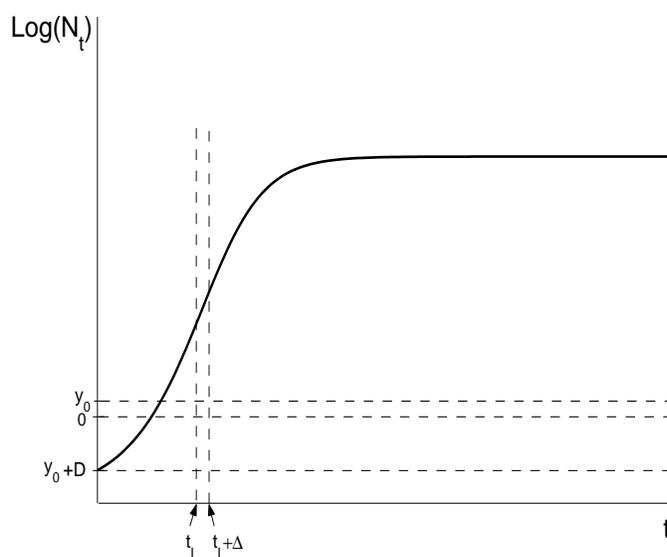


Figure 3.12: *An illustration of the three part model. Here N_t is the population size at time t , t_I is the inflection time point, Δ is the time span of the linear part, y_0 is given by (3.5), (3.6) or (3.7) (at $t = 0$), and $D = \log(N_0) - y_0$.*

Table 3.1: *The mean, minimum, maximum and coefficient of variation (%) of the initial OD values of all mutants and wild types in each run. The values are blank corrected and calibrated.*

| Run | Mean | Min | Max | CV (%) | Run | Mean | Min | Max | CV (%) |
|---------|------|------|------|--------|---------|------|------|------|--------|
| 39C0307 | 0.10 | 0.05 | 0.15 | 14 | MVC0413 | 0.06 | 0.03 | 0.21 | 34 |
| 39D0307 | 0.10 | 0.04 | 0.15 | 14 | MVD0413 | 0.07 | 0.02 | 0.11 | 22 |
| 39E0307 | 0.10 | 0.03 | 0.14 | 15 | MVE0413 | 0.07 | 0.04 | 0.14 | 22 |
| 39C0309 | 0.10 | 0.07 | 0.16 | 15 | MVC0417 | 0.06 | 0.03 | 0.09 | 21 |
| 39D0309 | 0.10 | 0.05 | 0.16 | 13 | MVD0417 | 0.06 | 0.02 | 0.10 | 23 |
| 39E0309 | 0.10 | 0.04 | 0.15 | 16 | MVE0417 | 0.06 | 0.03 | 0.08 | 19 |
| 41C0312 | 0.10 | 0.05 | 0.14 | 12 | MMC0408 | 0.07 | 0.03 | 0.36 | 42 |
| 41D0312 | 0.10 | 0.04 | 0.15 | 15 | MMD0408 | 0.08 | 0.03 | 0.15 | 30 |
| 41E0312 | 0.10 | 0.04 | 0.13 | 13 | MME0408 | 0.06 | 0.03 | 0.12 | 33 |
| 41C0314 | 0.10 | 0.06 | 0.19 | 14 | MMC0411 | 0.06 | 0.02 | 0.10 | 24 |
| 41D0314 | 0.10 | 0.05 | 0.15 | 21 | MMD0411 | 0.06 | 0.02 | 0.09 | 22 |
| 41E0314 | 0.09 | 0.04 | 0.12 | 14 | MME0411 | 0.06 | 0.03 | 0.11 | 20 |
| DNC0316 | 0.25 | 0.19 | 0.35 | 11 | NOC0305 | 0.08 | 0.04 | 0.16 | 23 |
| DND0316 | 0.30 | 0.22 | 0.44 | 12 | NOD0305 | 0.11 | 0.03 | 0.35 | 47 |
| DNE0316 | 0.29 | 0.19 | 0.48 | 13 | NOE0305 | 0.12 | 0.04 | 0.36 | 51 |
| DNC0319 | 0.22 | 0.16 | 0.31 | 13 | NOC0326 | 0.08 | 0.04 | 0.11 | 17 |
| DND0319 | 0.32 | 0.21 | 0.46 | 13 | NOD0326 | 0.08 | 0.04 | 0.20 | 28 |
| DNE0319 | 0.30 | 0.18 | 0.40 | 13 | NOE0326 | 0.09 | 0.04 | 0.23 | 26 |
| CAC0328 | 0.15 | 0.06 | 0.29 | 24 | NOC0406 | 0.06 | 0.03 | 0.08 | 20 |
| CAD0328 | 0.13 | 0.06 | 0.27 | 33 | NOD0406 | 0.06 | 0.03 | 0.09 | 23 |
| CAE0328 | 0.13 | 0.05 | 0.21 | 24 | NOE0406 | 0.06 | 0.03 | 0.09 | 18 |
| CAC0330 | 0.09 | 0.03 | 0.20 | 36 | NOC0426 | 0.07 | 0.01 | 0.15 | 48 |
| CAD0330 | 0.07 | 0.01 | 0.19 | 40 | NOD0426 | 0.07 | 0.02 | 0.17 | 40 |
| CAE0330 | 0.05 | 0.01 | 0.15 | 39 | NOE0426 | 0.06 | 0.02 | 0.18 | 40 |
| NAC0321 | 0.10 | 0.06 | 0.16 | 15 | | | | | |
| NAD0321 | 0.10 | 0.05 | 0.16 | 16 | | | | | |
| NAE0321 | 0.12 | 0.03 | 0.30 | 34 | | | | | |
| NAC0323 | 0.11 | 0.05 | 0.18 | 22 | | | | | |
| NAD0323 | 0.11 | 0.03 | 0.19 | 24 | | | | | |
| NAE0323 | 0.10 | 0.01 | 0.23 | 30 | | | | | |

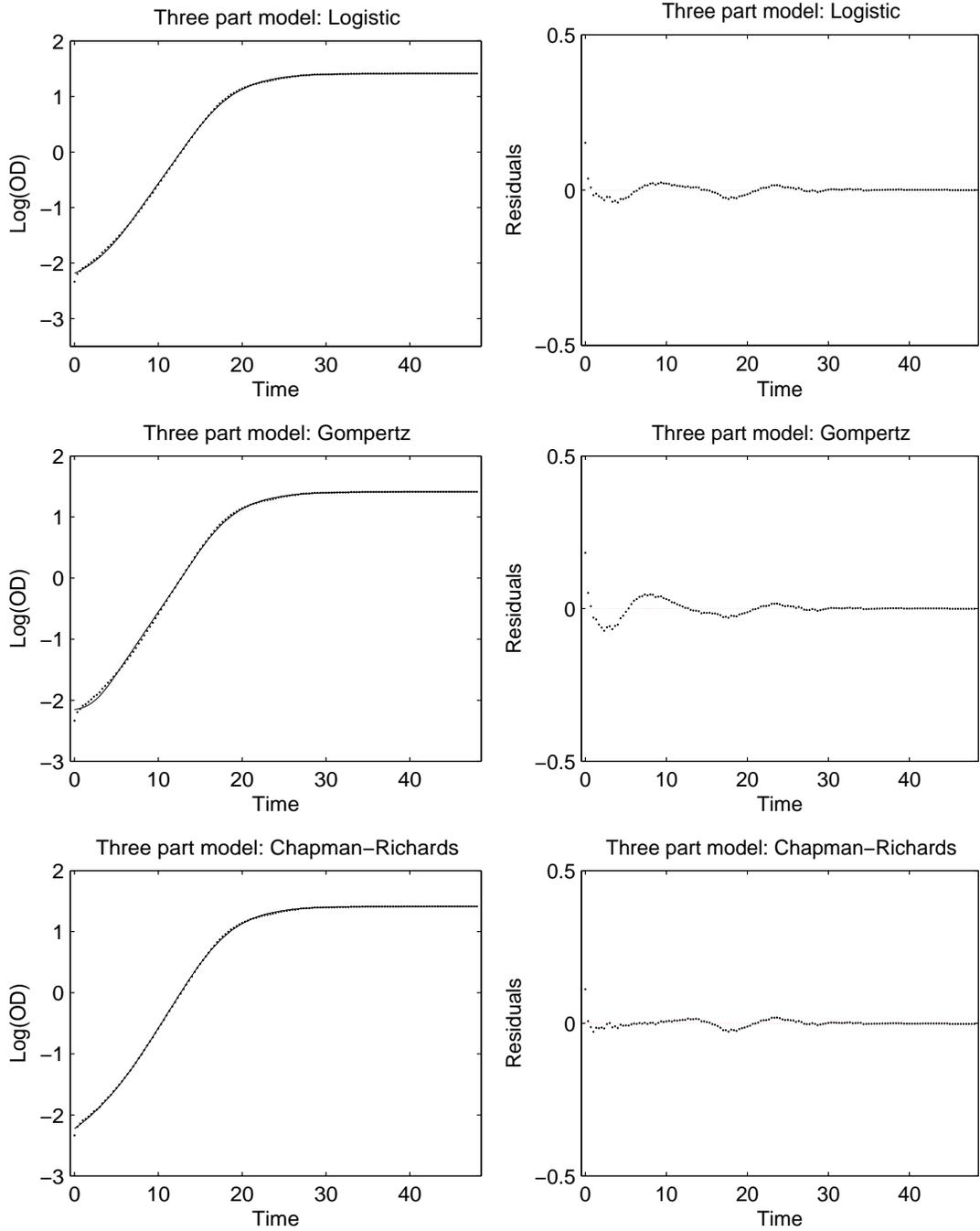


Figure 3.13: *The three part model with logistic, Gompertz and Chapman-Richards functions fitted to the data NOD0305, well 3. The estimates of Δ are 8.37 (Gompertz), 4.86 (Logistic) and 1.57 (Chapman-Richards).*

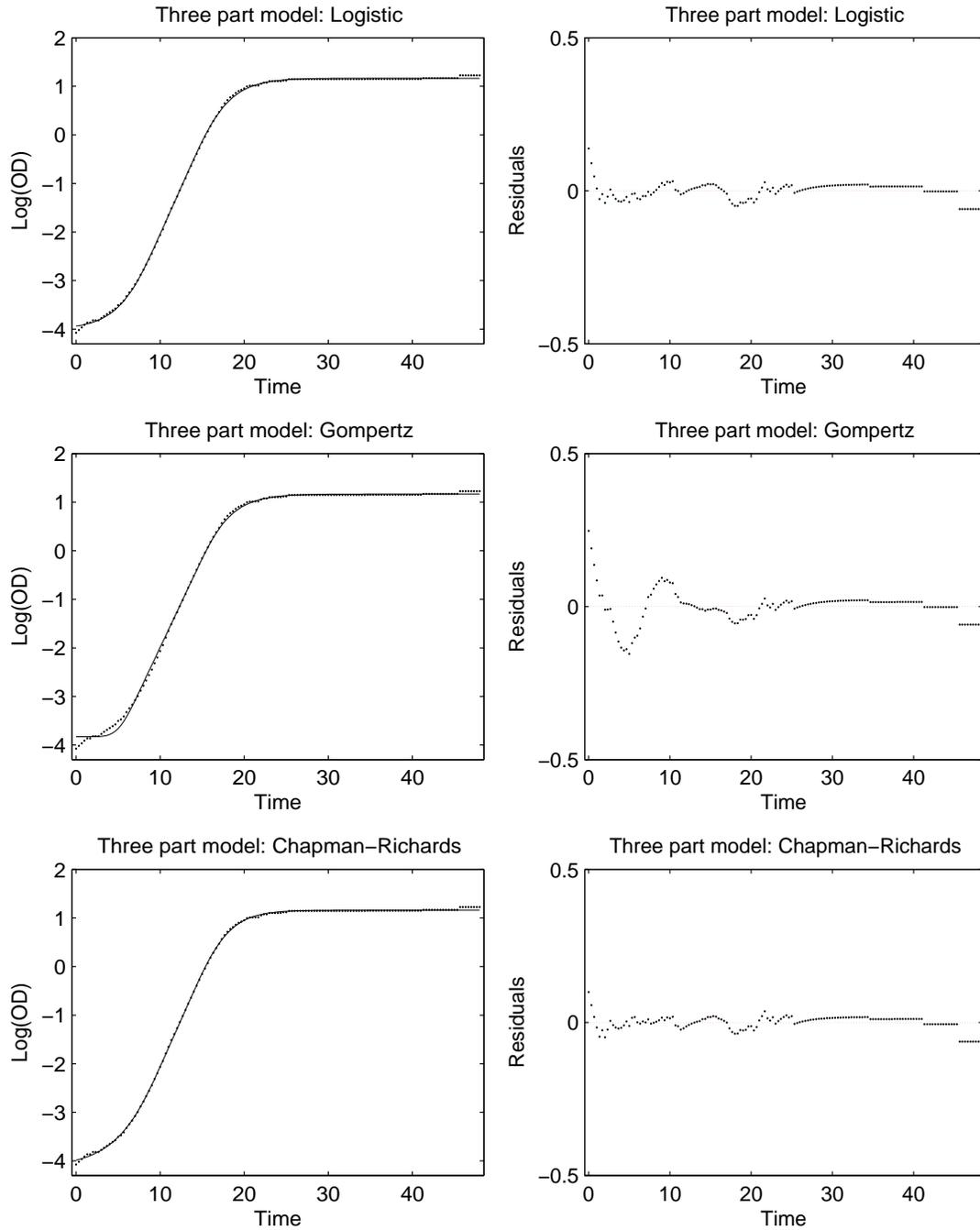


Figure 3.14: *The three part model with logistic, Gompertz and Chapman-Richards functions fitted to the data NOC0426, well 7. The estimates of Δ are 6.98 (Gompertz), 3.75 (Logistic) and 3.38 (Chapman-Richards).*

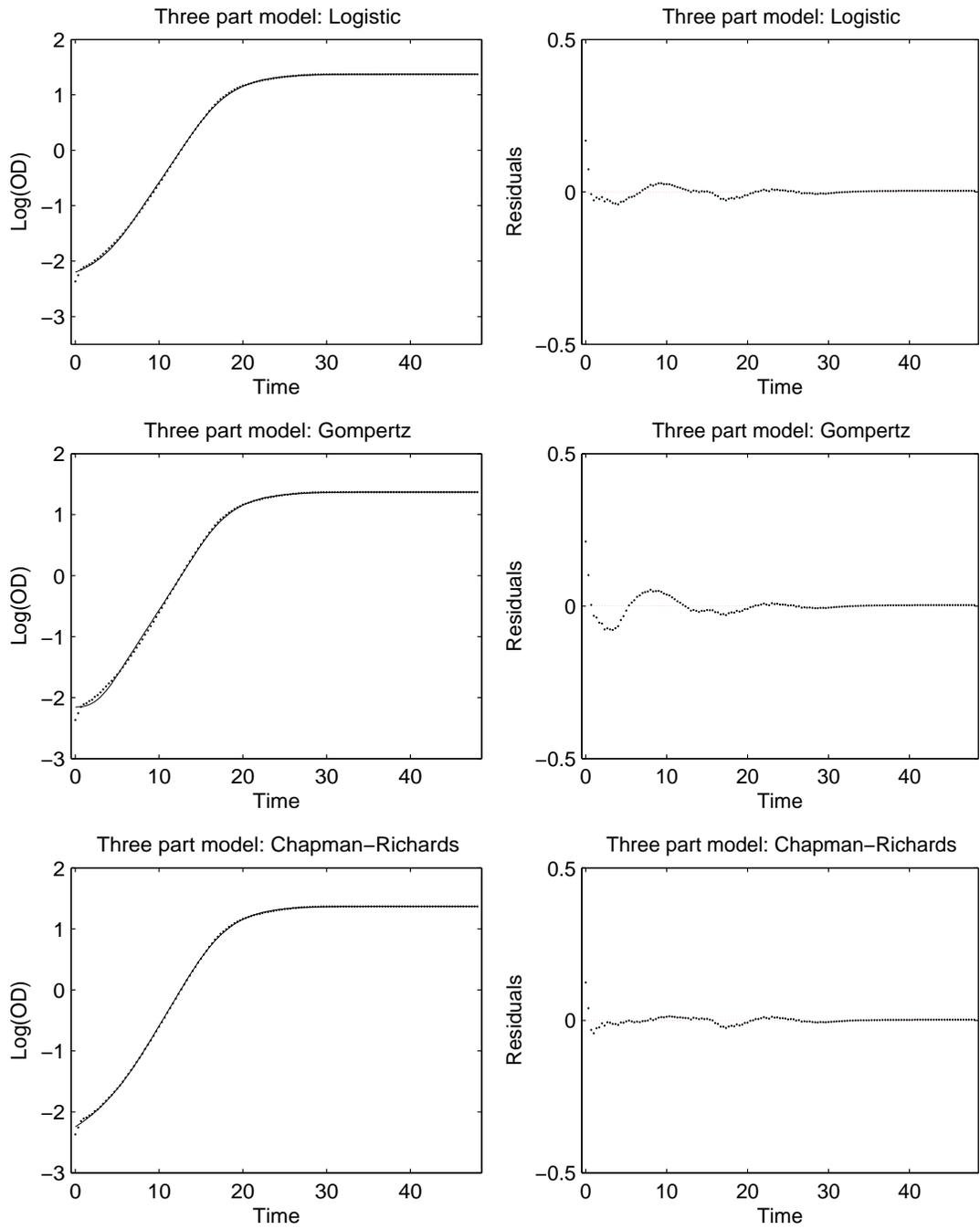


Figure 3.15: *The three part model with logistic, Gompertz and Chapman-Richards functions fitted to the data NOD0326, well 3. The estimates of Δ are 8.22 (Gompertz), 4.78 (Logistic) and 2.28 (Chapman-Richards).*

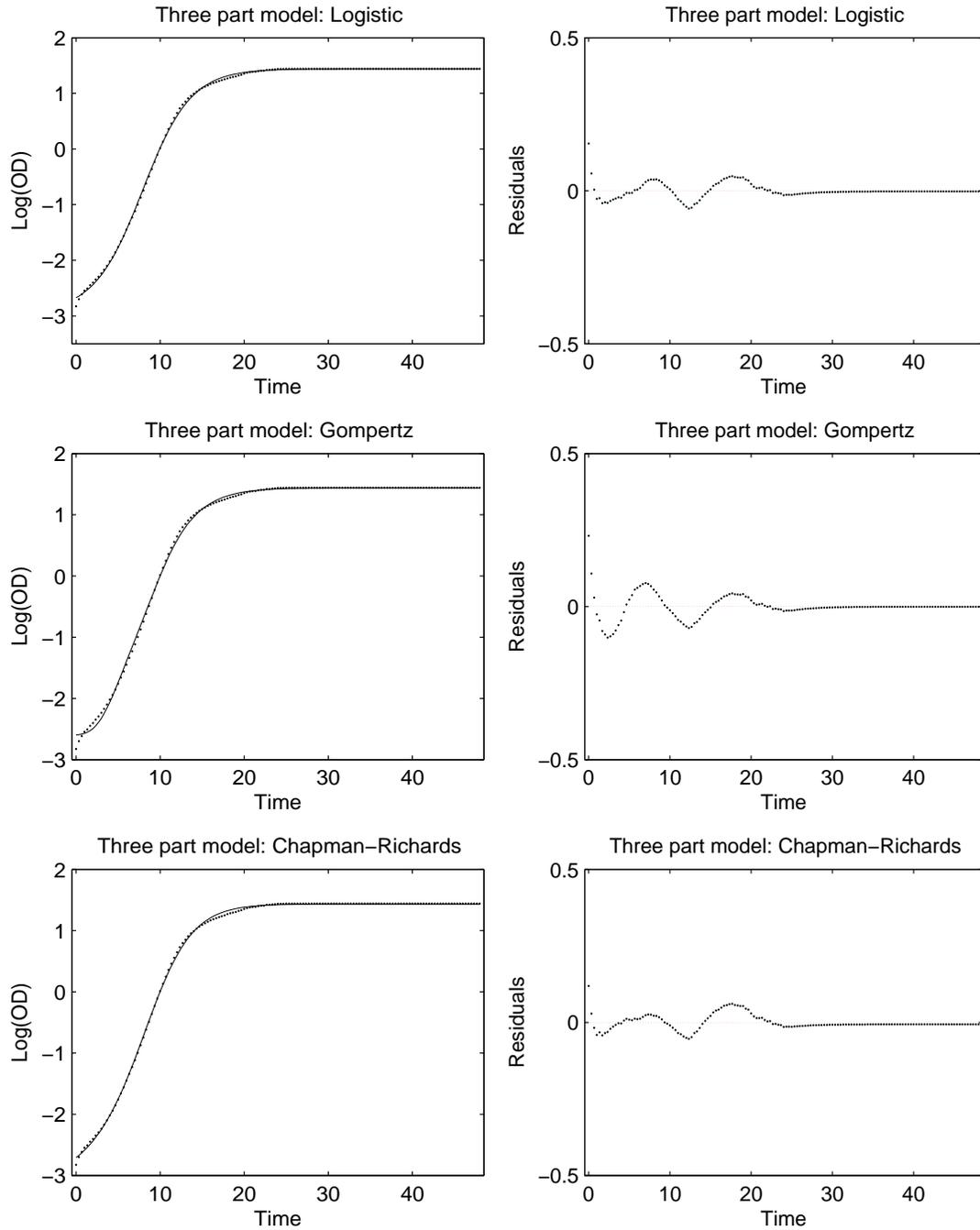


Figure 3.16: *The three part model with logistic, Gompertz and Chapman-Richards functions fitted to the data NOD0406, well 7. The estimates of Δ are 3.53 (Gompertz), 0 (Logistic) and 0 (Chapman-Richards).*

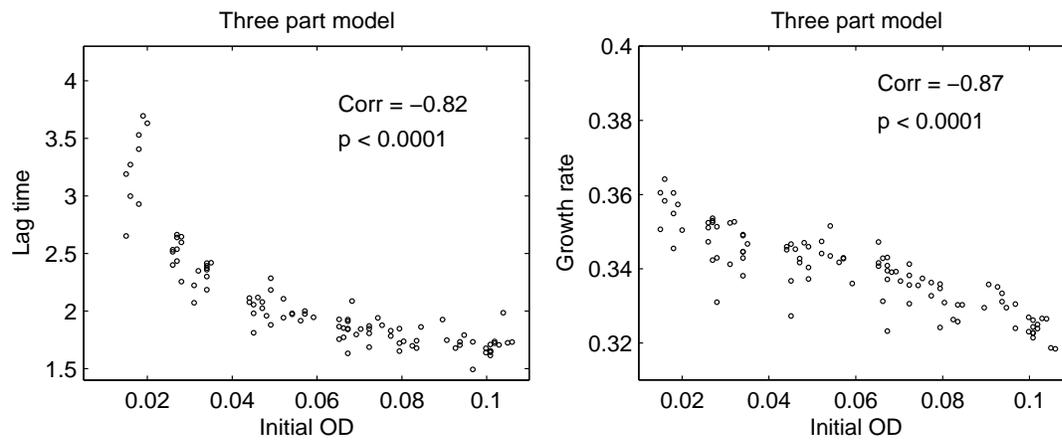


Figure 3.17: *The initial OD values of the 99 wild types in reference condition plotted against lag time and growth rate estimates from the three part model.*

Chapter 4

An alternative parameterization of the Chapman-Richards model

In this chapter we will see that the Chapman-Richards model growth curves presented in Chapter 3 can be expressed as a function of the initial population size s (on the non-logarithmic scale), the growth parameters λ, μ , and Y , and the derivative at time zero (on the logarithmic scale), denoted by d_0 . This last parameter is a natural complement to λ, μ , and Y in the phenotypic analysis of the mutants: if the fits of the models were perfect, d_0 would nicely reflect the initial adaptation behavior.

Although we cannot state the Chapman-Richards function explicitly in terms of the parameters s, d_0, λ, μ , and Y , it is still important to investigate the basic properties of this parameterization. It will, for example, be used in the construction of summary curves in Chapter 6.

Recall that the Chapman-Richards model is given by

$$g_t = \beta_0 \left[1 - \beta_1 e^{-\beta_2 t} \right]^{1/(1-\beta_3)} + D, \quad t \geq 0,$$

where either $\beta_0, \beta_2 > 0, 0 < \beta_3 < 1, 1 - \beta_3 < \beta_1 < 1$ or $\beta_0, \beta_2 > 0, \beta_3 > 1, \beta_1 < 1 - \beta_3$. The parameter D is always negative. The Chapman-Richards curves are not defined at $\beta_3 = 1$, but the limiting forms when β_3 tends to 1 and β_1 tends to 0 in a subordinated rate, are members of the Gompertz family.

The model we will study in this section is the Gompertz augmented Chapman-Richards model which is obtained from the above equation by writing $\beta_1 = e^b(1 - \beta_3)$, $b > 0$,

$$g_t = \beta_0 \left[1 - e^b(1 - \beta_3)e^{-\beta_2 t} \right]^{1/(1-\beta_3)} + D, \quad \text{for } \beta_3 \neq 1, \text{ and}$$

$$g_t = \beta_0 e^{-e^b - \beta_2 t} + D, \quad \text{for } \beta_3 = 1.$$

For more details, cf. Appendix C.

The parameterization properties of the model will be studied in Section 4.1. The parameter space (augmented with the parameters corresponding to the Gompertz curves) is also given explicitly. Section 4.2 investigates certain convexity properties of the parameter space.

4.1 Uniqueness

In this section we show that a hybrid parameterization (between the original and the new parameterization) with s , d_0 , λ , μ , and β_3 as parameters is unique. We then address the question of the uniqueness of the representation by the parameters s , d_0 , λ , μ , and Y . We formally check all but one of the steps of the proof. While no formal proof of the monotonicity of a certain implicit function stated in Conjecture 1 (on page 45) is available, we show through an extensive numerical investigation that the conjecture is likely to hold. Theorem 1 is the main result of the section.

We will need the following basic property of the Chapman-Richards model:

Proposition 1 *The $(\beta_0, \beta_1, \beta_2, \beta_3, D)$ -parameterization is unique.*

Proof. This uniqueness is probably well-known, but for completeness we give a proof in Appendix E.

The parameters of the new parameterization (in the case $\beta_3 \neq 1$) can be written as

$$s = e^{\beta_0(1-\beta_1)^{\frac{1}{1-\beta_3}+D}} \quad (4.1)$$

$$d_0 = \frac{\beta_0\beta_1\beta_2(1-\beta_1)^{\frac{\beta_3}{1-\beta_3}}}{1-\beta_3} \quad (4.2)$$

$$\lambda = \frac{(1-\beta_1)^{\frac{1}{1-\beta_3}} - \beta_3^{\frac{1}{1-\beta_3}} + \beta_3^{\frac{\beta_3}{1-\beta_3}} \log\left(\frac{\beta_1}{1-\beta_3}\right)}{\beta_2\beta_3^{\frac{\beta_3}{1-\beta_3}}} \quad (4.3)$$

$$\mu = \beta_0\beta_2\beta_3^{\frac{\beta_3}{1-\beta_3}} \quad (4.4)$$

$$\begin{aligned} Y &= e^{\beta_0+D} - e^{\beta_0(1-\beta_1)^{\frac{1}{1-\beta_3}+D}} \\ &= e^{\beta_0+D} - s, \\ &= e^A - s, \end{aligned} \quad (4.5)$$

where $A = \beta_0 + D$ is the asymptote of the curve (on the logarithmic scale).

We start with a lemma concerning the Gompertz model in the hybrid parameterization:

Lemma 1 *The Gompertz curve corresponding to any hybrid parameter combination $s > 0$, $0 < d_0 < \mu$, $\lambda > 0$, $\mu > 0$, and $\beta_3 = 1$ is unique. The parameter b is the solution of the equation $b + 1 - e^b = \log(\frac{d_0}{\mu})$, and the three other parameters are given by $\beta_0 = \frac{\mu\lambda}{\frac{b}{e} + e^{-e^b} - \frac{1}{e}}$, $\beta_2 = \frac{b + e^{-e^b + 1} - 1}{\lambda}$, and $D = \log(s) - \frac{\lambda\mu e^{-e^b}}{\frac{b}{e} + e^{-e^b} - \frac{1}{e}}$. Furthermore, the stationary phase OD increment is*

$$Y = e^{\left(\frac{1 - e^{-e^b}}{\frac{b-1}{e} + e^{-e^b}}\right)\lambda\mu + \log(s)} - s.$$

Proof. See Appendix C.

We will next state a series of technical lemmas and propositions formulated for a special Chapman-Richards sub-model, restricted by the assumptions $s = 1$, $0 < d_0 < 1$, $\lambda = 1$, $\mu = 1$. We will refer to this as the *unit-scaled model*.

In the unit-scaled model, the equations (4.1-4.4) are equivalent to the equations (4.6-4.9) below

$$d_0 = \frac{\beta_1(1 - \beta_1)^{\frac{\beta_3}{1 - \beta_3}}}{(1 - \beta_3)\beta_3^{\frac{\beta_3}{1 - \beta_3}}}, \quad (4.6)$$

$$\beta_0 = \frac{1}{\beta_3^{\frac{\beta_3}{1 - \beta_3}} \left[\log\left(\frac{\beta_1}{1 - \beta_3}\right) - \beta_3 \right] + (1 - \beta_1)^{\frac{1}{1 - \beta_3}}}, \quad (4.7)$$

$$\beta_2 = \frac{1}{\beta_0\beta_3^{\frac{\beta_3}{1 - \beta_3}}}, \quad (4.8)$$

$$D = -\beta_0(1 - \beta_1)^{\frac{1}{1 - \beta_3}}. \quad (4.9)$$

Recall that $\beta_0, \beta_2 > 0$, $0 < \beta_3 < 1$, $1 - \beta_3 < \beta_1 < 1$ or $\beta_0, \beta_2 > 0$, $\beta_3 > 1$, $\beta_1 < 1 - \beta_3$. The asymptote $A = \beta_0 + D$ can be written as

$$A = \frac{1 - (1 - \beta_1)^{\frac{1}{1 - \beta_3}}}{\beta_3^{\frac{\beta_3}{1 - \beta_3}} \left[\log\left(\frac{\beta_1}{1 - \beta_3}\right) - \beta_3 \right] + (1 - \beta_1)^{\frac{1}{1 - \beta_3}}}. \quad (4.10)$$

The following lemma addresses the hybrid parameterization in the unit-scaled case:

Lemma 2 *The (d_0, β_3) -parameterization is unique in the unit-scaled model, i.e. there is exactly one curve in the Chapman-Richards (Gompertz augmented) model for each combination of $0 < d_0 < 1$ and $\beta_3 > 0$, and $s = 1$, $\lambda = 1$, $\mu = 1$.*

Proof. The parameterization is unique if the equation (4.6) has at most one solution $\beta_1 < 1 - \beta_3$ for fixed $\beta_3 > 1$, or $1 - \beta_3 < \beta_1 < 1$ for fixed β_3 such that $0 < \beta_3 < 1$. Rewrite the equation (4.6) as

$$f(\beta_1) := \beta_1(1 - \beta_1)^{\frac{\beta_3}{1-\beta_3}} - d_0(1 - \beta_3)\beta_3^{\frac{\beta_3}{1-\beta_3}} = 0.$$

Differentiate f with respect to β_1 to obtain

$$\begin{aligned} f'(\beta_1) &= (1 - \beta_1)^{\frac{\beta_3}{1-\beta_3}} - \frac{\beta_1\beta_3}{(1 - \beta_3)} \frac{(1 - \beta_1)^{\frac{\beta_3}{1-\beta_3}}}{(1 - \beta_1)} \\ &= (1 - \beta_1)^{\frac{\beta_3}{1-\beta_3}} \left(1 - \frac{\beta_1\beta_3}{(1 - \beta_3)(1 - \beta_1)} \right). \end{aligned}$$

For $\beta_3 > 1$ and $\beta_1 < 1 - \beta_3$,

$$\begin{aligned} f'(\beta_1) &= (1 - \beta_1)^{\frac{\beta_3}{1-\beta_3}} \left(1 - \frac{\beta_1\beta_3}{(1 - \beta_3)(1 - \beta_1)} \right) \\ &> (1 - \beta_1)^{\frac{\beta_3}{1-\beta_3}} \left(1 - \frac{(1 - \beta_3)(1 - \beta_1)}{(1 - \beta_3)(1 - \beta_1)} \right) = 0, \end{aligned}$$

and for $0 < \beta_3 < 1$ and $1 - \beta_3 < \beta_1 < 1$,

$$\begin{aligned} f'(\beta_1) &= (1 - \beta_1)^{\frac{\beta_3}{1-\beta_3}} \left(1 - \frac{\beta_1\beta_3}{(1 - \beta_3)(1 - \beta_1)} \right) \\ &< (1 - \beta_1)^{\frac{\beta_3}{1-\beta_3}} \left(1 - \frac{(1 - \beta_3)\beta_3}{(1 - \beta_3)\beta_3} \right) = 0. \end{aligned}$$

The above monotonicity properties, the continuity and appropriate sign changes of f in the allowed β_1 intervals, prove the required existence and uniqueness of β_1 in both cases $\beta_3 < 1$ and $\beta_3 > 1$. (The uniqueness of the Gompertz curve when $\beta_3 = 1$ follows directly from Lemma 1). \square

It is time for a second result about the Gompertz augmented model:

Lemma 3 *Fix $0 < d_0 < 1$. In the unit-scaled model, the function A defined in (4.10) with the constraint (4.6) is continuous at $\beta_3 = 1$ as a function of $\beta_3 > 0$.*

Proof. See Appendix C.

The following lemma is used in the proof of Lemma 5(b).

Lemma 4 *Fix $\beta_3 > 0$. In the unit-scaled model, the function A defined in (4.10) with the constraint (4.6) is strictly increasing as a function of d_0 , $0 < d_0 < 1$.*

Proof. See Appendix E.

Lemma 5 *Fix $0 < d_0 < 1$. In the unit-scaled model, the function A defined in (4.10) with the constraint (4.6) satisfies*

$$(a) \lim_{\beta_3 \rightarrow 0} A = \infty$$

$$(b) \lim_{\beta_3 \rightarrow \infty} A = \frac{1-d_0}{d_0 - \log(d_0) - 1}.$$

Proof. See Appendix E.

Now, we are prepared to discuss the main alternative parameterization. As indicated earlier, we need the following monotonicity assumption:

Conjecture 1 *Fix $0 < d_0 < 1$. In the unit-scaled model, the function A defined in (4.10) with the constraint (4.6) is strictly decreasing as a function of $\beta_3 > 0$.*

Note that the conjecture is purely technical. Recall that the following restrictions also apply: $1 - \beta_3 < \beta_1 < 1$ for $0 < \beta_3 < 1$ or $\beta_1 < 1 - \beta_3$ for $\beta_3 > 1$. The conjecture is further discussed and numerically motivated in Appendix D.

In the sequel we assume that Conjecture 1 holds.

Proposition 2 *Provided that Conjecture 1 holds, the (d_0, A) -parameterization is unique in the unit-scaled model, and the Chapman-Richards (Gompertz augmented) curves exist if and only if $A > \frac{1-d_0}{d_0 - \log(d_0) - 1}$.*

Proof. Consider a model curve from the hybrid parameterization with $s = 1$, $\lambda = 1$, $\mu = 1$, and $0 < d_0 < 1$ and β_3 fixed. The asymptote of this curve is given by (4.10), where β_1 solves (4.6). Now consider A a function of β_3 . This function is obviously continuous at any $\beta_3 \neq 1$ and Lemma 3 states that it is also continuous at $\beta_3 = 1$. Conjecture 1 states that A is strictly decreasing and hence (4.10) has at most one solution β_3 , for β_1 and A fixed. Combining this with the two limits in Lemma 5 finally completes the proof.

Proposition 3 *Provided that Conjecture 1 holds, the (d_0, Y) -parameterization is unique in the unit-scaled model, and the Chapman-Richards (Gompertz augmented) curves exist only for $Y > e^{\frac{1-d_0}{d_0 - \log(d_0) - 1}} - 1$.*

Proof. Follows from Proposition 2 and equation (4.5).

For any Chapman-Richards (Gompertz augmented) model curve we can arbitrarily time scale, scale and translate the log-size dimension, and the resulting curve is still a Chapman-Richards (Gompertz augmented) model curve. This model invariance together with Lemma 2 and Proposition 3 will be used to prove:

Theorem 1 *Consider the Chapman-Richards Gompertz augmented model.*

(a) *Any model curve is uniquely determined by the parameters s , d_0 , λ , μ , and β_3 . The parameters are constrained by the inequalities $s > 0$, $0 < d_0 < \mu$, $\lambda > 0$, $\mu > 0$, and $\beta_3 > 0$. Curves with $\beta_3 = 1$ correspond to the Gompertz curves.*

(b) *Provided that Conjecture 1 holds, any model curve is uniquely determined by the parameters s , d_0 , λ , μ , and Y . The parameters are constrained by the inequalities $s > 0$, $0 < d_0 < \mu$, $\lambda > 0$, $\mu > 0$, and*

$$Y > \underline{Y}, \quad \underline{Y} := e^{\left(\frac{1 - \frac{d_0}{\mu}}{\frac{d_0}{\mu} - \log\left(\frac{d_0}{\mu}\right) - 1}\right)\lambda\mu + \log(s)} - s.$$

(c) *The unique Gompertz curve for each allowed parameter combination s , d_0 , λ , and μ corresponds to the stationary phase OD increment parameter*

$$Y = e^{\left(\frac{1 - e^{-e^b}}{\frac{b-1}{e} + e^{-e^b}}\right)\lambda\mu + \log(s)} - s,$$

where b is the solution of the equation

$$b + 1 - e^b = \log\left(\frac{d_0}{\mu}\right).$$

Proof.

(a) Take a Chapman-Richards model growth curve

$$g_t(\beta_0, \beta_1, \beta_2, \beta_3, D) = \beta_0 \left[1 - \beta_1 e^{-\beta_2 t}\right]^{1/(1-\beta_3)} + D$$

and transform it by multiplying t by some constant $c > 0$, by multiplying the whole curve by some constant $k > 0$, and by moving the curve (upwards or downwards) by some constant m . Then,

$$\begin{aligned} kg_{ct}(\beta_0, \beta_1, \beta_2, \beta_3, D) + m &= k \left(\beta_0 \left[1 - \beta_1 e^{-\beta_2 ct}\right]^{1/(1-\beta_3)} + D \right) + m \quad (4.11) \\ &= k\beta_0 \left[1 - \beta_1 e^{-\beta_2 ct}\right]^{1/(1-\beta_3)} + kD + m \\ &= g_t(k\beta_0, \beta_1, \beta_2 c, \beta_3, kD + m), \end{aligned}$$

so that the result is still a Chapman-Richards model curve.

Take a growth curve g_t with fixed $(\beta_0, \beta_1, \beta_2, \beta_3, D)$ corresponding to the hybrid parameters $(s, d_0, \lambda, \mu, \beta_3)$. Then using (4.11) with $k = \frac{1}{\lambda\mu}$, $c = \lambda$, $m = \frac{-\log(s)}{\lambda\mu}$, we get

$$\frac{1}{\lambda\mu}g_{t\lambda}(\beta_0, \beta_1, \beta_2, \beta_3, D) - \frac{\log(s)}{\lambda\mu} = \hat{g}_t(1, \frac{d_0}{\mu}, 1, 1, \beta_3), \quad (4.12)$$

where the \hat{g}_t refers to the unique curve with hybrid parameters known to exist in this case by Lemma 2. Inverting the relation (4.12) gives that

$$g_t(\beta_0, \beta_1, \beta_2, \beta_3, D) = \lambda\mu\hat{g}_t(1, \frac{d_0}{\mu}, 1, 1, \beta_3) + \log(s),$$

for any $s > 0$, $0 < d_0 < \mu$, $\lambda > 0$, $\mu > 0$, and $\beta_3 > 0$, so that g_t must also be uniquely determined by s , d_0 , λ , μ , and β_3 . Starting with an arbitrary combination of $s > 0$, $0 < d_0 < \mu$, $\lambda > 0$, $\mu > 0$, and $\beta_3 > 0$,

$$\lambda\mu\hat{g}_t(1, \frac{d_0}{\mu}, 1, 1, \beta_3) + \log(s)$$

is always a model curve with parameters s , d_0 , λ , μ , and β_3 . This motivates the parameter space restrictions.

(b) We may show the statements by showing that for s, d_0, λ , and μ fixed, β_3 is determined by Y , if Conjecture 1 holds. Using (4.12) we obtain

$$A(1, \frac{d_0}{\mu}, 1, 1, \beta_3) = \frac{A(s, d_0, \lambda, \mu, \beta_3)}{\lambda\mu} - \frac{\log(s)}{\lambda\mu},$$

and hence

$$Y(1, \frac{d_0}{\mu}, 1, 1, \beta_3) = \left(\frac{Y(s, d_0, \lambda, \mu, \beta_3)}{s} + 1 \right)^{\frac{1}{\lambda\mu}} - 1. \quad (4.13)$$

Now, suppose that several β_3 -choices yielded the same Y on the right side of (4.13). Then the same β_3 -choices would result in the same Y also on the left side, which would contradict Proposition 3, if the Conjecture 1 was true. Finally, by inverting (4.13), we also get the parameter space constraints from the restriction of Y in Proposition 3.

(c) Follows directly from Lemma 1. \square

4.2 Convexity properties

In this section we work under the assumption that Conjecture 1 is true and use the parameterization with s, d_0, λ, μ , and Y . The results in this section will be needed in Chapter 6 where two methods for constructing summary curves are discussed. More concretely, the existence of summary curves is equivalent to the convexity of the new parameter space (for all fixed s) or that of its logarithmic version (for all fixed $\log(s)$). As discussed in detail in Chapter 6, the method I summary curves do not always exist whereas the method II summary curves always exist.

Recall the notation \underline{Y} for the lower bound of Y (stated in Theorem 1),

$$\underline{Y}(s, d_0, \lambda, \mu) = s \left[e^{\left(\frac{1 - \frac{d_0}{\mu}}{\frac{d_0}{\mu} - \log\left(\frac{d_0}{\mu}\right) - 1} \right) \lambda \mu} - 1 \right].$$

The key to the proof is the following lemma:

Lemma 6 *The $\log(\underline{Y})$ is convex as a function of $\log(d_0)$, $\log(\lambda)$, and $\log(\mu)$ for any fixed $\log(s)$, where $s > 0$, $0 < d_0 < \mu$, $\lambda > 0$, and $\mu > 0$.*

Proof. The proof is rather technical and we have therefore chosen to give it in Appendix E.

Theorem 2 *Consider the parameter space $(s, d_0, \lambda, \mu, Y)$, where $s > 0$ is fixed, $0 < d_0 < \mu$, $\lambda > 0$, $\mu > 0$, and $Y > \underline{Y}$. Then the following holds*

- (a) *The parameter space is not convex for any fixed s .*
- (b) *The component-wise logarithmic version of the parameter space is convex for all fixed $\log(s)$.*

Proof.

- (a) Fix two parameter combinations with the same s :

$$(s, d_{0(1)}, \lambda_{(1)}, \mu_{(1)}, Y_{(1)}) \text{ and } (s, d_{0(2)}, \lambda_{(2)}, \mu_{(2)}, Y_{(2)}).$$

Take a convex combination of the parameters and let

$$d_{0\theta} = \theta d_{0(1)} + (1 - \theta)d_{0(2)},$$

$$\lambda_\theta = \theta \lambda_{(1)} + (1 - \theta)\lambda_{(2)},$$

$$\mu_\theta = \theta \mu_{(1)} + (1 - \theta)\mu_{(2)},$$

$$Y_\theta = \theta Y_{(1)} + (1 - \theta)Y_{(2)},$$

for some $0 < \theta < 1$. The parameter space is convex if and only if $\lambda_\theta > 0$, $\mu_\theta > 0$, $0 < d_{0\theta} < \mu_\theta$, and $Y_\theta > \underline{Y}_\theta(s, d_{0\theta}, \lambda_\theta, \mu_\theta)$. Since it is obvious that $\lambda_\theta > 0$, $\mu_\theta > 0$, $0 < d_{0\theta} < \mu_\theta$, the convexity of the parameter space is equivalent to proving that

$$Y_\theta > \underline{Y}_\theta(s, d_{0\theta}, \lambda_\theta, \mu_\theta). \quad (4.14)$$

We next construct a set of parameters for which the previous inequality is violated. For $s = 1$, take $\theta = 0.5$, $d_{0(1)} = 0.1$, $d_{0(2)} = 0.0001$, $\lambda_{(1)} = \lambda_{(2)} = 1$, $\mu_{(1)} = \mu_{(2)} = 1$, $Y_{(1)} = 0.91$ ($> \underline{Y}_{(1)} \approx 0.8997$), and $Y_{(2)} = 0.25$ ($> \underline{Y}_{(2)} \approx 0.1295$). Then

$$Y_\theta = \frac{0.25 + 0.91}{2} = 0.58$$

and

$$\underline{Y}_\theta(1, \frac{0.1 + 0.0001}{2}, 1, 1) \approx 0.5913,$$

which contradicts (4.14). For an arbitrary fixed s , multiply $Y_{(1)}$ and $Y_{(2)}$ by s and leave the other parameters unchanged.

(b) Fix again two parameter combinations with the same s :

$$(s, d_{0(1)}, \lambda_{(1)}, \mu_{(1)}, Y_{(1)}) \text{ and } (s, d_{0(2)}, \lambda_{(2)}, \mu_{(2)}, Y_{(2)}).$$

Take a convex combination of the parameters on the logarithmic level and denote the corresponding non-logarithmic parameters by

$$d_{0\theta} = d_{0(1)}^\theta d_{0(2)}^{1-\theta},$$

$$\lambda_\theta = \lambda_{(1)}^\theta \lambda_{(2)}^{1-\theta},$$

$$\mu_\theta = \mu_{(1)}^\theta \mu_{(2)}^{1-\theta},$$

$$Y_\theta = Y_{(1)}^\theta Y_{(2)}^{1-\theta},$$

for some $0 < \theta < 1$. Since $\lambda_\theta > 0$, $\mu_\theta > 0$, $0 < d_{0\theta} < \mu_\theta$, proving that

$$\begin{aligned} \log(Y_\theta) &= \theta \log(Y_{(1)}) + (1 - \theta) \log(Y_{(2)}) \\ &> \log [\underline{Y}_\theta (\log(s), \log(d_{0\theta}), \log(\lambda_\theta), \log(\mu_\theta))], \end{aligned} \tag{4.15}$$

will imply the convexity of the component-wise logarithmic version of the parameter space for all fixed $\log(s)$. The inequality (4.15) follows from Lemma 6 and the observation that

$$\log(Y_{(i)}) > \log [\underline{Y} (\log(s), \log(d_{0(i)}), \log(\lambda_{(i)}), \log(\mu_{(i)}))], \quad i = 1, 2.$$

□

Chapter 5

Standardizing curves

As was seen in Section 3.3, the lag time and growth rate depend strongly on the initial OD. However, in large-scale experiments with analysis of hundreds of mutants, it is hard to keep the initial OD constant between different experiments. Hence, it would be desirable to reduce the correlation and to make the curves more easily comparable by developing a method for standardizing growth curves with respect to the initial OD. Our approach has as a starting point the simultaneous fitting of a three part model curve (introduced in Chapter 3) in [12] (for more details, cf. Appendix F). In the optimal fit of the simultaneous model, one of the curves will typically be a Chapman-Richards curve while the other will be a three part model curve. However, this approach is not fully satisfactory as it does not neutralize the initial OD correlation with the lag time and growth rate.

The philosophy of the new approach we introduce in this chapter is as follows. Assume that the idealized model of a logarithmic growth curve consisting of a lag phase, an exponential phase and a stationary phase, is true. What difference should we expect between the curves starting from different population sizes, but with similar cell phase compositions? In the first phase, when there are plenty of nutrients available, we expect the same relative growth behavior. In the second phase, the time of the exponential growth will be shorter for a larger initial population. And finally, when the nutrient concentrations are "low enough", the entry into the stationary phase will take place with populations of approximately the same sizes and similar compositions, so that the logarithmic curves will have a similar shape also in this part.

In the type of data we have, a large proportion of the variability in the fitted curves comes from the initial population size. Can we predict what the behavior of a growth curve would have been, had the population had a standard initial OD? Can we reduce the sensitivity of the growth parameter estimates to initial OD? Essentially, the idealized model tells us to cut away a linear piece in the middle of one of the logarithmic growth curves in order to get the other. Moreover, it is natural to expect roughly the

same stationary phase OD increment. The total growth of the population has to do with how effectively the available energy is used. We do not know exactly what happens, but it is likely that some of the energy consumption in the beginning goes to initiating the growth process which would imply a slightly smaller stationary phase OD increment for larger populations, but we will ignore this. In the exponential phase the populations grow and consume nutrients similarly except that the population with a smaller initial OD grows for a longer time because there is more energy per cell available. The small difference in the population sizes when the lack of nutrients begins to slow down the growth, possibly also affects the relative growth rate slightly (the larger the population is, the faster it will consume the resources). However, this effect is probably quite small and therefore we will ignore it. Altogether these approximations motivate the assumption that the stationary phase OD increments should be approximately equal irrespective of the size of the initial OD (within certain limits of initial OD).

The idea behind the standardization is that we fix a standard initial OD and predict what would have happened, had we done the experiment with the standard initial OD and fitted the Chapman-Richards curve on these measurements. We hope that with the standardization, the curves from different runs and environments become more easily comparable. The standardization will also be useful for visualizing the data.

We first present a method for standardizing growth curves upwards, *i.e.* when the standard initial OD is larger than the observed initial OD. We begin by describing the method for standardizing one curve and then generalize it to obtain a standardized growth curve of two or more curves. Second, we present a method for standardizing growth curves downwards. The curves are standardized upwards or downwards depending on the relation between the chosen standard initial OD and their observed initial OD based on the ordinary Chapman-Richards curve fit.¹

5.1 Standardizing upwards

Here, we try to predict what would have happened had the initial OD been fixed to be larger than the observed initial OD. We use the three part model presented in Section 3.3 to fit the observed curve so that a standardized curve can be obtained by 'lifting' the fitted curve to start from $\log(s_0)$ and removing the linear piece from the middle (Figure 5.1). The growth parameters, *i.e.* λ (lag time), μ (maximum relative growth rate), and Y (stationary phase OD increment), are to be the same for the three part model curve and for the standardized curve. Since the time span of the linear part in the three part model is not modeled freely, the parameter values (and thus the fitted

¹The observed initial OD is defined as the value of the Chapman-Richards curve fit to the observed OD values at time zero.

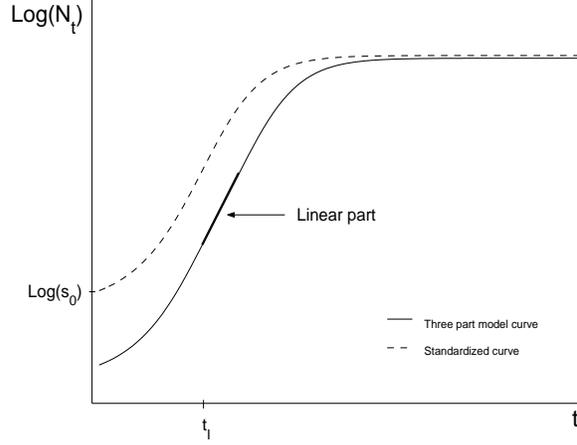


Figure 5.1: *An illustration of standardizing one curve when the standard initial OD s_0 is higher than the observed initial OD. Here N_t is the population size at time t and t_I is the inflection time point.*

curve) are not the same as they would be if the three part model was fitted without any constraints (in which case the time span of the linear part would be zero for most of the curves, as was concluded in Section 3.3).

Recall that a curve from the three part model is given by

$$g_t^* = \begin{cases} g_t, & t \leq t_I, \\ g_{t_I} + \mu(t - t_I), & t_I \leq t \leq t_I + \Delta, \\ g_{t_I - \Delta} + \mu\Delta, & t \geq t_I + \Delta, \end{cases} \quad (5.1)$$

where

$$g_t = \beta_0 \left[1 - \beta_1 e^{-\beta_2 t} \right]^{1/(1-\beta_3)} + D$$

is the Chapman-Richards function,

$$t_I = \frac{\log\left(\frac{\beta_1}{1-\beta_3}\right)}{\beta_2}$$

is the inflection time point (the time axis value where the linear part starts), and

$$\mu = \beta_0 \beta_2 \beta_3^{\frac{\beta_3}{1-\beta_3}}$$

is the (maximum relative) growth rate, and Δ is the time span of the linear part. We fit the three part model to the observed data with the constraint that by removing the linear piece in the middle and lifting the curve by $\tau \geq 0$, we obtain the standardized curve, which will be denoted by g_t^\bullet :

$$g_t^\bullet = \beta_0 \left[1 - \beta_1 e^{-\beta_2 t} \right]^{1/(1-\beta_3)} + D + \tau.$$

Let s_0 denote the standard initial OD (a fixed value). The logarithm of the initial OD of the standardized curve has to equal $\log(s_0)$, *i.e.*

$$\beta_0(1 - \beta_1)^{\frac{1}{1-\beta_3}} + D + \tau = \log(s_0). \quad (5.2)$$

Solving the equation (5.2) for τ gives

$$\tau = \log(s_0) - \beta_0(1 - \beta_1)^{\frac{1}{1-\beta_3}} - D.$$

The time span of the linear part, Δ , is adjusted so that the stationary phase OD increment of the three part model curve

$$Y = e^{\beta_0 + D + \mu \Delta} - e^{\beta_0(1-\beta_1)^{\frac{1}{1-\beta_3}} + D}$$

equals the stationary phase OD increment of the standardized curve

$$Y^\bullet = e^{\beta_0 + D + \tau} - e^{\beta_0(1-\beta_1)^{\frac{1}{1-\beta_3}} + D + \tau}.$$

This yields

$$\Delta = \frac{-\beta_0 - D + \log[e^{\beta_0 + D + \tau} - e^{\beta_0(1-\beta_1)^{\frac{1}{1-\beta_3}} + D + \tau} + e^{\beta_0(1-\beta_1)^{\frac{1}{1-\beta_3}} + D}]}{\mu}. \quad (5.3)$$

Model fitting procedure

We first fix a standard initial OD s_0 . The curves which initial OD according to the Chapman-Richards model fit is smaller than or equal to s_0 , will be standardized using this procedure:

1. An initial value of Δ is chosen.
2. The model (5.1) is fitted using a nonlinear least squares method, keeping Δ fixed.
3. The stationary phase OD increments, Y and Y^\bullet , are calculated. If $|Y - Y^\bullet| > c$, then a new value of Δ is calculated as given in (5.3). The constant c is the maximum allowed difference between Y and Y^\bullet (c is usually a very small real number).

The steps 2 and 3 are repeated until $|Y - Y^\bullet| < c$.

5.1.1 Standardizing two or more curves simultaneously

The method presented in the previous section can easily be generalized to obtain a single standardized curve for n curves. The three part models are fitted to the observed curves with a constraint that the standardized curve can be obtained by removing the linear pieces and lifting the curves to start at $\log(s_0)$. The lag time and growth rate are kept the same in all three part model curves and in the standardized curve. The stationary phase OD increment of the standardized curve is set to be the same as the average of the stationary phase OD increments of the three part model curves.

In the sequel, we discuss a standardization method for two curves. The generalization to n curves can be done analogously. The three part models with the Chapman-Richards function can be written as

$$g_t^{*(k)} = \begin{cases} g_t^{(k)}, & t \leq t_I, \\ g_{t_I}^{(k)} + \mu(t - t_I), & t_I \leq t \leq t_I + \Delta_k, \\ g_{t_I - \Delta_k}^{(k)} + \mu\Delta_k, & t \geq t_I + \Delta_k, \end{cases} \quad (5.4)$$

where

$$g_t^{(k)} = \beta_0 \left[1 - \beta_1 e^{-\beta_2 t} \right]^{1/(1-\beta_3)} + D_k,$$

and $k = 1, 2$. The standardized curve is

$$g_t^\bullet = \beta_0 \left[1 - \beta_1 e^{-\beta_2 t} \right]^{1/(1-\beta_3)} + D_k + \tau_k,$$

where $\tau_k \geq 0$.

The logarithm of the initial OD of the standardized curve is set to equal $\log(s_0)$, *i.e.*

$$\beta_0(1 - \beta_1)^{\frac{1}{1-\beta_3}} + D_1 + \tau_1 = \beta_0(1 - \beta_1)^{\frac{1}{1-\beta_3}} + D_2 + \tau_2 = \log(s_0). \quad (5.5)$$

Solving the equation (5.5) for τ_1 and τ_2 , gives

$$\begin{aligned} \tau_1 &= \log(s_0) - \beta_0(1 - \beta_1)^{\frac{1}{1-\beta_3}} - D_1, \\ \tau_2 &= \log(s_0) - \beta_0(1 - \beta_1)^{\frac{1}{1-\beta_3}} - D_2. \end{aligned}$$

The stationary phase OD increments of the three part model curves are

$$\begin{aligned} Y_1 &= e^{\beta_0 + D_1 + \mu\Delta_1} - e^{\beta_0(1-\beta_1)^{\frac{1}{1-\beta_3}} + D_1}, \\ Y_2 &= e^{\beta_0 + D_2 + \mu\Delta_2} - e^{\beta_0(1-\beta_1)^{\frac{1}{1-\beta_3}} + D_2}. \end{aligned}$$

The time spans of the linear parts, Δ_1 and Δ_2 , are adjusted so that the stationary phase OD increment of the standardized curve equals the average of the stationary phase OD increments of the three part model curves, i.e. $Y^\bullet = \frac{Y_1+Y_2}{2}$. The expressions for Δ_1 and Δ_2 become

$$\Delta_1 = \frac{-\beta_0 - D_2 - \tau_2 + \tau_1 + \log[\varpi]}{\mu}, \quad (5.6)$$

where

$$\varpi = e^{D_2+\tau_2} 2e^{\beta_0} - 2e^{\beta_0(1-\beta_1)^{1/(1-\beta_3)}} + e^{\beta_0(1-\beta_1)^{1/(1-\beta_3)}-\tau_1} + e^{\beta_0(1-\beta_1)^{1/(1-\beta_3)}-\tau_2} - e^{\beta_0+\Delta_2\mu-\tau_2},$$

and

$$\Delta_2 = \frac{-\beta_0 - D_2 + \log[\vartheta]}{\mu}, \quad (5.7)$$

where

$$\vartheta = e^{D_2+\tau_2} 2e^{\beta_0} - 2e^{\beta_0(1-\beta_1)^{1/(1-\beta_3)}} + e^{\beta_0(1-\beta_1)^{1/(1-\beta_3)}-\tau_1} + e^{\beta_0(1-\beta_1)^{1/(1-\beta_3)}-\tau_2} - e^{\beta_0+\Delta_1\mu-\tau_1}.$$

One possible variant of this standardization method would be to require that $\log(Y^\bullet) = \frac{\log(Y_1)+\log(Y_2)}{2}$ (instead of $Y^\bullet = \frac{Y_1+Y_2}{2}$), yielding a methodology closely related to the method II summary curves that will be presented in Chapter 6.

Model fitting procedure

The models are fitted using a nonlinear least squares method. First a standard initial OD s_0 is fixed. As in the standardization of a single curve, this procedure will be used only for curves which initial OD according to the Chapman-Richards model fit is smaller than or equal to s_0 :

1. Initial values for Δ_1 and Δ_2 are obtained by first standardizing each of the curves separately *i.e.* using the method presented in Section 5.1.
2. The Model (5.4) is fitted keeping Δ_1 and Δ_2 fixed.
3. The stationary phase OD increments, Y_1 , Y_2 , and Y^\bullet , are calculated. If $|\frac{Y_1+Y_2}{2} - Y^\bullet| > c$, a new value for Δ_1 is calculated using equation (5.6) and the model is fitted again keeping Δ_1 and Δ_2 fixed. The constant c is the maximum allowed difference between $\frac{Y_1+Y_2}{2}$ and Y^\bullet .
4. The stationary phase OD increments, Y_1 , Y_2 , and Y^\bullet , are calculated. If $|\frac{Y_1+Y_2}{2} - Y^\bullet| > c$, a new value for Δ_2 is calculated using equation (5.7) and the model is fitted again keeping Δ_1 and Δ_2 fixed.

The steps 3 and 4 are repeated until $|\frac{Y_1+Y_2}{2} - Y^\bullet| < c$. The model fitting procedure is illustrated in Figure 5.2.

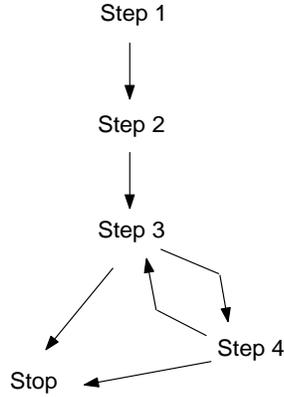


Figure 5.2: *An illustration of the model fitting procedure for standardizing two curves simultaneously.*

5.2 Standardizing downwards

When the standard initial OD is less than the observed initial OD, we cannot apply the same standardization method as before. It would be possible to fit a Chapman-Richards model to the observed data and then obtain a standardized curve by inserting a linear part in the inflection point of the curve fitted to the observed data and moving the curve to start at $\log(s_0)$. However, this method would have at least two problems. First, the model of the standardized curve would not be the Chapman-Richards model. Second, when standardizing downwards, we do not always know if the observed curve has reached the optimal growth rate. Only adding a linear part to the observed curve model might underestimate the slope.

For these reasons we will proceed differently. A Chapman-Richards model curve that lacks a part in the middle is fitted to the observed data. The standardized curve is then the Chapman-Richards model curve, including the part in the middle (that is missing in the curve fitted to the observed data). The stationary phase OD increment of the standardized curve is to be the same as of the observed curve, but the lag time and growth rate do not need to be the same.

The model of the observed curve can be written as

$$g_t^* = \begin{cases} g_t, & t \leq t_L, \\ g_{t+\Delta} - (g_{t_U} - g_{t_L}), & t \geq t_L, \end{cases} \quad (5.8)$$

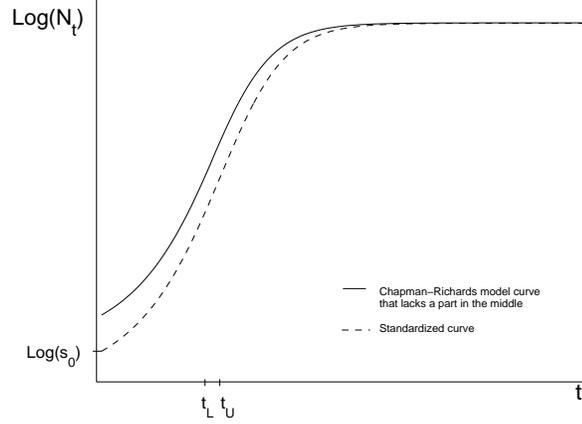


Figure 5.3: *An illustration of standardizing one curve when the standard initial OD s_0 is lower than the observed initial OD. Here N_t is the population size at time t , t_L is the inflection time point of the curve fitted to the observed data, and the inflection time point of the standardized curve is somewhere between t_L and t_U .*

where

$$g_t = \beta_0 \left[1 - \beta_1 e^{-\beta_2 t} \right]^{1/(1-\beta_3)} + D$$

is the Chapman-Richards function and

$$\Delta = t_U - t_L.$$

The derivatives of the curve at $t = t_L$ and $t = t_U$ have to be the same. The downwards standardization is illustrated in Figure 5.3.

The inflection time point of the standardized curve is somewhere between t_L and t_U , and the inflection time point of the curve fitted to the observed data is t_L . The t_L is obtained by setting the derivative of g_t ,

$$\frac{dg_t}{dt} = \frac{\beta_0 \beta_1 \beta_2 e^{-\beta_2 t} \left[1 - \beta_1 e^{-\beta_2 t} \right]^{1/(1-\beta_3)-1}}{1 - \beta_3},$$

at $t = t_L$ equal to the derivative at $t = t_U (=t_L + \Delta)$, and solving the equation with respect to t_L . Therefore,

$$t_L = \frac{1}{\beta_2} \log \left[\frac{\beta_1 e^{-\beta_2 \Delta} \left(e^{\beta_2 \Delta + \frac{\Delta \beta_2 (1-\beta_3)}{\beta_3}} - 1 \right)}{e^{\frac{\Delta \beta_2 (1-\beta_3)}{\beta_3}} - 1} \right].$$

The model of the standardized curve is written as

$$g_t^\bullet = \beta_0 \left[1 - \beta_1 e^{-\beta_2 t} \right]^{1/(1-\beta_3)} + D + \tau,$$

where τ (≤ 0) is obtained by setting the initial $\log(\text{OD})$ value of the standardized curve to $\log(s_0)$

$$\beta_0(1 - \beta_1)^{1/(1-\beta_3)} + D + \tau = \log(s_0) \quad (5.9)$$

and solving equation (5.9) with respect to τ which yields

$$\tau = \log(s_0) - \beta_0(1 - \beta_1)^{\frac{1}{1-\beta_3}} - D.$$

The stationary phase OD increment of the standardized curve

$$Y^\bullet = e^{\beta_0 + D + \tau} - e^{\beta_0(1-\beta_1)^{\frac{1}{1-\beta_3}} + D + \tau}$$

has to equal the stationary phase OD increment of the observed curve

$$\begin{aligned} Y &= e^{\beta_0 + D - (g_{t_U} - g_{t_L})} - e^{\beta_0(1-\beta_1)^{\frac{1}{1-\beta_3}} + D} \\ &= e^{\beta_0 + D - \beta_0 \left[(1 - \beta_1 e^{-\beta_2(t_L + \Delta)})^{\frac{1}{1-\beta_3}} - (1 - \beta_1 e^{-\beta_2 t_L})^{\frac{1}{1-\beta_3}} \right]} - e^{\beta_0(1-\beta_1)^{\frac{1}{1-\beta_3}} + D}. \end{aligned}$$

This gives

$$\Delta = -t_L + \frac{1}{\beta_2} \log \left[\frac{\beta_1}{1 - \left((1 - \beta_1 e^{-\beta_2 t_L})^{\frac{1}{1-\beta_3}} + [\varphi] \right)^{1-\beta_3}} \right], \quad (5.10)$$

where

$$\varphi = \frac{\beta_0 + D - \log \left[e^{\beta_0(1-\beta_1)^{\frac{1}{1-\beta_3}} + D} + e^{\beta_0 + D + \tau} - e^{\beta_0(1-\beta_1)^{\frac{1}{1-\beta_3}} + D + \tau} \right]}{\beta_0}.$$

Model fitting procedure

First the standard initial OD s_0 is fixed. The curves which initial OD according to the Chapman-Richards model fit is larger than or equal to s_0 will be standardized using this procedure:

1. An initial value of Δ is chosen.
2. The model (5.8) is fitted using a nonlinear least squares method, keeping Δ fixed.

3. The stationary phase OD increments, Y and Y^\bullet are calculated. If $|Y - Y^\bullet| > c$, then a new value of Δ is calculated as given in (5.10). The constant c is the maximum allowed difference between Y and Y^\bullet .

The steps 2 and 3 are repeated until $|Y - Y^\bullet| < c$.

Generalizing this method to two or more curves is not as trivial as in the case of standardizing upwards. The algorithms for simultaneous standardizations of curves downwards, or for simultaneous standardizations where some curves would be standardized upwards and some downwards, would become complicated but certainly not impossible. However, a standardized curve for two or more curves can easily be obtained by standardizing first each curve separately and then making a summary curve of them. The summarizing method will be presented in Chapter 6.

5.3 Fitting the standardization models to the data

We fitted the standardization models to hundreds of growth curves of the data described in Section 2.4.² The initial OD values vary between 0.01 and 0.48, and the average is 0.107. There are large differences in initial OD between different environments (Figures 3.9-3.10). A nonlinear regression model was fitted via least squares in the same way as in Section 3.2.3. The maximum allowed difference in the stationary phase OD increment between fitted and standardized curves was 0.001 (*i.e.* $c = 0.001$). The parameter estimates from the Chapman-Richards model fit were used as start values in the model fitting algorithms for β_0 , β_1 , β_2 , β_3 , and D . The start value for Δ was $20|s - s_0|$, where s is the initial OD according to the Chapman-Richards model fit to the observed data. The curve fit with different standard initial OD values was investigated visually and also using the coefficient of determination.

The fit is rather good when standardizing one curve, however, it is not as good as with the Chapman-Richards method. It is best for the curves with a small difference between the observed and standard initial OD. An example of the fit of a curve standardized upwards with $s_0 = 0.15$, $s_0 = 0.20$, and $s_0 = 0.30$ is given in Figure 5.4. The same curve is standardized downwards with $s_0 = 0.08$, $s_0 = 0.05$, and $s_0 = 0.03$ in Figure 5.5. Figure 3.2 shows the Chapman-Richards model fit of the curve.

For standardizing two curves the method works reasonably well when the curves have rather normal and similar shapes, see *e.g.* Figure 5.6. Also for obtaining a standardized curve of several curves the method works, given that the curves have rather normal and similar shapes (Figure 5.7). However, if that is not the case, the fit can become poor (Figure 5.8).

²The Matlab functions are available upon request.

Both when standardizing upwards or downwards, it is important that the standard initial OD does not differ too much from the observed OD. When the difference is large, the fit can become poor and the growth rate and lag time may be overestimated or underestimated. Two examples of a fit when s_0 is far from the observed initial OD are shown in Figure 5.9. The data are the same as in Figures 5.4-5.5. The growth parameter estimates from the standardized curves with different s_0 and the coefficient of determinations of the fitted curves are shown in Table 5.1.

We also compared the estimates of the growth parameters from the least squares fit of the Chapman-Richards model with the estimates from the standardized curves. The averages of the growth parameter estimates are nearly the same with both methods if the standard initial OD is close to the average of the observed initial OD values. The coefficient of variations of replicates' growth parameter estimates tend to be smaller with the standardization method.

In Section 3.3 we investigated the correlation of the initial OD with growth rate and lag time estimated with ordinary Chapman-Richard model and ordinary three part model, i.e. when the time span of the linear part is modeled freely. In the sequel, we investigate the mentioned correlations when the growth rate and lag time are estimated with the standardization method. Nine different standard initial OD values are used: 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, and 0.1. With all of them the correlation between growth rate and initial OD reduces remarkably, most with $s_0 = 0.04$, compared with the ordinary Chapman-Richards model or the ordinary three part model (Figures 3.8, 3.17, 5.10, and 5.11). The correlation between initial OD and lag time reduces also, however, it remains rather high with all values of s_0 .

5.4 Discussion

From a conceptional point of view, standardizing downwards proved to be more difficult than standardizing upwards. When standardizing downwards, if both the observed curve and the curve that we would have gotten from a culture with a standard initial OD, have reached the exponential phase, they should look similar both in the beginning and in the end, just as with standardizing upwards. However, if the observed curve has not reached the exponential phase, the whole population composition is different, and the transition mechanisms should give another curve form. If we have reached the exponential phase but our parametric model does not capture that, the shapes should again look similar. This observation makes the use of the cut-out approach slightly less ad-hoc.

In order not to have to model a long unknown part, the standard initial OD should not be too low compared to the observed initial OD. How large should the standard initial OD then be? It may be natural to use approximately the average of the observed initial OD values, or a value that is considered to be ideal. However,

more research on how to choose the standard initial OD is needed.

Besides enabling easy comparison of data from different experiments, the standardization method reduces the correlation between initial OD and growth rate and initial OD and lag time, compared to the ordinary Chapman-Richards method. It is possible that with the standardization method we have a systematic error in both lag time and growth rate. However, this systematic error will cancel out, at least partly, in the data analysis (in Chapter 8) when the mutant values are normalized using the wild type values in the same run.

The aims of using the ordinary Chapman-Richards method and the standardization method can be different. The standardization may be appropriate when the aim is to have comparable curves or to visualize data rather than to model the curves accurately.

Table 5.1: *The growth parameter estimates from the standardized curve and the coefficient of determination of the fitted curve (NOD0305, well 3). The growth parameter estimates and the coefficient of determination of the Chapman-Richards model fit to the same curve.*

| Method | λ | μ | Y | r^2 |
|--------------------------------|-----------|-------|-------|--------|
| Chapman-Richards | 2.897 | 0.227 | 4.007 | 0.9999 |
| Standardization, $s_0 = 0.015$ | 4.368 | 0.270 | 4.031 | 0.9998 |
| Standardization, $s_0 = 0.030$ | 3.674 | 0.251 | 4.019 | 0.9999 |
| Standardization, $s_0 = 0.050$ | 3.223 | 0.238 | 4.008 | 0.9999 |
| Standardization, $s_0 = 0.080$ | 2.835 | 0.228 | 3.995 | 0.9999 |
| Standardization, $s_0 = 0.150$ | 2.277 | 0.215 | 3.986 | 0.9999 |
| Standardization, $s_0 = 0.200$ | 2.148 | 0.212 | 3.981 | 0.9999 |
| Standardization, $s_0 = 0.300$ | 1.980 | 0.207 | 3.976 | 0.9998 |
| Standardization, $s_0 = 0.900$ | 1.711 | 0.198 | 3.962 | 0.9995 |

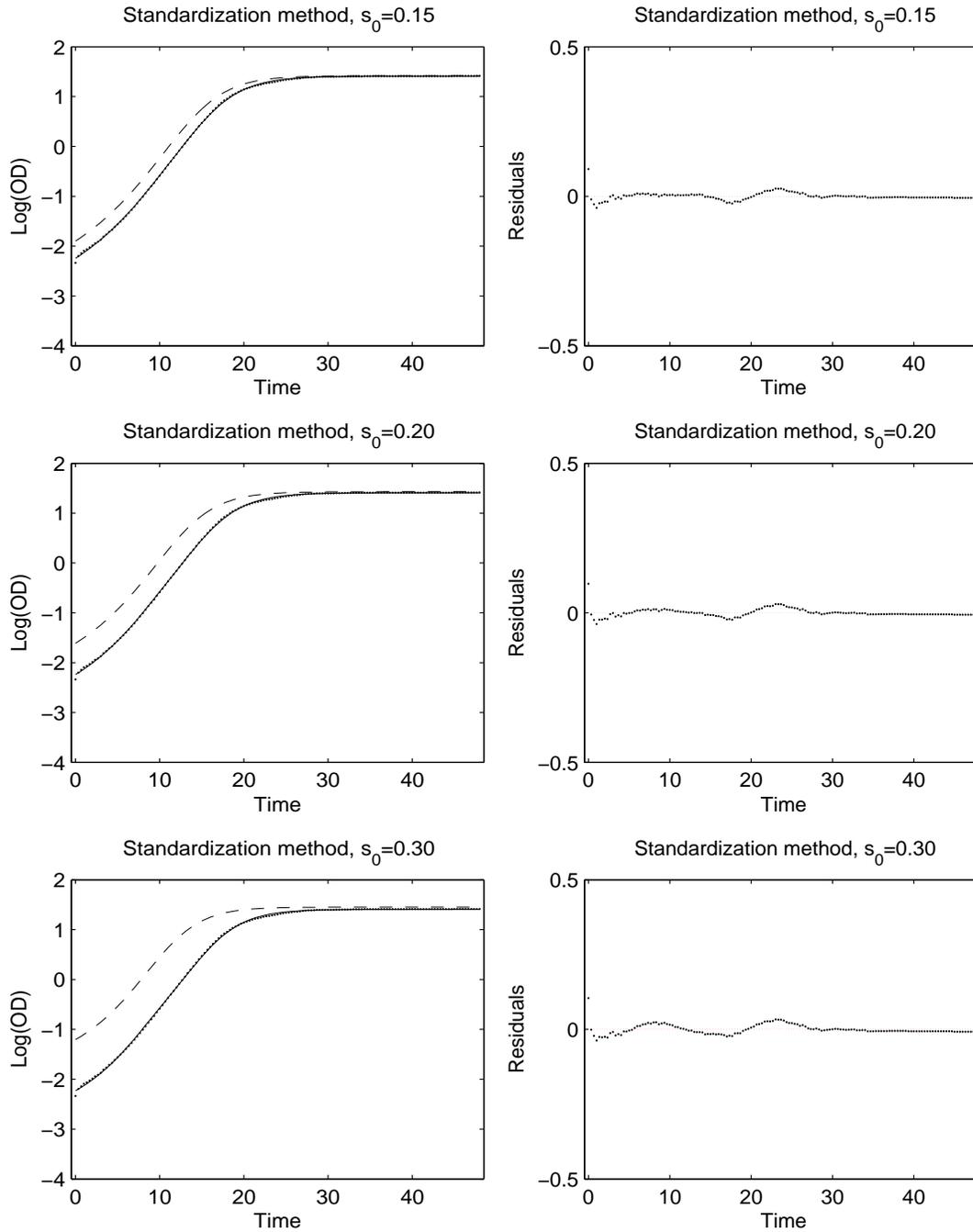


Figure 5.4: Standardizing one curve (NOD0305, well 3) upwards with different standard initial OD. The $\log(\text{OD})$ values (dotted), the fitted growth curve (solid) and the standardized growth curve (dashed). The corresponding residual plots of the fitted curves are on the right.

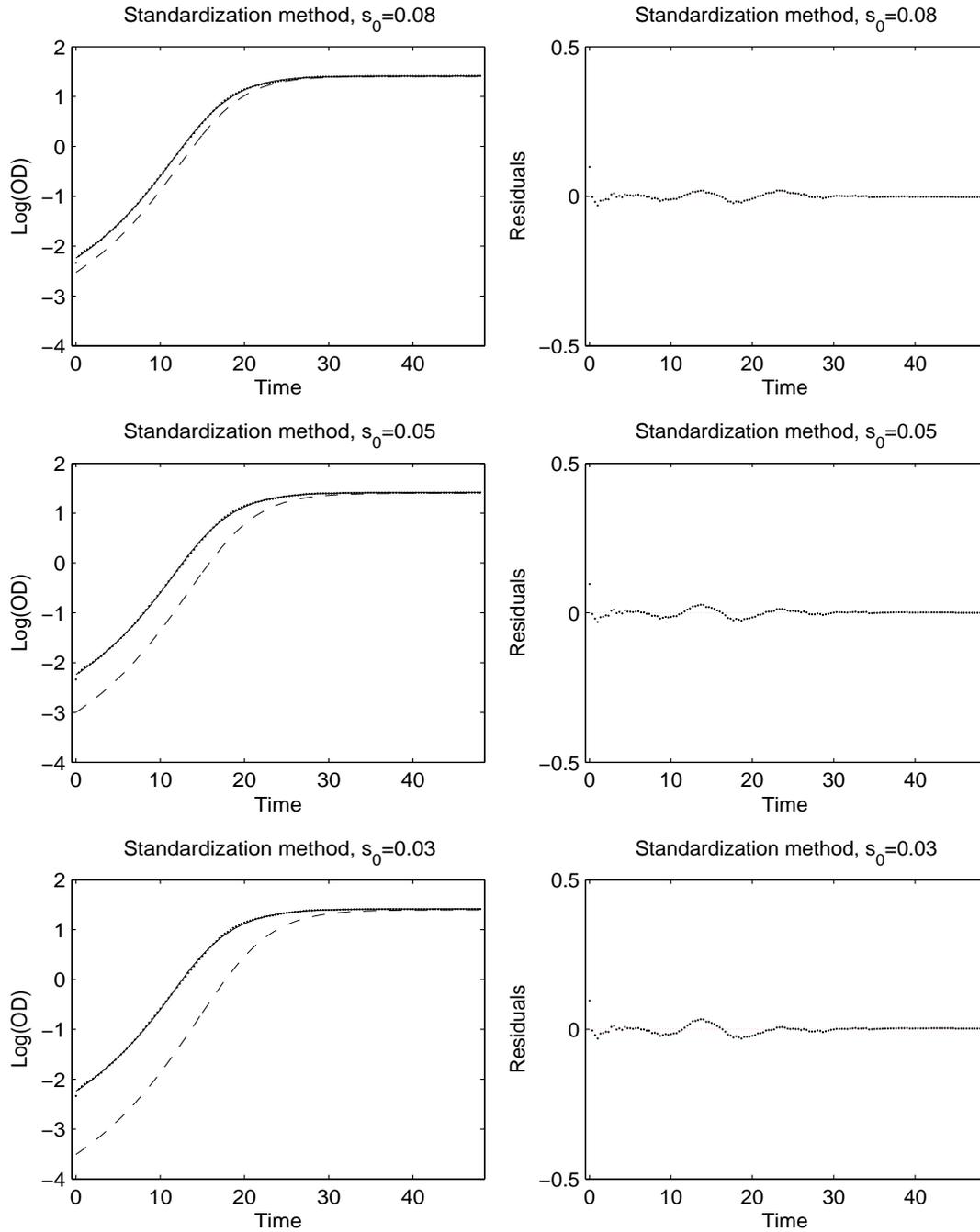


Figure 5.5: Standardizing one curve (NOD0305, well 3) downwards with different standard initial OD values. The $\log(OD)$ values (dotted), the fitted growth curves (solid) and the standardized growth curve (dashed). The corresponding residual plots of the fitted curves are on the right. 64

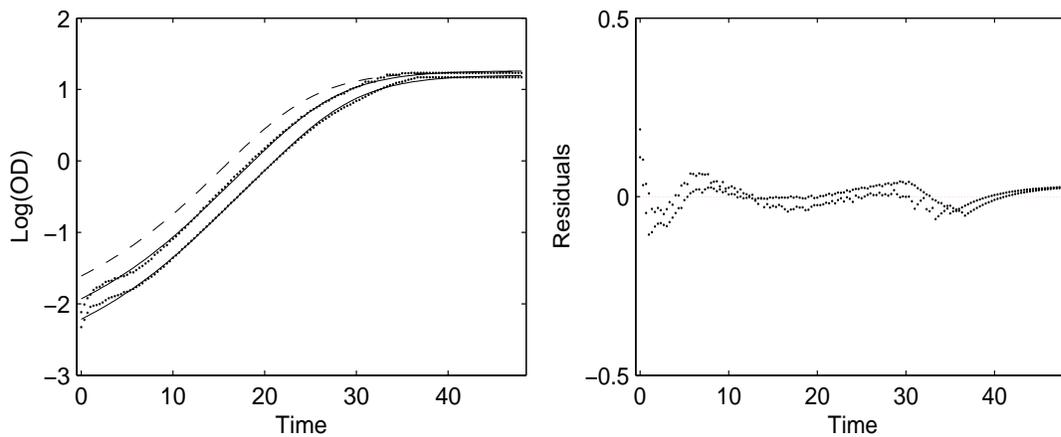


Figure 5.6: *Standardizing two curves (NAC0321 and NAC0323, well 88). The $\log(OD)$ values (dotted), the fitted growth curves (solid) and the standardized growth curve (dashed). The corresponding residual plot of the fitted curves is on the right.*

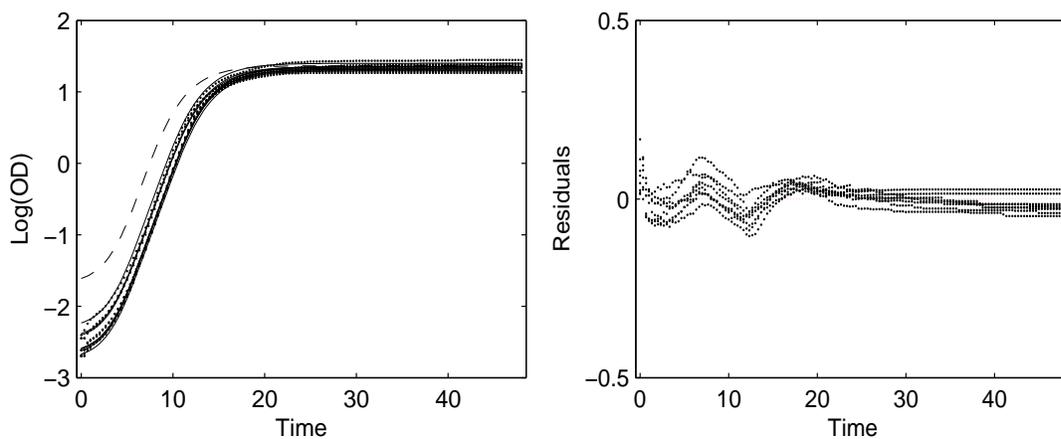


Figure 5.7: *Eight wild types in reference condition (NOC0326) fitted with the standardization method. The $\log(OD)$ values (dotted), the fitted growth curves (solid) and the standardized growth curve (dashed). The corresponding residual plot of the fitted curves is on the right.*

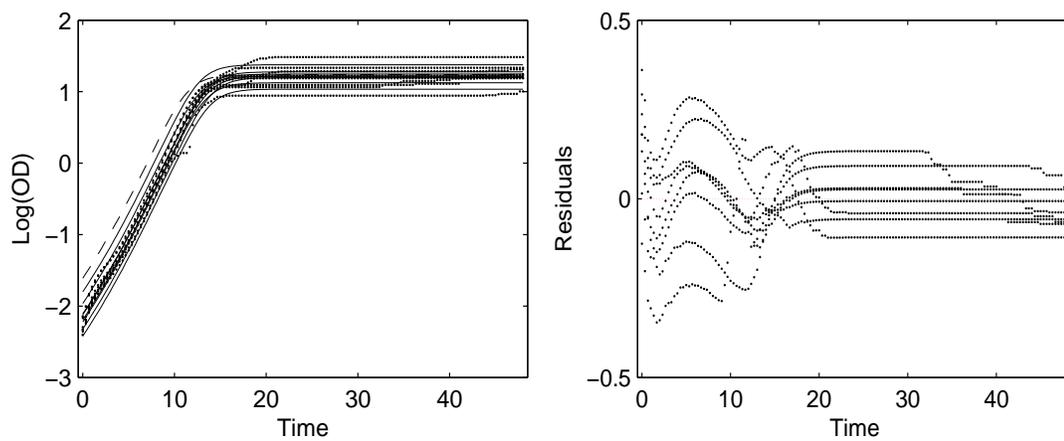


Figure 5.8: *Eight wild types in 39°C (39c0307) fitted with the standardization method. The $\log(OD)$ values (dotted), the fitted growth curves (solid) and the standardized growth curve (dashed). The corresponding residual plot of the fitted curves is on the right.*

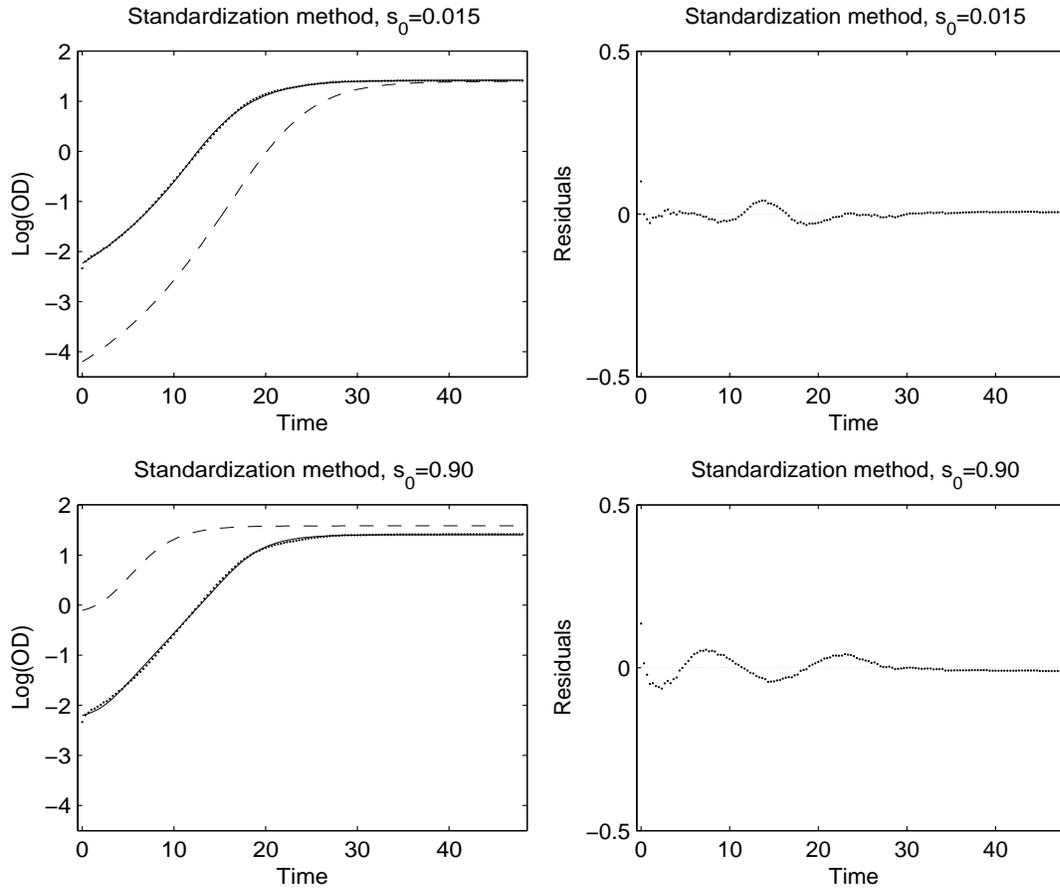


Figure 5.9: Standardizing one curve (NOD0305, well 3) with values of s_0 that differ greatly from the observed initial OD. The $\log(OD)$ values (dotted), the fitted growth curves (solid) and the standardized growth curve (dashed). The corresponding residual plots of the fitted curves are on the right.

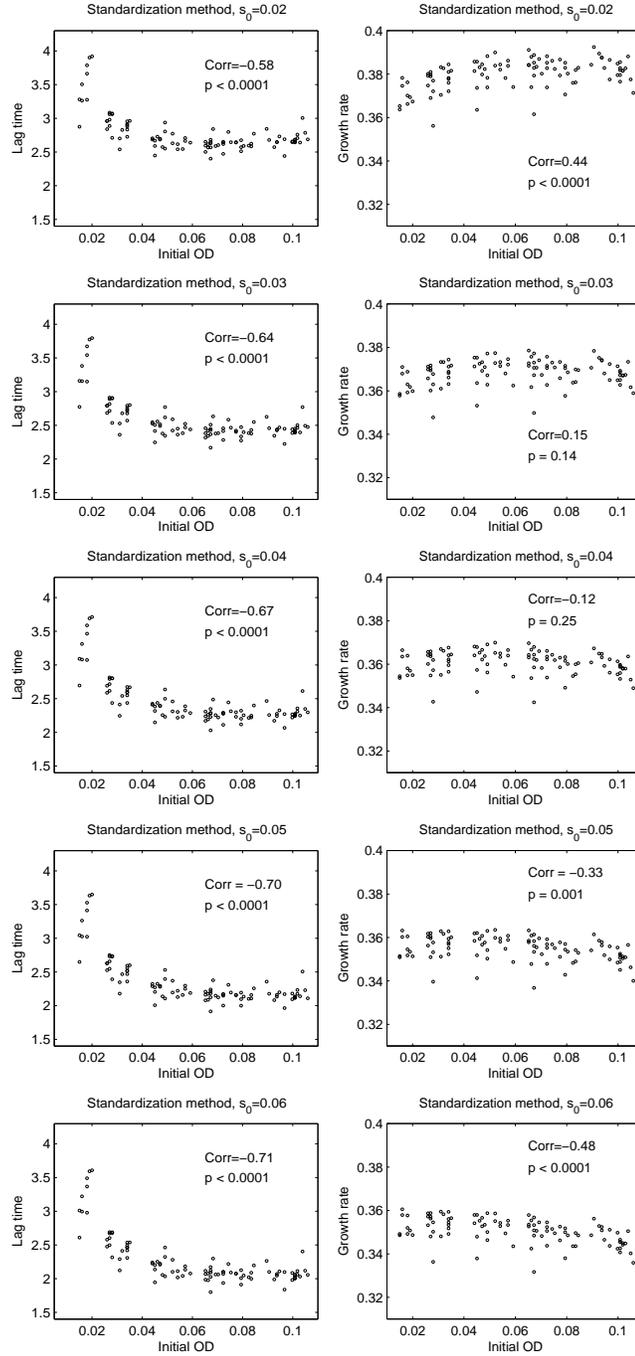


Figure 5.10: *The initial OD of the 99 wild types in reference condition plotted against lag time and growth rate estimates from the standardization method with different s_0 .*

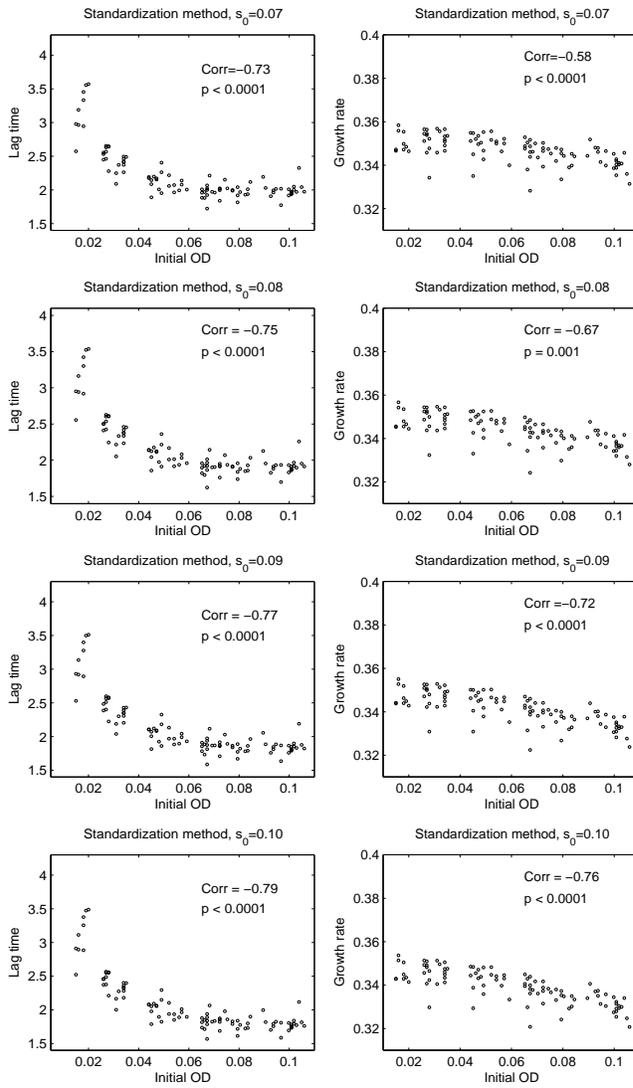


Figure 5.11: *The initial OD of the 99 wild types in reference condition plotted against lag time and growth rate estimates from the standardizing method with different s_0 .*

Chapter 6

Summarizing curves

In Chapter 5 we suggested a method by which the growth curves can be standardized with respect to the initial OD. It is possible to fit a standardized curve directly on all curves that we wish to have a representative curve for. However, if the curves do not have similar shapes, or if there are many curves to be standardized simultaneously, the fit can become poor and the estimated curves can get strange shapes. It may be better to standardize individual curves or groups of curves, and then summarize the results.

Two similar methods to summarize growth curves are presented. For these methods, the curves to be summarized have to be standardized first, *i.e.* they must have the same initial OD. The summary curves are based either on averages of the growth parameters d_0 , λ , μ , and Y , or on averages of $\log(d_0)$, $\log(\lambda)$, $\log(\mu)$, and $\log(Y)$, of the standardized curves. In this chapter we work under the assumption that Conjecture 1 is true, so that the growth parameter parameterization can be assumed to be unique according to Theorem 1.

6.1 Method I

In this method, the Chapman-Richards model is used for the summary curves. The d_0 , λ , μ , and Y of the summary curve are to equal the averages of the corresponding parameters of the standardized curves it summarizes, and the initial OD is to equal the standard initial OD.

The model parameter values are obtained as follows. Let n be the number of standardized curves to be summarized, and let

$$\bar{d}_0 = \frac{\sum_{i=1}^n d_{0(i)}}{n} \quad (6.1)$$

be their average derivative at time zero,

$$\bar{\lambda} = \frac{\sum_{i=1}^n \lambda(i)}{n} \quad (6.2)$$

their average lag time,

$$\bar{\mu} = \frac{\sum_{i=1}^n \mu(i)}{n} \quad (6.3)$$

their average growth rate, and

$$\bar{Y} = \frac{\sum_{i=1}^n Y(i)}{n} \quad (6.4)$$

their average stationary phase OD increment. Furthermore, s_0 is the standard initial OD. To find the parameters in the original parameterization it would be possible to use the nonlinear least squares method to minimize

$$\begin{aligned} f(\beta_0, \beta_1, \beta_2, \beta_3, D) &= \left[s_0 - e^{\beta_0(1-\beta_1)\frac{1}{1-\beta_3} + D} \right]^2 \\ &+ \left[\bar{d}_0 - \frac{\beta_0\beta_1\beta_2(1-\beta_1)^{\frac{1}{1-\beta_3}-1}}{1-\beta_3} \right]^2 \\ &+ \left[\bar{\lambda} - \frac{(1-\beta_1)^{\frac{1}{1-\beta_3}} - \beta_3^{\frac{1}{1-\beta_3}} + \beta_3^{\frac{\beta_3}{1-\beta_3}} \log\left(\frac{\beta_1}{1-\beta_3}\right)}{\beta_2\beta_3^{\frac{\beta_3}{1-\beta_3}}} \right]^2 \\ &+ \left[\bar{\mu} - \beta_0\beta_2\beta_3^{\frac{\beta_3}{1-\beta_3}} \right]^2 \\ &+ \left[\bar{Y} - \left(e^{\beta_0+D} - e^{\beta_0(1-\beta_1)\frac{1}{1-\beta_3} + D} \right) \right]^2, \end{aligned}$$

and provided that this minimum is approximately zero, the argmin vector would approximate the vector of the parameters $\beta_0, \beta_1, \beta_2, \beta_3$, and D . However, we have chosen to use the least squares method only to obtain estimates for β_1 and β_3 , and calculate the values of β_0, β_2 , and D explicitly.

In order to estimate β_1 and β_3 , we first translate the curve as shown in (4.12) so that the initial OD, growth rate, and lag time are all equal to one, the derivative at time zero is $\frac{\bar{d}_0}{\bar{\mu}}$, and the stationary phase OD increment is

$$\left(\frac{\bar{Y}}{s_0} + 1 \right)^{\frac{1}{\lambda\bar{\mu}}} - 1.$$

The translation does not affect β_1 and β_3 . They can be estimated by applying the nonlinear least squares method sketched above on the equations below, derived from (4.1-4.5) and the assumptions that $s = 1$, $\lambda = 1$, and $\mu = 1$:

$$\frac{\bar{d}_0}{\bar{\mu}} = \frac{\beta_1(1 - \beta_1)^{\frac{\beta_3}{1-\beta_3}}}{(1 - \beta_3)\beta_3^{\frac{\beta_3}{\beta_3-1}}}$$

$$\left(\frac{\bar{Y}}{s_0} + 1\right)^{\frac{1}{\lambda\bar{\mu}}} = e^{\left[\beta_3^{\frac{\beta_3}{1-\beta_3}} \log\left(\frac{\beta_1}{1-\beta_3}\right) + (1-\beta_1)^{\frac{1}{1-\beta_3}} - \beta_3^{\frac{1}{1-\beta_3}}\right]^{-1}} e^{1-(1-\beta_1)^{\frac{1}{1-\beta_3}}}.$$

Then we move back to the non-translated curve and obtain β_0 , β_2 , and D from the equations below, derived from (4.1-4.5),

$$\beta_2 = \frac{(1 - \beta_1)^{\frac{1}{1-\beta_3}} - \beta_3^{\frac{1}{1-\beta_3}} + \beta_3^{\frac{\beta_3}{1-\beta_3}} \log\left(\frac{\beta_1}{1-\beta_3}\right)}{\beta_3^{\frac{\beta_3}{1-\beta_3}} \bar{\lambda}},$$

$$\beta_0 = \frac{\bar{\mu}}{\beta_2 \beta_3^{\frac{\beta_3}{1-\beta_3}}},$$

$$D = \log(s_0) - \frac{\beta_0}{(1 - \beta_1)^{\frac{1}{\beta_3-1}}}.$$

There is a theoretical risk that the minimum zero cannot be reached, because the specific parameter vector is not permitted in the Chapman-Richards model (see Theorem 2), but the problem seems to be of minor practical relevance (see Section 6.3). This problem can be avoided by using method II, described in the next section.

6.2 Method II

Here, we construct a summary curve for which the logarithms of d_0 , λ , μ , and Y equal the averages of the logarithms of the corresponding parameters of the standardized curves that it summarizes. The model parameter values are obtained in the same way

as in method I, except that instead of \bar{d}_0 , $\bar{\lambda}$, $\bar{\mu}$, and \bar{Y} , as given in (6.1- 6.4),

$$\tilde{d}_0 = e^{\frac{\sum_{i=1}^n \log[d_{0(i)}]}{n}} = \sqrt[n]{e^{\sum_{i=1}^n \log[d_{0(i)}]},$$

$$\tilde{\lambda} = e^{\frac{\sum_{i=1}^n \log[\lambda_{(i)}]}{n}} = \sqrt[n]{e^{\sum_{i=1}^n \log[\lambda_{(i)}]},$$

$$\tilde{\mu} = e^{\frac{\sum_{i=1}^n \log[\mu_{(i)}]}{n}} = \sqrt[n]{e^{\sum_{i=1}^n \log[\mu_{(i)}]},$$

$$\tilde{Y} = e^{\frac{\sum_{i=1}^n \log[Y_{(i)}]}{n}} = \sqrt[n]{e^{\sum_{i=1}^n \log[Y_{(i)}]},$$

are used.

6.3 Fitting the data

Both summarizing methods were tested on hundreds of growth curves of the data described in Section 2.4.¹ There were no problems with the fit as long as the lag times were not close to zero. When this happened, method II was the more sensitive one. Although theoretically the method I summary curves do not always exist, this was never a problem in our data.

Figure 6.1 shows examples of summary curves of double measurements for mutants in 39°C. The two methods often result in almost the same curve, since the standardized curves of the double measurements tend to have similar shapes. In Figure 6.2 there are summary curves of a mutant in reference condition and in Caffeine, and a mutant in reference condition and in Dinitrophenol. It can be seen that when the shapes of the standardized curves are very different (which is natural in this case since they are curves from different environments), the summary curves from the two methods differ more. An example of summarizing several curves can be seen in Figure 6.3.

Figure 6.4 displays an example of three different ways to obtain a representative curve for the wild types in 39°C. In the first one, all 48 wild type curves are standardized simultaneously. In the second one, a method I summary curve of all the 48 individually standardized wild type curves is fitted. In the third one, a method I summary curve of the six runwisely standardized wild type curves is fitted. The three curves look similar. However, the lag time and growth rate differ quite a lot between the three methods (Table 6.1): the standardization method gives a remarkably

¹The Matlab functions are available upon request.

smaller slope and thus smaller lag time than the other two methods. For comparison, we look at the averages of the growth parameter estimates of the 48 curves from the Chapman-Richards method. The summary curve of individually standardized wild type curves gives growth parameter estimates closest to the averages from the Chapman-Richards method. However, especially in lag time, the differences between the estimates from the summary curve and the averages of the estimates from the Chapman-Richards method are large. Note however, that it is difficult to compare the summarizing and simultaneous standardization methods because in the simultaneous standardization the standard initial OD has to be larger than the observed initial OD values. Therefore, in this example it is also difficult to compare the Chapman-Richards method and summarizing method estimates, since the standard initial OD is higher than the observed OD values. If the standard initial OD was close to the average of the observed OD values, the Chapman-Richards and the summarizing method would produce rather similar results on average.

6.4 Discussion

Although theoretically method I summary curves do not always exist, this is not a problem in our data. The two summarizing methods produce often almost the same results. There are more computational problems with the method II when lag times are very close to zero.

It would have been possible to try other methods too, such as using averages for some parameters and averages of logarithms for some parameters. We have chosen to take logarithms of all parameters because we have previously used this type of measures in the calculation of logarithmic phenotypic indexes (LPI) in the analysis of the data [6][26]. Thus, this type of summary curves are natural because they can directly be used in the calculation and illustration of the LPI. The LPI will be discussed in more detail in Chapter 8.

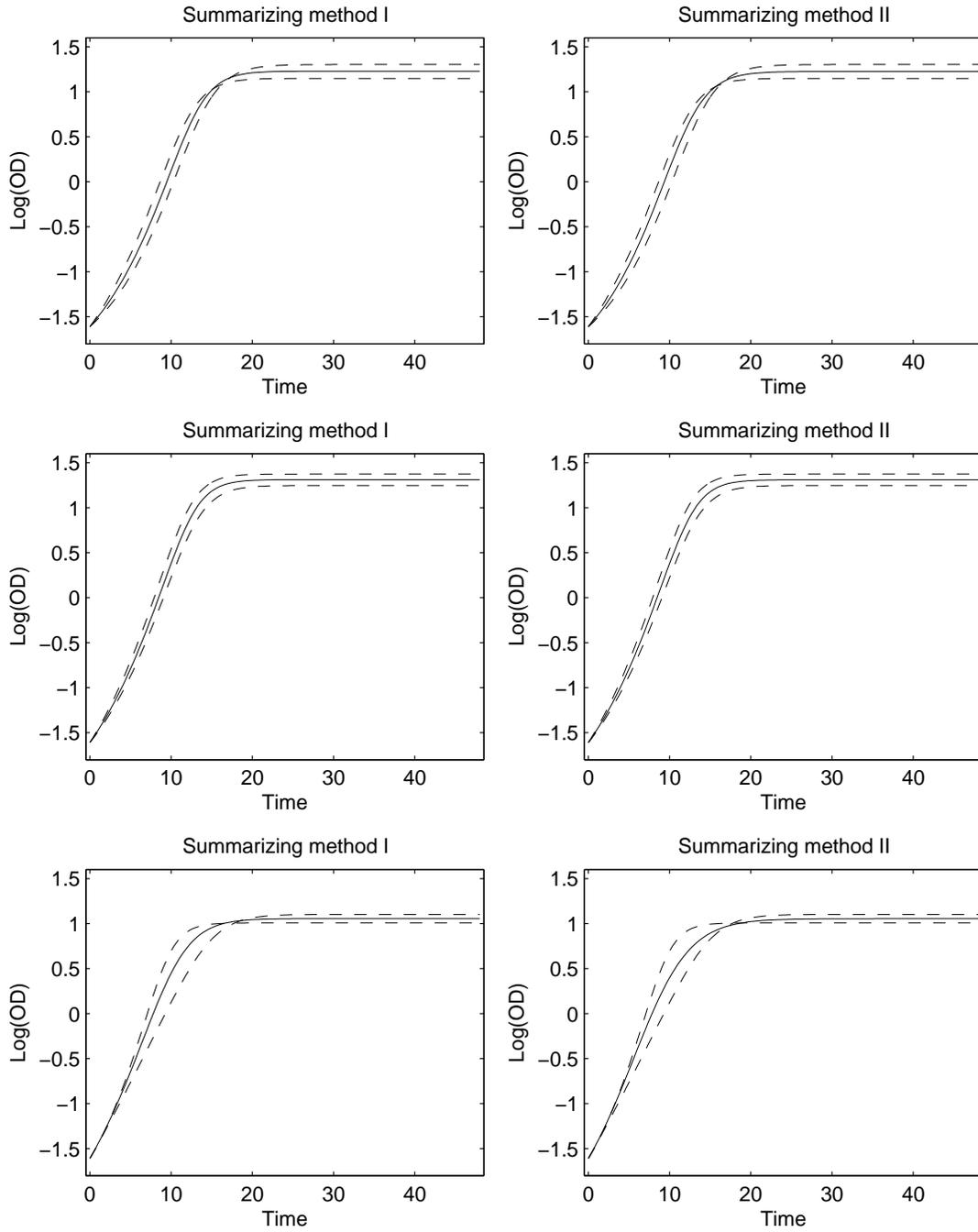


Figure 6.1: For three mutants: individually standardized curves (dashed) for both runs in 39°C (from top: 39E0307 and 39E0309, well 13; 39E0307 and 39E0309, well 25; 39C0307 and 39C0309, well 7) and their summary curves (solid).

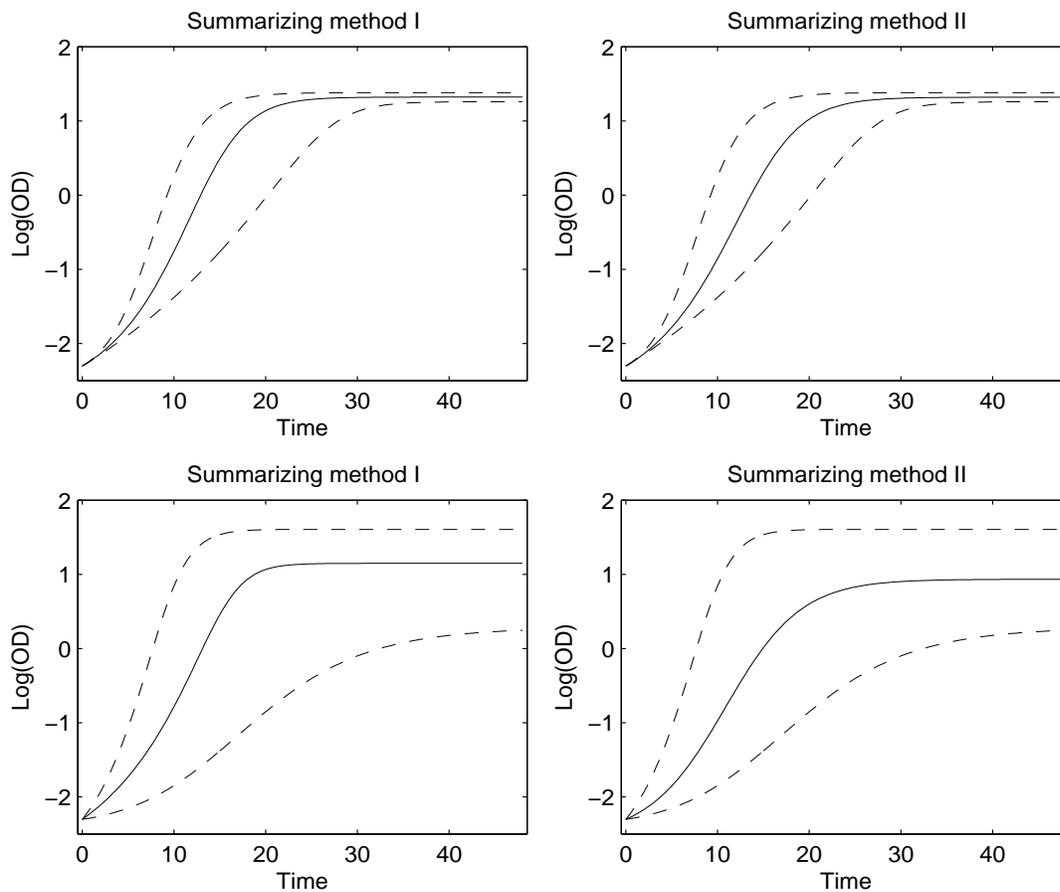


Figure 6.2: *Individually standardized curves (dashed) and their summary curves (solid). (Top) A mutant in caffeine and in reference condition (CAC0328 and NOC0305, well 4). (Bottom) A mutant in Dinitrophenol and in reference condition (DND0316 and NOD0305, well 8).*

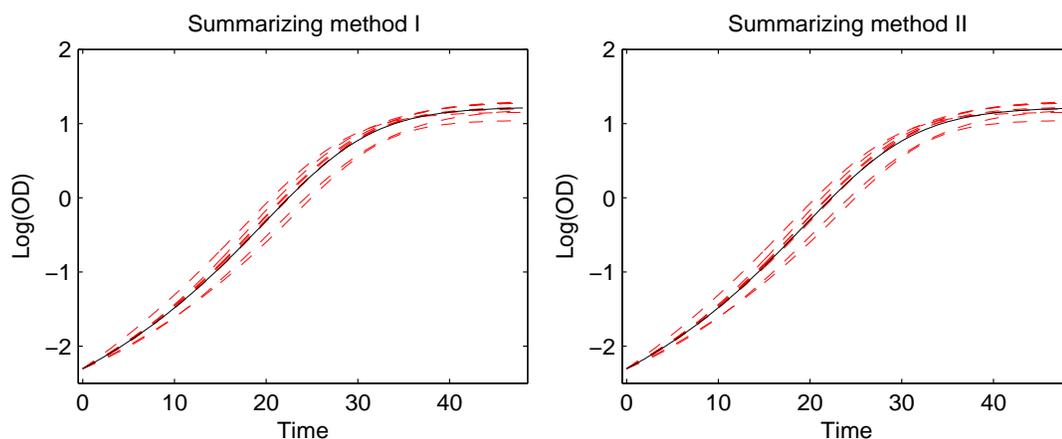


Figure 6.3: *Individually standardized curves (dashed) and their summary curve (solid) for the eight wild types in a run in Natrium chloride (NAC0323).*

Table 6.1: *Growth parameter values of the representative curves for the wild types in 39°C (see Figure 6.4).*

| Estimation method | λ | μ | Y |
|--|-----------|-------|-------|
| Standardized curve of all wild types | 0.238 | 0.241 | 3.326 |
| Summary curve (method I) of all individually standardized wild type curves | 1.484 | 0.275 | 3.384 |
| Summary curve (method I) of runwisely standardized wild type curves | 0.139 | 0.234 | 3.347 |
| Chapman-Richards method on each wild type, average value | 1.677 | 0.283 | 3.369 |

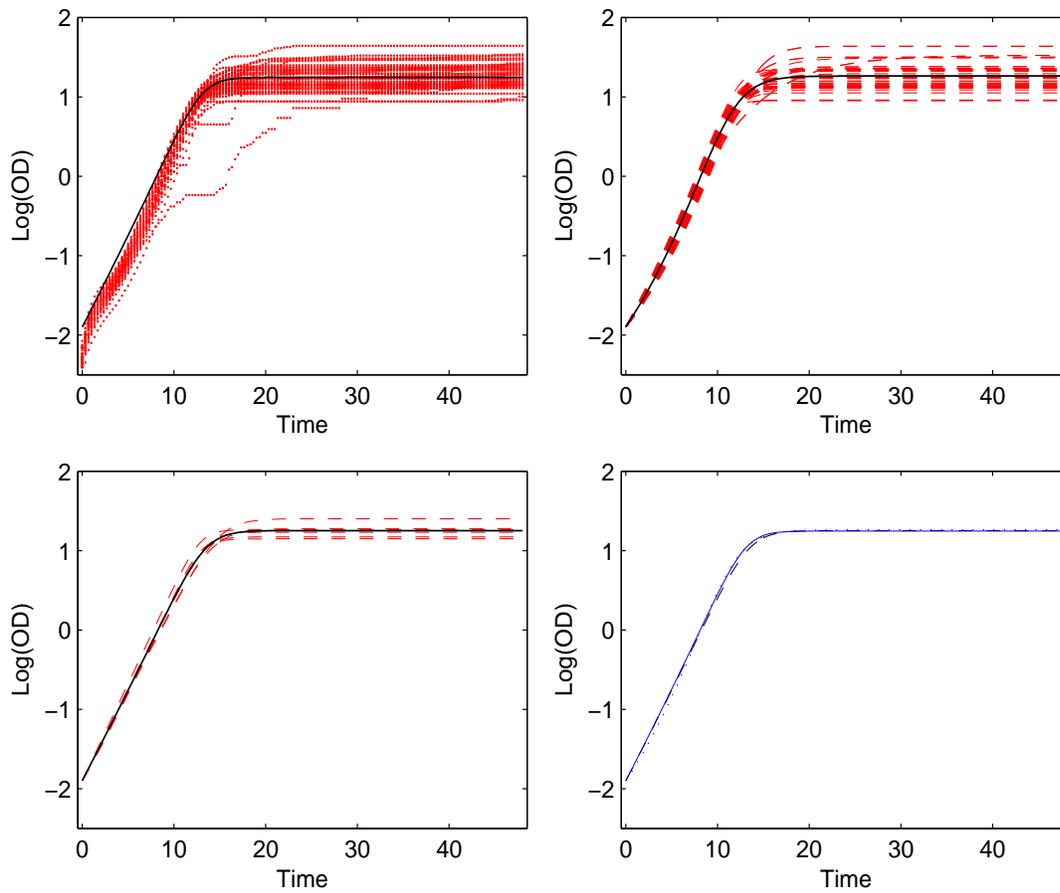


Figure 6.4: *Different ways to obtain a representative curve for all the wild types in 39°C. (Top, left): Wild type curves (dotted) and their standardized curve. (Top, right): Individually standardized wild type curves (dashed) and their summary curve (solid). (Bottom, left): Runwisely standardized wild type curves (dashed) and their summary curve (solid). (Bottom, right): In the same plot the standardized curve of all 48 wild types (solid), the summary curve of all individually standardized curves (dotted) and the summary curve of runwise wild type curves (dashed). The summary curves are fitted using method I.*

Chapter 7

Quality filters

In large-scale screenings, where hundreds of strains are measured in each run, a bad run may affect the results of hundreds of tests in the data analysis. Especially if the data from large-scale screenings are analyzed in an automatic way, it is of great importance to try to implement filters that automatically detect individual curves or whole runs that look atypical or spurious. In this chapter we will discuss the possibility to use the wild type controls in each run to identify dubious runs. We will also suggest a set of filtering methods that will address some of the problems with individual curves.

The motivation for having wild type controls in each run is twofold. First, we wish to neutralize the variability in the experimental conditions by comparing the behavior of the mutants with the behavior of the wild types in the same run. Second, the wild types are also there to control that the within run variability is reasonably stable. We will discuss how we can find dubious runs by using the growth parameters from the standardized wild type curves and by visually comparing their runwise summary curves.

In most of the data collected in PROPHECY, there are only two repeated measurements for each strain, so that it is rather hard to distinguish a bad behavior of a curve from the natural experimental variability of the two curves. However, *e.g.* the very fact that one of the curves may look nice and can be fitted by a standardized model curve, while the other cannot, is a sign of warning.

Coefficient of determination with a suitable threshold can be used to filter out individual curves that have atypical shapes and thus cannot be well described by the parametric model. This approach may also be applied to find collapsing curves. The OD values occasionally drop in successive time points long before the curve has entered the stationary phase. If this happens for several successive measurement time points, there is probably some aggregation of cells attaching to each other or to the wall of the well, and the measurements should not be trusted (an example is given in

Figure 7.1). However, if this only happens in single time points and after that the OD values are "normal" again, it is believed to be due to air or gas bubbles, and the rest of the measurements should not be too much affected. Most of the time the OD values drop in the end of the curves when they probably have reached the stationary phase (an example is given in Figure 7.2). In these cases the chosen smoothening (*i.e.* each OD value lower than the previous value is set to the previous value) will take care of this problem in a natural way. If the OD values drop before the curve has reached the stationary phase, the smoothening will typically make the estimated curve biased downwards. We will describe a simple filtering procedure to detect curves with this type of atypical behavior.

The samples that do not at least double in size are filtered away (an example is given in Figure 7.3). Some curves grow so slowly that at the last measurement time point they are still far from the stationary phase (an example is given in Figure 7.4). A simple filter to detect such curves will also be described. Yet another problem, which we will not treat in any formal way, is that in some experiments there seems to be no delay at all and the relative growth is maximal at time zero.

In the next chapter we will use the quality filtering techniques described here to compare the variability of so called logarithmic strain coefficients [6][26] for the growth parameters λ , μ , and Y estimated using ordinary Chapman-Richards, standardization, and summarizing methods. To do that we need to first define a set of quality filters and to discuss explicit choices for the thresholds.

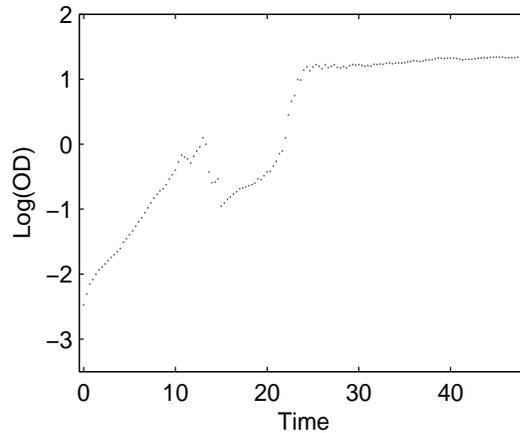


Figure 7.1: *An example of a curve which collapses before entering the stationary phase (39C0309, well 39). The OD values are calibrated and blank corrected, but they are not smoothened.*

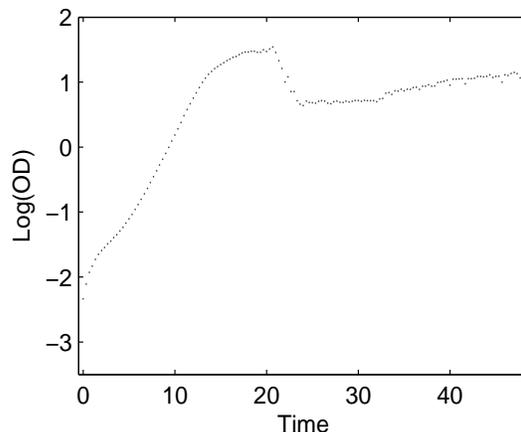


Figure 7.2: *An example of a curve which collapses after entering stationary phase (39C0309, well 83). The OD values are calibrated and blank corrected, but they are not smoothed.*

7.1 Quality filters for runs

The quality filters for the runs are based on the data of the eight wild types in each run.

Comparability of runs within environment

To investigate the comparability of the runs within specific environment, we first make runwise method I summary curves of the wild types. These summary curves are inspected visually. We also calculate coefficient of variations for λ , μ , and Y estimated from the summary curves. If at least one of the coefficient of variations is higher than a threshold, the runs are not considered comparable. In that case, either the deviating runs or the whole environment can be filtered away.

Within run variability

The within run variability is assessed by calculating the coefficient of variations of the λ , μ , and Y from the standardized wild type curves in each run. If the coefficient of variation for some run exceeds a threshold, the run will be filtered away.

7.1.1 Testing on data

The curves in Methylmethanesulfonate (MM) have so abnormal shapes that the Chapman-Richards model cannot describe them sufficiently well and thus no sum-

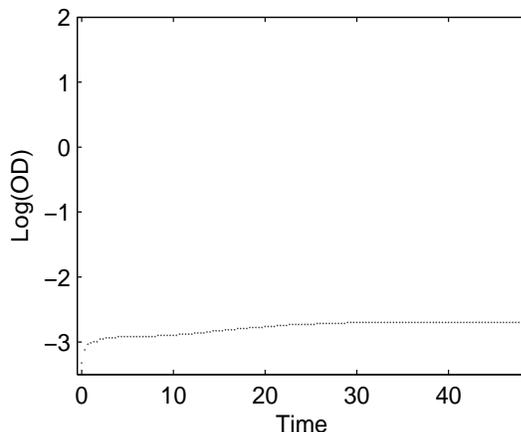


Figure 7.3: An example of a "non-growing" curve (41E0312, well 91). The OD values are calibrated and blank corrected, but they are not smoothed.

mary curves are fitted to the curves in this environment. The blank corrected and calibrated non-smoothened wild type curves in Methylmethanesulfonate are shown in Figure 7.5.

The runwise summary curves ($s_0 = 0.1$) for each environment except Methylmethanesulfonate are shown in Figure 7.6. In Dinitrophenol (DN) and Caffeine (CA) the stationary phase OD increments differ rather much between the runs, however, the shapes of the curves are similar. The coefficient of variations for the growth parameters from the summary curves are given in Table 7.1. The coefficient of variation for lag time is rather high in some environments, especially in $39^\circ C$, $41^\circ C$, and MV. In fact, in these environments there tends to be no delay, and the growth often slows down after a while. Thus calculating lag times in these environments may be questionable.

Appropriate thresholds for the coefficient of variations could be 10% for growth rate and 20% for stationary phase OD increment. Setting a threshold to the coefficient of variation for lag time is more complicated. The lag time itself is not a very robust measure and therefore either no threshold or a rather high threshold, *e.g.* 100%, for the coefficient of variation of lag time should be applied.

If at least one of the coefficient of variations exceeds the threshold, we conclude that there is something seriously wrong with the experiment and it should be redone or the results should be simply excluded from the analysis (or, at least, the parametric model should not be used). With these thresholds all the environments (except Methylmethanesulfonate which was excluded due to abnormal curve shapes) would pass the filter for the comparability of runs within environment.

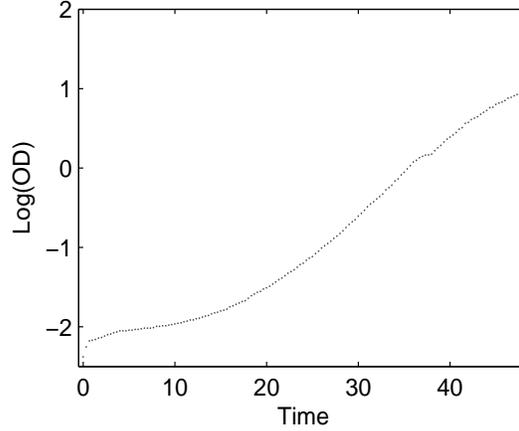


Figure 7.4: A growth curve (NAC0323, well 39) which has not reached the stationary phase at the last measurement time point. The OD values are calibrated and blank corrected, but they are not smoothed.

The coefficient of variations of the growth parameters from the standardized curves of the eight wild types in each run are given in Tables 7.2-7.3. Again, the coefficient of variations for lag time are high, especially in environments $39^{\circ}C$, $41^{\circ}C$, and MV. We will apply the following thresholds for the runwise coefficient of variations: 100% for the lag time, 15% for the growth rate, and 25% for the stationary phase OD increment. With these thresholds no runs are filtered out.

7.2 Quality filters for wild type curves

We try to find wild type curves that have not reached the stationary phase, collapse before reaching the stationary phase or cannot be sufficiently well described by the parametric model. We also filter out curves which deviate much from the others in the same run. All filters except (7.1) are based on the standardized curves. Appropriate thresholds for the different measures will be proposed in Section 7.2.1.

Curves that have not reached the stationary phase

We try to identify curves that have not reached the stationary phase by investigating the relation between the derivative of the fitted curve at the last time point (d_{end}) and the growth rate (μ). The derivative at the last time point is

$$d_{end} = \frac{\beta_0 \beta_1 \beta_2 e^{-\beta_2 t_{end}} (1 - \beta_1 e^{-\beta_2 t_{end}})^{\frac{1}{1-\beta_3}-1}}{1 - \beta_3},$$

where t_{end} is the last time point. Let $\Omega_{\frac{d_{end}}{\mu}}$ be the threshold for $\frac{d_{end}}{\mu}$. If $\frac{d_{end}}{\mu} \leq \Omega_{\frac{d_{end}}{\mu}}$, the curve is considered to have reached the stationary phase. The curves for which $\frac{d_{end}}{\mu} > \Omega_{\frac{d_{end}}{\mu}}$ are filtered out.

Curves that collapse before reaching the stationary phase

Curves that collapse before reaching the stationary phase are to be filtered away by first looking at the smoothed and non-smoothed OD-values (recall that all data are smoothed so that each OD value lower than previous value is corrected to equal the previous value). We calculate the absolute values of the differences of the OD values after and before smoothing, relative to the OD values after smoothing,

$$\omega = \left| \frac{OD_{\text{smoothened}} - OD_{\text{non-smoothened}}}{OD_{\text{smoothened}}} \right|, \quad (7.1)$$

until one hour after the stationary phase OD increment has been reached¹, ignoring the first five time points because the measurements tend to be shaky in the beginning. Of these ω 's, we take the third highest², and denote it by ω^* . Let Ω_{ω^*} be a threshold for ω^* . If $\omega^* > \Omega_{\omega^*}$ the curve is considered collapsing before reaching the stationary phase and it is filtered out. This method, however, fails to detect many collapsing curves. Therefore, we also investigate the coefficient of determination, r^2 , as given in (3.10).

Curves that cannot be well described with the parametric model

With the help of the coefficient of determination also curves that cannot be well described with the parametric model are detected.

Curves that deviate greatly from the other curves in the same run

The within run coefficient of variations of the growth rate (cv_{μ}) and stationary phase OD increment (cv_Y) of the remaining curves are investigated in order to detect curves that deviate greatly from the other curves in the same run. Let $\Omega_{cv_{\mu}}$ and Ω_{cv_Y} be the corresponding thresholds. When $cv_{\mu} > \Omega_{cv_{\mu}}$ or $cv_Y > \Omega_{cv_Y}$, the curve that deviates most from the others with respect to this parameter is removed. The coefficient of variations are calculated again, and the same procedure is repeated until both coefficient of variations are below the thresholds.

¹Defined as where the stationary phase OD increment according to the fit of the standardized curve has been obtained.

²The third highest value of ω is chosen so that the curves would not be filtered away because of a single collapsing OD value.

7.2.1 Testing on data

Before deciding on the thresholds for the quality filters for wild types, we tested how they would work on our data (*i.e.* all wild types in reference condition and in all environments except MM). We use a standard initial OD 0.1.

Most wild type curves have reached the stationary phase. An example of a wild type curve which may be considered not to have reached the stationary phase is shown in Figure 7.7. For this curve the $\frac{d_{end}}{\mu}$ is 0.138.

Figures 7.8 and 7.9 display the non-smoothened growth curves of the wild types of two runs in $39^{\circ}C$. The ω^* values (Table 7.4) might alarm about the wild type curve 7 in run 1. The wild type curve number 7 in run 2 that collapses already at an early stage might not be detected by investigating the ω^* . This curve can be detected by looking at its r^2 which is clearly smaller than the other wells' r^2 (Table 7.4). In fact, also the curve 7 in run 1 would have been detected by investigating its r^2 . Figure 7.10 shows another example of a curve that cannot be sufficiently well described by the model. For this curve the r^2 is 0.9066.

The previous steps filter out most of the deviating curves. The measures of cv_{μ} and cv_Y detect curves that deviate from the others even if they are otherwise rather "normal". These type of deviating curves are rare.

After having tested the filtering steps on our data, we propose the following thresholds: $\Omega_{\frac{d_{end}}{\mu}} = 0.08$, $\Omega_{r^2} = 0.995$, $\Omega_{\omega^*} = 0.3$, $\Omega_{cv_{\mu}} = 15\%$, $\Omega_{cv_Y} = 25\%$. The wild type curves in $39^{\circ}C$ which pass the quality filters using these thresholds are in black and the ones that do not are in grey in Figure 7.11. Roughly 96% of the wild type curves in all environments pass the quality filters.

7.3 Quality filters for mutant curves

With the quality filters for mutants, like for wild types, we try to find curves that have not reached the stationary phase, collapse before reaching the stationary phase or cannot be sufficiently well described by the parametric model. In addition, we try to find non-growing curves (this part is not included in the wild type quality filters because in our data there are no non-growing wild type curves). No replicate comparisons are done because there are only two replicates for each mutant (except in the reference condition). All filters, except (7.1) and when defining non-growing curves, are based on the standardized curves.

Non-growing curves

Non-growing curves are defined as the curves whose end OD value is less than twice the initial OD value.³ For the non-growing curves the lag time is set to 48 hours but no growth rate or stationary phase OD increment is calculated.

The curves that have not reached the stationary phase

The curves that have not reached the stationary phase are filtered out by investigating the derivative at the last time point, the same way as in case of wild types. The curves which are considered not to have reached the stationary phase are excluded from the analysis of stationary phase OD increment, but if they pass the other criteria of the quality assessment, they remain in the analysis of lag time and growth rate.

Curves that cannot be well described with the parametric model

Curves are to be excluded completely from the analysis of the data, if they cannot be fitted with the parametric model or collapse before reaching the stationary phase. These curves are detected in the same way as in case of wild types.

We tested the quality filters on all mutants in reference condition and in the six environments. The standard initial OD $s_0 = 0.1$ was used. The same thresholds as for wild types, *i.e.* $\Omega_{\frac{d_{end}}{\mu}} = 0.08$, $\Omega_{\omega^*} = 0.3$, and $\Omega_{r_2} = 0.995$ seem to work well.

Using the proposed thresholds, 5.6% of the mutant curves are filtered out totally, 4.7% are considered not to have reached the stationary phase but qualify for the analysis of lag time and growth rate while 0.14% are non-growing. These last ones are set to have a lag time 48 hours but are excluded from the analysis of growth rate and stationary phase OD increment. Note however, that when one of the two mutant curves is filtered out, the duplicate is not used in the analysis either. If at least one of the four mutant curves in the reference condition is filtered out, its replicates are not used in the analysis either (very few curves in the reference condition are filtered out). Taking also the duplicate/replicate exclusion into account, 9.5% of the curves are filtered out totally, 6.1% are filtered out from the analysis of the stationary phase OD increment but are included in the analysis of lag time and growth rate, and 0.2% are included in the analysis of lag time but excluded from the analysis of growth rate and stationary phase OD increment.

³The OD values are calibrated, blank corrected and smoothened.

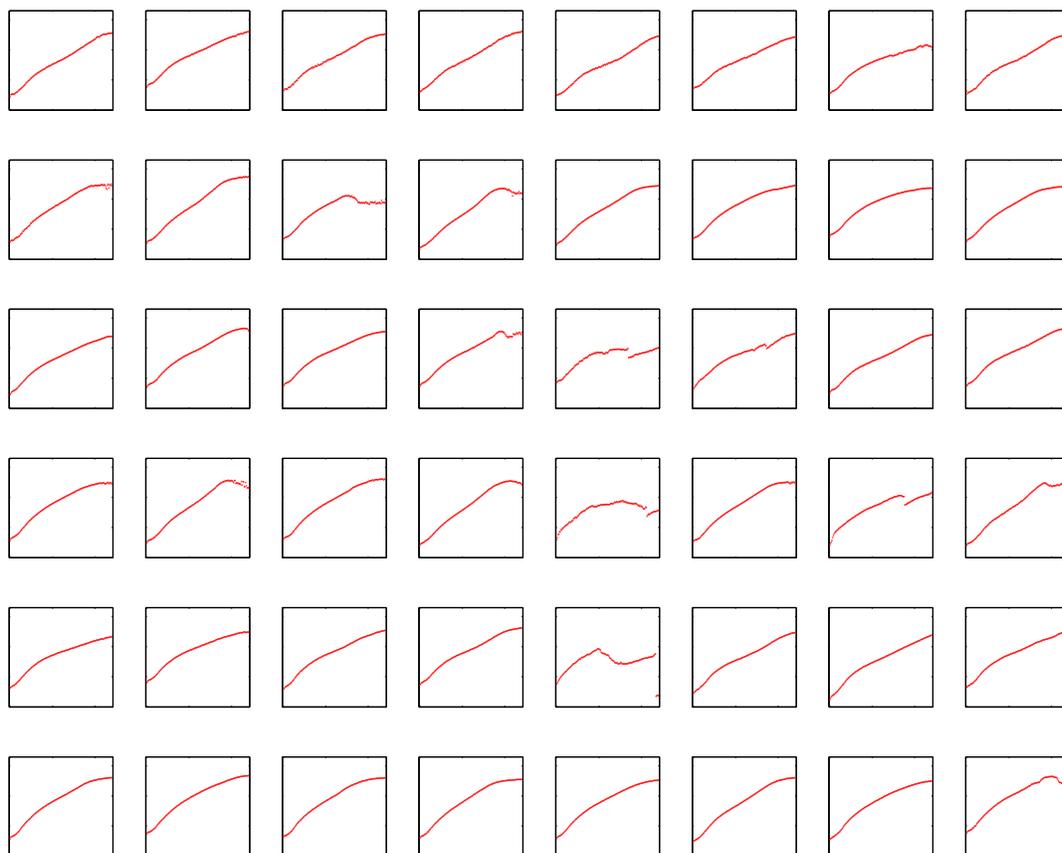


Figure 7.5: *Non-smoothed wild type curves in Methylmethanesulfonate (row-wise from the top: MMC0408, MMC0411, MMD0408, MMD0411, MME0408, MME0411.).*

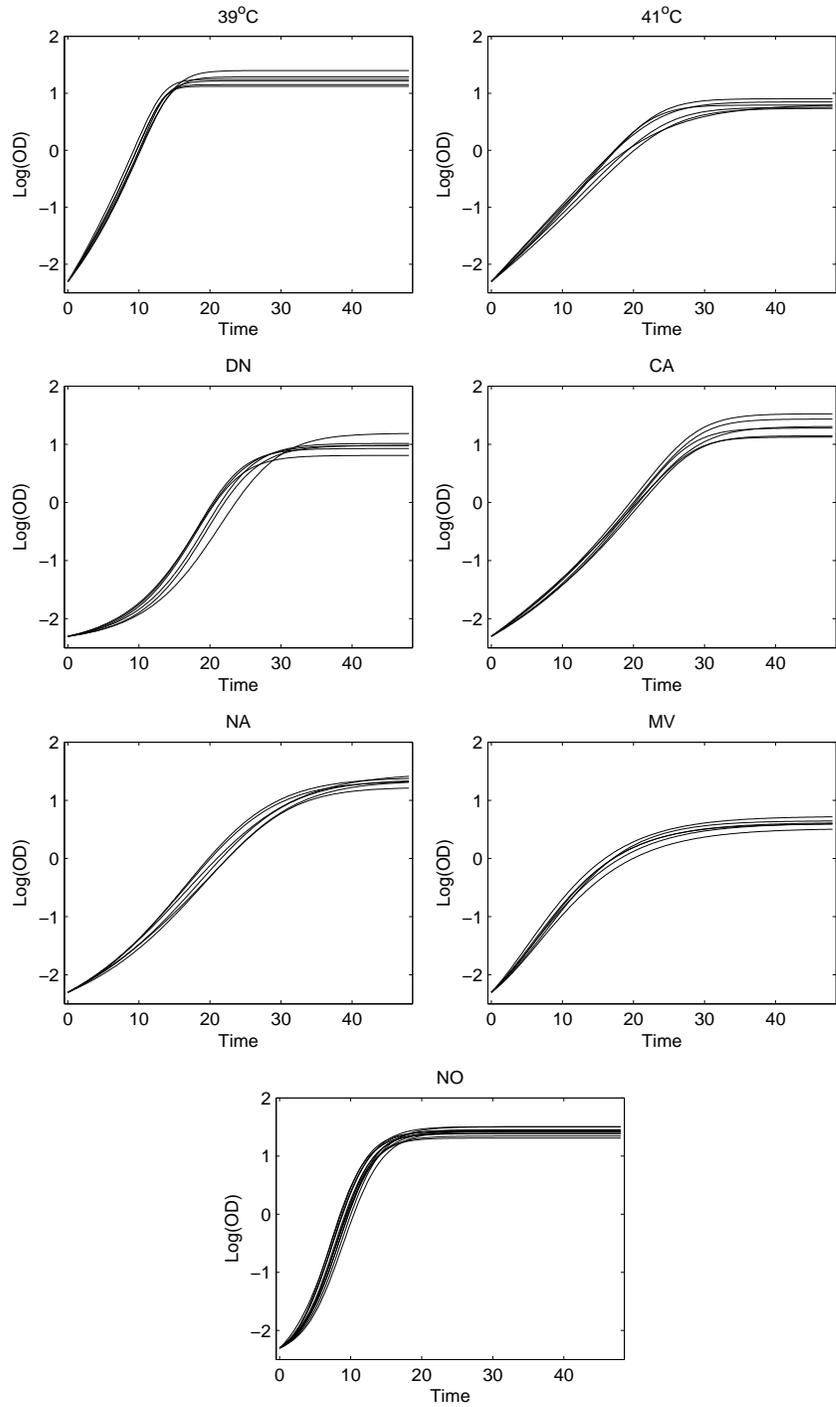


Figure 7.6: *Runwise summary curves of the eight wild types in each run.*

Table 7.1: *Coefficient of variations (%) for the growth parameters from the runwise summary curves of the eight wild types in each environment.*

| Environment | λ | μ | Y |
|-------------|-----------|-------|-------|
| 39°C | 24.24 | 3.26 | 10.51 |
| 41°C | 68.36 | 7.07 | 7.39 |
| DN | 9.52 | 3.75 | 13.25 |
| CA | 18.34 | 6.09 | 18.30 |
| NA | 15.87 | 4.18 | 8.85 |
| MV | 35.02 | 6.13 | 8.08 |
| NO | 12.17 | 3.61 | 4.67 |

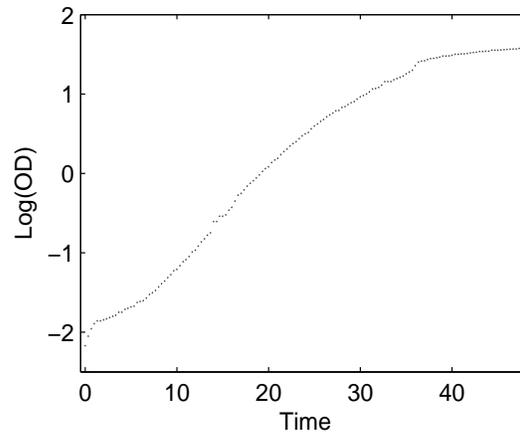


Figure 7.7: *A wild type growth curve (NAC0321, well 127) which may not have reached the stationary phase at the last measurement time point.*

Table 7.2: *Coefficient of variations (%) for the growth parameters from the eight standardized wild type curves in each run.*

| Run | λ | μ | Y |
|---------|-----------|-------|-------|
| 39C0307 | 42.82 | 8.19 | 16.62 |
| 39D0307 | 21.79 | 1.39 | 9.82 |
| 39E0307 | 10.72 | 1.75 | 8.28 |
| 39C0309 | 9.42 | 2.32 | 10.03 |
| 39D0309 | 12.84 | 2.76 | 16.30 |
| 39E0309 | 10.56 | 1.66 | 13.38 |
| 41C0312 | 27.20 | 3.21 | 11.27 |
| 41D0312 | 79.96 | 4.97 | 17.09 |
| 41E0312 | 85.15 | 2.31 | 6.99 |
| 41C0314 | 41.98 | 5.64 | 19.42 |
| 41D0314 | 56.52 | 4.05 | 15.60 |
| 41E0314 | 89.64 | 5.30 | 11.87 |
| DNC0316 | 3.81 | 3.50 | 21.43 |
| DND0316 | 13.03 | 9.05 | 17.89 |
| DNE0316 | 5.09 | 12.35 | 22.58 |
| DNC0319 | 7.24 | 4.86 | 17.82 |
| DND0319 | 11.05 | 10.11 | 19.17 |
| DNE0319 | 6.49 | 6.70 | 17.41 |
| CAC0328 | 12.10 | 5.01 | 15.36 |
| CAD0328 | 10.41 | 2.02 | 13.19 |
| CAE0328 | 14.23 | 4.15 | 13.49 |
| CAC0330 | 16.20 | 5.83 | 18.73 |
| CAD0330 | 11.09 | 2.07 | 23.27 |
| CAE0330 | 50.68 | 9.62 | 16.07 |
| NAC0321 | 12.85 | 1.24 | 7.44 |
| NAD0321 | 17.02 | 3.46 | 24.96 |
| NAE0321 | 22.37 | 8.28 | 9.72 |
| NAC0323 | 24.57 | 3.24 | 9.61 |
| NAD0323 | 11.97 | 3.38 | 5.15 |
| NAE0323 | 15.08 | 1.56 | 8.55 |

Table 7.3: *Coefficient of variations (%) for the growth parameters from the eight standardized wild type curves in each run.*

| Run | λ | μ | Y |
|---------|-----------|-------|-------|
| MVC0413 | 64.90 | 2.44 | 9.39 |
| MVD0413 | 46.50 | 0.83 | 6.03 |
| MVE0413 | 42.35 | 2.73 | 3.17 |
| MVC0417 | 42.49 | 3.78 | 8.48 |
| MVD0417 | 69.92 | 6.88 | 22.36 |
| MVE0417 | 59.00 | 2.70 | 12.29 |
| NOC0305 | 5.25 | 1.97 | 7.44 |
| NOD0305 | 16.41 | 7.43 | 6.43 |
| NOE0305 | 13.73 | 4.20 | 5.12 |
| NOC0326 | 2.38 | 1.28 | 5.22 |
| NOD0326 | 9.37 | 1.87 | 13.29 |
| NOE0326 | 9.37 | 1.87 | 13.30 |
| NOC0406 | 7.66 | 0.80 | 12.09 |
| NOD0406 | 3.56 | 0.94 | 4.61 |
| NOE0406 | 3.93 | 1.00 | 3.72 |
| NOC0426 | 10.77 | 1.63 | 12.70 |
| NOD0426 | 6.00 | 1.02 | 8.09 |
| NOE0426 | 3.24 | 1.07 | 4.51 |

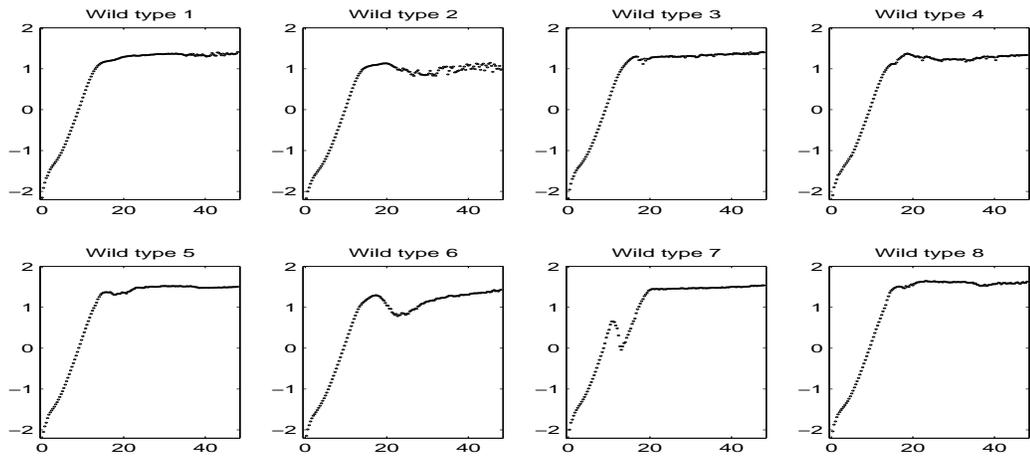


Figure 7.8: *The non-smoothened wild type growth curves of run in 39°C (39D0309).*

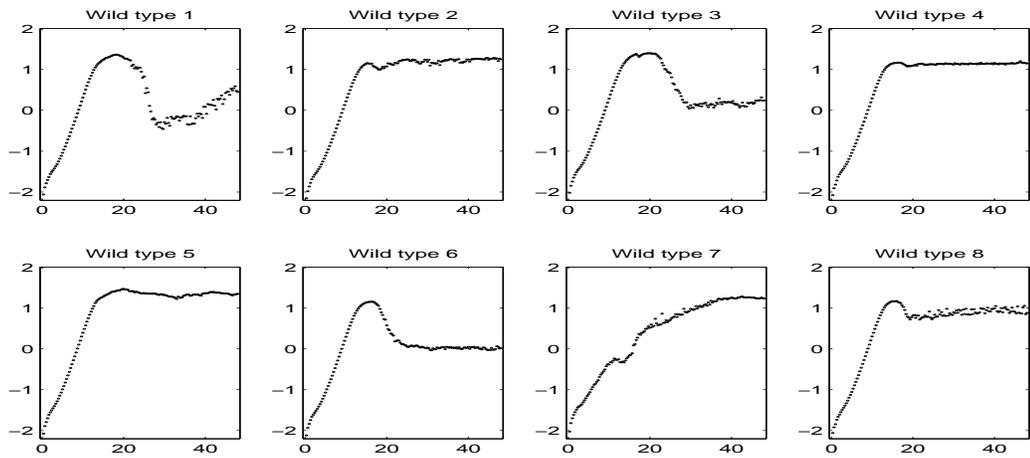


Figure 7.9: *The non-smoothened wild type growth curves of run in 39°C (39E0309).*

Table 7.4: The ω^* and the coefficient of determination (r^2) for the standardized wild type growth curves in two runs in 39°C.

| Wild type number | ω^* (39D0309) | ω^* (39E0309) | r^2 (39D0309) | r^2 (39E0309) |
|------------------|-------------------------|-------------------------|--------------------|--------------------|
| 1 | 0 | 0 | 0.9983 | 0.9997 |
| 2 | 0 | 0 | 0.9995 | 0.9983 |
| 3 | 0 | 0 | 0.9989 | 0.9997 |
| 4 | 0 | 0 | 0.9990 | 0.9997 |
| 5 | 0 | 0 | 0.9978 | 0.9996 |
| 6 | 0 | 0 | 0.9994 | 0.9998 |
| 7 | 0.4552 | 0.1876 | 0.9832 | 0.9901 |
| 8 | 0 | 0 | 0.9991 | 0.9997 |

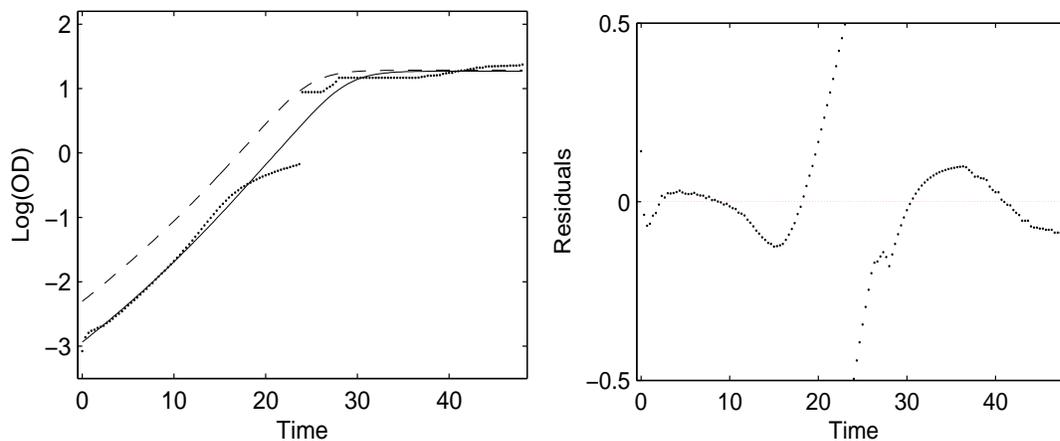


Figure 7.10: An example of a bad curve fit (MVD0417, well 171). The $\log(OD)$ values (dotted), the fitted growth curve (solid) and the standardized growth curve (dashed).

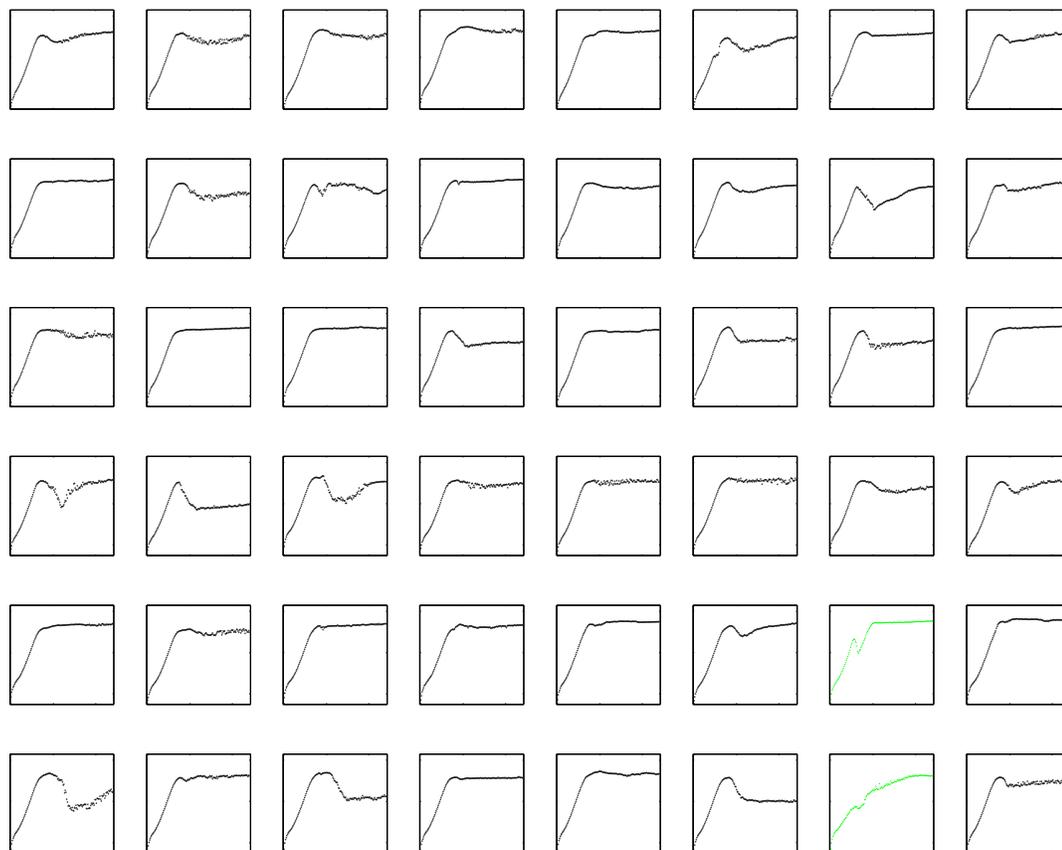


Figure 7.11: *The wild type growth curves in 39°C (each row representing a run, from the top: 39C0307, 39D0307, 39E0307, 39C0309, 39D0309, 39E0309). The ones that would pass the quality filters are in black and the ones that would not are in grey.*

Chapter 8

The effect of standardization and summarizing on logarithmic strain coefficients (LSC)

In the analysis of the data, the growth behavior of each mutant is related to the average behavior of the eight wild types in the same run, forming wild type normalized growth measures, termed runwise *logarithmic strain coefficients*, LSC_λ , LSC_μ , and LSC_Y .¹ The final LSC_λ , LSC_μ , and LSC_Y are the averages of the two (in environment) or four (in reference condition) runwise logarithmic strain coefficients. Furthermore, to provide quantitative measures of the specific gene-by-environment interactions and to compensate for general growth defects observed even under favorable growth conditions, LSC from environments are related to LSC from reference condition, forming *logarithmic phenotypic indexes*, LPI_λ , LPI_μ , and LPI_Y [26].

We are interested in whether standardization and summarizing have an effect on the logarithmic strain coefficients, and especially on the variance of the runwise LSC, *i.e.* the variance of the wild type normalized mutant replicates. We compare the LSC calculated based on the fitted (1) ordinary Chapman-Richards model curves, (2) standardized ($s_0 = 0.1$) curves for mutants and method I summary curves for wild types, and (3) standardized curves for mutants and method II summary curves for wild types. The LSC values are calculated on the data that pass the quality filters presented in Chapter 7. We see the replicates as a sample of size 2 (environments) or 4 (reference condition), which is motivated by that the repetitions are in different runs.

¹The terms used are $LSC_{\text{adaptation}}$, LSC_{rate} and $LSC_{\text{efficiency}}$ but we refer to these as LSC_λ , LSC_μ , and LSC_Y

8.1 LSC

The logarithmic strain coefficient for lag time for a specific mutant in a specific environment is calculated as

$$\begin{aligned} \text{LSC}_\lambda &= \frac{\sum_{r=1}^2 \left\{ \frac{1}{s} \sum_{k=1}^s \log(wt_k^{(r)}) - \log(x^{(r)}) \right\}}{2} \\ &= \frac{\text{LSC}_\lambda^{(1)} + \text{LSC}_\lambda^{(2)}}{2}, \end{aligned} \quad (8.1)$$

and in the reference condition as

$$\begin{aligned} \text{LSC}_{\lambda(0)} &= \frac{\sum_{r=1}^4 \left\{ \frac{1}{s} \sum_{k=1}^s \log(wt_{0,k}^{(r)}) - \log(x_0^{(r)}) \right\}}{4} \\ &= \frac{\text{LSC}_{\lambda(0)}^{(1)} + \text{LSC}_{\lambda(0)}^{(2)} + \text{LSC}_{\lambda(0)}^{(3)} + \text{LSC}_{\lambda(0)}^{(4)}}{4}, \end{aligned} \quad (8.2)$$

where $wt_k^{(r)}$ is the lag time of the k th wild type in the environment in the run r , s is the number of wild types (that remain in the data after the quality filtering, the maximum is eight) in the run, $wt_{0,k}^{(r)}$ is the lag time of the k th wild type in the reference condition in the run r , $x^{(r)}$ is the lag time of the mutant in the run r , and $x_0^{(r)}$ is the lag time of the mutant in the reference condition in the run r [6].

The logarithmic strain coefficients for growth rate and stationary phase OD increment are calculated analogously, except that for the LSC_μ , the doubling time, *i.e.* $\frac{\log(2)}{\mu}$, is used instead of the growth rate μ . The logarithmic phenotypic indexes for a specific mutant in a specific environment, are calculated as

$$\text{LPI}_\lambda = \text{LSC}_\lambda - \text{LSC}_{\lambda(0)} \quad (8.3)$$

$$\text{LPI}_\mu = \text{LSC}_\mu - \text{LSC}_{\mu(0)} \quad (8.4)$$

$$\text{LPI}_Y = \text{LSC}_{Y(0)} - \text{LSC}_Y. \quad (8.5)$$

8.1.1 The variance of runwise LSC

To investigate whether the standardization reduces the variance of the runwise logarithmic strain coefficients² we calculated the LSC in three different ways. First, using

²That is, the variance of $\text{LSC}^{(1)}$ and $\text{LSC}^{(2)}$, and the variance of $\text{LSC}_{(0)}^{(1)}$, $\text{LSC}_{(0)}^{(2)}$, $\text{LSC}_{(0)}^{(3)}$, and $\text{LSC}_{(0)}^{(4)}$, separately for each mutant in each environment and each growth parameter.

the growth parameters from the Chapman-Richards method in (8.1) and (8.2). Second and third, using the growth parameters from the standardized curves for the mutants as before, but for the wild types the method I and method II summary curve growth parameters. That is, instead of

$$\frac{1}{s} \sum_{r=1}^s \log(wt_r^{(k)}) \quad \text{and} \quad \frac{1}{s} \sum_{r=1}^s \log(wt_{0,r}^{(k)})$$

the logarithm of the specific growth parameter of the summary curve is taken. Note that the third way corresponds to using the growth parameters from the standardized curves in (8.1) and (8.2).

The averages of the LSC_λ , LSC_μ , and LSC_Y and the averages of the standard deviations of the runwise LSC_λ , LSC_μ , and LSC_Y in each environment and over all environments are shown in Table 8.1. The lag time estimation for the curves in environments $39^\circ C$ and $41^\circ C$ is questionable because there seems to be often almost no delay.

8.2 Discussion

Overall, the LSC-variances are slightly smaller with the standardizing and summarizing methods than with the direct Chapman-Richards approach. It is natural that the differences in LSC variances are not large between the three methods since the differences in initial OD values are rather small within runs (Table 3.1).

We used only $s_0 = 0.1$. It would have been interesting to do the LSC comparisons also with other values of s_0 . This will be the subject of further research.

Table 8.1: *Averages of the logarithmic strain coefficients and averages of the standard deviations of the runwise logarithmic strain coefficients.*

| Environment | Method | LSC_λ | \hat{s}_{LSC_λ} | LSC_μ | \hat{s}_{LSC_μ} | LSC_Y | \hat{s}_{LSC_Y} |
|-------------|------------|---------------|-------------------------|-----------|---------------------|---------|-------------------|
| 39°C | C-R | -0.13 | 0.15 | -0.07 | 0.03 | 0.02 | 0.12 |
| | Summary I | -0.09 | 0.22 | -0.07 | 0.03 | 0.02 | 0.12 |
| | Summary II | -0.17 | 0.30 | -0.07 | 0.03 | 0.02 | 0.12 |
| 41°C | C-R | 0.14 | 0.42 | -0.06 | 0.03 | 0.06 | 0.12 |
| | Summary I | 0.86 | 0.74 | -0.07 | 0.03 | 0.07 | 0.12 |
| | Summary II | 0.55 | 0.80 | -0.06 | 0.03 | 0.06 | 0.12 |
| DN | C-R | -0.05 | 0.07 | -0.02 | 0.07 | 0.06 | 0.24 |
| | Summary I | -0.06 | 0.06 | -0.01 | 0.07 | 0.07 | 0.24 |
| | Summary II | -0.06 | 0.06 | -0.01 | 0.07 | 0.06 | 0.24 |
| CA | C-R | -0.08 | 0.12 | -0.08 | 0.05 | -0.05 | 0.14 |
| | Summary I | -0.13 | 0.18 | -0.06 | 0.05 | -0.04 | 0.14 |
| | Summary II | -0.17 | 0.19 | -0.06 | 0.05 | -0.06 | 0.14 |
| NA | C-R | -0.10 | 0.09 | -0.01 | 0.03 | 0.01 | 0.08 |
| | Summary I | -0.12 | 0.08 | -0.01 | 0.03 | 0.01 | 0.08 |
| | Summary II | -0.13 | 0.08 | -0.01 | 0.03 | 0.01 | 0.08 |
| MV | C-R | -0.56 | 0.97 | -0.04 | 0.06 | -0.02 | 0.11 |
| | Summary I | -0.48 | 0.38 | -0.04 | 0.03 | -0.01 | 0.09 |
| | Summary II | -0.74 | 0.53 | -0.04 | 0.03 | -0.02 | 0.09 |
| NO | C-R | -0.07 | 0.13 | -0.04 | 0.05 | 0.01 | 0.09 |
| | Summary I | -0.05 | 0.12 | -0.04 | 0.04 | 0.02 | 0.09 |
| | Summary II | -0.05 | 0.12 | -0.04 | 0.04 | 0.02 | 0.09 |
| All | C-R | -0.14 | 0.29 | -0.05 | 0.05 | 0.01 | 0.13 |
| | Summary I | -0.14 | 0.28 | -0.04 | 0.04 | 0.01 | 0.13 |
| | Summary II | -0.05 | 0.24 | -0.04 | 0.04 | 0.02 | 0.13 |

Chapter 9

Conclusions

Modern genomics offers great opportunities for the study of measurement-related theoretical questions that are important in practice. In this thesis we focused on two problems related to the analysis of microbial growth: how to standardize the growth curves with respect to the initial population size, and how to estimate one curve from several experiments with different initial population sizes. We adopted the Chapman-Richards growth curves as our basic tool.

The Chapman-Richards model works well for a wide range of "normal" growth curves. However, for growth curves of atypical shapes the fit can be poor. Given the diversity of forms atypical curves assume, it is very difficult if not impossible to find a parametric model that fits sufficiently well all types of growth curves. One of the main causes of the bad fit with the Chapman-Richards model in our setting is that it assumes that the inflection point is after the first measurement time point, whereas in many atypical curves this does not seem to be the case. An inflection point before starting the measurements implies a negative lag time and that the maximum growth rate was obtained before starting the measurements. Hence, estimating lag time and growth rate from this type of curves is questionable.

Some of the concerns related to the growth parameter estimation do not directly depend on the model used. Warringer and Blomberg [25] 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 or not. 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 rapidly 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. Ericson *et al* [6] are currently working with another type of lag time definition than the one we have used.

The lag time and the growth rate depend strongly on the initial population size. However, in large scale experiments, it is difficult to have the same constant initial population size. We introduced a method to standardize growth curves with respect to the initial population size. We use a certain Chapman-Richards model to try to predict what the behavior of a growth curve would have been, had the population had a standard initial population size. The standardization reduces remarkably the initial population size correlation with the lag time and growth rate, compared to the ordinary Chapman-Richards method. It is also very useful for visualizing data: without standardization, it is difficult to know what the difference between the curves is. We found that the differences between the initial population sizes tend to be larger between environments than within environments (Table 3.1). Therefore, the standardization is important especially because it enables comparisons of curves from different environments. Furthermore, it will be of great value when clustering on the whole curves is desired.

We introduced two ways to construct a summary curve from standardized curves, in order to represent repetitions of similar growth experiments by a single curve. They are based on the averages of the growth parameters (method I), or on the averages of the logarithms of the growth parameters (method II) of the curves to be summarized. We showed that the method II summary curves always exist whereas the method I summary curves do not always exist, although the problem seems to be of minor practical relevance.

The standardized and the summary curves could be a natural complement to the phenotypic library Warringer *et al* [27], and Fernandez-Ricaud *et al* [7] have built. For example, a standardized curve for each mutant in each run and a summary curve for the eight wild types in each run could be made available in PROPHECY. Furthermore, a web tool to analyze the yeast growth data using the standardizing method, and to detect individual curves or whole runs that are atypical or spurious, could be developed.

More research on how to choose the standard initial population size is needed. One direction of study is to use the eight wild types in each run, and investigate the mean and the variance of their growth parameters from standardized curves with different standard initial population sizes. It would also be interesting to compare the standardizing method to the Warringer method [26] used today in PROPHECY, *e.g.* by comparing the LSC-values and their variances. This is the subject of future research.

The initial population size correlation with lag time and growth rate could be an artifact of the calibration curve function or the model. However, we do not believe that it is due to the calibration curve function, because the correlation reduces remarkably with the standardization. We do not believe that it is due to the model either, because the correlation is high also when using the Warringer method [26] to estimate the growth parameters. We do believe that it could be a biological effect, *i.e.* that the maximum growth rate cannot be reached if the initial population size is too large. This has not been tested properly. Therefore, studies with very small initial OD values in parallel with initial OD values of the size we have now should be done to verify whether this really is the case.

The quality filters presented in this thesis probably need to be developed further and complemented. Some of the problems related to the shapes of the growth curves may be due to a slightly false calibration curve function or due to a different (or varying) real blank than the one used in our subtraction. Both of these issues require further research. The measurements that are mostly affected by the blank are those in the very beginning of the logarithmic growth curve. Therefore, it may be relevant to study the effect of the first measurement points on the estimated growth parameters. This can be done for example by systematically comparing the estimated growth parameters and their variances, when the first measurement point is ignored, the first two measurement points are ignored, the first three measurement points are ignored, etc. One may also try to model the blank using a Bayesian type approach [18] so that it can differ from the fixed blank value with a penalty in the least squares procedure. Also other smoothing methods besides the one we used, where each OD value lower than the previous value is set to the previous value, may be considered. One alternative is to set each OD value lower than the previous value to the average of the logarithms of the previous and the next value.

It would be interesting to study the possibility to standardize growth curves using a non-parametric sigmoidal model. Standardizing upwards can probably be done approximately the same way as it was done in this thesis, but it may be more difficult to standardize downwards. Some attempts to estimate growth parameters using a non-parametric sigmoidal model are done by Warringer *et al* [24].

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] T.F. Coleman and Y. Li. On the Convergence of Reflective Newton Methods for Large-Scale Nonlinear Minimization Subject to Bounds. *Mathematical Programming*, 67(2):189–224, 1994.
- [4] T.F. Coleman and Y. Li. An Interior, Trust Region Approach for Nonlinear Minimization Subject to Bounds. *SIAM Journal on Optimization*, 6:418–445, 1996.
- [5] E. Ericson. *Large-scale phenotypic analysis of Saccharomyces cerevisiae deletion mutants*. Licentiate thesis, Department of Cell and Molecular Biology, Microbiology, Göteborg University, 2004.
- [6] E. Ericson, L. Fernandez, I. Pylvänäinen, J. Warringer, O. Nerman, and A. Blomberg. Analysis of 576 *Saccharomyces cerevisiae* Deletion Strains Revealed Significant Phenotypes for as Many Unknown as Known Genes. *In preparation*.
- [7] L. Fernandez-Ricaud, J. Warringer, E. Ericson, I. Pylvänäinen, G.J.L. Kemp, O. Nerman, and A. Blomberg. PROPHECY - a database for high-resolution phenomics. *Nucleic Acids Research*, 33:D1–D5, 2005.
- [8] W.E. Garthright. Refinements in the Prediction of Microbial Growth Curves. *Food Microbiology*, 8:239–248, 1991.
- [9] B. Gompertz. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philos. Trnas. R. Soc. London*, 115:513–585, 1825.
- [10] 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.

- [11] 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.
- [12] I. Pylvänäinen. *Modeling Yeast Growth Curves*. Licentiate Thesis, Department of Mathematical Statistics, Chalmers University of Technology and Göteborg University, 2003.
- [13] F.J. Richards. A Flexible Growth Function for Empirical Use. *Journal of Experimental Botany*, 10(29):290–300, 1959.
- [14] A. Ruepp, A. Zollner, D. Maier, K. Albermann, J. Hani, M. Mokrejs, I. Tetko, U. Güldener, G. Mannhaupt, M. Münsterkötter, and H.W. Mewes. The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. *Nucleic Acids Research*, 32(18):5539–5545, 2004.
- [15] L. Råde and B. Westergren. *Mathematics handbook for science and engineering*. Studentlitteratur, 2001.
- [16] J. Schnute. A versatile growth model with statistically stable parameters. *Can. J. Fish. Aquat Sci*, 38:1128–1140, 1981.
- [17] P. Singleton and D. Sainsbury, editors. *Dictionary of Microbiology and Molecular Biology*. John Wiley & Sons, 3rd edition, 2001.
- [18] D. Sorensen and D. Gianola. *Likelihood, Bayesian and MCMC methods in quantitative genetics*. Springer-Verlag, New York, 2002.
- [19] C.J. Stannard, A.P. Williams, and P.A. Gibbs. Temperature/Growth Relationship for Psychotrophic Food-Spoilage Bacteria. *Food Microbiology*, 2:115–122, 1985.
- [20] J.N. Strathern, E.W. Jones, and J.R. Broach. *The Molecular Biology of the Yeast Saccharomyces, Life Cycle and Inheritance*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1981.
- [21] L. von Bertalanffy. A quantitative theory of organic growth (Inquiries on growth laws. II). *Human Biology*, 10:181–213, 1938.
- [22] L. von Bertalanffy. *Teoretische Biologie*. A Franke AG Verlag, Bern, 1951.
- [23] J. Warringer. *Genetic robustness during environmental stress*. Doctoral Thesis, Department of Cell and Molecular Biology, Göteborg University, 2003.

- [24] J. Warringer, D. Anevski, and A. Blomberg. Quantitative Assessment of Complex Growth Dynamics during Environmental Stress. *In preparation*.
- [25] J. Warringer and A. Blomberg. Automated Screening in Environmental Arrays Allows Analysis of Quantitative Phenotypic Profiles in *Saccharomyces Cerevisiae*. *Yeast*, 20(1):53–67, 2003.
- [26] J. Warringer, E. Ericson, L. Fernandez, O. Nerman, and A. Blomberg. High-resolution yeast phenomics resolves different physiological features in the saline response. *PNAS*, 100(26):15724–15729, 2003.
- [27] J. Warringer, E. Ericson, I. Pylvänäinen, O. Nerman, and A. Blomberg. Revealing Quantitative Phenotypic Profiles. *Submitted*.
- [28] 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.
- [29] F. Zhang. *Matrix Theory: Basic Results and Techniques*. Springer-Verlag, New York, 1999.
- [30] 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

Figures

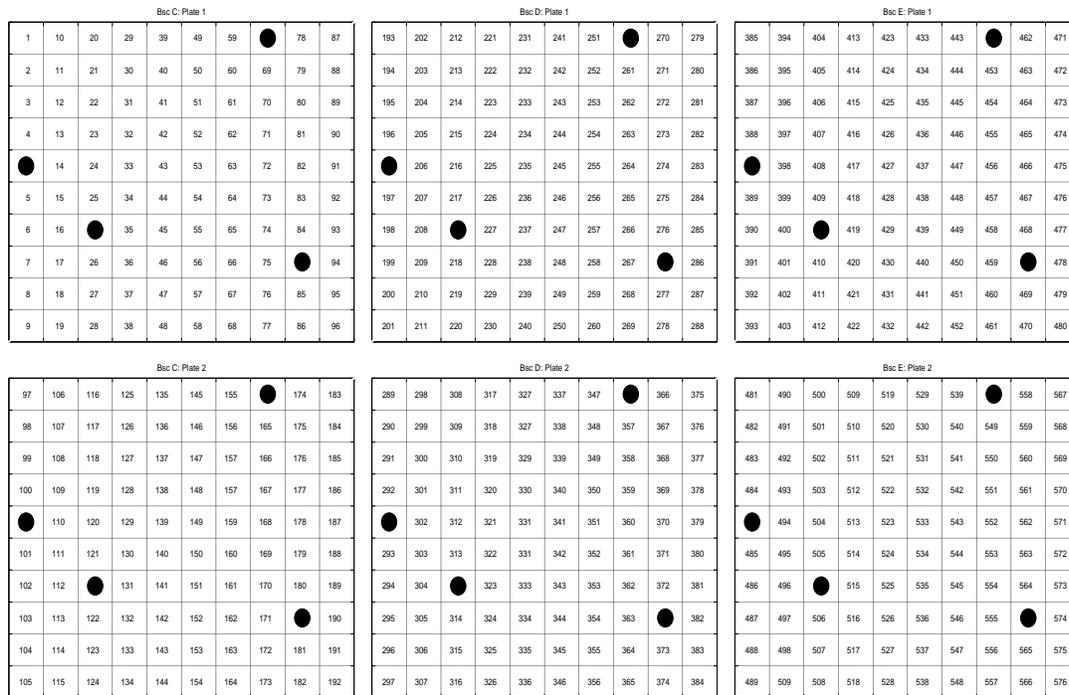


Figure A.1: *In the motivating dataset: the positioning of the mutants (numbered) and wild types (balls) on the plates and on the three different Bioscreen instruments (C, D and E).*

Appendix B

Tables

Table B.1: 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 subtracted from all the OD values 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 B.2: *Well specific means of the calibration curve function data. (Well specific blanks are subtracted from all the OD values 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 |

Appendix C

Gompertz augmented Chapman-Richards model

The Chapman-Richards curves are not defined at $\beta_3 = 1$, but the limiting forms when $\beta_3 \rightarrow 1$ and β_1 tends to 0 in a subordinated rate, are members of the Gompertz family; Writing $\beta_1 = e^b(1 - \beta_3)$, $b > 0$, we define the Gompertz augmented Chapman-Richards model as

$$g_t = \beta_0 \left[1 - e^b(1 - \beta_3)e^{-\beta_2 t} \right]^{1/(1-\beta_3)} + D, \quad \text{for } \beta_3 \neq 1, \quad (\text{C.1})$$

$$g_t = \beta_0 e^{-e^b - \beta_2 t} + D, \quad \text{for } \beta_3 = 1. \quad (\text{C.2})$$

It is straightforward to see that $\beta_3 \rightarrow 1$ implies that g_t defined by (C.1) converges to g_t defined by (C.2). The parameters s , d_0 , λ , μ and Y of the Gompertz function are:

$$s = e^{\beta_0 e^{-e^b} + D} \quad (\text{C.3})$$

$$d_0 = \beta_0 \beta_2 e^b e^{-e^b} \quad (\text{C.4})$$

$$\lambda = \frac{\frac{b}{e} + e^{-e^b} - \frac{1}{e}}{\frac{\beta_2}{e}} \quad (\text{C.5})$$

$$\mu = \frac{\beta_0 \beta_2}{e} \quad (\text{C.6})$$

$$Y = e^{\beta_0 + D} - e^s. \quad (\text{C.7})$$

We will use these equalities to prove Lemma 1.

Lemma 1 *The Gompertz curve corresponding to any hybrid parameter combination $s > 0$, $0 < d_0 < \mu$, $\lambda > 0$, $\mu > 0$, and $\beta_3 = 1$ is unique. The parameter b is the solution of the equation $b + 1 - e^b = \log(\frac{d_0}{\mu})$, and the three other parameters are given by $\beta_0 = \frac{\mu\lambda}{\frac{b}{e} + e^{-e^b} - \frac{1}{e}}$, $\beta_2 = \frac{b + e^{-e^b + 1} - 1}{\lambda}$, and $D = \log(s) - \frac{\lambda\mu e^{-e^b}}{\frac{b}{e} + e^{-e^b} - \frac{1}{e}}$. Furthermore, the stationary phase OD increment is*

$$Y = e^{\frac{1 - e^{-e^b}}{\frac{b-1}{e} + e^{-e^b}} \lambda \mu + \log(s)} - s.$$

Proof. Fix a parameter combination $s > 0$, $0 < d_0 < \mu$, $\lambda > 0$ and $\mu > 0$. From (C.4) and (C.6) we get

$$\begin{aligned} \frac{d_0}{\mu} &= \frac{\beta_0 \beta_2 e^b e^{-e^b}}{\frac{\beta_0 \beta_2}{e}} \\ &= e^{b+1} e^{-e^b}, \end{aligned}$$

which determines b (unique solution). We get from (C.5)

$$\beta_2 = \frac{b + e^{-e^b + 1} - 1}{\lambda}, \quad (\text{C.8})$$

from (C.8) and (C.6)

$$\beta_0 = \frac{\mu\lambda}{\frac{b}{e} + e^{-e^b} - \frac{1}{e}}, \quad (\text{C.9})$$

and from (C.3) and (C.9)

$$D = \log(s) - \frac{\lambda\mu e^{-e^b}}{\frac{b}{e} + e^{-e^b} - \frac{1}{e}}. \quad (\text{C.10})$$

Now, using (C.9) and (C.10) in (C.7), we get

$$Y = e^{\frac{1 - e^{-e^b}}{\frac{b-1}{e} + e^{-e^b}} \lambda \mu + \log(s)} - s.$$

□

The Lemma 1 implies that for the unit-scaled model

$$A = \frac{1 - e^{-e^b}}{\frac{b-1}{e} + e^{-e^b}}. \quad (\text{C.11})$$

Lemma 3 *Fix $0 < d_0 < 1$. In the unit-scaled model, the function A defined in (4.10) with the constraint (4.6) is continuous at $\beta_3 = 1$ as a function of $\beta_3 > 0$.*

Proof. Define b^* so that $e^{b^*+1}e^{-e^{b^*}} = d_0$, and assume that

$$\limsup_{\beta_3 \rightarrow 1} b = b^* + \gamma.$$

Fix a sequence $\beta_3^{(n)} \rightarrow 1$ such that, for the corresponding $b^{(n)}$ -sequence,

$$\lim_{n \rightarrow \infty} b^{(n)} = b^* + \gamma.$$

However, by rewriting (4.6) we conclude that (recall that $\beta_1 = e^b(1 - \beta_3)$)

$$\begin{aligned} d_0 &= \lim_{n \rightarrow \infty} d_0 \\ &= \lim_{n \rightarrow \infty} \frac{e^{b^{(n)}} \left(1 - e^{b^{(n)}}(1 - \beta_3^{(n)})\right)^{\frac{\beta_3^{(n)}}{1 - \beta_3^{(n)}}}}{\beta_3^{(n) \frac{\beta_3^{(n)}}{1 - \beta_3^{(n)}}}} \\ &= e^{b^* + \gamma + 1} e^{-e^{b^* + \gamma}}, \end{aligned}$$

which forces γ to equal 0 and thus $\limsup_{\beta_3 \rightarrow 1} b = b^*$. Analogously also $\liminf_{\beta_3 \rightarrow 1} b$ can be shown to equal b^* . Hence, $b \rightarrow b^*$ as $\beta_3 \rightarrow 1$. Using this and taking limit in (4.10), we get

$$\begin{aligned} \lim_{\beta_3 \rightarrow 1} A &= \lim_{\beta_3 \rightarrow 1} \frac{1 - (1 - e^b(1 - \beta_3))^{\frac{1}{1 - \beta_3}}}{\beta_3^{\frac{\beta_3}{1 - \beta_3}} \left[\log \left(\frac{e^b(1 - \beta_3)}{1 - \beta_3} \right) - \beta_3 \right] + (1 - e^b(1 - \beta_3))^{\frac{1}{1 - \beta_3}}} \\ &= \frac{1 - e^{-e^{b^*}}}{\frac{b^* - 1}{e} + e^{-e^{b^*}}}, \end{aligned}$$

which equals A in the Gompertz case (C.11). Hence, A is continuous at $\beta_3 = 1$. \square

We finish this appendix with two continuity remarks.

Remark 1. In the unit-scaled model for fixed $0 < d_0 < 1$ it is easy to see that also the parameters β_0 , β_2 and D converge when $\beta_3 \rightarrow 1$: setting (4.3), (4.4) and (4.1) to one and using $\beta_1 = e^b(1 - \beta_3)$, we obtain the equations for β_2 , β_0 and D , and taking limits we get

$$\begin{aligned} \lim_{\beta_3 \rightarrow 1} \beta_2 &= \lim_{\beta_3 \rightarrow 1} \frac{(1 - e^b(1 - \beta_3))^{\frac{1}{1-\beta_3}} - \beta_3^{\frac{1}{1-\beta_3}} + \beta_3^{\frac{\beta_3}{1-\beta_3}} \log\left(\frac{e^b(1-\beta_3)}{1-\beta_3}\right)}{\beta_3^{\frac{\beta_3}{1-\beta_3}}} \\ &= b^* + e^{-e^{b^*}+1} - 1, \\ \lim_{\beta_3 \rightarrow 1} \beta_0 &= \lim_{\beta_3 \rightarrow 1} \frac{1}{\beta_2 \beta_3^{\frac{\beta_3}{1-\beta_3}}} \\ &= \frac{1}{\frac{b^*}{e} + e^{-e^{b^*}} - \frac{1}{e}}, \\ \lim_{\beta_3 \rightarrow 1} D &= \lim_{\beta_3 \rightarrow 1} -\beta_0(1 - e^b(1 - \beta_3))^{\frac{1}{1-\beta_3}} \\ &= \frac{-e^{-e^{b^*}}}{\frac{b^*}{e} + e^{-e^{b^*}} - \frac{1}{e}}, \end{aligned}$$

where b^* again solves $d_0 = e^{b^*+1}e^{-e^{b^*}}$. Thus all the limiting parameters β_2 , β_0 and D converge to the Gompertz parameters as claimed.

Remark 2. Not only the growth parameters but also the whole Gompertz growth curves interpolate the Chapman-Richards. Consider the unit-scaled Chapman-Richards model and fix $0 < d_0 < 1$. Denote $\lim_{\beta_3 \rightarrow 1} \beta_0 = \beta_0^*$, $\lim_{\beta_3 \rightarrow 1} \beta_2 = \beta_2^*$ and $\lim_{\beta_3 \rightarrow 1} D = D^*$. Consider the Chapman-Richards curve

$$g_t = \beta_0 \left[1 - e^b(1 - \beta_3)e^{-\beta_2 t} \right]^{1/(1-\beta_3)} + D$$

for fixed t . Since

$$\left[1 - e^b(1 - \beta_3)e^{-\beta_2 t} \right]^{1/(1-\beta_3)} = e^{-e^{b^*}e^{-\beta_2^* t}} (1 + o(1)), \quad \text{for all } t,$$

where $o(1) \rightarrow 0$ as $\beta_3 \rightarrow 1$, we also get that

$$\lim_{\beta_3 \rightarrow 1} \left(\beta_0 \left[1 - e^b(1 - \beta_3)e^{-\beta_2 t} \right]^{1/(1-\beta_3)} + D \right) = \beta_0^* e^{-e^{b^*}e^{-\beta_2^* t}} + D^*,$$

which is the Gompertz curve evaluated in this t .

Appendix D

Discussion of Conjecture 1

Conjecture 1 *Fix $0 < d_0 < 1$. In the unit-scaled model, the function A defined in (4.10) with the constraint (4.6) is strictly decreasing as a function of $\beta_3 > 0$.*

Figure D.1 shows a graph where $\log(\log(A + 1) + 1)$ is plotted against β_3 between 0 and 100. Each curve corresponds to a different d_0 (between 0.01 to 0.99). It is easy to see that A is decreasing as a function of β_3 . The derivatives of A are plotted against β_3 in Figures D.2 (small values of β_3) and D.3 (large values of β_3) in order to see that A is strictly decreasing as a function of β_3 .

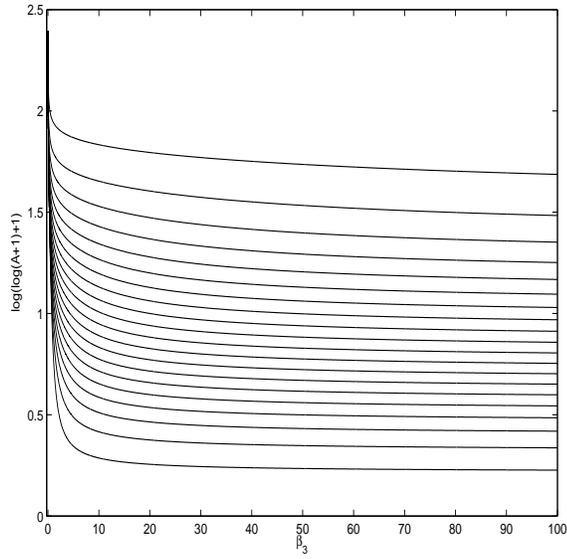


Figure D.1: $\log(\log(A + 1) + 1)$ is plotted against β_3 , each curve corresponds to a different d_0 (between 0.01 and 0.99). The larger d_0 , the larger value of $\log(\log(A + 1) + 1)$.

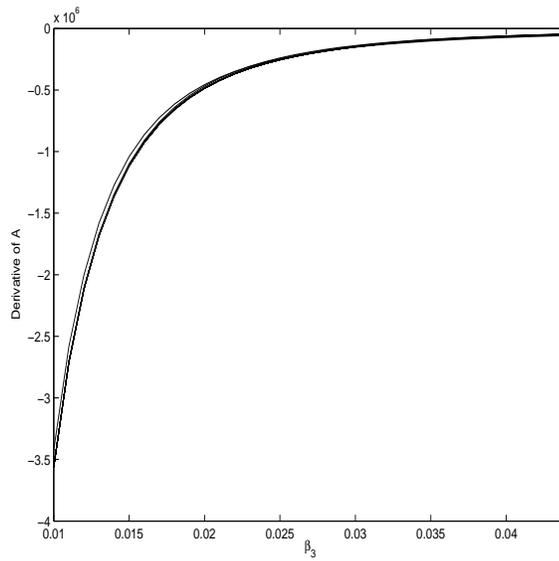


Figure D.2: Derivative of A plotted against small values of β_3 . Each curve corresponds to a different d_0 (between 0.01 and 0.99). The larger d_0 , the smaller derivative of A .

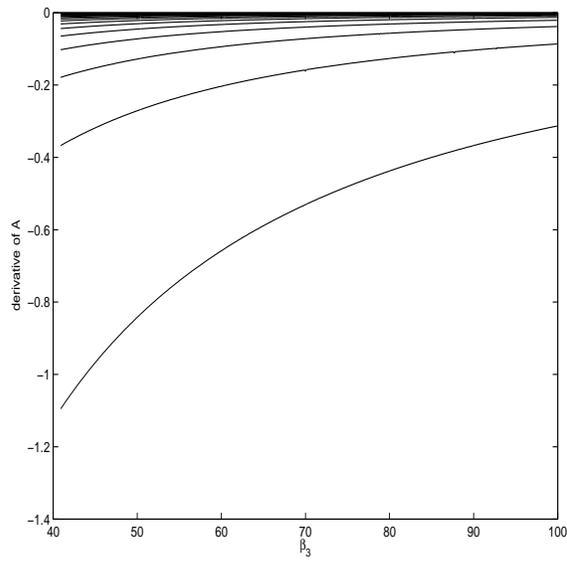


Figure D.3: Derivative of A plotted against large values of β_3 . Each curve corresponds to a different d_0 (between 0.01 and 0.99). The larger d_0 , the smaller derivative of A .

Appendix E

Proofs

Proposition 1 *The $(\beta_0, \beta_1, \beta_2, \beta_3, D)$ -parameterization is unique.*

Proof. First, look at only the part of the curve starting from the inflection time point t_I . Then, since

$$\beta_1 e^{-\beta_2 t_I} = 1 - \beta_3, \quad (\text{E.1})$$

we can write

$$\begin{aligned} g_{t_I+t} - g_{t_I} &= \beta_0 (1 - (1 - \beta_3) e^{-\beta_2 t})^{\frac{1}{1-\beta_3}} - \beta_0 \beta_3^{\frac{1}{1-\beta_3}} \\ &= \beta_0 (1 - e^{-\beta_2 t} [1 + o(1)]) - \beta_0 \beta_3^{\frac{1}{1-\beta_3}} \\ &= \beta_0 - \beta_0 e^{-\beta_2 t} [1 + o(1)] - \beta_0 \beta_3^{\frac{1}{1-\beta_3}}, \end{aligned}$$

where $o(1) \rightarrow 0$, as $t \rightarrow \infty$. Now, if two sets of parameters $(\beta_0, \beta_1, \beta_2, \beta_3, D)$ and $(\beta'_0, \beta'_1, \beta'_2, \beta'_3, D')$ pertain to the same curve g_t , then:

$$\begin{aligned} g_{t_I+t} - g_{t_I} + \beta_0 \beta_3^{\frac{1}{1-\beta_3}} - \beta_0 &= -\beta_0 e^{-\beta_2 t} [1 + o(1)] \\ g_{t_I+t} - g_{t_I} + \beta'_0 \beta'_3{}^{\frac{1}{1-\beta'_3}} - \beta'_0 &= -\beta'_0 e^{-\beta'_2 t} [1 + \tilde{o}(1)], \end{aligned}$$

where also $\tilde{o}(1) \rightarrow 0$ as $t \rightarrow \infty$. Since $\beta_0 e^{-\beta_2 t} [1 + o(1)] \rightarrow 0$ and $\beta'_0 e^{-\beta'_2 t} [1 + \tilde{o}(1)] \rightarrow 0$ as $t \rightarrow \infty$, it follows that

$$\beta_0 - \beta_0 \beta_3^{\frac{1}{1-\beta_3}} = \beta'_0 - \beta'_0 \beta'_3{}^{\frac{1}{1-\beta'_3}}, \quad (\text{E.2})$$

and therefore

$$\lim_{t \rightarrow \infty} \frac{\beta_0 e^{-\beta_2 t}}{\beta'_0 e^{-\beta'_2 t}} = 1. \quad (\text{E.3})$$

From (E.3) it follows that $\beta_0 = \beta'_0$ and $\beta_2 = \beta'_2$, and together with (E.2) we can conclude that

$$\beta_3^{\frac{1}{1-\beta_3}} = \beta'_3 \frac{1}{1-\beta'_3}$$

and hence that $\beta_3 = \beta'_3$. From (E.1) we also get that β_1 must equal β'_1 . Finally, the relation

$$g_0 = \beta_0 (1 - \beta_1)^{\frac{1}{1-\beta_3}} + D$$

shows that also D must equal D' . \square

Lemma 4 *Fix $\beta_3 > 0$. In the unit-scaled model, the function A defined in (4.10) with the constraint (4.6) is strictly increasing as a function of d_0 , $0 < d_0 < 1$.*

Proof. Let g_t be a unit-scaled curve corresponding to arbitrary fixed β_3 and d_0 , and denote the asymptotic parameter of this curve A . Consider instead this curve starting from time point $0 < T < t_I$, where t_I is the inflection time point of g_t , and re-scale and translate it to

$$g_t(T) = \frac{g_{T+t(1-T+g_T)} - g_T}{1 - T + g_T}.$$

This new curve's t -derivative at zero is $d_0(T) = g'_T > d_0$, and the re-scalings and translation were chosen so that the other four parameters are unchanged, so that $g_t(T)$ is again a unit-scaled curve. Denote the asymptote of the new curve

$$A(T) = \frac{A - g_T}{1 - T + g_T}.$$

It is straightforward to see that the derivative of $A(T)$ at $T = 0$ is

$$A'(0+) = -d_0 + (1 - d_0)A,$$

which is strictly positive if

$$A > \frac{d_0}{1 - d_0}.$$

By convexity of g_t ,

$$g_t > d_0 t \text{ for } t \leq t_I.$$

Furthermore, we have

$$d_0 t \geq t - 1 \text{ for } t \leq \frac{1}{1 - d_0}.$$

Thus if we assume that $t_I \leq \frac{1}{1 - d_0}$, we may conclude from the above inequalities that

$$g_{t_I} > t_I - 1.$$

But this would contradict the fact that g_t is unit-scaled, since this property implies that

$$g_{t_I} = t_I - 1,$$

and thus it follows that $t_I > \frac{1}{1 - d_0}$. This forces

$$g_{t_I} > \frac{d_0}{1 - d_0},$$

and therefore also $A > \frac{d_0}{1 - d_0}$, which proves $A'(0+) > 0$.

Consider the relation (using an obvious notation on the left side)

$$A(1, d_0(T), 1, 1, \beta_3) = A(T). \tag{E.4}$$

Recall that $d_0(T) = g'_T$, and differentiate and evaluate (E.4) at $T = 0$ to get

$$D_{d_0}[A(1, d_0, 1, 1, \beta_3)]d'_0(0) = D_{d_0}[A(1, d_0, 1, 1, \beta_3)]g''_0 = A'(0+).$$

Local convexity implies that $g''_0 > 0$, which proves $D_{d_0}[A(1, d_0, 1, 1, \beta_3)] > 0$. \square

Lemma 5 *Fix $0 < d_0 < 1$. In the unit-scaled model, the function A defined in (4.10) with the constraint (4.6) satisfies*

(a) $\lim_{\beta_3 \rightarrow 0} A = \infty$

(b) $\lim_{\beta_3 \rightarrow \infty} A = \frac{1 - d_0}{d_0 - \log(d_0) - 1}$.

Proof.

(a) Note that $1 - \beta_3 < \beta_1 < 1$ and $0 < \beta_3 < 1$, and

$$A = \beta_0 \left[1 - (1 - \beta_1)^{\frac{1}{1 - \beta_3}} \right].$$

We first look at the asymptotics of

$$\left[1 - (1 - \beta_1)^{\frac{1}{1-\beta_3}}\right].$$

From $1 - \beta_3 < \beta_1 < 1$ and $0 < \beta_3 < 1$, we get that

$$0 < (1 - \beta_1)^{\frac{1}{1-\beta_3}} < \beta_3^{\frac{1}{1-\beta_3}},$$

and since

$$\lim_{\beta_3 \rightarrow 0^+} \beta_3^{\frac{1}{1-\beta_3}} = 0,$$

it follows that

$$\lim_{\beta_3 \rightarrow 0^+} \left[(1 - \beta_1)^{\frac{1}{1-\beta_3}} \right] = 0 \quad (\text{E.5})$$

and hence

$$\lim_{\beta_3 \rightarrow 0^+} \left[1 - (1 - \beta_1)^{\frac{1}{1-\beta_3}} \right] = 1.$$

Next, we look at the asymptotics of

$$\beta_0 = \frac{1}{\beta_3^{\frac{\beta_3}{1-\beta_3}} \left[\log \left(\frac{\beta_1}{1-\beta_3} \right) - \beta_3 \right] + (1 - \beta_1)^{\frac{1}{1-\beta_3}}}.$$

From $1 - \beta_3 < \beta_1 < 1$ and $0 < \beta_3 < 1$ it follows that

$$-\beta_3 < \log \left(\frac{\beta_1}{1-\beta_3} \right) - \beta_3 < -\log(1 - \beta_3) - \beta_3,$$

so that

$$-\beta_3 \beta_3^{\frac{\beta_3}{1-\beta_3}} < \beta_3^{\frac{\beta_3}{1-\beta_3}} \left[\log \left(\frac{\beta_1}{1-\beta_3} \right) - \beta_3 \right] < -\beta_3^{\frac{\beta_3}{1-\beta_3}} [\log(1 - \beta_3) + \beta_3]. \quad (\text{E.6})$$

The limit of the left side of (E.6) is

$$\lim_{\beta_3 \rightarrow 0^+} \left(-\beta_3 \beta_3^{\frac{\beta_3}{1-\beta_3}} \right) = 0$$

and the limit of the right side of (E.6) is

$$\lim_{\beta_3 \rightarrow 0^+} -\beta_3^{\frac{\beta_3}{1-\beta_3}} [\log(1 - \beta_3) + \beta_3] = 0,$$

and hence

$$\lim_{\beta_3 \rightarrow 0^+} \beta_3^{\frac{\beta_3}{1-\beta_3}} \left[\log \left(\frac{\beta_1}{1-\beta_3} \right) - \beta_3 \right] = 0. \quad (\text{E.7})$$

Using (E.5) and (E.7) and the fact that $\beta_0 > 0$, we get

$$\begin{aligned} \lim_{\beta_3 \rightarrow 0^+} \beta_0 &= \lim_{\beta_3 \rightarrow 0^+} \frac{1}{\beta_3^{\frac{\beta_3}{1-\beta_3}} \left[\log \left(\frac{\beta_1}{1-\beta_3} \right) - \beta_3 \right] + (1-\beta_1)^{\frac{1}{1-\beta_3}}} \\ &= \infty, \end{aligned}$$

so that

$$\begin{aligned} \lim_{\beta_3 \rightarrow 0^+} A &= \left(\lim_{\beta_3 \rightarrow 0^+} \beta_0 \right) \left(\lim_{\beta_3 \rightarrow 0^+} \left[1 - (1-\beta_1)^{\frac{1}{1-\beta_3}} \right] \right) \\ &= \infty. \end{aligned}$$

(b) Let $A(1, d_0, 1, 1, \beta_3)$ denote A as a function of β_3 when $s = 1$, $\lambda = 1$, $\mu = 1$, and $0 < d_0 < 1$ is fixed. Note that $\beta_1 < 1 - \beta_3$. Take a fix β_3 and fix β_1 ,

$$\beta_1 = \frac{-\beta_3 e}{d_0^{\beta_3 - 1}}. \quad (\text{E.8})$$

Now using equation (E.8) in the equation of the derivative at time zero (4.6) we get

$$d_{\beta_3} = \frac{\left(\frac{-\beta_3 e}{d_0^{\beta_3 - 1}} \right) \left[1 - \left(\frac{-\beta_3 e}{d_0^{\beta_3 - 1}} \right)^{\frac{\beta_3}{1-\beta_3}} \right]}{(1-\beta_3) \beta_3^{\frac{\beta_3}{1-\beta_3}}},$$

where $d_{\beta_3} \rightarrow d_0$ as $\beta_3 \rightarrow \infty$,

$$\begin{aligned} \lim_{\beta_3 \rightarrow \infty} d_{\beta_3} &= \lim_{\beta_3 \rightarrow \infty} \frac{\left(\frac{-\beta_3 e}{d_0^{\beta_3 - 1}} \right) \left[1 - \left(\frac{-\beta_3 e}{d_0^{\beta_3 - 1}} \right)^{\frac{\beta_3}{1-\beta_3}} \right]}{(1-\beta_3) \beta_3^{\frac{\beta_3}{1-\beta_3}}} \\ &= \lim_{\beta_3 \rightarrow \infty} \frac{\left(\frac{\beta_3 e}{d_0^{\beta_3 - 1}} \right)^{\frac{1}{1-\beta_3}}}{\beta_3^{\frac{1}{1-\beta_3}}} \\ &= d_0, \end{aligned}$$

and

$$\begin{aligned} \lim_{\beta_3 \rightarrow \infty} A(1, d_{\beta_3}, 1, 1, \beta_3) &= \lim_{\beta_3 \rightarrow \infty} \frac{1 - \left(1 + \frac{\beta_3 e}{d_0^{\beta_3 - 1}}\right)^{\frac{1}{1 - \beta_3}}}{\beta_3^{\frac{1}{1 - \beta_3}} \left[\log \left(\frac{\left(\frac{-\beta_3 e}{d_0^{\beta_3 - 1}}\right)}{1 - \beta_3} \right) - \beta_3 \right] + \left(1 + \frac{\beta_3 e}{d_0^{\beta_3 - 1}}\right)^{\frac{1}{1 - \beta_3}}} \\ &= \frac{1 - d_0}{d_0 - \log(d_0) - 1}. \end{aligned}$$

Now, if $\tilde{d}_0 > d_0$, then $\tilde{d}_{\beta_3} > d_0$ when β_3 is large enough. Lemma 4 states that $A(1, d_0, 1, 1, \beta_3)$ is strictly increasing as a function of d_0 , and hence we have that

$$\begin{aligned} A(1, d_0, 1, 1, \beta_3) &< A(1, \tilde{d}_{\beta_3}, 1, 1, \beta_3) \text{ for } \beta_3 \text{ large enough} \\ \Rightarrow \limsup A(1, d_0, 1, 1, \beta_3) &\leq \limsup A(1, \tilde{d}_{\beta_3}, 1, 1, \beta_3) \\ &= \frac{1 - \tilde{d}_0}{\tilde{d}_0 - \log(\tilde{d}_0) - 1}. \end{aligned}$$

Analogously, for $\tilde{\tilde{d}}_0 < d_0$

$$\begin{aligned} \liminf A(1, d_0, 1, 1, \beta_3) &\geq \liminf A(1, \tilde{\tilde{d}}_{\beta_3}, 1, 1, \beta_3) \\ &= \frac{1 - \tilde{\tilde{d}}_0}{\tilde{\tilde{d}}_0 - \log(\tilde{\tilde{d}}_0) - 1}. \end{aligned}$$

Due to the continuity of $\frac{1 - d_0}{d_0 - \log(d_0) - 1}$ and the arbitrariness of $\tilde{d}_0 > d_0$ and $\tilde{\tilde{d}}_0 < d_0$

$$\lim_{\beta_3 \rightarrow \infty} A(1, d_0, 1, 1, \beta_3) = \frac{1 - d_0}{d_0 - \log(d_0) - 1}.$$

□

Remark to Lemma 5(b). In fact, it can be shown that the whole curve g_t converges to a curve which is defined as

$$\frac{d_0 (e^{(d_0 - \log(d_0) - 1)t} - 1)}{d_0 - \log(d_0) - 1}, \quad \text{when } t \leq 1 + \frac{1 - d_0}{d_0 - \log(d_0) - 1}$$

and

$$\frac{1 - d_0}{d_0 - \log(d_0) - 1}, \quad \text{when } t > 1 + \frac{1 - d_0}{d_0 - \log(d_0) - 1}.$$

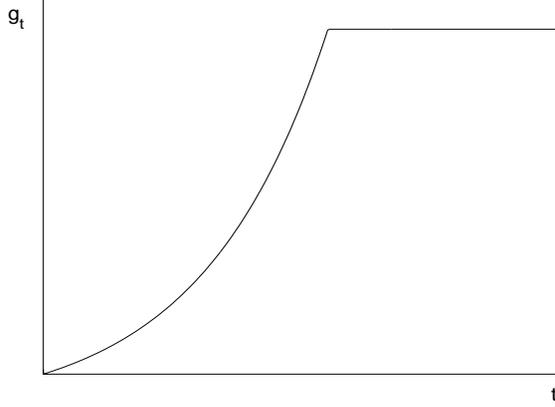


Figure E.1: A limiting curve of the Chapman-Richards model when $\beta_3 \rightarrow \infty$.

In general, the limiting curves have the form

$$\theta_1 \left(e^{-\theta_2 t} - 1 \right) + \theta_3, \quad \text{when } t \leq \frac{\log(\mu) - \log(\theta_1 \theta_2)}{\theta_2},$$

where $\theta_1, \theta_2, \theta_3 > 0$. Figure E.1 shows how these curves look like.

Lemma 6 *The $\log(\underline{Y})$ is convex as a function of $\log(d_0)$, $\log(\lambda)$, and $\log(\mu)$ for any fixed $\log(s)$, where $s > 0$, $0 < d_0 < \mu$, $\lambda > 0$, and $\mu > 0$.*

Proof. Take $x_1 = \log(\frac{d_0}{\mu}) = \log(d_0) - \log(\mu) < 0$, $x_2 = \log(\mu)$, $x_3 = \log(\lambda)$, and write $\log(\underline{Y})$ as a function of them

$$\begin{aligned} L\underline{Y}(\log(s), \mathbf{x}) &= \log \left[e^{\log(s)} \left(e^{\frac{1-e^{x_1}}{e^{x_1}-x_1-1} e^{x_2+x_3}} - 1 \right) \right] & (E.9) \\ &= \log(s) + \log \left[e^{\frac{1-e^{x_1}}{e^{x_1}-x_1-1} e^{x_2+x_3}} - 1 \right], \end{aligned}$$

where $\mathbf{x} = (x_1, x_2, x_3)$. Define

$$k(\mathbf{x}) = \frac{1 - e^{x_1}}{e^{x_1} - x_1 - 1} e^{x_2+x_3}$$

and

$$\begin{aligned} m(\mathbf{x}) &= \log \left[e^{\frac{1-e^{x_1}}{e^{x_1}-x_1-1} e^{x_2+x_3}} - 1 \right] \\ &= \log[e^{k(\mathbf{x})} - 1]. \end{aligned}$$

The function $LY(\log(s), \mathbf{x})$ is convex for any fixed $\log(s)$, if $m(\mathbf{x})$ is convex, *i.e.* if the Hessian

$$D_{\mathbf{x}}^2[m(\mathbf{x})] = \frac{e^{k(\mathbf{x})}}{e^{k(\mathbf{x})} - 1} \left[D_{\mathbf{x}}^2[k(\mathbf{x})] - \frac{1}{e^{k(\mathbf{x})} - 1} D_{\mathbf{x}}[k(\mathbf{x})] D_{\mathbf{x}}[k(\mathbf{x})]^T \right] \quad (\text{E.10})$$

is positive semidefinite. Observe that $k(\mathbf{x}) > 0$ since $x_1 < 0$, and thus

$$\frac{e^{k(\mathbf{x})}}{e^{k(\mathbf{x})} - 1} > 0.$$

Therefore, to prove that (E.10) is positive semidefinite, it is enough to prove that

$$D_{\mathbf{x}}^2[k(\mathbf{x})] - \frac{1}{e^{k(\mathbf{x})} - 1} D_{\mathbf{x}}[k(\mathbf{x})] D_{\mathbf{x}}[k(\mathbf{x})]^T \quad (\text{E.11})$$

is positive semidefinite.

The function $k(\mathbf{x})$ can be written as a product of

$$f(\mathbf{x}) = \frac{1 - e^{x_1}}{e^{x_1} - x_1 - 1} \quad \text{and} \quad g(\mathbf{x}) = e^{x_2 + x_3}.$$

Now, using that $\frac{\partial g(\mathbf{x})}{\partial x_2} = \frac{\partial g(\mathbf{x})}{\partial x_3} = g(\mathbf{x})$, we can write

$$D_{\mathbf{x}}^2[k(\mathbf{x})] = g \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_1} & f & f \\ \frac{\partial f}{\partial x_1} & f & f \end{pmatrix},$$

$$D_{\mathbf{x}}[k(\mathbf{x})] D_{\mathbf{x}}[k(\mathbf{x})]^T = g^2 \begin{pmatrix} \left(\frac{\partial f}{\partial x_1} \right)^2 & \frac{\partial f}{\partial x_1} f & \frac{\partial f}{\partial x_1} f \\ \frac{\partial f}{\partial x_1} f & f^2 & f^2 \\ \frac{\partial f}{\partial x_1} f & f^2 & f^2 \end{pmatrix},$$

and (E.11) can be written as

$$g \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} - \left(\frac{\partial f}{\partial x_1} \right)^2 \frac{g}{e^{gf} - 1} & \frac{\partial f}{\partial x_1} - \frac{\partial f}{\partial x_1} \frac{fg}{e^{gf} - 1} & \frac{\partial f}{\partial x_1} - \frac{\partial f}{\partial x_1} \frac{fg}{e^{gf} - 1} \\ \frac{\partial f}{\partial x_1} - \frac{\partial f}{\partial x_1} \frac{fg}{e^{gf} - 1} & f - \frac{f^2 g}{e^{gf} - 1} & f - \frac{f^2 g}{e^{gf} - 1} \\ \frac{\partial f}{\partial x_1} - \frac{\partial f}{\partial x_1} \frac{fg}{e^{gf} - 1} & f - \frac{f^2 g}{e^{gf} - 1} & f - \frac{f^2 g}{e^{gf} - 1} \end{pmatrix},$$

which is positive semidefinite if all the submatrices have a non-negative determinant [29], and that is what we will prove next. We have that $g > 0$, we will prove that

$$f - \frac{f^2 g}{e^{gf} - 1} > 0 \quad (\text{E.12})$$

$$\frac{\partial f}{\partial x_1} - \frac{\partial f}{\partial x_1} \frac{fg}{e^{gf} - 1} > 0 \quad (\text{E.13})$$

$$\frac{\partial^2 f}{\partial x_1^2} - \left(\frac{\partial f}{\partial x_1} \right)^2 \frac{g}{e^{gf} - 1} > 0 \quad (\text{E.14})$$

$$\left[\frac{\partial^2 f}{\partial x_1^2} - \left(\frac{\partial f}{\partial x_1} \right)^2 \frac{g}{e^{gf} - 1} \right] \left[f - \frac{f^2 g}{e^{gf} - 1} \right] - \left[\frac{\partial f}{\partial x_1} - \frac{\partial f}{\partial x_1} \frac{fg}{e^{gf} - 1} \right]^2 > 0, \quad (\text{E.15})$$

the determinants of all the other submatrices can immediately be seen to equal zero.

To prove (E.12) and (E.13), we will use that

$$e^{gf} - 1 > gf \Rightarrow \frac{g}{e^{gf} - 1} < \frac{1}{f}. \quad (\text{E.16})$$

Now,

$$f - \frac{f^2 g}{e^{gf} - 1} > f - \frac{f^2}{f} = 0,$$

and

$$\frac{\partial f}{\partial x_1} - \frac{\partial f}{\partial x_1} \frac{fg}{e^{gf} - 1} > \frac{\partial f}{\partial x_1} - \frac{\partial f}{\partial x_1} \frac{f}{f} = 0.$$

To prove (E.14) is a slightly more elaborate task. We will use the inequality

$$x_1 > e^{x_1/2} - e^{-x_1/2}, \text{ if } x_1 < 0, \quad (\text{E.17})$$

which in turn can be seen by defining

$$h(x_1) = x_1 - (e^{x_1/2} - e^{-x_1/2}),$$

and taking derivative

$$\begin{aligned} h'(x_1) &= 1 - \frac{1}{2}(e^{x_1/2} + e^{-x_1/2}) \\ &< 1 - \sqrt{e^{x_1/2} e^{-x_1/2}} = 0, \text{ for } x_1 < 0, \end{aligned}$$

where in the end we used the well known inequality between arithmetic and geometric means [15]. From this and $h(0) = 0$ it follows that $h(x_1) > 0$ for all $x_1 < 0$, which is equivalent to (E.17).

Now, from (E.16) we first get

$$\frac{\partial^2 f}{\partial x_1^2} - \left(\frac{\partial f}{\partial x_1}\right)^2 \frac{g}{e^{gf} - 1} > \frac{\partial^2 f}{\partial x_1^2} - \left(\frac{\partial f}{\partial x_1}\right)^2 \frac{1}{f}, \quad (\text{E.18})$$

and then we study the sign of this bound

$$\frac{\partial^2 f}{\partial x_1^2} - \left(\frac{\partial f}{\partial x_1}\right)^2 \frac{1}{f} = \frac{e^{3x_1}x_1 - 2e^{3x_1} + 5e^{2x_1} - e^{x_1}x_1^2 - e^{x_1}x_1 - 4e^{x_1} + 1}{(-e^{x_1} + x_1 + 1)^3(e^{x_1} - 1)}, \quad (\text{E.19})$$

where $(-e^{x_1} + x_1 + 1)^3 < 0$ and $e^{x_1} - 1 < 0$, for $x_1 < 0$, so the denominator is positive. We will prove (E.14) from positivity of the nominator, which we may first rewrite as

$$(e^{x_1} - 1)^2 - e^{x_1}x_1^2 + e^{x_1} \{(e^{x_1} - 1)[x_1(e^{x_1} + 1) - 2(e^{x_1} - 1)]\},$$

where $e^{x_1} > 0$, $e^{x_1} - 1 < 0$, and then from (E.17) we see that

$$\begin{aligned} (e^{x_1} - 1)^2 - e^{x_1}x_1^2 &> (e^{x_1} - 1)^2 - e^{x_1}(e^{\frac{x_1}{2}} - e^{-\frac{x_1}{2}})^2 \\ &= (e^{x_1} - 1)^2 - (e^{x_1} - 1)^2 \\ &= 0, \text{ for } x_1 < 0. \end{aligned}$$

We get (E.14) if we also show that $x_1(e^{x_1} + 1) - 2(e^{x_1} - 1) < 0$ for $x_1 < 0$. Write

$$v(x_1) = x_1(e^{x_1} + 1) - 2(e^{x_1} - 1).$$

Then

$$\begin{aligned} v'(x_1) &= 1 + e^{x_1}(x_1 - 1) \\ &> 1 + e^{x_1}(e^{\frac{x_1}{2}} - e^{-\frac{x_1}{2}} - 1) \quad (\text{by E.17}) \\ &= e^{\frac{x_1}{2}}(e^{x_1} - 1) - (e^{x_1} - 1) \\ &= (e^{x_1} - 1)(e^{\frac{x_1}{2}} - 1) \\ &> 0, \end{aligned}$$

and $v(0) = 0$. Thus

$$v(x_1) = x_1(e^{x_1} + 1) - 2(e^{x_1} - 1) < 0, \quad \text{for } x_1 < 0,$$

and the right-hand side of (E.19) is therefore positive, which together with (E.18) proves (E.14).

To prove (E.15) we first observe that

$$\frac{\partial^2 f}{\partial x_1^2} f > \left(\frac{\partial f}{\partial x_1} \right)^2, \quad (\text{E.20})$$

since the right-hand side of (E.19) is positive by the proof of (E.14). Using (E.20) we get

$$\begin{aligned} & \left[\frac{\partial^2 f}{\partial x_1^2} - \left(\frac{\partial f}{\partial x_1} \right)^2 \frac{g}{e^{gf} - 1} \right] \left[f - \frac{f^2 g}{e^{gf} - 1} \right] - \left[\frac{\partial f}{\partial x_1} - \frac{\partial f}{\partial x_1} \frac{fg}{e^{gf} - 1} \right]^2 \\ & > \left[\frac{\partial^2 f}{\partial x_1^2} - \frac{\partial^2 f}{\partial x_1^2} f \frac{g}{e^{gf} - 1} \right] \left[f - \frac{f^2 g}{e^{gf} - 1} \right] - \left[\frac{\partial f}{\partial x_1} - \frac{\partial f}{\partial x_1} \frac{fg}{e^{gf} - 1} \right]^2 \\ & = \left[\frac{\partial^2 f}{\partial x_1^2} \left(1 - \frac{fg}{e^{gf} - 1} \right) \right] \left[f \left(1 - \frac{fg}{e^{gf} - 1} \right) \right] - \left[\frac{\partial f}{\partial x_1} \left(1 - \frac{fg}{e^{gf} - 1} \right) \right]^2 \\ & = \left[1 - \frac{fg}{e^{gf} - 1} \right]^2 \left[\frac{\partial^2 f}{\partial x_1^2} f - \left(\frac{\partial f}{\partial x_1} \right)^2 \right] > 0. \end{aligned}$$

The transformation between $\log(d_0)$, $\log(\lambda)$, $\log(\mu)$ and x_1, x_2, x_3 is affine so that the convexity of $\log(\underline{Y})$ holds also in the former coordinates and the lemma is proved.

□

Appendix F

Simultaneous models for two growth curves

Often in case of double samples, the initial OD values of the two samples vary, but the end OD values are almost the same, and the growth curves have approximately the same shape except for the length of the exponential phase. This is natural, 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 double samples, and they can have different initial and final OD values, but the shapes of the growth curves (apart from the length of the exponential phase) still tend to be nearly the same. In such cases modeling the growth curves simultaneously would possibly give a better estimate of the growth behavior than *e.g.* taking averages of growth parameters of two separately modeled curves.

F.1 Model I

We tried to model two growth curves simultaneously using the three part model presented in Section 3.3 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 initial OD and the logarithm of the end OD) is

$$g_t^{*(1)} = \begin{cases} g_t^{(1)}, & t \leq t_I, \\ g_{t_I}^{(1)} + \mu(t - t_I), & t_I \leq t \leq t_I + \Delta_1, \\ g_{t_I - \Delta_1}^{(1)} + \mu\Delta_1, & t \geq t_I + \Delta_1, \end{cases} \quad (\text{F.1})$$

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

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

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

$$g_t^{*(2)} = \begin{cases} g_t^{(2)}, & t \leq t_I, \\ g_{t_I}^{(2)} + \mu(t - t_I), & t_I \leq t \leq t_I + \Delta_2, \\ g_{t_I - \Delta_2}^{(2)} + \mu\Delta_2, & t \geq t_I + \Delta_2, \end{cases} \quad (\text{F.2})$$

where

$$g_t^{(2)} = \beta_0 \left[1 - \beta_1 e^{-\beta_2 t} \right]^{1/(1-\beta_3)} + D_2. \quad (\text{F.3})$$

This model as well as model II below can easily be generalized to more than two samples.

F.2 Model II

We also tried to fit a model where the asymptotes of the curves (F.1) and (F.2) were forced to be the same. Now

$$\begin{aligned} \beta_0 + \Delta_1 \beta_0 \beta_2 \beta_3^{\frac{\beta_3}{1-\beta_3}} + D_1 &= \beta_0 + \Delta_2 \beta_0 \beta_2 \beta_3^{\frac{\beta_3}{1-\beta_3}} + D_2 \\ \Rightarrow D_2 &= D_1 + (\Delta_1 - \Delta_2) \beta_0 \beta_2 \beta_3^{\frac{\beta_3}{1-\beta_3}} \\ &= D_1 + (\Delta_1 - \Delta_2) \mu. \end{aligned}$$

Thus, the Chapman-Richards function (F.3) in the model of the curve with a lower initial OD can be written as

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

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.

F.3 Fitting the simultaneous models to the data

We fitted the simultaneous models on duplicate measurements of the data presented in Section 3.2.3. The least squares method was used the same way as in Section 3.2.3.

With both of the models the estimates of Δ_1 were nearly always zero. With model I the fit was rather good when the shapes of the two curves were almost the same. However, if the shapes differed much, the fit was not good. When the asymptotes of the two curves were nearly the same, the fits of the two models were similar, see *e.g.* Figure F.1. When the asymptotes really differed, the fit was naturally better with model I, see *e.g.* Figure F.2. Also if the difference between the time spans of the exponential phases is very large, the fit can become poor.

It might be useful to be able to model the eight wild types in each run simultaneously. Although it is possible to generalize the simultaneous models for more than two curves, the computations would become rather complicated. Moreover, the simultaneous models do not enable easy comparison of all the curves in the experiment.

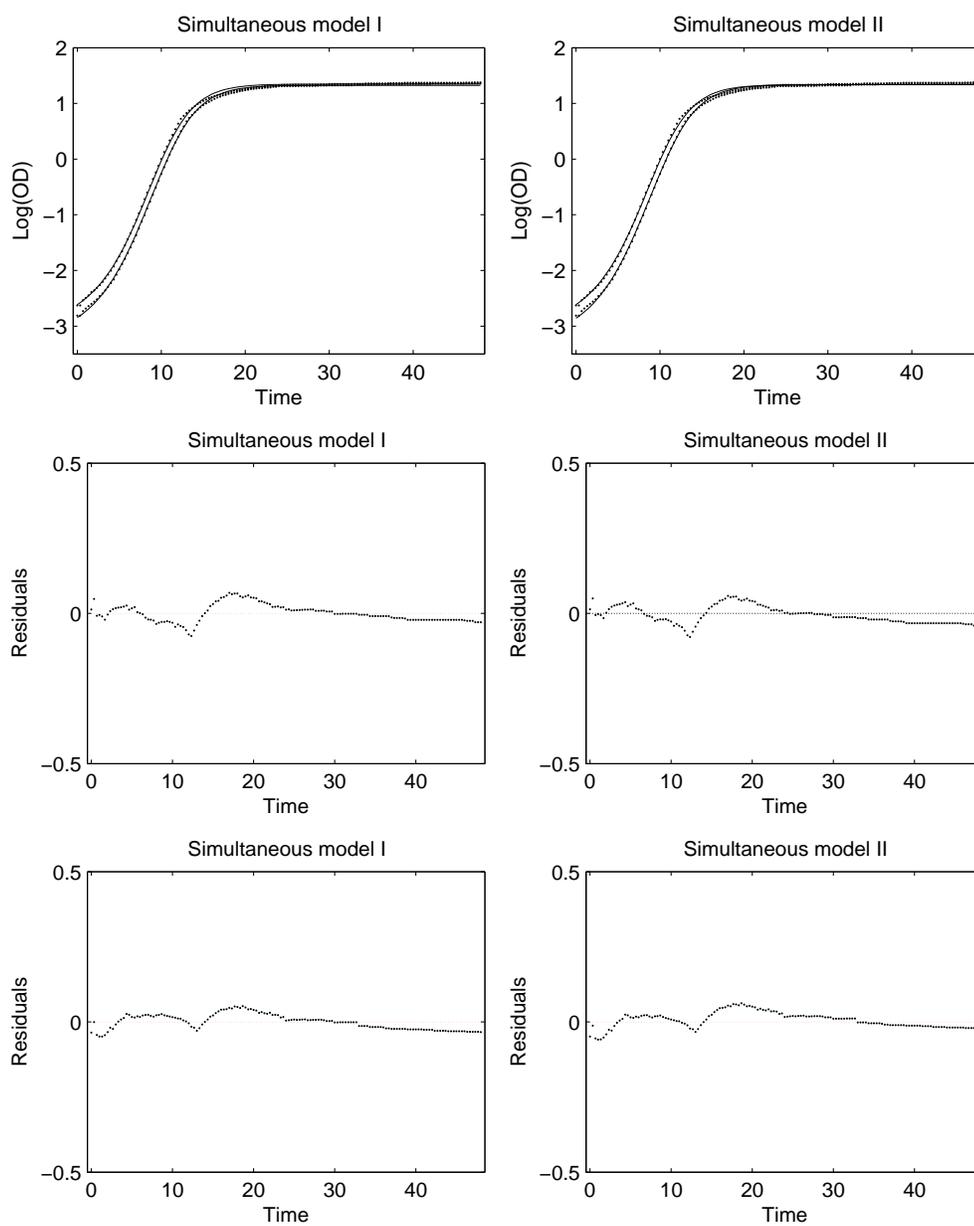


Figure F.1: *Two curves are fitted using the simultaneous models. The residual plots of the fit of the upper curve are in the middle and the residual plots of the fit of lower curve are at the bottom.*

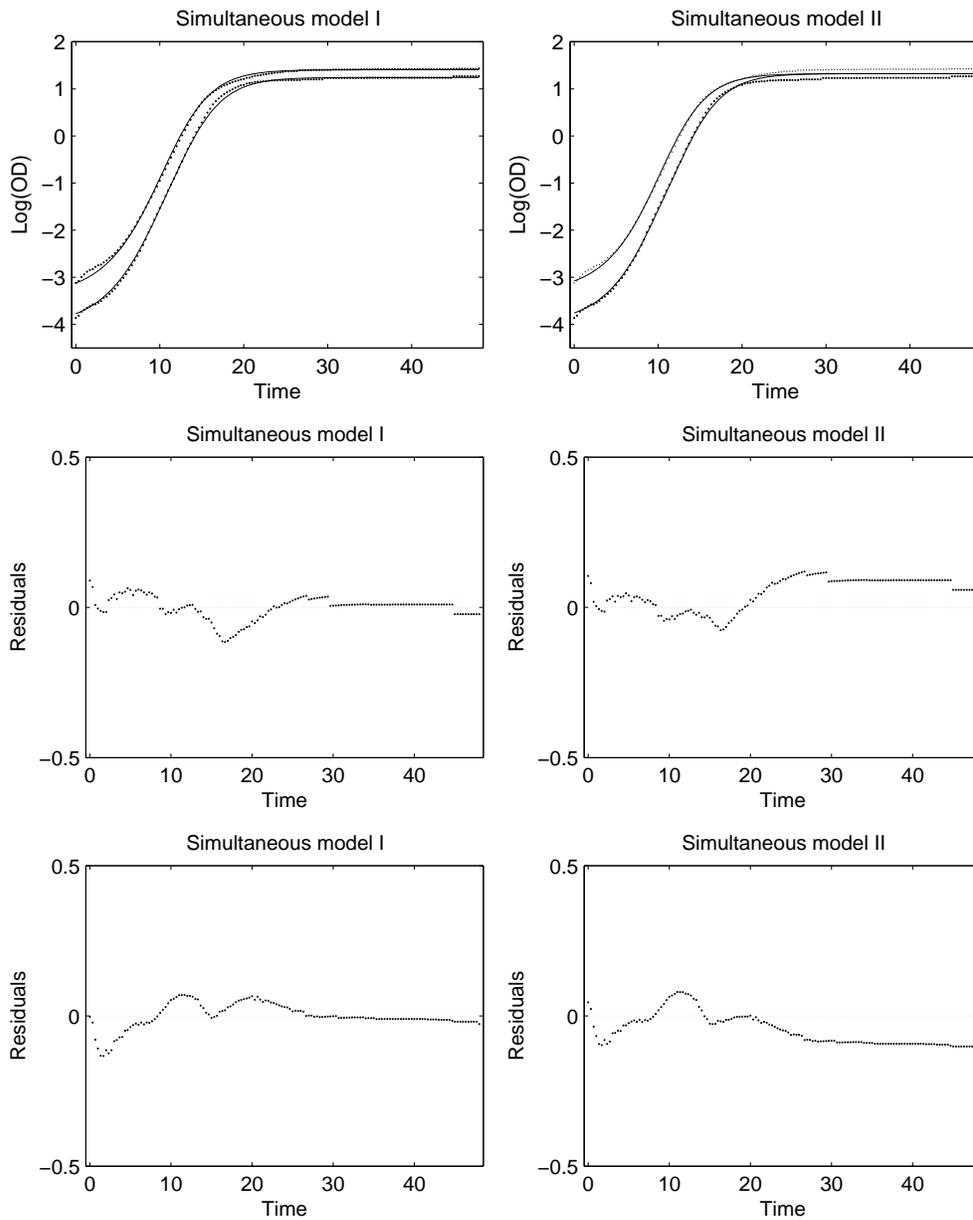


Figure F.2: *Two curves are fitted using the simultaneous models. The residual plots of the fit of the upper curve are in the middle and the residual plots of the fit of lower curve are at the bottom.*